Digging for gold nuggets: uncovering novel candidate genes for variation in gastrointestinal nematode burden in a wild bird species

Marius A. Wenzel* & Stuart B. Piertney

February 13, 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

Abstract

The extent to which genotypic variation at a priori identified candidate genes can explain variation in complex phenotypes is a major debate in evolutionary biology. Whilst some high-profile genes such as the MHC or MC1R clearly do account for variation in ecologically relevant characters, many complex phenotypes such as response to parasite infection may well be underpinned by a large number of genes, each of small and effectively undetectable effect. Here, we characterise a suite of novel candidate genes for variation in gastrointestinal nematode (Trichostongylus tenuis) burden among red grouse (Lagopus lagopus scotica) individuals across a network of moors in north-east Scotland. We test for associations between parasite load and genotypic variation in twelve genes previously identified to be differentially expressed in experimentally infected red grouse or genetically differentiated among red grouse populations with naturally different parasite loads. These genes are associated with a broad physiological response including immune system processes. Based on individual-level generalized linear models, genotypic variants in nine genes were significantly associated with parasite load, with effect sizes accounting for differences of 514–666 worms per bird. All but one of these variants were synonymous or untranslated, suggesting that these may be linked to protein-coding variants or affect regulatory processes. In contrast, population-level analyses revealed few and inconsistent associations with parasite load, and little evidence of...
signatures of natural selection. We discuss the broader significance of these contrasting re-
sults in the context of the utility of population genomics and landscape genomics approaches
in detecting adaptive genomic signatures.

Keywords: complex phenotypes, genetic architecture, infinitesimal theory, candidate genes, par-
asite susceptibility, large-effect polymorphisms

Running title: Novel candidate genes for parasite load in red grouse

Introduction

A fundamental goal in molecular ecology and evolutionary biology is to identify how different
eco-evolutionary processes influence the genetic variation that underpins adaptation in natural
populations (Mitchell-Olds et al., 2007; Ellegren & Sheldon, 2008; Andrew et al., 2013). Such
efforts, however, have been hampered because only rarely is the genetic architecture of phe-
otypic characters of ecological and adaptive importance properly known, which precludes the
identification of appropriate genomic targets through which gene dynamics in relation to adap-
tation can be followed (Ellegren & Sheldon, 2008; Allendorf et al., 2010). This problem can be
obviated, and the genetic basis of ecologically relevant characters resolved, in two conceptually
different ways. One strategy is to use approaches such as genome-wide association (GWAS) or
quantitative-trait loci (QTL) mapping, where the genetic architecture of a phenotypic charac-
ter is explored and derived de novo from statistical association between phenotypic states and
marker alleles in the study system (Ellegren & Sheldon, 2008). The alternative strategy focuses
on exploiting a set of candidate genes with a priori evidence for a functional link to the ecological
character of interest (Hoffmann & Willi, 2008; Piertney & Webster, 2010).

Candidate genes in the strict sense are typically derived from functional assays in model
systems and are usually also applicable to non-model systems (Fitzpatrick et al., 2005; Piertney
& Webster, 2010). Classic examples for such “bottom-up” candidate genes (sensu Piertney
& Webster, 2010) are the MC1R and MHC genes, which have proven to be extremely useful
paradigms for molecular selection and phenotypic adaptation (Hoekstra, 2006; Piertney & Oliver,
2006). In contrast, candidate genes in the broad sense are novel genes that are discovered directly
from transcriptomic or genomic assays in the target species (“top-down” candidate genes, sensu
Piertney & Webster, 2010) and thus provide a more comprehensive perspective than classic
bottom-up candidate genes (Hoffmann & Willi, 2008; Piertney & Webster, 2010). Differential
gene expression between individuals or populations that either differ naturally in the character
of interest or are subject to experimental intervention can highlight functionally relevant loci
that then become targets to be screened for SNPs (e.g., Orsini et al., 2011; Webster et al., 2011a;
Wang et al., 2012; De Wit & Palumbi, 2013; Gossner et al., 2013). Further, genome-wide scans
for locus-specific signatures of selection without a priori assumptions of phenotypic links or
causal environmental factors can reveal outlier loci that are then examined for associations with
environmental factors or phenotypic characters (e.g., Manel et al., 2009; Hess & Narum, 2011;
Matala et al., 2011; Pespeni & Palumbi, 2013; Milano et al., 2014). Similarly, locus-specific
genetic differentiation among samples of individuals that differ in the character of interest may
indicate adaptive significance with respect to character divergence, irrespective of the magnitude
of differentiation relative to the whole genome (e.g., Brown et al., 2013).

Notwithstanding whether genome-wide or candidate gene approaches are used, a perceived
problem is that the genetic architecture of the character of interest may involve a multitude
of genotypic variants with minute, possibly undetectable, effect sizes consistent with Fisher’s
infinitesimal theory (Fisher, 1919; Rockman, 2012). Clearly, the broad literature that has linked
specific genotypic variants to key phenotypic traits using genome-wide or candidate gene ap-
proaches indicates that large-effect “gold nuggets” (sensu Rockman, 2012) do exist, particularly
for simple, Mendelian phenotypes (Stern & Orgogozo, 2008; Rockman, 2012). Recent studies
using bottom-up candidate gene approaches have revealed such causal variants for phenotypes of
varying complexity, for example, body colouration and colour perception (Lehtonen et al., 2011;
Walsh et al., 2012; Poelstra et al., 2013), thermal responses (Shimada et al., 2011; Bedulina
et al., 2013), osmoregulation (Bedulina et al., 2013), growth and reproduction (Hemmer-Hansen
et al., 2011; Bedulina et al., 2013), immune function (Turner et al., 2012), and response to air
pollution (Bashalkhanov et al., 2013). Nevertheless, there remains substantial concern over the
generality of large-effect causal variants and in particular whether they exist for more complex
phenotypes (Rockman, 2012).

Here, we inform this debate by characterising a suite of novel candidate genes that were
developed using a top-down strategy to examine the genetic basis of an exemplary complex
phenotype, namely host response to chronic parasite burden. As a model, we use the interaction
between red grouse (Lagopus lagopus scotica Latham) and its primary parasite, the gastroin-
testinal nematode Trichostrongylus tenuis Mehlis. L. l. scotica is an economically important
subspecies of the willow ptarmigan (Lagopus lagopus) endemic to the heather moorlands of
upland Scotland and northern England (Martínez-Padilla et al., 2014). *T. tenuis* is highly prevalent in red grouse where it maintains a direct life cycle and imposes major fitness costs (Wilson, 1983). Infective larvae are ingested with heather shoots (Saunders et al., 1999) and establish in the caecum where adult parasites cause haemorrhaging that results in poor physiological condition and compromised survival and fecundity (Watson et al., 1987; Hudson et al., 1992; Delahay et al., 1995; Delahay & Moss, 1996). More than 90% of birds in a population are typically infected (Wilson, 1983) and, although some parasite-directed responses are mounted (Webster et al., 2011a), grouse typically cannot purge the infection such that they continue to bear parasite burdens for life (Shaw & Moss, 1989).

Long-term defence against chronic parasite insult has been demonstrated in several species to result in multifactorial effects on host behaviour and physiology (Sadd & Schmid-Hempel, 2009; Thomas et al., 2010; Biron & Loxdale, 2013; Poulin, 2013). These effects may be underpinned by substantial numbers of genes with small individual effect sizes (Wilfert & Schmid-Hempel, 2008; Rockman, 2012). Although heritability of parasite susceptibility and tolerance is well documented (Gauly & Erhardt, 2001; Stear et al., 2007; Mazé-Guilmo et al., 2014) and a range of bottom-up candidate immune system genes, such as the MHC (Oliver et al., 2009; Oppelt et al., 2010; Sin et al., 2014), interferon gamma (Coltman et al., 2001; Stear et al., 2007), Toll-like receptors (Downing et al., 2010) and cytokines (Luikart et al., 2008; Downing et al., 2010; Turner et al., 2012), have provided some insight, identification of novel top-down candidate genes for parasite infection in animals beyond immune system genes has proven to be a major challenge. For example, no conclusive associations between allele frequencies and parasite prevalence were found in *Daphnia* (Orsini et al., 2012) and Soay sheep populations (Brown et al., 2013), despite employing transcriptomic assays for candidate gene discovery (Orsini et al., 2011; Pemberton et al., 2011). These case studies cast doubt onto the existence of novel large-effect genotypic variants for parasite infection.

However, for red grouse there is the smoking gun of tell-tale signs that the response to parasite infection may indeed be influenced by some large effect genes. First, assays of exonic polymorphisms among two geographically distant grouse populations that differ in parasite load have indicated heterogeneity among locus-specific estimates of genetic differentiation (Wenzel et al., 2014). This suggests that significantly differentiated genes may be associated with parasite-driven selection processes rather than demographic isolation. Second, comparative transcriptomic analysis following experimental manipulation of parasite load has highlighted a
number of differentially expressed genes, some of which are involved in immune system pro-
cesses and unspecific tissue damage responses (Webster et al., 2011a,b). Third, an exploratory
genome-wide epigenomics study has highlighted an epigenetic signature of parasite load at spe-
cific genomic regions across a network of wild grouse populations (Wenzel & Piertney, 2014).
In concert, these studies demonstrate detectable functional and genetic effects associated with
parasite infection and parasite burden in the red grouse system. However, the key question
remaining to be answered is whether genotypic variation at such genes explains variance in
parasite load in natural red grouse populations.

Here, we examine this issue directly in red grouse individuals from a network of grouse moors
in north-east Scotland with parasite loads ranging nearly across five orders of magnitude (Wen-
zel & Piertney, 2014). We genotype these grouse at a previously developed suite of candidate
genes for parasite response and at a set of anonymous non-coding control markers that will
facilitate interpretation of adaptive patterns (Wenzel et al., 2014; Wenzel & Piertney, 2015).
Capitalising on extensive insight from previous ecological, physiological and molecular studies
on red grouse, we test for associations between genotypic variation and parasite load and explore
signatures of selection while accounting for confounding factors conferred by heterogeneity in
physiological condition, grouse moor management and shared environments due to social struc-
turing. Our analysis strategy combines population genomics and landscape genomics approaches
that together allow for investigating these patterns at the population and individual scale. We
hypothesize that grouse carrying different genotypic variants across these candidate genes will
der differ significantly in parasite load.

Materials and Methods

Study system

Our study system consists of 21 sampling sites (hereafter: populations) in a well-studied land-
scape of grouse moors in north-east Scotland near Deeside, Aberdeenshire (Fig. 1; Wenzel &
Piertney, 2014). One common management action on these grouse moors is parasite control to
improve grouse fitness and population growth (Martínez-Padilla et al., 2014). This is achieved
by dispensing quartz grit coated with an anthelmintic drug across the moors and allowing grouse
to self-medicate by ingesting this medicated grit alongside natural grit during normal feeding
behaviour (Newborn & Foster, 2002; Webster et al., 2008; Cox et al., 2010). Among the sites
sampled for the present study, medicated grit had been used for at least two years at fifteen sites, and not been used for at least 10 years at six sites (Table 1). As such, the presence or absence of medicated grit is an important confounding factor expected to impact parasite load.

Morphological measurements and tissue biopsies were taken from shot grouse following driven or walked-up sporting shoots in autumn 2012. As grouse populations on these moors typically display a degree of social and genetic structure in space and time, conferred by philopatry and territoriality by males (Watson et al., 1994; MacColl et al., 2000; Pierrtney et al., 1998, 1999, 2000, 2008), individuals were aged (“young”: < 1 year; “old”: > 1 year) and, where possible, old birds were preferentially sampled to minimise sampling bias through over-representation of kin groups (Wenzel & Pierrtney, 2014). As measures of physiological condition, body weight was measured to the nearest 10 g with a spring balance and supra-orbital comb size (width and length) was measured to the nearest mm. Carotenoid-based supra-orbital combs in both males and female grouse act as testosterone-dependent signals that are modulated by parasite load through impact on immune function (Mougeot & Redpath, 2004; Mougeot et al., 2004; Mougeot, 2008), oxidative status (Mougeot et al., 2009, 2010a) or physiological stress (Bortolotti et al., 2009; Mougeot et al., 2010b). As a result, comb size reflects an interaction between condition, age, sex and, to some extent, parasite load (Mougeot et al., 2004, 2005, 2009; Martínez-Padilla et al., 2010; Martinez-Padilla et al., 2011; Vergara et al., 2012a,b; Wenzel & Pierrtney, 2014).

Liver samples were taken for DNA extraction and caecum samples were taken for parasite load estimation from faecal parasite egg counts using the standard McMaster chamber slide method (Seivwright et al., 2004). DNA was extracted from 2–3 c. 2 mm³ shreds of liver tissue following the method of Hogan et al. (2008). Each bird was sexed genetically by amplification and electrophoresis of the gonosome-linked CHD genes (Griffiths et al., 1998) as described in Wenzel et al. (2012).

DNA sequencing and assembly

We selected 12 candidate genes for response to parasite infection from a suite of genes that were previously developed for red grouse from transcriptomic and genomic data (Lls_CG01–Lls_CG12; Wenzel et al., 2014). These genes are located in seven chromosomes in the chicken genome (Wenzel et al., 2014) and capture a broad physiological response, including immune system, xenobiotic detoxification, oxidative stress and metabolism processes (Table 2). Genetic differentiation among grouse populations with different parasite loads suggested that these
genes may be under natural selection, and genes Lls_CG01–Lls_CG07 also changed levels of
gene transcription in infected birds compared to control birds (Wenzel et al., 2014). To facilitate
interpretation of genotypic patterns and association with parasite load in these putatively adap-
tive candidate genes, we also selected four anonymous non-coding sequence markers (ANMs) as
control loci, located in four other chicken chromosomes to minimise likelihood of linkage dise-
equilibrium (Wenzel & Piertney, 2015). These control loci were designed in large unannotated
genomic regions remote from exonic annotations, and are therefore expected to be selectively
neutral and not functionally linked to parasite load (Wenzel & Piertney, 2015).

Candidate genes and control loci were amplified following the PCR conditions detailed in
Wenzel et al. (2014) and (Wenzel & Piertney, 2015), respectively. PCR amplicons were purified
by adding 1 U each of Exonuclease I (ExoI) and Antarctic phosphatase in a final concentration
of 1X CutSmart reaction buffer and 1X Antarctic phosphatase reaction buffer (all New England
Biolabs), and incubating at 37 °C for 45 min followed by enzyme heat deactivation at 80 °C for 20
min. Amplicon sequences were obtained from single-end Sanger sequencing on an ABI 3070XL
automatic capillary sequencer (The GenePool, University of Edinburgh, UK) and supplemented
with sequences from a 454 amplicon sequencing run on Roche 454 GS FLX+ (Eurofins Genomics,
Ebersberg, Germany) (details in supplementary materials S1).

Sequences were quality-controlled, assembled and aligned in GENEIOUS R7 (Drummond et al.,
2014). Heterozygous nucleotide sites in Sanger sequences and in individual-specific consensus se-
quences of 454 read contigs were coded as IUPAC degenerate bases and mapped to locus-specific
Sanger reference sequences (Wenzel et al., 2014; Wenzel & Piertney, 2015). Polymorphic sites
in these alignments were identified by eye and constant sites were removed. All sequences were
then subjected to haplotype reconstruction and imputation of missing data using the software
PHASE 2.1.1 (Stephens et al., 2001; Stephens & Scheet, 2005) with 1000 iterations, a thinning
interval of 10 and a burn-in of 1000. To optimise the reconstruction process, known haplotypes
were extracted from high-coverage 454 contigs of 14–62 (median: 32) individuals and included
as anonymous reference haplotype panels (dataDryad doi: TBC).

**Statistical analysis**

Summary statistics for sequence diversity (polymorphic sites, nucleotide diversity, haplotype
diversity) and test statistics for the neutral equilibrium population model (Tajima’s $D$, Fu & Li’s
$D$ and $F$) were obtained from reconstructed haplotype alignments using DNASP v5 (Librado &
Rozas, 2009). Global population differentiation was estimated with $G_{ST}$ and $N_{ST}$ statistics using the software SPADS (Dellicour & Mardulyn, 2014) with 1,000 permutations to obtain estimates of statistical significance. To examine signatures of natural selection, loci with disproportionate population differentiation ($F_{ST}$-outliers) were identified using BAYESCAN2 (Foll & Gaggiotti, 2008) with haplotype data as input. Additionally, to test whether haplotype reconstruction introduced bias, BAYESCAN was then run on 500 dataset replicates that were created by drawing one random SNP from each locus, thus retaining independence among loci. All runs consisted of $10^5$ iterations with a thinning interval of 20 after 20 pilot runs ($10^4$ iterations each) and a burn-in of $5 \cdot 10^5$. Outliers were selected at a significance threshold of $q \leq 0.05$.

**Population-level analysis**

Population-level associations between genotypic variation at each locus and parasite load were examined using genetic differentiation statistics based on population haplotype frequencies and medians of parasite load. To ascertain whether the loci can be treated as independent for multi-locus analyses, haplotypes were converted to diploid multi-locus allele frequencies using SPADS and PGDSPIDER (Lischer & Excoffier, 2012), and linkage disequilibrium among combinations of all 16 loci within each population was tested for in GENEPOP 4.2.1 (Raymond & Rousset, 1995; Rousset, 2008) with 10,000 MCMC dememorisations, 100 batches of 5000 MCMC iterations and a significance threshold of $\alpha = 0.05$.

Locus-by-locus hierarchical analysis of molecular variance (AMOVA; Excoffier et al., 1992) was used to test whether groups of populations with similar median parasite loads are significantly genetically differentiated. The AMOVA method partitions total genetic variance across three hierarchical components: among groups of populations ($\phi_{CT}$), among populations within groups ($\phi_{SC}$) and within populations ($\phi_{ST}$). The 21 populations (sampling sites) were divided into three groups of seven populations with broadly similar median parasite loads (4, 100–800, >800 worms per bird; Figure 1). Point estimates and statistical significance for $\phi_{CT}$, $\phi_{SC}$ and $\phi_{ST}$ were obtained using SPADS with 1,000 permutations.

Relationships between population genetic differentiation and median parasite load, corrected for isolation by distance and anthelmintic medication regime at sampling locations, were examined using isolation-by-stressor analysis based on non-parametric partial Mantel tests (Smouse et al., 1986) and a Bayesian parametric regression method that allows for testing the effect of multiple combinations of predictors on genetic differentiation (Foll & Gaggiotti, 2006). Pairwise
differentiation statistics among populations ($G_{ST}$, $G^{'ST}$, $D_{est}$) were computed for each locus individually and also combining all candidate genes or all control loci, using the package *diveRsity* (Keenan et al., 2013) in R 3.0.3 (R Core Team, 2014). Relationships between patterns of population differentiation between candidate genes and control loci were examined using scatter plots and linear models. Locus-by-locus Mantel tests with 9,999 permutations were then carried out to estimate the correlation between pairwise population matrices of linearized genetic differentiation and either logarithmic geographic distance (= isolation by distance) or median parasite load conditioned by binary differences in medication regime (0: same regime, 1: different regime) and logarithmic geographic distance (= isolation by stressor), using the R package *ecodist* (Goslee & Urban, 2007). Finally, Bayesian locus-by-locus models that regress population-specific local $F_{ST}$ estimates (Gaggiotti & Foll, 2010) on all possible combinations of median parasite load and three covariates (longitude, latitude and presence/absence of medicated grit) were fitted using the software *geste* (Foll & Gaggiotti, 2006). The posterior likelihoods of all model were then used to identify those predictor combinations that best explain variation in genetic differentiation. All models comprised a run length of $10^5$ with a thinning interval of 20, following 20 pilot runs of $10^4$ iterations each and a burn-in of $5 \cdot 10^5$.

**Individual-level analysis**

Given the substantial heterogeneity of parasite load within populations, we then focussed on individual-based generalized linear model analysis of parasite load and genotypic variants while accounting for additional factors that confound this relationship (e.g., Manel et al., 2009; Oliver et al., 2009; Sin et al., 2014; Wenzel & Piertney, 2014). In red grouse, three confounding factors are important to address. First, parasite load may be correlated within sampling sites (epidemiological neighbourhood effect; Hubbard et al., 2010), due to transmission dynamics conferred by a shared environment and social structuring of kin groups (Piertney et al., 1998, 1999; Martinez-Padilla et al., 2012). Second, parasite control through medicated grit reduces typical parasite load compared to sites where parasites are not controlled (Newborn & Foster, 2002). Third, individual parasite loads are expected to vary with sex, age and environment-dependent physiological condition (Mougeot et al., 2004, 2005, 2009; Martínez-Padilla et al., 2010; Martínez-Padilla et al., 2011; Vergara et al., 2012a,b).

To address these factors, parasite load was modelled as an over-dispersed Poisson-distributed variable in a generalised estimating equations framework (GEE) using the R package *geepack*
Observations were clustered by populations (21 clusters) and an exchangeable within-cluster correlation structure was applied to account for spatial correlation of parasite load caused by neighbourhood effects. This model choice provides easily interpretable population-average effect sizes and robust standard errors to facilitate significance testing of predictors (Hubbard et al., 2010). Individuals from medicated sites had significantly lower parasite loads than those from non-medicated sites (medians: 79 and 980 worms per bird; Wilcoxon’s \( W = 4802; P \ll 0.001 \)), so the presence or absence of medicated grit at a sampling site was included as a binary covariate. Relationships among morphological variables were examined using linear models to ascertain which covariates to include without causing collinearity or over-fitting the model (Graham, 2003; Oliver et al., 2009). As expected, comb area was strongly associated with sex \( (F_{1,166} = 101.1; P \ll 0.001) \), age \( (F_{1,166} = 28.1; P \ll 0.001) \) and weight \( (F_{1,166} = 11.4; P < 0.001) \), and was therefore included as a proxy variable for sex-, age- and condition-specific covariance in parasite load. More complex models did not yield substantially different results, indicating that the simpler model with comb area is both biologically and statistically appropriate.

Rather than relying on reconstructed haplotypes, original SNP genotypes were extracted from unphased sequence alignments and coded as categorical factors with up to three levels (homozygote for allele 1, homozygote for allele 2, heterozygote). Factor levels with a frequency below 0.05 were removed and the model was fitted on the reduced dataset if two levels remained. Further, each SNP was re-coded as up to three binary variables that represent the presence of allele 1, allele 2 or heterozygosity, respectively, and models were then re-run for alleles and heterozygosity independently. This approach is robust because it examines consistency across three different ways of subdividing observations during model fitting and also allows for intuitive separation of allele-specific effects and heterozygote advantage (Oliver et al., 2009; Oppelt et al., 2010). To aid the identification of such heterozygote effects, each SNP was tested for global deviation from Hardy-Weinberg equilibrium by calculating Wright’s inbreeding coefficient \( F_{IS} = 1 - \frac{H_O}{H_E} \) and testing significance with the \( \chi^2 \) test. Finally, for comparison with SNP-based analysis, the same models were re-run with each haplotype coded as a binary variable.

GEE model \( P \)-values were corrected for multiple testing within each type of genotypic variant (genotypes, alleles, heterozygosity, haplotypes) using the false discovery-rate approach (Benjamini & Hochberg, 1995). Significant \( (q \leq 0.1) \) model coefficient estimates \( (\beta_1) \) were linearised \( (e^{\beta_1}) \) and expressed as a percentage change in parasite load. Absolute effect sizes were calculated
from differences in predicted least-square population mean estimates between genotypic variants when keeping all other variables constant at mean values, using the R package *doBy* (Højsgaard & Halekoh, 2013). To provide an indication of how well genotypic variation explains parasite load, the difference between the quasi-likelihood based information criterion (QICu; Pan, 2001) estimates was calculated between the full model and an equivalent model omitting the genetic term.

Finally, to identify population-genomics signatures of parasite load in those SNP alleles identified through individual-based GEE models, we examined whether population-level allele frequencies were associated with median parasite load when accounting for the presence or absence of anthelmintic medication in populations. Allele frequencies were calculated from SNP genotype data and used as a predictor of median parasite load in a generalised linear model with negative binomial error structure and including a binary covariate representing anthelmintic medication, using the R package *MASS* (Venables & Ripley, 2002).

### Results

Estimated parasite loads among the 173 individuals ranged from 4 to 9283 worms per bird and population medians ranged from 4 to 2236 worms per bird (Table 1). Across all loci, between 98 and 173 individuals were sequenced at more than 50% of all polymorphic sites. The twelve candidate genes were sequenced in 125–173 individuals and contained 2–17 polymorphic sites that defined between 4–63 reconstructed haplotypes (Table 3). Sequencing failure was higher for control loci, where 98–171 individuals were successfully sequenced. Those sequences contained 6–15 polymorphic sites and defined 8–51 reconstructed haplotypes (Table 3). Estimated haplotype reconstruction fidelity based on small reference haplotype panels was 80±18%, indicating some uncertainty for some loci caused by few SNPs with disproportionate amounts of missing genotypes. Consensus sequences for all loci with genic annotations and SNP locations are available from GENBANK accessions KM236217–KM236228 (candidate genes) and KP210037–KP210040 (control loci). Significant deviation from the neutral equilibrium population model was detected in four candidate genes and three control loci, coinciding with significant genetic differentiation among populations (Table 3). No evidence for linkage disequilibrium among the 16 loci was detected based on reconstructed haplotypes, indicating that the loci can be treated as independent.
Population genetics analyses provided weak and inconsistent evidence for association between genetic differentiation and parasite load. Following hierarchical AMOVA, population groups with similar median parasite loads were marginally significantly differentiated at locus Lls_CG11 ($\phi_{CT} = 0.015; P = 0.053$), but not at any other locus (Table 3). Pairwise genetic differentiation estimates were typically higher for control loci ($G_{ST} = -0.032–0.123$) compared to candidate genes ($G_{ST} = -0.015–0.027$), even within population groups with similar parasite loads (Fig. 2), highlighting a discrepancy between neutral and adaptive genetic structure. When accounting for isolation by distance and differences in anthelmintic medication regime, differences in parasite load were consistently associated with three measures of genetic differentiation for two loci (Table 4), though only locus Lls_CG05 remained significant after correction for multiple testing using the false-discovery-rate method. Finally, Bayesian regression of population-specific local $F_{ST}$ estimates highlighted parasite load as a significant predictor (alongside latitude) for locus Lls_CG06 only. For most loci, no combination of predictors was more likely than a constant model, though medicated grit was a significant predictor for two loci (Table 4).

In contrast, individual-level GEE models provided consistent evidence for several positive and negative associations between genotypic variants and parasite load in nine candidate genes, including those highlighted by population genomics analyses. Of 131 SNPs in total, 19 were excluded because they were monomorphic after removing factor levels with frequencies below 0.05. All loci were represented by at least two SNPs after exclusion (supplementary materials S2). Parasite load was significantly ($q \leq 0.1$) associated with eight genotypes, seven alleles and five heterozygosity states in candidate gene SNPs, and with four genotypes, two alleles and four heterozygosity states in control locus SNPs (Fig. 3). Taken together, these associations cover eight SNPs in seven out of twelve candidate genes and six SNPs in three out of four control loci (Table 5). Of all SNPs, 70 (53 %) did not significantly deviate from Hardy-Weinberg equilibrium, 55 (42 %) displayed significant heterozygote deficiency and 6 (5 %) displayed significant heterozygote excess (supplementary materials S2). Haplotype-based analysis was impeded by low sample sizes for most haplotypes (only 86 out of 386 haplotypes were tested), but highlighted significant negative associations for four haplotypes in four candidate genes, of which two genes were not identified in SNP-based analysis (Lls_CG04 and Lls_CG09; supplementary materials S2). Significant models consistently provided a substantially better fit than equivalent models omitting the genetic term ($\Delta QICu = 4 – 108$), with the exception of one control locus SNP (Table 5). The relative predicted effect sizes of candidate gene SNP variants on parasite load

12
ranged from a reduction by 59 % to an increase by 196 % (Table 5) and the absolute predicted
effect on population means of the present study system ranged from –514 to +666 worms per
bird (Fig. 4). The highlighted candidate gene SNPs represent three mutation types, though
only a single SNP was non-synonymous, two SNPs were synonymous and all other SNPs were
in untranslated mRNA regions (Table 5).

The observed associations in candidate genes were predominantly driven by the presence of
particular alleles rather than heterozygosity. Although congruent associations across genotypic
variants within SNPs did not always allow for separating allele effects from heterozygosity effects
(particularly when not all genotypes were sampled; supplementary materials S3), there was no
evidence of deviation from Hardy-Weinberg equilibrium in these cases (Table 5). The single
exception is Lls_CG05 SNP 15, which displayed heterozygote deficiency and a positive associ-
ation between heterozygosity and parasite load, consistent with heterozygote disadvantage. In
all cases where an individual allele had a significant effect, the complementary allele either had
the opposite but non-significant effect or the model was not run because of low allele frequency
(<0.05) (supplementary materials S2). In spite of these highlighted, well supported individual-
based associations between alleles and parasite load, these same alleles were not associated with
parasite load on a population level. The population-level frequencies of the seven identified
candidate gene alleles ranged from 0.00 to 0.44, and no allele was significantly associated with
population median parasite load (Fig. 5), though allele C at Lls_CG06 SNP 1 displayed a
marginally significant negative association ($\beta_1 = -7.01; P = 0.10$), consistent with its negative
effect in individual-level GEE models (Table 5).

$F_{ST}$ outlier tests suggested that a subset of candidate genes that contained genotypic variants
significantly associated with parasite load may be under natural selection (Table 6). In contrast,
loci Lls_CG07, Lls_CG10 and Lls_CG11 were identified as outliers, but there was no association
with parasite load. However, the identified outliers all displayed disproportionately small genetic
differentiation and those loci with the greatest genetic differentiation were not identified as
outliers (Table 6). These patterns remained when candidate genes were analysed alone and
were therefore not driven by the control loci, whose simulated and observed $F_{ST}$ values were
overall considerably higher than those of the candidate genes (Table 6, Figure 2). Patterns
were similar between haplotypes and SNPs, but only a subset of significant outliers when using
haplotypes remained significant when using SNPs (Table 6). Results from $F_{ST}$ outlier tests and
tests for deviation from the neutral equilibrium population model were only weakly congruent
Discussion

We examined whether genotypic variation in a set of candidate genes that are associated with functionally diverse responses to parasitic gastrointestinal nematode infection in red grouse explains variation in actual nematode burden among individuals in natural grouse populations. Our findings highlight significant associations of parasite load with genotypic variants of substantial effect sizes in most candidate genes on an individual, but not population level. These results validate these genes as candidates for nematode infection among individuals of a wild bird species and suggest that this complex phenotype may, to some extent, be underpinned by large-effect genes.

Across all analyses, significant associations between parasite load and genotypic variants in at least one SNP or haplotype were detected in nine out of twelve candidate genes. These nine genes are involved in a range of physiological functions that effectively represent a multi-factorial response to long-term parasite insult (Thomas et al., 2010; Poulin, 2013). CYP2K4, GSTK1 and UGT1A1 represent key modification and conjugation enzymes involved in the two main stages of xenobiotics metabolism and detoxification (Guillemette, 2003; Tew & Townsend, 2012; Bock, 2014). In infected red grouse, such xenobiotics may originate from primary T. tenuis infection or from secondary pathogen infection facilitated through caecal damage and haemorrhaging (Watson et al., 1987). Moreover, CYP2K4 and GSTK1 enzymes are involved in producing and regulating reactive oxygen species, respectively (Lewis, 2002; Symons & King, 2003; Hellou et al., 2012), and may therefore be involved in parasite defence during an immune response or response to oxidative stress caused by other parasite defence mechanisms (Mougeot et al., 2009, 2010a). GAL9 is a chicken homologue of β-defensin 9 involved in innate antimicrobial immune responses and interactions with adaptive immune system processes (Mukhopadhyaya et al., 2010; Semple & Dorin, 2012). Further adaptive immune system components are represented by TCB and MFI2, a chicken homologue of the EOS47 eosinophil surface antigen (McNagny et al., 1996; Rahmanto et al., 2012). SUMO3 may be involved in regulation of cell cycle and gene expression, possibly as a response to physiological stress (Yang & Paschen, 2009; Sang et al., 2011). Similarly, CAPRIN1 and CCNL1 are principally involved in cell-cycle control (Wang et al., 2005; Lim & Kaldis, 2013) and epigenetic regulation of gene transcription and
RNA processing (Lim & Kaldis, 2013).

Genotypic variants in these genes may affect the efficacy of these diverse parasite defence mechanisms in a number of ways, all of which may translate into compromised physiological condition and ability to resist parasite infection (Sheldon & Verhulst, 1996; Loehmiller & Deerenberg, 2000; Wilfert & Schmid-Hempel, 2008; Sadd & Schmid-Hempel, 2009). The functional effects of most genotypic variants are cryptic in most cases, because all but one genotypic variant were either a synonymous mutation in a coding region or a mutation located in an untranslated mRNA region. As such, these variants will not directly affect the translated amino acid sequence of the gene product, but may instead be involved in regulating gene expression or epigenetic mechanisms or be linked to a variant with any such effect. Synonymous mutations and non-coding mutations may affect gene transcription through cis-regulatory elements, such as splicing sites or binding sites for epigenetic mechanisms such as miRNAs or transcription factors, whereas translation can be affected by altered mRNA stability, impeded ribosome binding efficiency and codon bias (Chamary et al., 2006; Sauna & Kimchi-Sarfaty, 2011; Hunt et al., 2014). Such cryptic effects could also explain the few inconsistent associations between parasite load and genotypic variants in neutral, non-coding control loci, which may otherwise be fortuitous statistical artefacts. Intriguingly, we previously identified genome-wide cytosine methylation states that may be under selection and were also associated with parasite load in this same study system (Wenzel & Piertney, 2014). Although none of these identified methylation states were linked to any of the candidate genes of the present study, the presence of an epigenetic signature of parasite load in concert with a genetic signature in epigenetic factors such as CAPRIN1, CCNL1 and SUMO3 suggest that epigenetic processes may indeed play an important role in the host-parasite interactions of this study system (Poulin & Thomas, 2008; Gómez-Díaz et al., 2012).

These findings provide novel evidence for both beneficial and detrimental effects of genotypic variation in top-down candidate genes for gastrointestinal nematode abundance among individuals in natural populations. However, there was no evidence for heterozygote advantage, though heterozygote disadvantage may be operating at one gene, contrary to previous reports for nematodes (Luikart et al., 2008) and other parasites (Oliver et al., 2009; Oppelt et al., 2010). The genetic basis of parasite susceptibility has been notoriously difficult to uncover (Wilfert & Schmid-Hempel, 2008), although some insight has been gained through bottom-up candidate gene approaches involving MHC (e.g., Oliver et al., 2009; Oppelt et al., 2010; Sin et al., 2014),
interferon gamma (Coltman et al., 2001; Stear et al., 2007), Toll-like receptors (Downing et al., 2010; Gavan et al., in press) and cytokines (Luikart et al., 2008; Downing et al., 2010; Turner et al., 2012). Landscape-scale application of novel top-down candidate genes for parasite susceptibility in *Daphnia* revealed differential patterns of selection at these genes among populations in different stress environments, but no conclusive associations between allele frequencies and parasite prevalence were found (Orsini et al., 2011, 2012). Susceptibility to nematode infection in sheep species has previously been found to be associated with alleles in the MHC and interferon gamma genes (Stear et al., 2007), and also with heterozygosity in a T-cell receptor gene (TCRG4) and a cytokine regulator (ADCYAP1) (Luikart et al., 2008), but none of these genes explained nematode abundance in a large-scale study on 960 Soay sheep (Brown et al., 2013). The same study also tested a panel of 123 candidate SNPs derived from genomic and top-down transcriptomic studies on various sheep species, but revealed no more significant associations with nematode abundance than expected by chance (Brown et al., 2013).

Identifying novel candidate genes for the genetic basis of complex phenotypes can be hampered by relying on transcriptomic assays among different phenotypes to establish functional relevance (e.g., Orsini et al., 2011; Pemberton et al., 2011; Wang et al., 2012; De Wit & Palumbi, 2013; Gossner et al., 2013). This is because phenotypic diversity may be mediated by trans-regulated variation in gene transcript number rather than sequence polymorphisms in transcripts or tightly linked cis-regulatory elements (Stern & Orgogozo, 2008; Duncan et al., 2014). Integrating transcriptomic data with genomic assays, as we did for identifying the candidate genes for red grouse (Wenzel et al., 2014), or focussing on genome scans alone (e.g., Manel et al., 2009; Hess & Narum, 2011; Matala et al., 2011; Pespeni & Palumbi, 2013; Milano et al., 2014) is likely to reduce the risk of false positive identification. Great potential also lies in shifting focus from targeting sequence polymorphisms to gene dynamics and epigenetics. For example, Schneider et al. (2014) recently demonstrated a purely transcriptomic application of the candidate gene approach to reveal a transcriptomic basis for diet-induced morphological plasticity in cichlids. The candidate genes were previously discovered using large-scale comparative transcriptomics of divergent phenotypes following diet manipulation (Gunter et al., 2013). This approach avoids the conceptual disjoint between the genome and the transcriptome and is likely to become an important complement to whole-genome population transcriptomics approaches (Ouborg et al., 2010; Debes et al., 2012; Matzkin, 2012). In the same vein, genome scans for epigenetic patterns such as cytosine methylation may well facilitate identification of candidate regulatory regions.
that are associated with phenotypic responses to environmental factors (Duncan et al., 2014; Wenzel & Piertney, 2014).

In spite of our ability to detect a genetic signature of individual-level parasite susceptibility in red grouse, such a signature was not manifest at a population level, irrespective of whether SNP or haplotype data were used. This discrepancy highlights that population genomics approaches may fail to identify adaptive differences among individuals, unless the examined populations are both genetically and phenotypically differentiated to substantial degrees. The red grouse populations examined here were only subtly genetically differentiated with substantial variability of parasite load within populations, which is likely to impede statistical inference when reducing this variation to population-level statistics. Moreover, selection on parasite defence mechanisms may be weakened when an anthelmintic is administered through medicated grit (Newborn & Foster, 2002), which may further dampen parasite-associated adaptive population-level signatures. Classic tests for the neutral equilibrium model highlighted all loci with strong population structure (particularly the control loci), indicating that demographic history may confound detection of selection in this system (Nielsen, 2001). Indeed, tests for signatures of natural selection highlighted almost all genes to be under balancing selection, which is likely to be a statistical artefact caused by low population differentiation or an inappropriate population model (Lotterhos & Whitlock, 2014), illustrating a key issue with current outlier detection software applied on landscape scale systems. Not least, signatures of selection may well be due to unknown environmental factors other than parasites, because three genes were identified as $F_{ST}$ outliers but were not associated with parasite load.

In consequence, population genomics signatures may be difficult to detect for complex polygenic phenotypes unless the phenotype is underpinned by at least some large-effect polymorphisms under strong selection pressure. Classic systems such as MC1R alleles that are perfectly associated with melanism in mouse populations (Hoekstra et al., 2004; Mullen & Hoekstra, 2008) are unlikely to be the norm for most population genomics scenarios given Fisher’s infinitesimal model (Fisher, 1919). Nevertheless, our results show that this does not preclude the detection of fine-scale patterns among individuals. As such, an individual-based landscape genomics approach to link alleles with environmental or phenotypic variables is likely to be more powerful in detecting adaptive genetic discontinuities on continuous landscapes than population genomics approaches (Joost et al., 2007, 2013). Landscape genomics remains a challenging field for the multitude of confounding factors that can conspire to mask overall signatures of adaptive pro-
cesses (Manel et al., 2010; Joost et al., 2013), but its concepts may be paramount to examining spatio-temporal dynamics of novel polymorphisms in the field.

In summary, our results provide evidence for substantial effects of genotypic variants on gastrointestinal parasite load in a wild bird species. This demonstrates that carefully chosen candidate genes have the potential to link specific mutations of large effect to complex phenotypes, which contrasts with the prevailing view that no such “gold nuggets” generally exist (Rockman, 2012). The key advantages of the candidate gene approach are its statistical power compared to genome-wide approaches (Amos et al., 2011) and its a priori assembly of theoretical and empirical evidence for a phenotypic link. Motivated by an enhanced ability to amalgamate genomic, transcriptomic, epigenomic and proteomic data, the candidate gene approach is likely to remain a mainstay in the toolbox of molecular ecologists endeavouring to uncover the ’omics of phenotypic variation.

**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 Marianne James, Mario Röder and Keliya Bai for fieldwork assistance, Lucy M.I. Webster and Steve Paterson for help during prior development of genetic markers, Heather Ritchie for helpful comments on manuscript drafts, and all estate owners, factors and keepers for access to field sites, most particularly MJ Taylor and Mike Nisbet (Airlie), Neil Brown (Allargue), RR Gledson and David Scrimgeour (Delnadamph), Andrew Salvesen and John Hay (Dinnet), Stuart Young and Derek Calder (Edinglassie), Kirsty Donald and David Busfield (Glen Dye), Neil Hogbin and Ab Taylor (Glen Muick), Alistair Mitchell (Glenlivet), Simon Blackett, Jim Davidson and Liam Donald (Invercauld), Richard Cooke and Fred Taylor (Invermark), Shaila Rao and Christopher Murphy (Mar Lodge), and Ralph Peters and Philip Astor (Tillypronie).

**Conflicts of interest**

No conflicts of interest have been declared.
Figure 1: Sites in Aberdeenshire, Angus and Moray that were sampled following grouse sporting shoots in autumn 2012. Median parasite load for each site is indicated by three categories. Detailed locations, sample sizes and parasite loads are presented in table 1.

Figure 2: Relationships between genetic differentiation ($G_{ST}$) among population pairs estimated across candidate genes or control loci. Each data point represents one population pair. Solid lines indicate linear relationships for all data points (grey line) and three subsets representing population pairs within three parasite load categories (coloured symbols and lines). The dotted line represents a hypothetical 1:1 relationship. Patterns were similar for $G'_{ST}$ and $D_{est}$ (not shown).
Figure 3: Graphical summary of regression model coefficients and statistical significance of generalised estimating equations (GEE) predicting parasite load by genotypic variation (diploid SNP genotype or presence or absence of an individual allele or heterozygosity) at 131 SNPs in twelve candidate genes and four neutral control loci. Each vertical line represents one genotypic variant comparison within a SNP (up to two comparisons for SNP genotype and alleles; single comparison for heterozygosity). Line height represents single-test statistical significance (–log10 $P$-value). Line colours indicates statistical significance levels before and after correction for multiple testing using the false-discovery-rate approach (red: $q \leq 0.05$, orange: $q \leq 0.1$ and black: $P \leq 0.05$). Line symbols indicate the sign of the regression coefficient (“up”: positive, “down”: negative).
Figure 4: Predicted effect sizes (mean ± SE) of genotypic variants in candidate gene SNPs significantly associated with parasite load. Estimates were derived from differences in GEE-predicted least-square population means between genotypic variants when keeping all other variables constant at mean values. When two genotype effects were significant, only the largest effect is plotted.
Figure 5: Population-level associations between median parasite load and frequencies of seven candidate gene SNP alleles with significant associations in individual-level GEE models. Populations are ordered by median parasite load. Allele frequencies are represented by circle area and colours indicate the presence or absence of medicated grit. Regression statistics of negative binomial models fitted for each allele and accounting for medication regime are presented atop each column.
Table 1: Locations, sample sizes (M=male, F=female, Y=young) and parasite loads (median number of worms per bird with 25 % and 75 % quantiles) of 21 sampling sites.

<table>
<thead>
<tr>
<th>Site</th>
<th>Estate</th>
<th>Long.</th>
<th>Lat.</th>
<th>Anthelmintic grit</th>
<th>Total</th>
<th>M</th>
<th>F</th>
<th>Y</th>
<th>25 %</th>
<th>Median</th>
<th>75 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glenlivet</td>
<td>57.29</td>
<td>-3.18</td>
<td>Yes</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>981</td>
</tr>
<tr>
<td>2</td>
<td>Glenlivet</td>
<td>57.25</td>
<td>-3.28</td>
<td>Yes</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>Edinglassie</td>
<td>57.24</td>
<td>-3.20</td>
<td>Yes</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Edinglassie</td>
<td>57.21</td>
<td>-3.19</td>
<td>Yes</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>Allargue</td>
<td>57.19</td>
<td>-3.29</td>
<td>Yes</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Allargue</td>
<td>57.19</td>
<td>-3.23</td>
<td>Yes</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Delnadamph</td>
<td>57.16</td>
<td>-3.26</td>
<td>No</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>348</td>
<td>930</td>
<td>1616</td>
</tr>
<tr>
<td>8</td>
<td>Delnadamph</td>
<td>57.14</td>
<td>-3.30</td>
<td>No</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>937</td>
<td>1237</td>
<td>1837</td>
</tr>
<tr>
<td>9</td>
<td>Invercauld</td>
<td>57.10</td>
<td>-3.29</td>
<td>Yes</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>608</td>
<td>1826</td>
</tr>
<tr>
<td>10</td>
<td>Invercauld</td>
<td>57.08</td>
<td>-3.35</td>
<td>Yes</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>99</td>
<td>422</td>
<td>2856</td>
</tr>
<tr>
<td>11</td>
<td>Dinnet</td>
<td>57.12</td>
<td>-3.11</td>
<td>Yes</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>12</td>
<td>Dinnet</td>
<td>57.11</td>
<td>-3.06</td>
<td>Yes</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>282</td>
<td>682</td>
</tr>
<tr>
<td>13</td>
<td>Tillypronie</td>
<td>57.18</td>
<td>-2.94</td>
<td>Yes</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>103</td>
<td>380</td>
</tr>
<tr>
<td>14</td>
<td>Mar Lodge</td>
<td>56.95</td>
<td>-3.66</td>
<td>No</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>264</td>
<td>525</td>
<td>1244</td>
</tr>
<tr>
<td>15</td>
<td>Invercauld</td>
<td>56.87</td>
<td>-3.40</td>
<td>Yes</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>146</td>
<td>541</td>
<td>1006</td>
</tr>
<tr>
<td>16</td>
<td>Airlie</td>
<td>56.81</td>
<td>-3.08</td>
<td>No</td>
<td>13</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>786</td>
<td>2120</td>
<td>2977</td>
</tr>
<tr>
<td>17</td>
<td>Glen Muick</td>
<td>56.99</td>
<td>-3.01</td>
<td>Yes</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>908</td>
<td>2236</td>
<td>3812</td>
</tr>
<tr>
<td>18</td>
<td>Invermark</td>
<td>56.94</td>
<td>-2.89</td>
<td>Yes</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>552</td>
<td>1084</td>
<td>1386</td>
</tr>
<tr>
<td>19</td>
<td>Invermark</td>
<td>56.89</td>
<td>-2.89</td>
<td>Yes</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>429</td>
<td>603</td>
<td>650</td>
</tr>
<tr>
<td>20</td>
<td>Glen Dye</td>
<td>56.95</td>
<td>-2.72</td>
<td>No</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>448</td>
<td>916</td>
<td>1448</td>
</tr>
<tr>
<td>21</td>
<td>Glen Dye</td>
<td>56.96</td>
<td>-2.69</td>
<td>No</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>358</td>
<td>841</td>
<td>1509</td>
</tr>
</tbody>
</table>

| 173 | 130 | 43 | 30 |

Table 2: Characterisation of candidate genes for response to parasitic nematode infection in red grouse. Gene names and descriptors are given alongside an indication of physiological function. Full characterisation is detailed in Wenzel et al. (2014).

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Descriptor</th>
<th>Putative function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lls_CG01</td>
<td>TCB</td>
<td>T-cell receptor beta chain T17T-22</td>
<td>Innate immune response</td>
</tr>
<tr>
<td>Lls_CG02</td>
<td>CYP2K4</td>
<td>Cytochrome P450 2K4</td>
<td>Detoxification; oxidative balance</td>
</tr>
<tr>
<td>Lls_CG03</td>
<td>GAL9</td>
<td>Gallinacin-9</td>
<td>Innate immune response</td>
</tr>
<tr>
<td>Lls_CG04</td>
<td>GSTK1</td>
<td>Glutathione S-transferase kappa 1</td>
<td>Detoxification; oxidative balance</td>
</tr>
<tr>
<td>Lls_CG05</td>
<td>CAPRIN1</td>
<td>Caprin-1</td>
<td>Cell cycle; gene expression</td>
</tr>
<tr>
<td>Lls_CG06</td>
<td>UGT1A1</td>
<td>UDP-glucuronosyltransferase 1-1</td>
<td>Detoxification</td>
</tr>
<tr>
<td>Lls_CG07</td>
<td>ATP1A1</td>
<td>Sodium/potassium-transporting ATPase subunit alpha-1</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>Lls_CG08</td>
<td>MFI2</td>
<td>Melanotransferrin</td>
<td>Immune system</td>
</tr>
<tr>
<td>Lls_CG09</td>
<td>CCNL1</td>
<td>Cyclin-L1</td>
<td>Cell cycle; gene expression</td>
</tr>
<tr>
<td>Lls_CG10</td>
<td>SPCS2</td>
<td>Signal peptidase complex subunit 2</td>
<td>Peptide translocation</td>
</tr>
<tr>
<td>Lls_CG11</td>
<td>MIOS</td>
<td>WD repeat-containing protein mio</td>
<td>Protein modification</td>
</tr>
<tr>
<td>Lls_CG12</td>
<td>SUMO3</td>
<td>Small ubiquitin-related modifier 3</td>
<td>Cell cycle; gene expression</td>
</tr>
</tbody>
</table>
Table 3: Summary statistics for candidate genes and neutral control loci based on reconstructed haplotypes. The number of sequenced individuals (n), total polymorphic sites (P), transitions (Ti), transversions (Tv), synonymous sites (S), non-synonymous sites (NS), non-coding sites (NC), average pairwise nucleotide differences (k) and nucleotide diversity (π) are given alongside number of haplotypes (H), haplotype diversity (Hd) and Tajima’s D (DT), Fu & Li’s D (DFL) and F (FFL) test statistics for the equilibrium neutral population model. Global population differentiation is presented as GST, NST and hierarchical AMOVA statistics (φCT represents differentiation among the three population groups indicated in Figure 1). Significant statistics are emboldened and significance is indicated with asterisks.

<table>
<thead>
<tr>
<th>Locus</th>
<th>n</th>
<th>Size (bp)</th>
<th>P</th>
<th>Ti</th>
<th>Tv</th>
<th>S</th>
<th>NS</th>
<th>NC</th>
<th>k</th>
<th>π</th>
<th>H</th>
<th>Hd</th>
<th>DT</th>
<th>DFL</th>
<th>FFL</th>
<th>GST</th>
<th>NST</th>
<th>φCT</th>
<th>φSC</th>
<th>φST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lls_CG01</td>
<td>171</td>
<td>312</td>
<td>13</td>
<td>10</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>3.18</td>
<td>0.010</td>
<td>58</td>
<td>0.910</td>
<td>1.325</td>
<td>0.024</td>
<td>0.629</td>
<td>-0.001</td>
<td>-0.006</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.005</td>
</tr>
<tr>
<td>Lls_CG02</td>
<td>173</td>
<td>191</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2.35</td>
<td>0.012</td>
<td>21</td>
<td>0.709</td>
<td>3.619***</td>
<td>0.939</td>
<td>2.228***</td>
<td>-0.012</td>
<td>-0.016</td>
<td>-0.003</td>
<td>-0.012</td>
<td>-0.015</td>
</tr>
<tr>
<td>Lls_CG03</td>
<td>168</td>
<td>158</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>2.08</td>
<td>0.013</td>
<td>13</td>
<td>0.779</td>
<td>2.310*</td>
<td>1.025</td>
<td>1.766*</td>
<td>0.021*</td>
<td>0.046**</td>
<td>0.004</td>
<td>0.040**</td>
<td>0.044**</td>
</tr>
<tr>
<td>Lls_CG04</td>
<td>169</td>
<td>206</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.78</td>
<td>0.004</td>
<td>4</td>
<td>0.591</td>
<td>1.876</td>
<td>0.606</td>
<td>1.209</td>
<td>-0.008</td>
<td>-0.005</td>
<td>0.005</td>
<td>-0.007</td>
<td>-0.002</td>
</tr>
<tr>
<td>Lls_CG05</td>
<td>140</td>
<td>454</td>
<td>15</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>4.17</td>
<td>0.009</td>
<td>49</td>
<td>0.763</td>
<td>1.785</td>
<td>1.550*</td>
<td>1.977*</td>
<td>0.017</td>
<td>0.037*</td>
<td>0.007</td>
<td>0.035*</td>
<td>0.028*</td>
</tr>
<tr>
<td>Lls_CG06</td>
<td>166</td>
<td>336</td>
<td>17</td>
<td>10</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>0</td>
<td>2.49</td>
<td>0.007</td>
<td>63</td>
<td>0.707</td>
<td>-0.163</td>
<td>1.622*</td>
<td>1.106</td>
<td>0.082***</td>
<td>0.110***</td>
<td>0.004</td>
<td>0.109***</td>
<td>0.112***</td>
</tr>
<tr>
<td>Lls_CG07</td>
<td>169</td>
<td>182</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1.30</td>
<td>0.007</td>
<td>11</td>
<td>0.652</td>
<td>1.198</td>
<td>0.941</td>
<td>1.231</td>
<td>0.013</td>
<td>0.005</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td>Lls_CG08</td>
<td>172</td>
<td>250</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>3.43</td>
<td>0.014</td>
<td>42</td>
<td>0.904</td>
<td>1.615</td>
<td>0.735</td>
<td>1.293</td>
<td>0.007</td>
<td>0.014</td>
<td>0.000</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>Lls_CG09</td>
<td>172</td>
<td>250</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1.24</td>
<td>0.004</td>
<td>10</td>
<td>0.604</td>
<td>1.060</td>
<td>0.939</td>
<td>1.174</td>
<td>0.048**</td>
<td>0.052***</td>
<td>0.002</td>
<td>0.045**</td>
<td>0.047**</td>
</tr>
<tr>
<td>Lls_CG10</td>
<td>173</td>
<td>226</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1.23</td>
<td>0.005</td>
<td>15</td>
<td>0.677</td>
<td>0.252</td>
<td>1.098</td>
<td>0.957</td>
<td>0.000</td>
<td>-0.016</td>
<td>0.001</td>
<td>-0.016</td>
<td>-0.015</td>
</tr>
<tr>
<td>Lls_CG11</td>
<td>151</td>
<td>184</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.76</td>
<td>0.004</td>
<td>4</td>
<td>0.629</td>
<td>1.792</td>
<td>0.612</td>
<td>1.178</td>
<td>0.013</td>
<td>0.005</td>
<td>0.015</td>
<td>-0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>Lls_CG12</td>
<td>125</td>
<td>124</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0.91</td>
<td>0.007</td>
<td>5</td>
<td>0.666</td>
<td>1.320</td>
<td>0.757</td>
<td>1.117</td>
<td>-0.059</td>
<td>-0.027</td>
<td>-0.005</td>
<td>0.004</td>
<td>-0.001</td>
</tr>
<tr>
<td>Lls_ANM_18_1</td>
<td>133</td>
<td>371</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>2.30</td>
<td>0.006</td>
<td>20</td>
<td>0.877</td>
<td>2.100</td>
<td>1.118</td>
<td>1.752*</td>
<td>0.071***</td>
<td>0.106***</td>
<td>-0.004</td>
<td>0.098***</td>
<td>0.095***</td>
</tr>
<tr>
<td>Lls_ANM_20_2</td>
<td>98</td>
<td>401</td>
<td>10</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>4.43</td>
<td>0.011</td>
<td>12</td>
<td>0.730</td>
<td>3.728***</td>
<td>1.331</td>
<td>2.632**</td>
<td>0.149***</td>
<td>0.223***</td>
<td>-0.034</td>
<td>0.199***</td>
<td>0.172***</td>
</tr>
<tr>
<td>Lls_ANM_22_2</td>
<td>98</td>
<td>355</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>1.28</td>
<td>0.004</td>
<td>8</td>
<td>0.475</td>
<td>0.509</td>
<td>1.066</td>
<td>1.041</td>
<td>0.024</td>
<td>0.038</td>
<td>-0.010</td>
<td>0.040*</td>
<td>0.040</td>
</tr>
<tr>
<td>Lls_ANM_6_1</td>
<td>171</td>
<td>405</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>4.75</td>
<td>0.012</td>
<td>51</td>
<td>0.878</td>
<td>2.480*</td>
<td>0.883</td>
<td>1.818*</td>
<td>0.009</td>
<td>0.022*</td>
<td>0.006</td>
<td>0.024*</td>
<td>0.030**</td>
</tr>
</tbody>
</table>

*: P ≤ 0.05; **: P ≤ 0.01; ***: P ≤ 0.001
Table 4: Population-level association tests between genetic differentiation and parasite load, accounting for geography and anthelmintic medication regime. Mantel test correlation coefficients ($r$) are presented for three differentiation statistics ($G_{ST}$, $G'_{ST}$, $D_{est}$) versus logarithmic geographic distance (isolation by distance) and versus logarithmic differences in median parasite load conditioned by logarithmic geographic distance and differences in anthelmintic medication regime (isolation by stressor). Alongside, posterior likelihoods are presented for latitude, longitude, medicated grit and parasite load predictors in Bayesian regression of local $F_{ST}$ estimates (GESTE software) and for the most likely model containing a constant factor and any combination of these four predictors. Emboldened values represent statistically significant regression coefficients and variables included in the most likely GESTE models.

<table>
<thead>
<tr>
<th>Locus</th>
<th>$G_{ST}$</th>
<th>$G'_{ST}$</th>
<th>$D$</th>
<th>$G_{ST}$</th>
<th>$G'_{ST}$</th>
<th>$D$</th>
<th>Lat.</th>
<th>Long.</th>
<th>Grit</th>
<th>Parasite load</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lls_CG01</td>
<td>0.038</td>
<td>0.042</td>
<td>0.041</td>
<td>-0.065</td>
<td>-0.074</td>
<td>-0.074</td>
<td>0.297</td>
<td>0.394</td>
<td>0.331</td>
<td>0.321</td>
<td>0.205</td>
</tr>
<tr>
<td>Lls_CG02</td>
<td>0.001</td>
<td>0.008</td>
<td>0.011</td>
<td>-0.123</td>
<td>-0.120</td>
<td>-0.117</td>
<td>0.457</td>
<td>0.466</td>
<td>0.467</td>
<td>0.472</td>
<td>0.082</td>
</tr>
<tr>
<td>Lls_CG03</td>
<td>0.071</td>
<td>0.065</td>
<td>0.067</td>
<td><strong>0.205</strong></td>
<td><strong>0.195</strong></td>
<td><strong>0.194</strong></td>
<td>0.322</td>
<td>0.341</td>
<td>0.449</td>
<td>0.333</td>
<td>0.173</td>
</tr>
<tr>
<td>Lls_CG04</td>
<td>0.152</td>
<td>0.107</td>
<td>0.094</td>
<td>-0.002</td>
<td>0.012</td>
<td>0.014</td>
<td>0.113</td>
<td>0.071</td>
<td>0.097</td>
<td>0.171</td>
<td>0.684</td>
</tr>
<tr>
<td>Lls_CG05</td>
<td>-0.001</td>
<td>-0.037</td>
<td>-0.038</td>
<td><strong>0.251</strong></td>
<td><strong>0.305</strong></td>
<td><strong>0.311</strong></td>
<td>0.256</td>
<td>0.315</td>
<td>0.343</td>
<td>0.191</td>
<td>0.314</td>
</tr>
<tr>
<td>Lls_CG06</td>
<td><strong>0.173</strong></td>
<td><strong>0.237</strong></td>
<td><strong>0.249</strong></td>
<td>-0.016</td>
<td>-0.097</td>
<td>-0.092</td>
<td><strong>0.629</strong></td>
<td>0.147</td>
<td>0.198</td>
<td><strong>0.771</strong></td>
<td>0.395</td>
</tr>
<tr>
<td>Lls_CG07</td>
<td>0.095</td>
<td>0.112</td>
<td>0.113</td>
<td>-0.144</td>
<td>-0.159</td>
<td>-0.160</td>
<td>0.361</td>
<td>0.348</td>
<td>0.382</td>
<td>0.389</td>
<td>0.167</td>
</tr>
<tr>
<td>Lls_CG08</td>
<td>-0.147</td>
<td>-0.159</td>
<td>-0.160</td>
<td>0.152</td>
<td>0.131</td>
<td>0.130</td>
<td>0.362</td>
<td>0.388</td>
<td>0.355</td>
<td>0.326</td>
<td>0.171</td>
</tr>
<tr>
<td>Lls_CG09</td>
<td>0.092</td>
<td>0.102</td>
<td>0.107</td>
<td>-0.043</td>
<td>-0.011</td>
<td>-0.001</td>
<td>0.394</td>
<td>0.378</td>
<td>0.394</td>
<td>0.409</td>
<td>0.136</td>
</tr>
<tr>
<td>Lls_CG10</td>
<td>0.092</td>
<td>0.070</td>
<td>0.068</td>
<td>0.046</td>
<td>0.042</td>
<td>0.040</td>
<td>0.279</td>
<td>0.274</td>
<td>0.302</td>
<td>0.330</td>
<td>0.259</td>
</tr>
<tr>
<td>Lls_CG11</td>
<td><strong>0.256</strong></td>
<td><strong>0.260</strong></td>
<td><strong>0.260</strong></td>
<td>-0.041</td>
<td>-0.033</td>
<td>-0.030</td>
<td>0.280</td>
<td>0.364</td>
<td><strong>0.488</strong></td>
<td>0.268</td>
<td>0.186</td>
</tr>
<tr>
<td>Lls_CG12</td>
<td>0.077</td>
<td>0.069</td>
<td>0.068</td>
<td>0.083</td>
<td>0.116</td>
<td>0.131</td>
<td>0.327</td>
<td>0.221</td>
<td>0.249</td>
<td>0.257</td>
<td>0.345</td>
</tr>
<tr>
<td>Lls_ANM_18_1</td>
<td>0.016</td>
<td>0.028</td>
<td>0.031</td>
<td><strong>0.172</strong></td>
<td>0.099</td>
<td>0.091</td>
<td>0.472</td>
<td>0.249</td>
<td>0.321</td>
<td>0.215</td>
<td>0.247</td>
</tr>
<tr>
<td>Lls_ANM_20_2</td>
<td>0.055</td>
<td>0.015</td>
<td>0.008</td>
<td>-0.121</td>
<td>-0.026</td>
<td>-0.052</td>
<td>0.158</td>
<td>0.180</td>
<td>0.305</td>
<td>0.225</td>
<td>0.398</td>
</tr>
<tr>
<td>Lls_ANM_22_2</td>
<td>-0.103</td>
<td>-0.064</td>
<td>-0.050</td>
<td>-0.082</td>
<td>-0.060</td>
<td>-0.051</td>
<td>0.447</td>
<td>0.460</td>
<td>0.457</td>
<td>0.465</td>
<td>0.087</td>
</tr>
<tr>
<td>Lls_ANM_6_1</td>
<td>0.104</td>
<td>0.088</td>
<td>0.087</td>
<td>-0.096</td>
<td>-0.060</td>
<td>-0.061</td>
<td>0.320</td>
<td>0.218</td>
<td>0.255</td>
<td>0.249</td>
<td>0.348</td>
</tr>
</tbody>
</table>

*: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$
Table 5: SNPs significantly associated with parasite load based on individual-level generalised estimating equations (GEE). SNP mutation type and Wright’s inbreeding coefficient (\(F_{IS}\) with significance indicated by asterisks) are presented alongside model coefficients and \(P\)-values for significant genotype comparisons, the presence of a particular allele and the presence of heterozygosity. Mean predicted effect size is expressed as relative change in parasite load. \(\Delta QICu\) indicates the changes of QICu compared to a model without the genetic term (positive \(\Delta QICu\) indicates improvement).

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNP</th>
<th>Type</th>
<th>(F_{IS})</th>
<th>Comparison</th>
<th>Coefficient ± SE</th>
<th>(P)-value</th>
<th>Effect</th>
<th>(\Delta QICu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lls_CG01</td>
<td>SNP 9</td>
<td>non-synonymous</td>
<td>-0.065</td>
<td>C/T vs. C/C</td>
<td>-0.894 ± 0.216</td>
<td>&lt;0.001</td>
<td>-59%</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele T</td>
<td>-0.894 ± 0.216</td>
<td>&lt;0.001</td>
<td>-59%</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.894 ± 0.216</td>
<td>&lt;0.001</td>
<td>-59%</td>
<td>103</td>
</tr>
<tr>
<td>Lls_CG02</td>
<td>SNP 2</td>
<td>untranslated</td>
<td>-0.157</td>
<td>C/C vs. A/A</td>
<td>1.087 ± 0.301</td>
<td>&lt;0.001</td>
<td>+196%</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A/C vs. A/A</td>
<td>0.784 ± 0.259</td>
<td>0.002</td>
<td>+119%</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele C</td>
<td>0.947 ± 0.283</td>
<td>0.001</td>
<td>+158%</td>
<td>39</td>
</tr>
<tr>
<td>Lls_CG03</td>
<td>SNP 5</td>
<td>untranslated</td>
<td>-0.064</td>
<td>A/G vs. A/A</td>
<td>0.354 ± 0.115</td>
<td>0.002</td>
<td>+42%</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele G</td>
<td>0.382 ± 0.113</td>
<td>0.001</td>
<td>+47%</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>0.331 ± 0.118</td>
<td>0.005</td>
<td>+39%</td>
<td>4</td>
</tr>
<tr>
<td>Lls_CG05</td>
<td>SNP 10</td>
<td>untranslated</td>
<td>0.344***</td>
<td>Allele G</td>
<td>0.472 ± 0.165</td>
<td>0.004</td>
<td>+60%</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>SNP 15</td>
<td>untranslated</td>
<td>0.372***</td>
<td>Heterozygosity</td>
<td>0.532 ± 0.166</td>
<td>0.001</td>
<td>+70%</td>
<td>29</td>
</tr>
<tr>
<td>Lls_CG06</td>
<td>SNP 1</td>
<td>synonymous</td>
<td>-0.095</td>
<td>C/T vs. T/T</td>
<td>-0.497 ± 0.166</td>
<td>0.003</td>
<td>-39%</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele C</td>
<td>-0.407 ± 0.166</td>
<td>0.003</td>
<td>-39%</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.407 ± 0.166</td>
<td>0.003</td>
<td>-39%</td>
<td>42</td>
</tr>
<tr>
<td>Lls_CG08</td>
<td>SNP 10</td>
<td>untranslated</td>
<td>0.088</td>
<td>G/G vs. A/A</td>
<td>1.078 ± 0.337</td>
<td>0.001</td>
<td>+194%</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A/G vs. A/A</td>
<td>1.022 ± 0.325</td>
<td>0.002</td>
<td>+178%</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele G</td>
<td>1.048 ± 0.318</td>
<td>0.001</td>
<td>+185%</td>
<td>68</td>
</tr>
<tr>
<td>Lls_CG12</td>
<td>SNP 3</td>
<td>synonymous</td>
<td>0.162</td>
<td>C/T vs. C/C</td>
<td>-0.711 ± 0.181</td>
<td>&lt;0.001</td>
<td>-51%</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele T</td>
<td>-0.586 ± 0.155</td>
<td>&lt;0.001</td>
<td>-44%</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.703 ± 0.180</td>
<td>&lt;0.001</td>
<td>-50%</td>
<td>15</td>
</tr>
<tr>
<td>Lls_ANM_18_1</td>
<td>SNP 1</td>
<td>untranslated</td>
<td>0.230*</td>
<td>C/G vs. C/C</td>
<td>-0.586 ± 0.190</td>
<td>0.002</td>
<td>-44%</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.669 ± 0.155</td>
<td>&lt;0.001</td>
<td>-49%</td>
<td>102</td>
</tr>
<tr>
<td>SNR7</td>
<td></td>
<td>untranslated</td>
<td>0.575***</td>
<td>Allele G</td>
<td>-0.617 ± 0.188</td>
<td>0.001</td>
<td>-46%</td>
<td>-33</td>
</tr>
<tr>
<td>Lls_ANM_20_2</td>
<td>SNP 5</td>
<td>untranslated</td>
<td>0.653***</td>
<td>Heterozygosity</td>
<td>-0.786 ± 0.246</td>
<td>0.001</td>
<td>-54%</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>SNP 6</td>
<td>untranslated</td>
<td>0.653***</td>
<td>C/T vs. C/C</td>
<td>-0.836 ± 0.286</td>
<td>0.004</td>
<td>-57%</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.876 ± 0.246</td>
<td>0.001</td>
<td>-54%</td>
<td>9</td>
</tr>
<tr>
<td>SNR7</td>
<td></td>
<td>untranslated</td>
<td>0.653***</td>
<td>A/G vs. A/A</td>
<td>-0.836 ± 0.286</td>
<td>0.004</td>
<td>-57%</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heterozygosity</td>
<td>-0.876 ± 0.246</td>
<td>0.001</td>
<td>-54%</td>
<td>9</td>
</tr>
<tr>
<td>Lls_ANM_6_1</td>
<td>SNP 1</td>
<td>untranslated</td>
<td>0.214</td>
<td>T/T vs. C/C</td>
<td>-0.588 ± 0.197</td>
<td>0.003</td>
<td>-44%</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allele C</td>
<td>0.505 ± 0.169</td>
<td>0.003</td>
<td>+66%</td>
<td>43</td>
</tr>
</tbody>
</table>

*: \(P \leq 0.05\); **: \(P \leq 0.01\); ***: \(P \leq 0.001\)
Table 6: $F_{ST}$-outlier test results using BAYESCAN2 with either multi-locus haplotype data or 500 replicates generated by drawing one random SNP from each locus. Haplotype-based $F_{ST}$ estimates are presented with an indication of statistical significance after false-discovery-rate correction for multiple testing ($q$-values). SNP-based $F_{ST}$ estimates and $q$-values are summarised as quantiles and proportions of replicates below two significance thresholds. Statistically significant $F_{ST}$ estimates and $q$-values are emboldened and annotated with asterisks.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Haplotypes</th>
<th>$F_{ST}$</th>
<th>25 %</th>
<th>Median</th>
<th>75 %</th>
<th>Median</th>
<th>$q \leq 0.1$</th>
<th>$q \leq 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lls_CG01</td>
<td>0.008***</td>
<td>0.035</td>
<td>0.048</td>
<td>0.112</td>
<td>0.032</td>
<td>62.4%</td>
<td>57.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG02</td>
<td>0.011***</td>
<td>0.033</td>
<td>0.035</td>
<td>0.039</td>
<td>0.004</td>
<td>99.8%</td>
<td>98.6%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG03</td>
<td>0.023***</td>
<td>0.066</td>
<td>0.109</td>
<td>0.123</td>
<td>0.196</td>
<td>99.8%</td>
<td>29.2%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG04</td>
<td>0.019***</td>
<td>0.040</td>
<td>0.042</td>
<td>0.044</td>
<td>0.012</td>
<td>98.8%</td>
<td>9.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG05</td>
<td>0.016***</td>
<td>0.079</td>
<td>0.105</td>
<td>0.136</td>
<td>0.219</td>
<td>21.2%</td>
<td>4.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG06</td>
<td>0.060</td>
<td>0.121</td>
<td>0.135</td>
<td>0.147</td>
<td>0.411</td>
<td>99.8%</td>
<td>6.2%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG07</td>
<td>0.020***</td>
<td>0.035</td>
<td>0.045</td>
<td>0.132</td>
<td>0.022</td>
<td>98.8%</td>
<td>60.0%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG08</td>
<td>0.013***</td>
<td>0.057</td>
<td>0.077</td>
<td>0.114</td>
<td>0.119</td>
<td>96.0%</td>
<td>27.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG09</td>
<td>0.063</td>
<td>0.113</td>
<td>0.127</td>
<td>0.136</td>
<td>0.328</td>
<td>44.8%</td>
<td>2.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG10</td>
<td>0.015***</td>
<td>0.033</td>
<td>0.104</td>
<td>0.128</td>
<td>0.184</td>
<td>39.4%</td>
<td>42.6%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG11</td>
<td>0.039*</td>
<td>0.050</td>
<td>0.052</td>
<td>0.057</td>
<td>0.043</td>
<td>33.8%</td>
<td>61.8%</td>
<td></td>
</tr>
<tr>
<td>Lls_CG12</td>
<td>0.032**</td>
<td>0.043</td>
<td>0.119</td>
<td>0.131</td>
<td>0.278</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>Lls_ANM_18_1</td>
<td>0.078</td>
<td>0.121</td>
<td>0.129</td>
<td>0.142</td>
<td>0.405</td>
<td>0.2%</td>
<td>34.4%</td>
<td></td>
</tr>
<tr>
<td>Lls_ANM_20_2</td>
<td>0.094</td>
<td>0.137</td>
<td>0.143</td>
<td>0.150</td>
<td>0.495</td>
<td>0.0%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>Lls_ANM_22_2</td>
<td>0.067</td>
<td>0.120</td>
<td>0.129</td>
<td>0.138</td>
<td>0.365</td>
<td>1.4%</td>
<td>0.0%</td>
<td></td>
</tr>
<tr>
<td>Lls_ANM_6_1</td>
<td>0.014***</td>
<td>0.068</td>
<td>0.113</td>
<td>0.127</td>
<td>0.206</td>
<td>22.0%</td>
<td>29.8%</td>
<td></td>
</tr>
</tbody>
</table>

*: $q \leq 0.05$; **: $q \leq 0.01$; ***: $q \leq 0.001$
Supplementary data

- Document S1: Description of 454 amplicon sequencing run used to supplement Sanger sequence data
- Document S2: Complete GEE and Hardy-Weinberg equilibrium test results
- Document S3: Summary plots of parasite load in genotypic variants of all SNPs significantly associated with parasite load

Data Accessibility

- Alignment consensus sequences: GENBANK accessions KM236217–KM236228
- Full alignments: DataDryad doi:TBC
- Haplotype reference panel for PHASE: DataDryad doi:TBC
- Phenotypic data: DataDryad doi:TBC

References


enigmatic amphipod species (*Eulimnogammarus cyaneus* and *E. verrucosus*) from Lake Baikal.


Delahay, R.J. & Moss, R. 1996. Food intake, weight changes and egg production in captive red
606 grouse before and during laying: effects of the parasitic nematode Trichostrongylus tenuis.

Delahay, R.J., Speakman, J.R. & Moss, R. 1995. The energetic consequences of parasitism:
effects of a developing infection of Trichostrongylus tenuis (Nematoda) on red grouse (Lagopus

Dellicour, S. & Mardulyn, P. 2014. SPADS 1.0: a toolbox to perform spatial analyses on DNA


Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled,
J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., S., S., Thierer, T. & Wilson, A.

Duncan, E.J., Gluckman, P.D. & Dearden, P.K. 2014. Epigenetics, plasticity, and evolution:
208–220.


from metric distances among DNA haplotypes: application to human mitochondrial DNA

Fisher, R.A. 1919. The Correlation between Relatives on the Supposition of Mendelian Inheri-


Foll, M. & Gaggiotti, O. 2006. Identifying the environmental factors that determine the genetic


affects recruitment success in young male red grouse. Oikos 90: 261–270.

Manel, S., Conord, C. & Després, L. 2009. Genome scan to assess the respective role of host-
plant and environmental constraints on the adaptation of a widespread insect. BMC Evol. 
Biol. 9: 288.

Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenberg, M.S., Scribner, 
K.T., Bonin, A. & Fortin, M.J. 2010. Perspectives on the use of landscape genetics to detect 

Martínez-Padilla, J., Redpath, S.M., Zeineddine, M. & Mougeot, F. 2014. Insights into popu-
83: 85–98.

Martinez-Padilla, J., Vergara, P., Mougeot, F. & Redpath, S.M. 2012. Parasitized mates increase 

Martinez-Padilla, J., Vergara, P., Pérez-Rodríguez, L., Mougeot, F., Casas, F., Ludwig, S., 

Testing the interactive effects of testosterone and parasites on carotenoid-based ornamentation 

among Chinook salmon populations in the Columbia River Basin. Trans. Amer. Fish. Soc. 
140: 783–807.


host tolerance and resistance inferred from a wild host–parasite system. Proc. R. Soc. Lond. 


39


