Genome-wide association and genome partitioning reveal novel genomic regions underlying variation in gastrointestinal nematode burden in a wild bird

Marius A. Wenzel*, Marianne C. James§, Alex Douglas & Stuart B. Piertney

July 6, 2015

Institute of Biological and Environmental Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

* corresponding author. email address: marius.a.wenzel.08@aberdeen.ac.uk. Phone number: +44 1224 272395. Fax number: +44 (0)1224 272396

§ current address: Food Standards Scotland, Aberdeen AB11 5RL, UK

Keywords: host-parasite interactions; life-history trade-offs; genome-wide association; GWAS; genome partitioning; missing heritability

Running title: Genetic basis of nematode burden in red grouse

Abstract

Identifying the genetic architecture underlying complex phenotypes is a notoriously difficult problem that often impedes progress in understanding adaptive eco-evolutionary processes in natural populations. Host-parasite interactions are fundamentally important drivers of evolutionary processes, but a lack of understanding of the genes involved in the host's response to chronic parasite insult makes it particularly difficult to understand the mechanisms of host life-history trade-offs and adaptive dynamics involved. Here we examine the genetic basis of gastrointestinal nematode (*Trichostrongylus tenuis*) burden in 695 red grouse (*Lagopus lagopus scotica*) individuals genotyped at 384 genome-wide SNPs. We first use genome-wide association to identify individual SNPs associated with nematode burden. We then partition genome-wide heritability to identify chromosomes with greater heritability than expected from gene content, due to harbouring a multitude of additive SNPs with individually undetectable effects. We identified five SNPs on five chromosomes that accounted for differences of up to 556 worms per bird, but together explained at best 4.9 % of the phenotypic variance. These SNPs were closely linked to genes representing a range of physiological processes including the immune system, protein degradation and energy metabolism. Genome partitioning

indicated genome-wide heritability of up to 29 % and three chromosomes with excess heritability of up to 4.3 % (total 8.9 %). These results implicate SNPs and novel genes underlying nematode burden in this system and suggest that this phenotype is somewhere between being based on few large-effect genes (oligogenic) and based on a large number of genes with small individual but large combined effects (polygenic).

Introduction

Host-parasite interactions are widely recognised as fundamentally important drivers of evolutionary and ecological processes in natural populations. The Red-Queen dynamics of the host-parasite co-evolutionary arms race dictate that the ability of a host to cope with parasite insult is a major component of individual fitness (van Valen, 1973; Brockhurst, 2011). This selective pressure on the host can influence how energetic resources are allocated to immune defence over other vital cellular functions and life-history components, and thus how parasite-driven life-history trade-offs are resolved (Sheldon & Verhulst, 1996; Zuk & Stoehr, 2002; Schmid-Hempel, 2003). Such decisions affect individual mating behaviour and sexual selection processes (Hamilton & Zuk, 1982; Hill & Farmer, 2005) as well as population demography and dynamics (Lochmiller, 1996; Hudson *et al.*, 2006).

A proper understanding of host-parasite eco-evolutionary dynamics requires knowledge of the genetic basis of individual host resistance or susceptibility (Paterson & Piertney, 2011). Traditionally, and perhaps intuitively, most focus has been placed on the immune system (Lochmiller & Deerenberg, 2000; Schmid-Hempel, 2003; Sadd & Schmid-Hempel, 2009; Owen *et al.*, 2010). As such, a range of classic immunological candidate genes have been implicated, including the Major Histocompatibility Complex MHC (Piertney & Oliver, 2006; Tobler *et al.*, 2014), interferon gamma IFNG (Coltman *et al.*, 2001; Stear *et al.*, 2007), Toll-like receptors TLR (Downing *et al.*, 2010; Gavan *et al.*, in press) and cytokines (Downing *et al.*, 2010; Turner *et al.*, 2012). However, a focus on the immune system alone may be misleading in systems where hosts do not attain complete parasite resistance and instead suffer chronic infection (Schmid-Hempel, 2003; Sadd & Schmid-Hempel, 2009). These cases might necessitate a complex constitutive response by the host, which may involve a multitude of genes across broad physiological processes rather than one or few particular key immune genes. Indeed, parasites in general and nematodes in particular often evade or manipulate the host's immune response and have extensive multidimensional effects on host behaviour and physiology (Thomas *et al.*, 2010; Biron & Loxdale, 2013; Poulin, 2013). If the host response is then mediated through changes in the allocation of resources among the different life-history categories and cellular processes (Sheldon & Verhulst, 1996; Lochmiller & Deerenberg, 2000), this will inevitably involve a broad suite of genotypic mechanisms with a range of physiological functions divorced from the immune system (Hill, 2011).

In order to identify novel genes that affect a host's response to parasites, functional transcriptomic

assays can provide extensive suites of gene targets, which can then be screened for genotypic variation (Hoffmann & Willi, 2008; Piertney & Webster, 2010; Orsini *et al.*, 2011; Pemberton *et al.*, 2011; Gossner *et al.*, 2013; Wenzel & Piertney, 2015). However, this strategy does not necessarily provide a robust understanding of the genetic basis of a phenotype because transcriptomic responses are context dependent and are not always linked to genotypic variation in homologues of different species (Orsini *et al.*, 2012; Brown *et al.*, 2013; Wenzel & Piertney, 2015). A more robust genome-scale understanding of genetic architecture can be gained by quantitative-trait-loci (QTL) mapping or genome-wide association (GWA) analysis directly in the target species (Slate, 2005; Hill, 2012). These approaches are based on associations between genome-wide genetic markers and phenotypic traits, usually without *a priori* marker selection based on putative function. QTL mapping and GWA analysis have been undertaken extensively for gastrointestinal nematode burden in sheep or cattle breeds and have highlighted that nematode burden is associated with small numbers of predominantly anonymous loci, some of which are located near classic immunological candidate genes (e.g., Beh *et al.*, 2002; Davies *et al.*, 2006; Beraldi *et al.*, 2007; Silva *et al.*, 2012; Riggio *et al.*, 2013). These results are consistent with insights from QTL studies in other host-parasite systems and suggest that parasite burden may after all be primarily based on small numbers of large-effect immune genes (Wilfert & Schmid-Hempel, 2008).

Notwithstanding, a major issue with these methods is that those genotypic variants that are identified through QTL or GWA usually explain only a small fraction of the phenotypic variance even if the phenotype is highly heritable (Manolio *et al.*, 2009). This "missing heritability" issue fuels an ongoing controversy over the power of candidate gene or QTL/GWA approaches to detect small-effect polymorphisms and whether the endeavour to identify individual genotypic variants linked to complex phenotypes is inherently misguided (Amos *et al.*, 2011; Rockman, 2012; Robinson *et al.*, 2013). The central tenet of the argument is that polygenic phenotypes that are underpinned by the joint effect of a multitude of loci each of small effect, corresponding to Fisher's infinitesimal model (Fisher, 1919), are intractable with SNP-by-SNP association approaches because only loci with individually large effects can be detected. As such, these loci may be misleading about the distribution and effect size of the genetic variants that truly underpin phenotypic evolution (Rockman, 2012). The high incidence of missing heritability suggests that polygenic architectures may be common and thus reinforces the argument to abandon classic methods involving candidate genes (Yang *et al.*, 2011b; Hill, 2012; Rockman, 2012). Parasite susceptibility may well be such a polygenic phenotype if parasites have extensive physiological and behavioural effects on their host (Thomas *et al.*, 2010; Biron & Loxdale, 2013; Poulin, 2013). Although heritability of parasite burden is well established, for example in sheep, cattle and poultry (e.g., Stear *et al.*, 2007; Kaufmann *et al.*, 2011), and previous candidate gene and QTL/GWA studies have indeed identified some large effect genes, these approaches may well have failed to identify genes with subtle individual, but large joint effects.

One way forward in addressing this issue is to compare results obtained from GWA with emerging "genome partitioning" methods based on quantitative genetics. These approaches do not identify individual SNPs or haplotypes, rather than linkage groups of SNPs (usually chromosomes) that together explain more phenotypic variance than expected from a polygenic null model where all markers contribute equally (Yang *et al.*, 2011b,a; Hill, 2012). Genome partitioning has proven extremely useful for retrieving some of the missing heritability unaccounted for by genotypic variants identified through GWA. This has been well illustrated in complex phenotypes, for example in human height (Yang *et al.*, 2010, 2011b) and disorders such as schizophrenia, Tourette syndrome and obsessive-compulsive disorder (Lee *et al.*, 2012; Davis *et al.*, 2013). Genome partitioning can also be useful in non-model systems with limited genomic resources. Gastrointestinal nematode burden in sheep has recently been shown to be primarily polygenic, but five chromosomes explained disproportionate amounts of phenotypic variance, implying particular genomic regions as targets for fine-mapping causal variants (Al-Kalaldeh *et al.*, 2013). The utility of comparing across QTL/GWA and genomic partitioning approaches is further illustrated by contrasting results in recent studies that examined the genetic basis of wing length in birds. In spite of multiple QTLs identified in zebra finches (Schielzeth *et al.*, 2012) and reed warblers (Tarka *et al.*, 2010), genome partitioning in great tits with substantial numbers of samples and markers instead suggests that wing length may be a purely polygenic additive phenotype (Robinson *et al.*, 2013). This highlights that genomic architectures may be idiosyncratic among different species and that genome partitioning may be a useful method for reconciling differences in the numbers and identities of large-effect genetic polymorphisms.

Here, we examine the genetic basis of chronic nematode infection in a wild bird species using GWA and genome partitioning approaches to identify genotypic variants that explain detectable amounts of phenotypic variance and genomic regions that harbour missing heritability. We focus on red grouse (*Lagopus lagopus scotica*) and its primary gastrointestinal nematode parasite *Trichostrongylus tenuis*. This host-parasite system has become an important model for understanding parasite-driven demographic and ecological dynamics in the absence of a classic immune response (Hudson *et al.*, 1998; Redpath *et al.*, 2006a; Webster *et al.*, 2011b). Red grouse are an economically important game bird endemic to the upland heather moors of Scotland and northern England (Martínez-Padilla *et al.*, 2014). *T. tenuis* displays a direct life cycle where infective larvae are ingested with heather shoots, adults establish in the caeca and eggs are shed with faeces (Hudson & Dobson, 1989). Adult worms burrow into the caecal walls where they cause haemorrhaging and necrosis with substantial impacts on grouse fitness (Hudson, 1986; Watson *et al.*, 1987; Hudson *et al.*, 1992; Delahay *et al.*, 1995; Delahay & Moss, 1996). Prevalence of infection with *T. tenuis* in grouse populations is typically greater than 90 % and grouse bear chronic nematode burdens due to an inability to purge the infection despite mounting directed responses (Wilson, 1983; Shaw & Moss, 1989; Webster *et al.*, 2011a). This long-term exposure to nematode insult has marked effects on grouse life-history trade-offs and population dynamics. Interactions between nematode burden and

testosterone-mediated territorial contest behaviour and sexual selection by females have been identified as key drivers of life-history trade-offs in male grouse in particular (MacColl *et al.*, 2000; Mougeot *et al.*, 2003, 2004, 2005c, 2009; Piertney *et al.*, 2008; Bortolotti *et al.*, 2009; Wenzel *et al.*, 2013). Further, the direct impact of nematode infection on grouse fecundity and survival together with intrinsic density-dependent aggression and stress-mediated immuno-suppression implicates *T. tenuis* as a major component in regulating grouse population dynamics (Hudson *et al.*, 1998; Mougeot *et al.*, 2005b; Seivwright *et al.*, 2005; Redpath *et al.*, 2006a; Webster *et al.*, 2011b).

Motivated by this importance of chronic nematode infection for red grouse ecology, substantial effort has gone into characterising the physiological categories involved in the molecular response to infection and into identifying genomic regions associated with variation in nematode burden among individual grouse. Transcriptomic assays following experimental nematode infection have highlighted a broad range of physiological categories beyond the immune system and have identified testosterone-dependent transcription dynamics that are consistent with trade-offs involving depression of parasite defence mechanisms (Webster *et al.*, 2011a,b; Wenzel *et al.*, 2013). These insights have been channelled into developing novel candidate genes for nematode susceptibility (Wenzel *et al.*, 2015), which have been confirmed to explain variation in nematode burden among grouse individuals in a network of moors in north-east Scotland (Wenzel & Piertney, 2015). In that same study system, DNA methylation patterns at specific genomic regions have been linked to nematode burden, suggesting that parasites may affect epigenetic mechanisms impacting the regulations of specific genes (Wenzel & Piertney, 2014). In concert, this previous work has established that the host-parasite interactions in this system may be linked to multiple large-effect candidate genes involved in immune system, oxidative stress, energy metabolism and cell cycle processes. However, it is unknown whether previously undetected genome-wide large-effect genotypic variants exist, and to what extent linkage groups of variants with individually undetectable effects contribute to variation in nematode burden.

We extend and expand on these issues by carrying out genome-wide association and genome partitioning of variation in nematode burden in a sample of 695 red grouse individuals from five locations in Scotland and England across twelve years and genotyped at 384 genome-wide SNPs. First, we fit custom generalized linear models to identify individual SNPs that are associated with individual nematode burden. Second, we ascertain putative functions of these identified SNPs via linked genes identified through homology with the chicken genome. Finally, we estimate genome-wide and chromosome-specific heritabilities of nematode burden, and identify chromosomes that contribute disproportionately to heritability through multiple additive genotypic variants of individually small effects.

Materials and Methods

Phenotypic data

A total of 695 red grouse individuals were sampled at four sites near Deeside in north-east Scotland and one site in Catterick, northern England from 1995–2012 (Table 1). The core sampling effort was carried out at Invermark and Invercauld in 2011–2012 and was supplemented with birds that were sampled as part of previous studies investigating grouse population ecology (e.g., MacColl *et al.*, 2000; Redpath *et al.*, 2006b), behavioural ecology (e.g., Mougeot *et al.*, 2003, 2009) and physiological responses to nematode infection (e.g., Webster *et al.*, 2011b; Wenzel *et al.*, 2013; Wenzel & Piertney, 2014, 2015). Birds were either captured at night using lamping and netting techniques or collected following driven or walked-up sporting shoots. *T. tenuis* burdens were estimated from faecal egg counts using the standard McMaster chamber slide method and empirical prediction functions for worm burden (Moss *et al.*, 1990; Seivwright *et al.*, 2004). For shot birds, caecum samples were stored cold immediately after sampling and processed within one week to ensure reliable estimates (Seivwright *et al.*, 2004). Sex and age were determined morphologically using plumage and supra-orbital comb size. Birds that fledged in the year of sampling were classed as "young", and all other birds were classed as "old" (>1 year), allowing for statistical control of typically lower nematode burden in young birds (Shaw & Moss, 1989; Hudson & Dobson, 1997). Body mass was measured to the nearest 10 g using a spring balance and wing length was measured to the nearest mm. As a measure of physiological condition, the scaled mass condition index (Peig & Green, 2009) was then calculated for each individual (i) as:

$$CI_i = m_i \left(\frac{w_0}{w_i} \right)^b$$

where m_i is body mass, w_i is wing length, w_0 is mean wing length and b is the slope of standardised major axis regression of $\ln m$ on $\ln w$ across the data set.

Genetic data

DNA was extracted from feather calamus or liver tissue samples using a standard salting out method (Hogan *et al.*, 2008). Individuals were genotyped at 384 genome-wide SNPs using the Illumina GoldenGate BeadXPress platform (NBAF-S, University of Sheffield). Based on orthology with the chicken genome (*galGal4* assembly) following BLAT searches (Kent, 2002), these SNPs were designed such that the number of SNPs per chromosome was approximately proportional to chromosome size. As such, chromosomes 1–15, 17–26, 28 and Z were covered by 1–82 SNPs (median 6) with a median distance between consecutive SNPs of 1.9 Mbp (supplementary materials S1).

Quality filtering and calculation of summary statistics were carried out in PLINK 1.9 (Chang *et al.*,

2014). SNPs with calling rate below 0.90 and minor allele frequency (MAF) below 0.05 were removed. The inbreeding coefficient (F_{IS}) for each SNP and individual was computed from observed and expected heterozygosities, and Hardy-Weinberg equilibrium exact tests were carried out to identify SNPs with genotyping errors. Quality filtering resulted in a remaining total of 271 SNPs. All 695 individuals had a genotyping rate larger than 80 %, with an overall genotype coverage of 99.6 %. Linkage disequilibrium (r^2) was estimated among SNP pairs within each chromosome. The degree of relatedness among individuals was ascertained from the realized relationship matrix based on genetic identity-by-state (Yang *et al.*, 2011a). Genetic structure was ascertained by computing global F_{ST} among sampling sites and graphically examining the first 20 principal components of the genetic relationship matrix (Johnston *et al.*, 2014). The minimum number of principal components that capture most of the genetic structure was then identified by examining the eigenvalues and selecting a principal component cut-off such that the differences in eigenvalues across subsequent principal components were minimal and relatively stable. To aid the identification of this cut-off, an *ad hoc* statistic ΔE was calculated that relates the changes in eigenvalues from each principal component to the following and the preceding component ($\Delta E_{pc} = \frac{|E_{pc} - E_{pc+1}|}{|E_{pc} - E_{pc-1}|}$), and peaks in ΔE were examined.

Genome-wide association analysis

GWA analysis was carried out by implementing implement custom generalized linear models with negative binomial error structure in the statistical software R 3.0.3 (R Core Team, 2014) and the package *MASS* (Venables & Ripley, 2002). The discrete worm counts were not normally distributed and therefore inappropriate for analysis with common GWA analysis packages. Exploratory modelling including all sampling-related and phenotypic variables and applying a Poisson error structure indicated substantial over-dispersion ($\phi > 1000$), and diagnostic residuals plots confirmed that a negative-binomial error structure provided the best fit for these data.

Relationships among nematode burden and sampling-related or phenotypic variables were ascertained by graphical exploration and linear modelling to identify confounding variables. Nematode burden varied considerably among sampling sites, years and months, but sample sizes were very low for several years and months (supplementary materials S2). To avoid statistical issues with small sample sizes, these two variables were simplified by combining factor levels. Years were binned into three time periods with substantially different nematode burdens (1995–1999; 2000–2006; 2010–2012; supplementary materials S2). Similarly, sampling activity was predominantly limited to spring (April) and autumn (August, September, October), reflecting breeding and territory establishment seasons respectively (Mougeot *et al.*, 2005a; Redpath *et al.*, 2006b). In consequence, months were binned into two seasons ("spring": January to June; "autumn": July to December) that captured seasonal differences in nematode burden (supplementary materials S2). Nematode burden was higher in males ($\beta_1 = 0.29 \pm 0.16$; $z_{693} = 1.86$, $P = 0.06$) and old

birds ($\beta_1 = 0.61 \pm 0.15$; $z_{693} = 4.22$, $P < 0.001$), confirming expectations from previous studies (Mougeot *et al.*, 2004, 2005d, 2009; Martínez-Padilla *et al.*, 2010; Vergara *et al.*, 2012) and justifying inclusion of sex and age as explanatory variables. Similarly, significant variation in nematode burden was explained by body mass ($\beta_1 = 0.003 \pm 0.001$; $z_{691} = 2.553$, $P = 0.011$), wing length ($\beta_1 = 0.035 \pm 0.008$; $z_{648} = 4.679$, $P < 0.001$) and condition index ($\beta_1 = -0.003 \pm 0.001$; $z_{647} = -3.149$, $p = 0.001$) individually, though all of these variables were stratified by sex and age (supplementary materials S2).

Five GWA models were constructed with different biological and statistical complexity to balance out biological consistency and statistical power. The base-line model contained *site*, *period*, *season*, *sex* and *age* variables (model 1; $n = 695$). Two additional models were designed to account for condition-specific covariance in nematode burden. First, by including *body mass* as an additional variable, omitting two observations with missing data (model 2; $n = 693$). Second, by substituting the *condition index* for body mass, omitting 44 further observations with missing data (model 3; $n = 649$). These two models introduce some collinearity among sex, age and body mass or condition index. Although variance inflation factors for these variables were low (< 2), an alternative model was constructed to resolve collinearity. Mixed correspondence analysis (MCA) was applied on the *sex*, *age* and *condition index* variables using the R package *ade4* (Dray & Dufour, 2007) and all three independent principal components were then fitted instead of the original variables (model 4; $n = 649$). This model was also implemented in an alternative version with full sample size (model 5; $n = 695$), where missing condition indices were imputed before MCA, using the multiple imputation method implemented in the R package *missMDA* (Josse & Husson, 2012).

Associations between individual SNPs and nematode burden were then examined by extending these five models to include an additional categorical explanatory variable that represented the three possible genotypes. To check for allele-specific effects, all models were also implemented for each individual allele independently, by including a single binary explanatory variable representing allele presence. To account for genetic relatedness and population structure, the first eight principal components of the genetic relationship matrix were included as explanatory variables. Although multiple elbow points were apparent when plotting eigenvalues, the greatest change in eigenvalues followed by comparatively steady decrease was after the eighth component (Figure 1). Including more components into models provided qualitatively similar results but incurred a disproportionate power penalty through overfitting.

Coefficient estimates and P-values for the genetic term were extracted from each model, and P-values were corrected for multiple testing using the Benjamini-Hochberg false-discovery-rate (FDR) method (Benjamini & Yekutieli, 2005) genome-wise within each model. Results were visualised using Manhattan plots and associations were deemed significant at FDR-corrected $q \leq 0.1$ and highly significant at $q \leq 0.05$. Predicted nematode burdens for each SNP genotype were calculated for significant SNPs as least-square population means from GWA model 2 when keeping all other variables constant at mean values, using the

R package *doBy* (Højsgaard & Halekoh, 2013). Effect size was then calculated as the absolute difference between the largest and the smallest estimate, and standard errors were propagated. Finally, to estimate the proportion of phenotypic variance explained by each individual SNP, the SNP genotypes were fitted as the sole predictor of $\log_{10}$-transformed nematode burden in a simple linear model. The OLS coefficient of determination (r^2) then gives a liberal estimate of the heritability accounted for by a single SNP.

Identified SNPs were mapped to the chicken genome to ascertain genomic context and putative physiological functions of genes linked to the identified polymorphism. The SNP probe sequences were mapped to the *galGal4* assembly using BLAT (Kent, 2002) and genomic locations were visualised using the UCSC genome browser (Kent *et al.*, 2002). Proximal characterised ENSEMBL chicken genes were then explored in ENSEMBL BIOMARTS (Kinsella *et al.*, 2011), identifying orthologues and extracting GENEONTOLOGY (The Gene Ontology Consortium, 2000) annotations.

Genome partitioning

Genome-wide and chromosome-specific narrow-sense heritabilities (h^2) of nematode burden were estimated using genome partitioning analysis as implemented in the software GCTA (Yang *et al.*, 2011a). This method is based on partitioning phenotypic variance across multiple additive genetic variance components to test whether these components explain significant amounts of phenotypic variance. Each genetic component represents an arbitrary group of genetic markers, usually chromosomes or smaller linkage groups. Additive genetic variance components are estimated from identity-by-state genetic relationship matrices (GRMs) rather than from pedigree data (Yang *et al.*, 2011b; Hill, 2012). In order to satisfy the requirement for a Gaussian quantitative phenotype in GCTA analysis, nematode burden was $\log_{10}$-transformed (Okada *et al.*, 2010; Brown *et al.*, 2013).

Genome-wide heritability was estimated by generating a single GRM in PLINK using all SNPs and fitting the GRM as a single genetic variance component in linear mixed models in GCTA. Statistical significance of the genetic variance was obtained from a log-likelihood ratio test in comparison to a null model without a genetic variance component. Two types of sensitivity analysis were undertaken. First, to ascertain sensitivity to the number of explanatory variables included, three nested analyses were carried out and compared: 1) including sex and age explanatory variables and eight PCA components to account for genetic stratification; 2) adding sampling site and period explanatory variables; and 3) adding sampling season and body mass explanatory variables. This final model contained the same explanatory variables as model 2 in the GWA analysis. Second, the sensitivity to SNP density was examined by generating 500 replicate GRMs from random genome-wide draws of 5–95 % (in 5 % increments) of all SNPs and estimating heritability from each GRM independently. This was carried out for the base-line model without control variables and for the three models with nested control variables.

Chromosome-specific heritability was estimated by generating one GRM for each chromosome using

only SNPs on that chromosome and fitting this GRM alongside a second GRM based on all genome-wide SNPs apart from those on that chromosome. A likelihood-ratio test was then carried out between this full model and a model excluding the chromosome-specific component. Given a null-model of polygenic architecture, chromosome-specific heritability should scale linearly with chromosomal gene content (Davis *et al.*, 2013; Robinson *et al.*, 2013). To test this hypothesis, heritability estimates were regressed on numbers of RefSeq genes in chicken chromosomes (*galGal4* assembly), retrieved from the UCSC Table Browser (Karolchik *et al.*, 2004). Given the imbalanced SNP design across chromosomes, heritability was also regressed on numbers of genotyped SNPs and an interaction between both predictors in independent models. Finally, to test whether the heritability of some chromosomes is larger than expected given chromosomal gene content, another GCTA analysis was carried out where the chromosome-specific component was fitted alongside a GRM using all genome-wide SNPs (Robinson *et al.*, 2013). All analyses were carried out for the four nested models as detailed above for genome-wide heritability estimation.

Results

Summary statistics

Observed and expected heterozygosities at each of 271 SNPs ranged from 0.086–0.534 (median 0.348) and 0.107–0.500 (median 0.349) respectively. A total of 20 SNPs on 10 chromosomes displayed significant departures from Hardy-Weinberg equilibrium ($P \leq 0.05$). However, only three SNPs on chromosome 5 remained significant after FDR-correction for multiple testing ($q \leq 0.05$) and none of these displayed extreme genotype distributions or P-values that would suggest genotyping errors. All SNPs were in effective linkage equilibrium, with r^2 ranging from 0.000 to 0.088 (median: 0.001; 99th percentile: 0.014) (supplementary materials S3).

Inbreeding coefficients (F_{IS}) of individual samples were normally distributed, ranging from –0.251 to 0.247 with a mean of 0.011 ± 0.085 SD (supplementary materials S3). Relatedness among individuals was approximately centred on zero (median: –0.003; 1st percentile: –0.144; 99th percentile: 0.156) (supplementary materials S3). This distribution is consistent with a simulated distribution of genetic relatedness estimates for unrelated red grouse from neutral polymorphisms (Piertney *et al.*, 1999), and as such suggests that the genotyped individuals are effectively unrelated. Population genetic differentiation among the five sampling sites was very low ($F_{ST} = 0.005$), though the English site was more strongly differentiated from the Scottish sites ($F_{ST} = 0.020$), as would be expected under isolation-by-distance. Accordingly, little genetic clustering was apparent in PCA eigenvector plots (Figure 1).

Genome-wide association analysis

GWA analysis yielded support for associations between nematode burden and genotypes or alleles of five SNPs on five chromosomes (Figure 2). Congruence among models incorporating sex and age (model 1) or sex, age and body mass (model 2) variables was very high, highlighting genotypes of all five SNPs and alleles of three SNPs (Table 2). However, models incorporating sex, age and condition index or equivalent principal components of mixed correspondence analysis (models 3 and 4) highlighted weaker associations ($q \leq 0.1$) for alleles or genotypes in only three of these five SNPs (Table 2). These models probably suffered from decreased power due to reduced sample sizes, given that the equivalent model 5 with imputed principal components (and hence full sample size) yielded almost identical results to models 1 and 2 (Table 2). Notwithstanding, each of these five SNPs is highlighted by at least three models, indicating that these are robust results despite the potential for reduced power in some models.

Of these, SNPs X1407, X5104 and X1375 were best supported, displaying consistently strong associations ($q \leq 0.05$) of genotypes or alleles in the three full models, and weaker associations ($q \leq 0.1$) of genotypes or alleles in the two reduced models (Table 2). Genotypes in two further SNPs X2277 and X2298 were moderately supported across models, but there was no support for alleles in any model. Some models additionally highlighted weak associations ($q \leq 0.1$) for two further SNPs on chromosome 1 and 2 (not shown), but these were inconsistent and as such were disregarded. Predicted nematode burdens among genotypes of the five highlighted SNPs indicated additive allelic effects in most cases, and absolute effect sizes ranged from 313–556 worms per bird (Figure 3). None of these SNPs deviated from Hardy-Weinberg equilibrium and hence there was no evidence for heterozygosity advantage or disadvantage. The proportion of phenotypic variation explained by these SNPs individually in simple linear models ranged from 0.1–2.8 % (sum: 4.9 %), which should be taken as liberal best-case estimates (Table 2). Full association statistics for all models and SNPs alongside tests for Hardy-Weinberg equilibrium are available in supplementary materials S4–S5.

Homology with the chicken genome indicated that SNPs X1407 and X1375 were located in introns of genes *MAPKBP1* (mitogen-activated protein kinase binding protein 1) and *KLHL34* (kelch-like family member 34) on chromosomes 5 and 1 respectively (Table 2). All other SNPs were located in non-coding regions on chromosomes 10, 13 and 20, approximately 6–540 kbp (equivalent to 0.02–1.51 cM assuming 2.8 cM/Mbp, Hillier *et al.*, 2004) remote from the following upstream and downstream genes: *RAPGEF6* (Rap guanine nucleotide exchange factor 6), *FNIP1* (folliculin interacting protein 1), *CEBPB* (CCAAT/enhancer binding protein β), *PTPN1* (protein tyrosine phosphatase, non-receptor type 1), *NR2F2* (nuclear receptor subfamily 2, group F, member 2) and *MCTP2* (multiple C2 domains, transmembrane 2). The putative functions of these linked genes involve innate and adaptive immune system processes, protein degradation and energy metabolism (Table 2; supplementary materials S6).

Genome partitioning

Estimated genome-wide heritability of nematode burden based on all SNPs together was $h^2 = 0.294$. This estimate decreased rapidly when increasing numbers of explanatory variables were added to the model (Table 3). Sensitivity analysis by randomly sampling subsets of genome-wide SNPs and repeating heritability estimation identified an asymptotic relationship between heritability and SNP density, though this was much weaker when additional explanatory variables were included (supplementary materials S7). Although these heritability estimates are likely to increase at higher SNP densities, these SNPs capture a statistically significant amount of heritability that is consistently larger than the heritability estimates of the five SNPs identified using GWA.

Partitioning genome-wide heritability across chromosomes was limited by variation in SNP density and power reduction when multiple explanatory variables were included in the models. Chromosomes 21, 22, 24, 26 and 28 could not be analysed because of insufficient SNP density. For other chromosomes, some models failed to converge for particular combinations of variables. Estimates of chromosome-specific heritability across all models ranged from 10^{-6} to 0.099, of which 1–11 estimates were significantly different from zero (Figure 4). The summed chromosome-specific heritabilities within models were very similar to genome-wide heritabilities, despite the reduced number of chromosomes analysed (Table 3), suggesting some covariance across chromosomes that may inflate individual estimates. Heritability estimates were significantly positively associated with chromosomal gene content and number of genotyped SNPs in simple models only, presumably due to reduced power in more complex models (Table 3). Although these two predictors were highly correlated as a consequence of the SNP design (Spearman's $\rho = 0.93$; $P < 0.001$), there was a significant interaction among them (Table 3), indicating an effect beyond the mere number of genotyped SNPs per chromosome. All these relationships were primarily driven by chromosomes 1 and 2, which had particularly large heritability, size and SNP density (Figure 4). When these two chromosomes were removed, the relationships remained significant for the model without control variables only (not shown). Chromosomes 5 and 17 displayed significantly ($p < 0.05$) greater heritability than expected from their gene contents in some models (excess $h^2 = 0.029 \pm 0.020$ and $h^2 = 0.017 \pm 0.017$ respectively; Figure 4). Some weaker evidence for excess heritability was also apparent for chromosomes 1 ($h^2 = 0.043 \pm 0.034$), 13 and 14 ($h^2 = 0.010 \pm 0.012$ in both cases). Total excess heritability across all chromosomes was $h^2 = 0.066 - 0.089$.

These results were partially consistent with the GWA results. Chromosomes 1 and 5 contained the best supported SNPs X1407 and X1375, and chromsome 1 also contained other suggestive SNPs (Figure 2). SNPs on chromosomes 10, 13 and 20 were highlighted in the GWA analysis, but these chromosomes showed no significant excess heritability. In contrast, chromosome 17 displayed evidence for excess heritability, but no SNP on this chromosome was highlighted in the GWA analysis.

Discussion

This study examined the broad genome-wide basis of variation in gastrointestinal nematode burden among red grouse individuals. Genome-wide association identified five SNPs that were closely linked to novel genes putatively involved in multiple physiological processes beyond the immune system, consistent with the broad functions of previously identified genes using candidate gene approaches (Wenzel *et al.*, 2015; Wenzel & Piertney, 2015) and epigenetic DNA methylation analysis (Wenzel & Piertney, 2014). Genome partitioning indicated moderate genome-wide heritability of nematode burden, and highlighted that some chromosomes contain additive genotypic variants of individually small effects that represent disproportionate amounts of heritability. These findings contribute to elucidating the broad genomic basis of parasite susceptibility (Wilfert & Schmid-Hempel, 2008) and suggest that nematode burden in grouse may be somewhere between a purely polygenic phenotype corresponding to Fisher's infinitesimal model (Fisher, 1919) and a purely oligogenic phenotype where few loci explain a large proportion of phenotypic variance (Rockman, 2012).

T. tenuis infection is localised to the caecum where adult worms cause haemorrhaging and tissue necrosis with substantial impact on grouse condition and fitness (Watson *et al.*, 1987; Hudson *et al.*, 1992; Delahay *et al.*, 1995; Delahay & Moss, 1996). The majority of genes closely linked to the five SNPs identified through GWA are putatively involved in immune system processes that may be associated with chronic intestinal *T. tenuis* infection: MAPKBP1 is involved in signal transduction during inflammation and in intestinal homeostasis (Lecat *et al.*, 2012). In humans, MAPKBP1 (alias JNKBP1) is an antagonist to the NOD2 receptor which recognises bacterial cell wall components and helps orchestrate an innate and adaptive immune response (Lecat *et al.*, 2012). Failure to regulate expression of NOD2 in intestinal tissue by MAPKBP1 disrupts cytokine signal transduction and is implicated in Crohn's disease, a degenerative disorder of the intestine (Hugot *et al.*, 2001; Lecat *et al.*, 2012). Similarly, CEBPB is involved in signal transduction following an inflammatory immune response (Ramji & Foka, 2002), and may specifically regulate antibacterial activity of macrophages and repair of necrotic tissue (Ruffell *et al.*, 2009). Other genes represent broader immunoregulatory factors: RAPGEF6 is a guanine nucleotide exchange factor for RAP GTPases (Kuiperij *et al.*, 2003), which are involved in regulating signalling interactions between antigen-presenting cells and T-cells and in regulating leukocyte integrin activation (Katagiri *et al.*, 2002; Scheele *et al.*, 2007). FNIP1 is an adapter protein that interacts with a range of factors that regulate cellular energy metabolism (Park *et al.*, 2012). Disruption of FNIP1 function arrests development of B lymphocytes and invariant natural killer T-cells due to an inability to regulate metabolic homeostasis during cell proliferation, particularly under metabolic stress (Park *et al.*, 2012, 2014). MCTP2 is involved in intercellular signal transduction and has been implicated in various neurological disorders such as schizophrenia and autism (Shin *et al.*, 2005; Djurovic *et al.*, 2009). Its link to parasite infection is unclear, but transcriptomic studies suggest that it may be cryptically involved in immune system function (Verner

et al., 2012), possibly through regulation of T-helper cell differentiation and function (Äijö *et al.*, 2012).

The remaining identified genes represent a range of other physiological pathways that may play a key role in underpinning host-parasite interactions in red grouse. KLHL34 is part of the evolutionarily highly conserved kelch-like protein family whose functions are not well understood, though they are associated with a range of human diseases (Dhanoa *et al.*, 2013). The best studied function is a role in protein ubiquitination (Gupta & Beggs, 2014), which links kelch-like proteins to a broad range of cellular processes including inflammatory immune response and protein degradation, potentially as part of a xenobiotics detoxification response (Ben-Neriah, 2002; von Mikecz, 2006). In infected red grouse, such xenobiotics may originate from primary *T. tenuis* infection or from secondary pathogen infection following caecal haemorrhaging (Watson *et al.*, 1987). Further, both NR2F2 (alias COUP-TFII) and PTPN1 are involved in regulating glucose homeostasis, energy expenditure and adipogenesis (Li *et al.*, 2009; Tsou *et al.*, 2012), which could plausibly impact physiological condition with broad consequences for the ability of red grouse to cope with parasite infection (Sheldon & Verhulst, 1996; Svensson *et al.*, 1998).

Although GWA provides no direct evidence for a mechanistic involvement of these genes, their putative physiological functions are broadly consistent with novel candidate genes previously identified through comparative transcriptomics between infected and nematode-free grouse (Wenzel *et al.*, 2015; Wenzel & Piertney, 2015). These candidate genes represent innate and adaptive immune system processes, in particular eosinophil-mediated parasite expulsion and antimicrobial peptides, alongside key enzymes in the xenobiotics detoxification and oxidative stress pathways, and broad cell proliferation regulators (Wenzel & Piertney, 2015). Together, these insights reinforce the notion that host-parasite interactions in red grouse are not a primarily immunological matter. This provides an intriguing perspective on the classic immunological paradigm of host life-history trade-offs (Sheldon & Verhulst, 1996; Lochmiller & Deerenberg, 2000; Zuk & Stoehr, 2002; Schmid-Hempel, 2003; Owen *et al.*, 2010). In spite of well-established immunological links in red grouse (Mougeot & Redpath, 2004; Webster *et al.*, 2011a; Wenzel *et al.*, 2013; Wenzel & Piertney, 2015) and sheep or cattle species (e.g., Beh *et al.*, 2002; Davies *et al.*, 2006; Beraldi *et al.*, 2007; Silva *et al.*, 2012; Riggio *et al.*, 2013), the cost of chronic parasite infection, particularly in the case of nematodes, may well extend to other physiological and perhaps even behavioural categories (Thomas *et al.*, 2010; Biron & Loxdale, 2013; Poulin, 2013). In red grouse, for example, DNA methylation states at specific loci linked to genes involved in immune system, metabolism, cell cycle regulation and epigenetic mechanisms have been associated with parasite burden in the field (Wenzel & Piertney, 2014), suggesting that chronic parasite infection may have a broad epigenetic component (Poulin & Thomas, 2008). This perspective is still consistent with the classic idea of life-history trade-offs (Hamilton & Zuk, 1982; Sheldon & Verhulst, 1996; Lochmiller & Deerenberg, 2000), but suggests that physiological condition is an amalgam of a multitude of cellular functions that may either directly

contribute to parasite defence or affect its efficacy indirectly by broad effects on energy metabolism or molecular signalling cascades (Hill, 2011).

The genome partitioning analysis broadly supported this perspective. In spite of low heritability of the five individual SNPs, genome-wide heritability using all SNPs together was moderate ($h^2 = 0.07 - 0.29$) and within the range of reported heritability estimates for nematode burden in sheep (Dominik, 2005; Stear *et al.*, 2007) and chicken (Kaufmann *et al.*, 2011; Wongrak *et al.*, 2014). This could imply that parasite susceptibility may to some extent be a polygenic phenotype involving large numbers of genes with individually small, but large joint effects (Yang *et al.*, 2011b; Hill, 2012; Rockman, 2012). However, chromosomes 1, 5 and 17 contributed disproportionate amounts to total heritability despite different SNP densities, suggesting that nematode burden may not be primarily based on a large number of genome-wide polymorphisms with small effects, rather than on moderate numbers of small effect genes in these three chromosomes. In fact, chromosomes 1 and 5 contain two SNPs highlighted by the present GWA analysis and seven previously identified novel candidate genes for response to parasite infection (Wenzel *et al.*, 2015), most of which have been confirmed to explain variation in parasite load in the field (Wenzel & Piertney, 2015). The two identified SNPs are not in the vicinity of any of these genes (40–80 Mbp and 7–8 Mbp distance on chromosomes 1 and 5 respectively), highlighting their novelty. This chromosomal congruence across candidate genes, GWA and genome partitioning provides support for a primarily oligogenic basis of parasite susceptibility in chromosomes 1 and 5, making them priority targets for fine-mapping novel genomic candidate regions in addition to those already discovered. Chromosome 17 may also be a good target on the quest for the "missing heritability", though it has not been implicated in GWA or previous candidate genes studies to contain large-effect genes. The genome partitioning results are consistent with a study in sheep, where five chromosomes explained disproportionate amounts of variance in gastrointestinal nematode burden, despite an overall relationship between chromosomal heritability and gene content (Al-Kalaldeh *et al.*, 2013).

Despite some congruence in numbers and function of identified genes across red grouse studies with different approaches, there are some inconsistencies that deserve consideration. First, some SNPs highlighted by the GWA analysis are on chromosomes that do not display excess heritability or contain any previously identified candidate genes. Similarly, many previously identified candidate genes are on chromosomes that are not highlighted by GWA or genome partitioning (Wenzel *et al.*, 2015; Wenzel & Piertney, 2015). Each of these three techniques highlights a different set of genomic regions overall, reflecting different ways of statistical inference with different power and potential for false positives (Ellegren & Sheldon, 2008; Amos *et al.*, 2011). In the present study, one caveat to consider with respect to genome partitioning in particular is that SNP density was low and the study design was based on comparatively few samples. Although few chromosomes displayed significant total heritability, we were able to detect three chromosomes with more heritability than expected, suggesting that power was sufficient to detect

large heritability effects but may have been insufficient to detect more subtle effects. A large-scale study with thousands of samples and markers will be required to fully reconcile these differences and resolve the underlying genomic architecture (Lee *et al.*, 2012).

Second, in spite of the putative functional importance of the linked genes identified from orthology with chicken, the actual molecular effects of the identified SNPs are unclear. None of the identified SNPs were in exonic regions, suggesting that the association of alleles with parasite load may not be due to amino acid substitution in the gene product, though these SNPs may be linked to unknown exonic SNPs with such effect. Instead, intronic SNPs may affect transcript splicing or downstream mRNA processing with effect on translation dynamics (Pagani & Baralle, 2004; Hunt *et al.*, 2014). Similarly, non-coding SNPs may affect gene expression or chromatin organisation through epigenetic mechanisms such as miRNAs or transcription factors (Pagani & Baralle, 2004; Chamary *et al.*, 2006). Epigenetic processes may play a particularly important role in the red grouse system because DNA methylation at a range of non-coding genomic regions is associated with parasite load (Wenzel & Piertney, 2014), SNPs in previously identified candidate genes associated with parasite load were predominantly non-synonymous, intronic or untranslated (Wenzel & Piertney, 2015) and some of these genes are themselves involved in regulating gene expression (Wenzel & Piertney, 2015).

In summary, genome-wide association and genome partitioning have added new insight to our previous work on red grouse that highlighted a range of genetic polymorphisms in or close to candidate genes that are linked to nematode burden among grouse individuals, accounting for differences of as much as 666 worms per bird (Wenzel & Piertney, 2015). By combining a range of strategies for examining the genetic basis of a complex phenotype, we have reinforced the idea that parasite susceptibility may involve a considerable number of genes involved in a range of physiological categories beyond the immune system. We further identified particular chromosomes as priority targets for fine-mapping polymorphisms of small effect in the quest for the missing heritability of nematode burden. These findings have painted a consistent picture of the genetic basis of nematode burden in red grouse and illustrate that genome partitioning is a powerful addition to the classic strategies for studying the genetic basis of complex phenotypes.

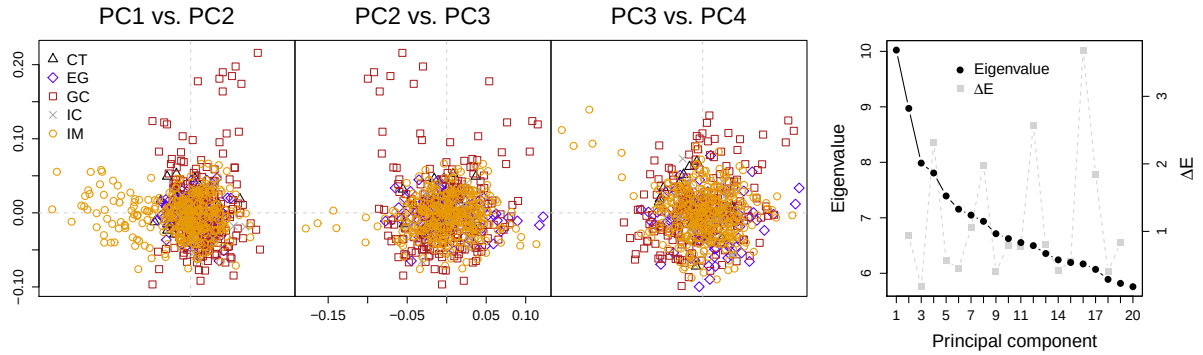

Figure 1: Principal components analysis of the genome-wide genetic relationship matrix based on 271 SNPs. The first three panels are scatterplots of sequential combinations of the first four principal components (eigenvectors), highlighting sampling sites (CT = Catterick; EG = Edinglassie; GC = Glas Choille; IC = Invercauld; IM = Invermark). The last panel plots the eigenvalues of the first 20 principal components and an *ad hoc* statistic ΔE (quotient of the eigenvalue change to the following versus to the preceding principal component) whose peaks aid identification of elbow points.

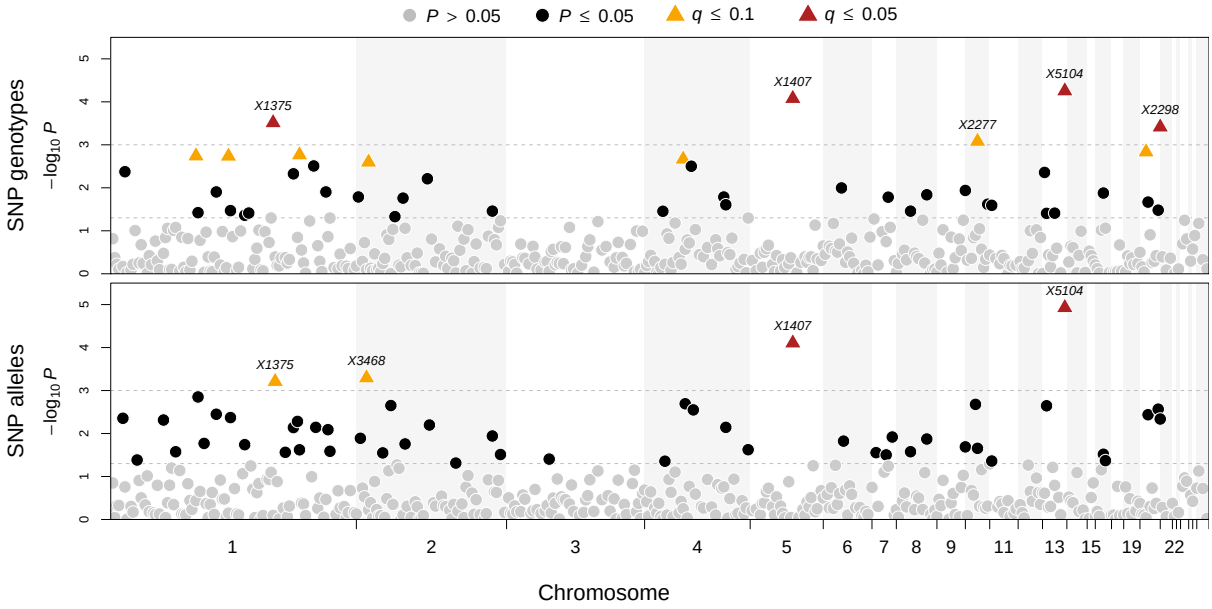

Figure 2: Genome-wide association results (model 2 with full set of control variables, Materials & Methods) for nematode burden using genotypes (top) or alleles (bottom) of 271 SNPs ordered along chromosomes based on orthology with the chicken genome. Each symbol represents statistical significance (P-value) of a single test (genotype comparison or allele presence). Statistical significance before and after correction for multiple testing (false discovery rate correction; q-values) is indicated by symbol shape and colours. Two common single-test significance thresholds ($P = 0.05$ and $P = 0.001$) are indicated by dashed lines for illustrative purposes.

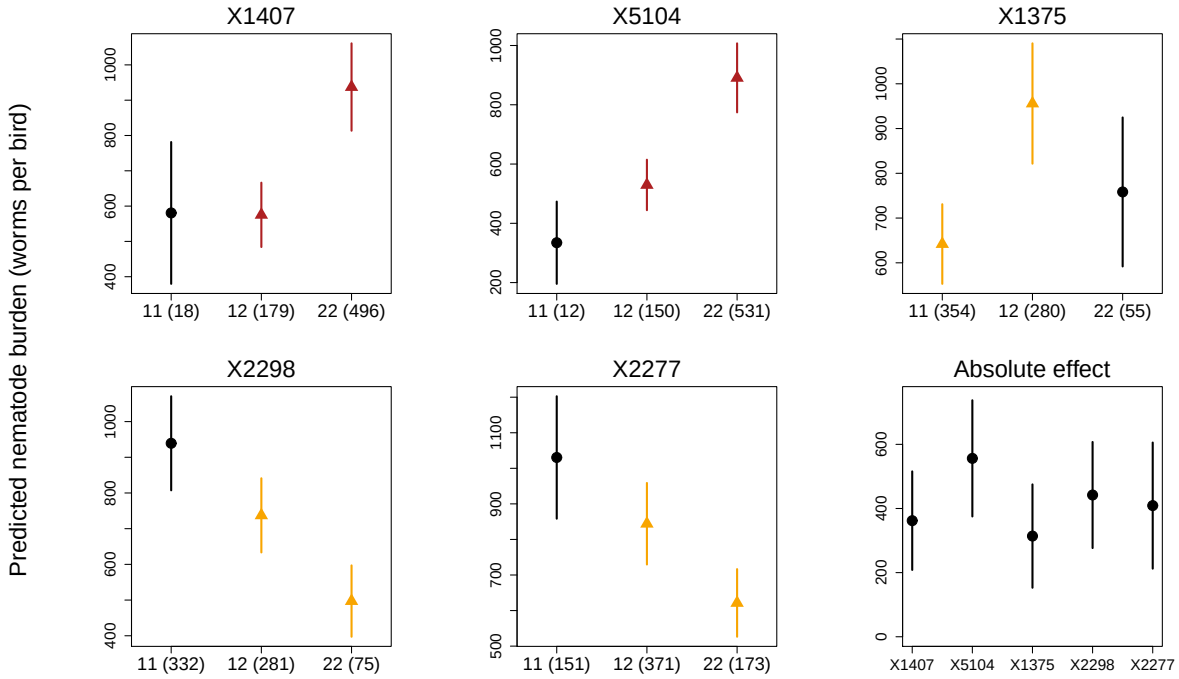

Figure 3: Predicted nematode burden according to three possible genotypes (11 and 22: homozygote for allele 1 or 2; 12: heterozygote) at five SNPs identified through genome-wide association. Least-square estimates (± SE) of model 2 (all other variables kept constant at mean values) are presented in the first five panels. Coloured triangles denote statistically significant ($q \leq 0.1$) genotype comparisons in the fitted model as in Figure 2. Numbers in brackets indicate sample sizes. The bottom right panel indicates overall effect size as the largest absolute difference in predicted estimates (± propagated SE).

Table 1: Sampling locations, sampling years and sample sizes by sex and age (young: hatched in same year; or old: >1 year) of 695 red grouse samples. Detailed sample information is presented in supplementary materials S2.

				Sex		Age		
Site	Latitude	Longitude	Years	F	M	Y	O	Total
Edinglassie (EG)	57.22 º	–3.19 º	2000–2012	13	86	31	68	99
Glas Choille (GC)	57.12 º	–3.32 º	1995–2003	63	136	131	68	199
Invercauld (IC)	57.08 º	–3.30 º	2012	15	16	10	21	31
Invermark (IM)	56.90 º	–2.88 º	2011–2012	109	238	76	271	347
Catterick (CT)	54.33 º	–1.87 º	2006	0	19	12	7	19
Total				200	495	260	435	695

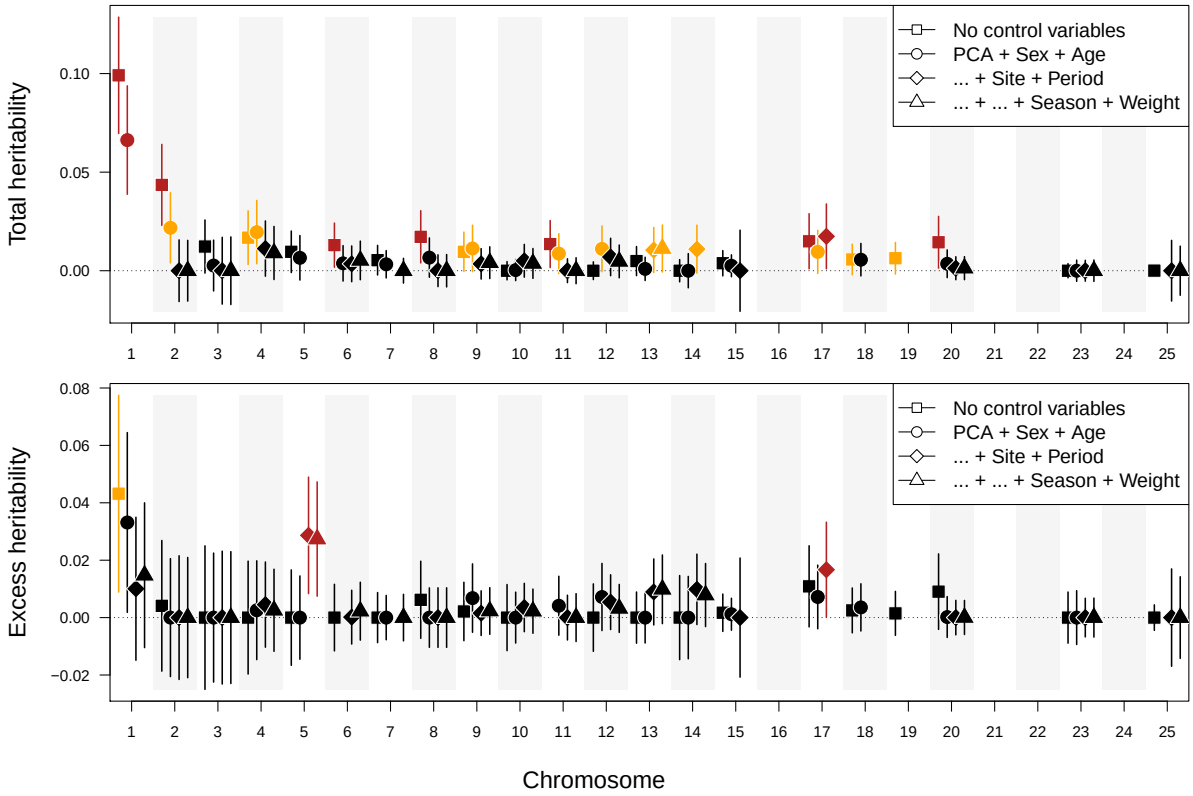

Figure 4: Chromosome-specific total heritability (top panel) and excess heritability of nematode burden unaccounted for by gene content (bottom panel). Heritability was partitioned across chromosomes using linear mixed models incorporating chromosome-specific genetic variance and three nested sets of control variables. Both panels display heritability estimates ($\pm$ SE) with colour-coded statistical significance (black: $p > 0.1$; orange: $p \leq 0.1$; red: $p \leq 0.05$).

Table 2: Characterisation of SNPs significantly associated with nematode burden using genome-wide association. Statistical significance is indicated by asterisks for genotypes and alleles across five negative binomial models. Each model incorporates sampling site, period and season alongside eight principal components of the genetic relationship matrix as control variables. Additional control variables are indicated (MC: principal component following multiple correspondence analysis on age, sex and condition index). A liberal heritability estimate is indicated as the coefficient of determination (r^2) of a simple OLS model with the SNP as the sole predictor. Minor allele frequency (MAF), observed (H_O) and expected (H_E) heterozygosities are presented with the P-value for Hardy-Weinberg equilibrium (P_{HWE}). Chicken chromosome number and functional information on proximal ENSEMBL genes are given below.

		SNPs (genotype/allele)				
Model	n	X1407	X5104	X1375	X2298	X2277
1: $+sex + age$	695	** / **	** / **	** / **	** / n.s.	* / n.s.
2: $+sex + age + mass$	693	**/ **	** / **	** / **	** / n.s.	* / n.s.
3: $+sex + age + condition$	649	* / *	n.s./ *	n.s. / n.s.	* / n.s.	n.s. / n.s.
4: $+MC1 + MC2 + MC3$	649	* / *	n.s./ *	n.s. / n.s.	* / n.s.	n.s. / n.s.
5: $+MC1 + MC2 + MC3$ (imputed)	695	**/ **	** / **	** / *	** / n.s.	* / n.s.
Heritability (OLS r^2)		0.028	0.006	0.010	0.004	0.001
MAF		0.155	0.126	0.283	0.313	0.484
H_O		0.258	0.217	0.406	0.408	0.534
H_E		0.262	0.220	0.406	0.430	0.500
P_{HWE}		0.665	0.728	1	0.184	0.081
Chromosome		5	13	1	20	10
Type of mutation		intronic	non-coding	intronic	non-coding	non-coding
Proximal genes (distance)		MAPKBP1	RAPGEF6 (6 kbp); FNIP1 (23 kbp)	KLHL34	CEBPB (30 kbp); PTPN1 (127 kbp)	NR2F2 (540 kbp); MCTP2 (215 kbp)
Ensembl ENSGALG gene ID		ENSGALG00000008692	ENSGALG00000006569; ENSGALG00000017462	ENSGALG00000016388	ENSGALG00000008014; ENSGALG00000008010	ENSGALG00000007000; ENSGALG00000006981
Key GENEONTOLOGY term		GO:0005515 protein binding	GO:0007264 small GTPase mediated signal transduction; GO:0002327 immature B cell differentiation	GO:0005515 protein binding	GO:0006954 inflammatory response; GO:0008286 insulin receptor signaling pathway	GO:0006629 lipid metabolic process; GO:0019722 calcium-mediated signaling
Physiological process		immune system	immune system	protein ubiquitination	immune system; energy metabolism	energy metabolism; immune system

**: $q \leq 0.05$; *: $q \leq 0.1$; n.s.: $q > 0.1$

Table 3: Heritability estimates of nematode burden based on additive genetic variance derived from identity-by-state genetic relationship matrices. Genome-wide heritability estimates (h^2) are presented with standard errors (SE), and statistical significance is indicated by the likelihood-ratio test statistic LRT (compared to a model without the genetic variance component) and the associated P-value. Alongside, the sum of chromosome-specific heritability estimates is given with results of three independent linear regressions on chromosome-specific numbers of genes, numbers of genotyped SNPs, or the interaction among both predictors. All results are presented for four nested models with increasing numbers of control variables.

	Genome-wide h^2				Chromosome-specific h^2			
Model	h^2	SE	LRT	P	$\sum h^2$	Number of genes	Number of SNPs	Genes-SNPs interaction
No control variables	0.294	0.041	108.643	$\ll 0.001$	0.290	$F_{1,18} = 55.39$; $P \ll 0.001$	$F_{1,18} = 60.62$; $P \ll 0.001$	$F_{3,16} = 51.21$; $P < 0.001$
$+PCA + sex + age$	0.182	0.044	25.347	$\ll 0.001$	0.184	$F_{1,17} = 37.58$; $P \ll 0.001$	$F_{1,17} = 38.84$; $P \ll 0.001$	$F_{3,15} = 35.93$; $P < 0.001$
$+PCA + sex + age + site + period$	0.075	0.040	4.558	0.016	0.070	$F_{1,14} = 0.16$; $P = 0.697$	$F_{1,14} = 0.23$; $P = 0.639$	$F_{3,12} = 0.45$; $P = 0.722$
$+PCA + sex + age + site + period + season + mass$	0.070	0.040	3.970	0.023	0.039	$F_{1,12} = 0.01$; $P = 0.999$	$F_{1,12} = 0.01$; $P = 0.951$	$F_{3,10} = 0.901$; $P = 0.474$

Acknowledgements

This study was funded by a BBSRC studentship (MA Wenzel) and NERC grants NE/H00775X/1 and NE/D000602/1 (SB Piertney). The authors are grateful to Fiona Leckie, Andrew MacColl, Jesús Martínez-Padilla, François Mougeot, Steve Redpath, Pablo Vergara† and Lucy M.I. Webster for samples; Keliya Bai, Daisy Brickhill, Edward Graham, Alyson Little, Daniel Mifsud, Lizzie Molyneux and Mario Röder for fieldwork assistance; Gillian Murray-Dickson and Laura Watt for laboratory assistance; Heather Ritchie for helpful comments on manuscript drafts; and all estate owners, factors and keepers for access to field sites, most particularly Stuart Young and Derek Calder (Edinglassie), Simon Blackett, Jim Davidson and Liam Donald (Invercauld and Glas Choille), Richard Cooke and Fred Taylor† (Invermark) and T. Helps (Catterick).

Supplementary materials

- S1: Characterisation of 384 SNPs based on orthology with the chicken genome
- S2: Distribution of sample sizes and parasite load across sampling-related and phenotypic categories
- S3: SNP linkage disequilibrium and distributions of sample inbreeding coefficients (F_{IS}) and relatedness.
- S4: GWA model statistics and tests for SNP Hardy-Weinberg equilibrium
- S5: Manhattan plots of all five models used for GWA analysis
- S6: Full GENEONTOLOGY terms in *biological process* ontology of genes linked to SNPs identified through GWA
- S7: Sensitivity analysis of genome-wide heritability estimation to SNP density

Data accessibility

- Genotype data (DataDryad: doi 10.5061/dryad.02pr5)
- Metadata (sampling/phenotype information) (DataDryad: doi 10.5061/dryad.02pr5)

Author contributions

SBP conceived and designed the study. MAW and MCJ performed field and lab work. MCJ and AD developed SNP markers. MAW analysed the data. MAW and SBP wrote the manuscript.

References

Äijö T, Edelman SM, Lönnberg T *et al.* (2012) An integrative computational systems biology approach identifies differentially regulated dynamic transcriptome signatures which drive the initiation of human T helper cell differentiation. *BMC Genomics*, **13**, 572.

Al-Kalaldeh M, Gibson J, Van der Werf J, Gondro C, Villalobos N *et al.* (2013) Partitioning the genetic variance into genomic and pedigree components for parasite resistance in sheep. In *Proceedings of the Twentieth Conference of the Association for the Advancement of Animal Breeding and Genetics, Translating Science into Action, Napier, New Zealand, 20th-23rd October 2013.*, 412–415, Association for the Advancement of Animal Breeding and Genetics.

Amos W, Driscoll E, Hoffman J (2011) Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1183–1188.

Beh K, Hulme D, Callaghan M *et al.* (2002) A genome scan for quantitative trait loci affecting resistance to *Trichostrongylus colubriformis* in sheep. *Animal Genetics*, **33**, 97–106.

Ben-Neriah Y (2002) Regulatory functions of ubiquitination in the immune system. *Nature Immunology*, **3**, 20–26.

Benjamini Y, Yekutieli D (2005) Quantitative trait loci analysis using the false discovery rate. *Genetics*, **171**, 783–790.

Beraldi D, McRae AF, Gratten J *et al.* (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.

Biron DG, Loxdale HD (2013) Host–parasite molecular cross-talk during the manipulative process of a host by its parasite. *The Journal of Experimental Biology*, **216**, 148–160.

Bortolotti GR, Mougeot F, Martínez-Padilla J, Webster LMI, Piertney SB (2009) Physiological stress mediates the honesty of social signals. *PLoS ONE*, **4**, e4983.

Brockhurst MA (2011) Sex, death, and the red queen. *Science*, **333**, 166–167.

Brown E, Pilkington J, Nussey D *et al.* (2013) Detecting genes for variation in parasite burden and immunological traits in a wild population: testing the candidate gene approach. *Molecular Ecology*, **22**, 757–773.

Chamary J, Parmley JL, Hurst LD (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, **7**, 98–108.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2014) Second-generation PLINK: rising to the challenge of larger and richer datasets. *arXiv*, 1410.4803v1.

Coltman D, Wilson K, Pilkington J, Stear M, Pemberton J (2001) A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in a naturally-parasitized population of Soay sheep. *Parasitology*, **122**, 571–582.

Davies G, Stear M, Benothman M *et al.* (2006) Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. *Heredity*, **96**, 252–258.

Davis LK, Yu D, Keenan CL *et al.* (2013) Partitioning the heritability of tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genetics*, **9**, e1003864.

Delahay RJ, Moss R (1996) Food intake, weight changes and egg production in captive red grouse before and during laying: effects of the parasitic nematode *Trichostrongylus tenuis*. *The Condor*, **98**, 501–511.

Delahay RJ, Speakman JR, Moss R (1995) The energetic consequences of parasitism: effects of a developing infection of *Trichostrongylus tenuis* (Nematoda) on red grouse (*Lagopus lagopus scoticus*) energy balance, body weight and condition. *Parasitology*, **110**, 473–482.

Dhanoa BS, Cogliati T, Satish AG, Bruford EA, Friedman JS (2013) Update on the Kelch-like (KLHL) gene family. *Human Genomics*, **7**, 13.

Djurovic S, Le Hellard S, Kähler AK *et al.* (2009) Association of *MCTP2* gene variants with schizophrenia in three independent samples of Scandinavian origin (SCOPE). *Psychiatry Research*, **168**, 256–258.

Dominik S (2005) Quantitative trait loci for internal nematode resistance in sheep: a review. *Genetics Selection Evolution*, **37**, S83–S96.

Downing T, Lloyd AT, O'Farrelly C, Bradley DG (2010) The differential evolutionary dynamics of avian cytokine and TLR gene classes. *Journal of Immunology*, **184**, 6993–7000.

Dray S, Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–175.

Fisher RA (1919) The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.

Gavan MK, Oliver MK, Douglas A, Piertney SB (in press) Gene dynamics of Toll-like receptor 4 through a population bottleneck in an insular population of water voles (*Arvicola amphibius*). *Conservation Genetics*.

Gossner A, Wilkie H, Joshi A, Hopkins J (2013) Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. *Veterinary Research*, **44**, 68.

Gupta VA, Beggs AH (2014) Kelch proteins: emerging roles in skeletal muscle development and diseases. *Skeletal Muscle*, **4**, 11.

Hamilton W, Zuk M (1982) Heritable true fitness and bright birds: a role for parasites? *Science*, **218**, 384–387.

Hill GE (2011) Condition-dependent traits as signals of the functionality of vital cellular processes. *Ecology Letters*, **14**, 625–634.

Hill GE, Farmer KL (2005) Carotenoid-based plumage coloration predicts resistance to a novel parasite in the house finch. *Naturwissenschaften*, **92**, 30–34.

Hill WG (2012) Quantitative genetics in the genomics era. *Current Genomics*, **13**, 196.

Hillier LW, Miller W, Birney E *et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, **432**, 695–716.

Højsgaard S, Halekoh U (2013) *doBy - Groupwise summary statistics, LSmeans, general linear contrasts, various utilities*. R package version 4.5-10.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Hogan FE, Cooke R, Burridge CP, Norman JA (2008) Optimizing the use of shed feathers for genetic analysis. *Molecular Ecology Resources*, **8**, 561–567.

Hudson P, Dobson A (1989) Population biology of *Trichostrongylus tenuis*, a parasite of economic importance for red grouse management. *Parasitology Today*, **5**, 283–291.

Hudson PJ (1986) The effect of a parasitic nematode on the breeding production of red grouse. *Journal of Animal Ecology*, **55**, 85–92.

Hudson PJ, Dobson AP (1997) Transmission dynamics and host-parasite interactions of *Trichostrongylus tenuis* in red grouse (*Lagopus lagopus scoticus*). *Journal of Parasitology*, **83**, 194–202.

Hudson PJ, Dobson AP, Lafferty KD (2006) Is a healthy ecosystem one that is rich in parasites? *Trends in Ecology & Evolution*, **21**, 381–385.

Hudson PJ, Dobson AP, Newborn D (1992) Do parasites make prey vulnerable to predation? Red grouse and parasites. *Journal of Animal Ecology*, **61**, 681–692.

Hudson PJ, Dobson AP, Newborn D (1998) Prevention of population cycles by parasite removal. *Science*, **282**, 2256–2258.

Hugot JP, Chamaillard M, Zouali H *et al.* (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411**, 599–603.

Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C (2014) Exposing synonymous mutations. *Trends in Genetics*, **30**, 308–321.

Johnston SE, Orell P, Pritchard VL *et al.* (2014) Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **23**, 3452–3468.

Josse J, Husson F (2012) Handling missing values in exploratory multivariate data analysis methods. *Journal de la Société Française de Statistique*, **153**, 79–99.

Karolchik D, Hinrichs AS, Furey TS *et al.* (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, **32**, D493–D496.

Katagiri K, Hattori M, Minato N, Kinashi T (2002) Rap1 functions as a key regulator of T-cell and antigen-presenting cell interactions and modulates T-cell responses. *Molecular and Cellular Biology*, **22**, 1001–1015.

Kaufmann F, Daş G, Preisinger R, Schmutz M, König S, Gauly M (2011) Genetic resistance to natural helminth infections in two chicken layer lines. *Veterinary Parasitology*, **176**, 250–257.

Kent WJ (2002) BLAT–The BLAST-Like Alignment Tool. *Genome Research*, **12**, 656–664.

Kent WJ, Sugnet CW, Furey TS *et al.* (2002) The Human Genome Browser at UCSC. *Genome Research*, **12**, 996–1006.

Kinsella RJ, Kähäri A, Haider S *et al.* (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, **2011**, bar030.

Kuiperij HB, de Rooij J, Rehmann H *et al.* (2003) Characterisation of PDZ-GEFs, a family of guanine nucleotide exchange factors specific for Rap1 and Rap2. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, **1593**, 141–149.

Lecat A, Di Valentin E, Somja J *et al.* (2012) The c-Jun N-terminal kinase (JNK)-binding protein (JNKBP1) acts as a negative regulator of NOD2 protein signaling by inhibiting its oligomerization process. *Journal of Biological Chemistry*, **287**, 29213–29226.

Lee SH, DeCandia TR, Ripke S *et al.* (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics*, **44**, 247–250.

Li L, Xie X, Qin J *et al.* (2009) The nuclear orphan receptor COUP-TFII plays an essential role in adipogenesis, glucose homeostasis, and energy metabolism. *Cell Metabolism*, **9**, 77–87.

Lochmiller RL (1996) Immunocompetence and animal population regulation. *Oikos*, 594–602.

Lochmiller RL, Deerenberg C (2000) Trade-offs in evolutionary immunology: just what is the cost of immunity? *Oikos*, **88**, 87–98.

MacColl ADC, Piertney SB, Moss R, Lambin X (2000) Spatial arrangement of kin affects recruitment success in young male red grouse. *Oikos*, **90**, 261–270.

Manolio TA, Collins FS, Cox NJ *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

Martínez-Padilla J, Redpath SM, Zeineddine M, Mougeot F (2014) Insights into population ecology from long-term studies of red grouse *Lagopus lagopus scoticus*. *Journal of Animal Ecology*, **83**, 85–98.

Martínez-Padilla J, Mougeot F, Webster LMI, Pérez-Rodríguez L, Piertney SB (2010) Testing the interactive effects of testosterone and parasites on carotenoid-based ornamentation in a wild bird. *Journal of Evolutionary Biology*, **23**, 902–913.

Moss R, Trenholm IB, Watson A, Parr R (1990) Parasitism, predation and survival of hen red grouse *Lagopus lagopus scoticus* in spring. *Journal of Animal Ecology*, **59**, 631–642.

Mougeot F, Dawson A, Redpath SM, Leckie F (2005a) Testosterone and autumn territorial behavior in male red grouse *Lagopus lagopus scoticus*. *Hormones and Behavior*, **47**, 576–584.

Mougeot F, Evans SA, Redpath SM (2005b) Interactions between population processes in a cyclic species: Parasites reduce autumn territorial behaviour of male red grouse. *Oecologia*, **144**, 289–298.

Mougeot F, Irvine JR, Seivwright L, Redpath SM, Piertney S (2004) Testosterone, immunocompetence, and honest sexual signaling in male red grouse. *Behavioral Ecology*, **15**, 930–937.

Mougeot F, Martínez-Padilla J, Webster LMI, Blount JD, Pérez-Rodríguez L, Piertney SB (2009) Honest sexual signalling mediated by parasite and testosterone effects on oxidative balance. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 1093–1100.

Mougeot F, Piertney SB, Leckie F *et al.* (2005c) Experimentally increased aggressiveness reduces population kin structure and subsequent recruitment in red grouse *Lagopus lagopus scoticus*. *Journal of Animal Ecology*, **74**, 488–497.

Mougeot F, Redpath SM (2004) Sexual ornamentation relates to immune function in male red grouse *Lagopus lagopus scoticus*. *Journal of Avian Biology*, **35**, 425–433.

Mougeot F, Redpath SM, Leckie F (2005d) Ultra-violet reflectance of male and female red grouse, *Lagopus lagopus scoticus*: Sexual ornaments reflect nematode parasite intensity. *Journal of Avian Biology*, **36**, 203–209.

Mougeot F, Redpath SM, Moss R, Matthiopoulos J, Hudson PJ (2003) Territorial behaviour and population dynamics in red grouse *Lagopus lagopus scoticus*. I. Population experiments. *Journal of Animal Ecology*, **72**, 1073–1082.

Okada Y, Kamatani Y, Takahashi A *et al.* (2010) Common variations in PSMD3–CSF3 and PLCB4 are associated with neutrophil count. *Human Molecular Genetics*, **19**, 2079–2085.

Orsini L, Jansen M, Souche EL, Geldof S, De Meester L (2011) Single nucleotide polymorphism discovery from expressed sequence tags in the waterflea *Daphnia magna*. *BMC Genomics*, **12**, 309.

Orsini L, Spanier KI, De Meester L (2012) Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: validation in space, time and experimental evolution. *Molecular Ecology*, **21**, 2160–2175.

Owen JP, Nelson AC, Clayton DH (2010) Ecological immunology of bird-ectoparasite systems. *Trends in Parasitology*, **26**, 530–539.

Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nature Reviews Genetics*, **5**, 389–396.

Park H, Staehling K, Tsang M *et al.* (2012) Disruption of *Fnip1* Reveals a Metabolic Checkpoint Controlling B Lymphocyte Development. *Immunity*, **36**, 769–781.

Park H, Tsang M, Iritani BM, Bevan MJ (2014) Metabolic regulator Fnip1 is crucial for iNKT lymphocyte development. *Proceedings of the National Academy of Sciences*, **111**, 7066–7071.

Paterson S, Piertney SB (2011) Frontiers in host–parasite ecology and evolution. *Molecular Ecology*, **20**, 869–871.

Peig J, Green AJ (2009) New perspectives for estimating body condition from mass/length data: the scaled mass index as an alternative method. *Oikos*, **118**, 1883–1891.

Pemberton J, Beraldi D, Craig B, Hopkins J (2011) Digital gene expression analysis of gastrointestinal helminth resistance in Scottish blackface lambs. *Molecular Ecology*, **20**, 910–919.

Piertney S, Oliver M (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity*, **96**, 7–21.

Piertney SB, Lambin X, Maccoll ADC *et al.* (2008) Temporal changes in kin structure through a population cycle in a territorial bird, the red grouse *Lagopus lagopus scoticus*. *Molecular Ecology*, **17**, 2544–2551.

Piertney SB, MacColl ADC, Lambin X, Moss R, Dallas JF (1999) Spatial distribution of genetic relatedness in a moorland population of red grouse (*Lagopus lagopus scoticus*). *Biological Journal Of The Linnean Society*, **68**, 317–331.

Piertney SB, Webster LMI (2010) Characterising functionally important and ecologically meaningful genetic diversity using a candidate gene approach. *Genetica*, **138**, 419–432.

Poulin R (2013) Parasite manipulation of host personality and behavioural syndromes. *The Journal of Experimental Biology*, **216**, 18–26.

Poulin R, Thomas F (2008) Epigenetic effects of infection on the phenotype of host offspring: parasites reaching across host generations. *Oikos*, **117**, 331–335.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramji D, Foka P (2002) CCAAT/enhancer-binding proteins: structure, function and regulation. *Biochemical Journal*, **365**, 561–575.

Redpath SM, Mougeot F, Leckie FM, Elston DA, Hudson PJ (2006a) Testing the role of parasites in driving the cyclic population dynamics of a gamebird. *Ecology Letters*, **9**, 410–418.

Redpath SM, Mougeot F, Leckie FM, Evans SA (2006b) The effects of autumn testosterone on survival and productivity in red grouse, *Lagopus lagopus scoticus*. *Animal Behaviour*, **71**, 1297–1305.

Riggio V, Matika O, Pong-Wong R, Stear M, Bishop S (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 MR, Santure AW, DeCauwer I, Sheldon BC, Slate J (2013) Partitioning of genetic variation across the genome using multimarker methods in a wild bird population. *Molecular Ecology*, **22**, 3963–3980.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution*, **66**, 1–17.

Ruffell D, Mourkioti F, Gambardella A *et al.* (2009) A CREB-C/EBPβ cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. *Proceedings of the National Academy of Sciences*, **106**, 17475–17480.

Sadd BM, Schmid-Hempel P (2009) PERSPECTIVE: principles of ecological immunology. *Evolutionary Applications*, **2**, 113–121.

Scheele JS, Marks RE, Boss GR (2007) Signaling by small GTPases in the immune system. *Immunological Reviews*, **218**, 92–101.

Schielzeth H, Forstmeier W, Kempenaers B, Ellegren H (2012) QTL linkage mapping of wing length in zebra finch using genome-wide single nucleotide polymorphisms markers. *Molecular Ecology*, **21**, 329–339.

Schmid-Hempel P (2003) Variation in immune defence as a question of evolutionary ecology. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 357–366.

Seivwright LJ, Redpath SM, Mougeot F, Leckie F, Hudson PJ (2005) Interactions between intrinsic and extrinsic mechanisms in a cyclic species: testosterone increases parasite infection in red grouse. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 2299–2304.

Seivwright LJ, Redpath SM, Mougeot F, Watt L, Hudson PJ (2004) Faecal egg counts provide a reliable measure of *Trichostrongylus tenuis* intensities in free-living red grouse *Lagopus lagopus scoticus*. *Journal of Helminthology*, **78**, 69–76.

Shaw JL, Moss R (1989) Factors affecting the establishment of the caecal threadworm *Trichostrongylus tenuis* in red grouse (*Lagopus lagopus scoticus*). *Parasitology*, **99**, 259–264.

Sheldon BC, Verhulst S (1996) Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution*, **11**, 317–321.

Shin OH, Han W, Wang Y, Südhof TC (2005) Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca2+ binding properties. *Journal of Biological Chemistry*, **280**, 1641–1651.

Silva M, Sonstegard T, Hanotte O *et al.* (2012) Identification of quantitative trait loci affecting resistance to gastrointestinal parasites in a double backcross population of Red Maasai and Dorper sheep. *Animal Genetics*, **43**, 63–71.

Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology*, **14**, 363–379.

Stear MJ, Fitton L, Innocent GT, Murphy L, Rennie K, Matthews L (2007) The dynamic influence of genetic variation on the susceptibility of sheep to gastrointestinal nematode infection. *Journal of the Royal Society Interface*, **4**, 767–776.

Svensson E, Råberg L, Koch C, Hasselquist D (1998) Energetic stress, immunosuppression and the costs of an antibody response. *Functional Ecology*, **12**, 912–919.

Tarka M, Åkesson M, Beraldi D *et al.* (2010) A strong quantitative trait locus for wing length on chromosome 2 in a wild population of great reed warblers. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 2361–2369.

The Gene Ontology Consortium (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25–29.

Thomas F, Poulin R, Brodeur J (2010) Host manipulation by parasites: a multidimensional phenomenon. *Oikos*, **119**, 1217–1223.

Tobler M, Plath M, Riesch R *et al.* (2014) Selection from parasites favours immunogenetic diversity but not divergence among locally adapted host populations. *Journal of Evolutionary Biology*, **27**, 960–974.

Tsou RC, Zimmer DJ, De Jonghe BC, Bence KK (2012) Deficiency of PTP1B in leptin receptor-expressing neurons leads to decreased body weight and adiposity in mice. *Endocrinology*, **153**, 4227–4237.

Turner AK, Begon M, Jackson JA, Paterson S (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology*, **21**, 1632–1646.

van Valen L (1973) A new evolutionary law. *Evolutionary Theory*, **1**, 1–30.

Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. Springer, New York, fourth edition, iSBN 0-387-95457-0.

Vergara P, Mougeot F, Martínez-Padilla J, Leckie F, Redpath SM (2012) The condition dependence of a secondary sexual trait is stronger under high parasite infection level. *Behavioral Ecology*, **23**, 502–511.

Verner J, Kabathova J, Tomancova A *et al.* (2012) Gene expression profiling of acute graft-vs-host disease after hematopoietic stem cell transplantation. *Experimental Hematology*, **40**, 899–905.

von Mikecz A (2006) The nuclear ubiquitin-proteasome system. *Journal of Cell Science*, **119**, 1977–1984.

Watson H, Lee DL, Hudson PJ (1987) The effect of *Trichostrongylus tenuis* on the caecal mucosa of young, old and anthelmintic-treated wild red grouse, *Lagopus lagopus scoticus*. *Parasitology*, **94**, 405–411.

Webster LMI, Mello LV, Mougeot F, Martínez-Padilla J, Paterson S, Piertney SB (2011a) Identification of genes responding to nematode infection in red grouse. *Molecular Ecology Resources*, **11**, 305–313.

Webster LMI, Paterson S, Mougeot F, Martínez-Padilla J, Piertney SB (2011b) Transcriptomic response of red grouse to gastro-intestinal nematode parasites and testosterone: Implications for population dynamics. *Molecular Ecology*, **20**, 920–931.

Wenzel MA, Piertney SB (2014) Fine-scale population epigenetic structure in relation to gastro-intestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, **23**, 4256–4273.

Wenzel MA, Piertney SB (2015) Digging for gold nuggets: uncovering novel candidate genes for variation in gastrointestinal nematode burden in a wild bird species. *Journal of Evolutionary Biology*, **28**, 807–825.

Wenzel MA, Webster LM, Paterson S, Mougeot F, Martínez-Padilla J, Piertney SB (2013) A transcriptomic investigation of handicap models in sexual selection. *Behavioral Ecology and Sociobiology*, **67**, 221–234.

Wenzel MA, Webster LMI, Paterson S, Piertney SB (2015) Identification and characterisation of 17 polymorphic candidate genes for response to parasitic nematode (*Trichostrongylus tenuis*) infection in red grouse (*Lagopus lagopus scotica*). *Conservation Genetics Resources*, **7**, 23–28.

Wilfert L, Schmid-Hempel P (2008) The genetic architecture of susceptibility to parasites. *BMC Evolutionary Biology*, **8**, 187.

Wilson GR (1983) The prevalence of caecal threadworms (*Trichostrongylus tenuis*) in red grouse (*Lagopus lagopus scoticus*). *Oecologia*, **58**, 265–268.

Wongrak K, Daş G, Von Borstel UK, Gauly M (2014) Genetic variation for worm burdens in laying hens naturally infected with gastrointestinal nematodes. *British Poultry Science*, **56**.

Yang J, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565–569.

Yang J, Lee SH, Goddard ME, Visscher PM (2011a) GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, **88**, 76–82.

Yang J, Manolio TA, Pasquale LR *et al.* (2011b) Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, **43**, 519–525.

Zuk M, Stoehr AM (2002) Immune defense and host life history. *American Naturalist*, **160**, S9–S22.