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 (*Trichostrongylus 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 results 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, parasite 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 phenotypic characters of ecological and adaptive importance properly known, which precludes the identification of appropriate genomic targets through which gene dynamics in relation to adaptation 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 character 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

58 gene expression between individuals or populations that either differ naturally in the character
59 of interest or are subject to experimental intervention can highlight functionally relevant loci
60 that then become targets to be screened for SNPs (e.g., Orsini et al., 2011; Webster et al., 2011a;
61 Wang et al., 2012; De Wit & Palumbi, 2013; Gossner et al., 2013). Further, genome-wide scans
62 for locus-specific signatures of selection without *a priori* assumptions of phenotypic links or
63 causal environmental factors can reveal outlier loci that are then examined for associations with
64 environmental factors or phenotypic characters (e.g., Manel et al., 2009; Hess & Narum, 2011;
65 Matala et al., 2011; Pespeni & Palumbi, 2013; Milano et al., 2014). Similarly, locus-specific
66 genetic differentiation among samples of individuals that differ in the character of interest may
67 indicate adaptive significance with respect to character divergence, irrespective of the magnitude
68 of differentiation relative to the whole genome (e.g., Brown et al., 2013).

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

84 Here, we inform this debate by characterising a suite of novel candidate genes that were
85 developed using a top-down strategy to examine the genetic basis of an exemplary complex
86 phenotype, namely host response to chronic parasite burden. As a model, we use the interaction
87 between red grouse (*Lagopus lagopus scotica* Latham) and its primary parasite, the gastroin-
88 testinal nematode *Trichostrongylus tenuis* Mehlis. *L. l. scotica* is an economically important
89 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

122 number of differentially expressed genes, some of which are involved in immune system pro-
123 cesses and unspecific tissue damage responses (Webster et al., 2011a,b). Third, an exploratory
124 genome-wide epigenomics study has highlighted an epigenetic signature of parasite load at spe-
125 cific genomic regions across a network of wild grouse populations (Wenzel & Piertney, 2014).
126 In concert, these studies demonstrate detectable functional and genetic effects associated with
127 parasite infection and parasite burden in the red grouse system. However, the key question
128 remaining to be answered is whether genotypic variation at such genes explains variance in
129 parasite load in natural red grouse populations.

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

143 Materials and Methods

144 Study system

145 Our study system consists of 21 sampling sites (hereafter: populations) in a well-studied land-
146 scape of grouse moors in north-east Scotland near Deeside, Aberdeenshire (Fig. 1; Wenzel &
147 Piertney, 2014). One common management action on these grouse moors is parasite control to
148 improve grouse fitness and population growth (Martínez-Padilla et al., 2014). This is achieved
149 by dispensing quartz grit coated with an anthelmintic drug across the moors and allowing grouse
150 to self-medicate by ingesting this medicated grit alongside natural grit during normal feeding
151 behaviour (Newborn & Foster, 2002; Webster et al., 2008; Cox et al., 2010). Among the sites

152 sampled for the present study, medicated grit had been used for at least two years at fifteen
153 sites, and not been used for at least 10 years at six sites (Table 1). As such, the presence or
154 absence of medicated grit is an important confounding factor expected to impact parasite load.
155 Morphological measurements and tissue biopsies were taken from shot grouse following driven
156 or walked-up sporting shoots in autumn 2012. As grouse populations on these moors typically
157 display a degree of social and genetic structure in space and time, conferred by philopatry and
158 territoriality by males (Watson et al., 1994; MacColl et al., 2000; Piertney et al., 1998, 1999,
159 2000, 2008), individuals were aged ("young": < 1 year; "old": > 1 year) and, where possible,
160 old birds were preferentially sampled to minimise sampling bias through over-representation of
161 kin groups (Wenzel & Piertney, 2014). As measures of physiological condition, body weight
162 was measured to the nearest 10 g with a spring balance and supra-orbital comb size (width and
163 length) was measured to the nearest mm. Carotenoid-based supra-orbital combs in both males
164 and female grouse act as testosterone-dependent signals that are modulated by parasite load
165 through impact on immune function (Mougeot & Redpath, 2004; Mougeot et al., 2004; Mougeot,
166 2008), oxidative status (Mougeot et al., 2009, 2010a) or physiological stress (Bortolotti et al.,
167 2009; Mougeot et al., 2010b). As a result, comb size reflects an interaction between condition,
168 age, sex and, to some extent, parasite load (Mougeot et al., 2004, 2005, 2009; Martínez-Padilla
169 et al., 2010; Martinez-Padilla et al., 2011; Vergara et al., 2012a,b; Wenzel & Piertney, 2014).
170 Liver samples were taken for DNA extraction and caecum samples were taken for parasite
171 load estimation from faecal parasite egg counts using the standard McMaster chamber slide
172 method (Seivwright et al., 2004). DNA was extracted from 2–3 c. 2 mm^3 shreds of liver tissue
173 following the method of Hogan et al. (2008). Each bird was sexed genetically by amplification
174 and electrophoresis of the gonosome-linked CHD genes (Griffiths et al., 1998) as described in
175 Wenzel et al. (2012).

176 DNA sequencing and assembly

177 We selected 12 candidate genes for response to parasite infection from a suite of genes that
178 were previously developed for red grouse from transcriptomic and genomic data (Lls_CG01–
179 Lls_CG12; Wenzel et al., 2014). These genes are located in seven chromosomes in the chicken
180 genome (Wenzel et al., 2014) and capture a broad physiological response, including immune
181 system, xenobiotic detoxification, oxidative stress and metabolism processes (Table 2). Ge-
182 netic 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 adaptive 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 disequilibrium (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 (*Exo*I) 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 ºC for 45 min followed by enzyme heat deactivation at 80 ºC 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 sequences 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 &

214 Rozas, 2009). Global population differentiation was estimated with G_{ST} and N_{ST} statistics using
215 the software SPADS (Dellicour & Mardulyn, 2014) with 1,000 permutations to obtain estimates
216 of statistical significance. To examine signatures of natural selection, loci with disproportionate
217 population differentiation (F_{ST}-outliers) were identified using BAYESCAN2 (Foll & Gaggiotti,
218 2008) with haplotype data as input. Additionally, to test whether haplotype reconstruction
219 introduced bias, BAYESCAN was then run on 500 dataset replicates that were created by drawing
220 one random SNP from each locus, thus retaining independence among loci. All runs consisted
221 of 10^5 iterations with a thinning interval of 20 after 20 pilot runs (10^4 iterations each) and a
222 burn-in of $5 \cdot 10^5$. Outliers were selected at a significance threshold of $q \leq 0.05$.

223 **Population-level analysis**

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

232 Locus-by-locus hierarchical analysis of molecular variance (AMOVA; Excoffier et al., 1992)
233 was used to test whether groups of populations with similar median parasite loads are signifi-
234 cantly genetically differentiated. The AMOVA method partitions total genetic variance across
235 three hierarchical components: among groups of populations (φ_{CT}), among populations within
236 groups (φ_{SC}) and within populations (φ_{ST}). The 21 populations (sampling sites) were divided
237 into three groups of seven populations with broadly similar median parasite loads (4, 100–800,
238 >800 worms per bird; Figure 1). Point estimates and statistical significance for φ_{CT}, φ_{SC} and
239 φ_{ST} were obtained using SPADS with 1,000 permutations.

240 Relationships between population genetic differentiation and median parasite load, corrected
241 for isolation by distance and anthelmintic medication regime at sampling locations, were exam-
242 ined using isolation-by-stressor analysis based on non-parametric partial Mantel tests (Smouse
243 et al., 1986) and a Bayesian parametric regression method that allows for testing the effect of
244 multiple combinations of predictors on genetic differentiation (Foll & Gaggiotti, 2006). Pairwise

245 differentiation statistics among populations (G_{ST}, G'_{ST}, D_{est}) were computed for each locus in-
246 dividually and also combining all candidate genes or all control loci, using the package *diveRsity*
247 (Keenan et al., 2013) in R 3.0.3 (R Core Team, 2014). Relationships between patterns of popu-
248 lation differentiation between candidate genes and control loci were examined using scatter plots
249 and linear models. Locus-by-locus Mantel tests with 9,999 permutations were then carried out to
250 estimate the correlation between pairwise population matrices of linearized genetic differentia-
251 tion and either logarithmic geographic distance (= isolation by distance) or median parasite load
252 conditioned by binary differences in medication regime (0: same regime, 1: different regime) and
253 logarithmic geographic distance (= isolation by stressor), using the R package *ecodist* (Goslee
254 & Urban, 2007). Finally, Bayesian locus-by-locus models that regress population-specific local
255 F_{ST} estimates (Gaggiotti & Foll, 2010) on all possible combinations of median parasite load and
256 three covariates (longitude, latitude and presence/absence of medicated grit) were fitted using
257 the software GESTE (Foll & Gaggiotti, 2006). The posterior likelihoods of all model were then
258 used to identify those predictor combinations that best explain variation in genetic differentia-
259 tion. All models comprised a run length of 10^5 with a thinning interval of 20, following 20 pilot
260 runs of 10^4 iterations each and a burn-in of $5 \cdot 10^5$.

261 **Individual-level analysis**

262 Given the substantial heterogeneity of parasite load within populations, we then focussed on
263 individual-based generalized linear model analysis of parasite load and genotypic variants while
264 accounting for additional factors that confound this relationship (e.g., Manel et al., 2009; Oliver
265 et al., 2009; Sin et al., 2014; Wenzel & Piertney, 2014). In red grouse, three confounding
266 factors are important to address. First, parasite load may be correlated within sampling sites
267 (epidemiological neighbourhood effect; Hubbard et al., 2010), due to transmission dynamics
268 conferred by a shared environment and social structuring of kin groups (Piertney et al., 1998,
269 1999; Martinez-Padilla et al., 2012). Second, parasite control through medicated grit reduces
270 typical parasite load compared to sites where parasites are not controlled (Newborn & Foster,
271 2002). Third, individual parasite loads are expected to vary with sex, age and environment-
272 dependent physiological condition (Mougeot et al., 2004, 2005, 2009; Martínez-Padilla et al.,
273 2010; Martinez-Padilla et al., 2011; Vergara et al., 2012a,b).

274 To address these factors, parasite load was modelled as an over-dispersed Poisson-distributed
275 variable in a generalised estimating equations framework (GEE) using the R package *geepack*

(Halekoh et al., 2006). 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 χ^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 (β_1) were linearised (e^{β_1}) and expressed as a percentage change in parasite load. Absolute effect sizes were calculated

308 from differences in predicted least-square population mean estimates between genotypic variants
309 when keeping all other variables constant at mean values, using the R package *doBy* (Højsgaard
310 & Halekoh, 2013). To provide an indication of how well genotypic variation explains parasite
311 load, the difference between the quasi-likelihood based information criterion (QICu; Pan, 2001)
312 estimates was calculated between the full model and an equivalent model omitting the genetic
313 term.

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

321 Results

322 Estimated parasite loads among the 173 individuals ranged from 4 to 9283 worms per bird and
323 population medians ranged from 4 to 2236 worms per bird (Table 1). Across all loci, between
324 98 and 173 individuals were sequenced at more than 50 % of all polymorphic sites. The twelve
325 candidate genes were sequenced in 125–173 individuals and contained 2–17 polymorphic sites
326 that defined between 4–63 reconstructed haplotypes (Table 3). Sequencing failure was higher for
327 control loci, where 98–171 individuals were successfully sequenced. Those sequences contained
328 6–15 polymorphic sites and defined 8–51 reconstructed haplotypes (Table 3). Estimated haplo-
329 type reconstruction fidelity based on small reference haplotype panels was 80±18 %, indicating
330 some uncertainty for some loci caused by few SNPs with disproportionate amounts of miss-
331 ing genotypes. Consensus sequences for all loci with genic annotations and SNP locations are
332 available from GENBANK accessions KM236217–KM236228 (candidate genes) and KP210037–
333 KP210040 (control loci). Significant deviation from the neutral equilibrium population model
334 was detected in four candidate genes and three control loci, coinciding with significant genetic
335 differentiation among populations (Table 3). No evidence for linkage disequilibrium among the
336 16 loci was detected based on reconstructed haplotypes, indicating that the loci can be treated
337 as independent.

338 Population genetics analyses provided weak and inconsistent evidence for association be-
339 tween genetic differentiation and parasite load. Following hierarchical AMOVA, population
340 groups with similar median parasite loads were marginally significantly differentiated at locus
341 Lls_CG11 ($\phi_{CT} = 0.015$; $P = 0.053$), but not at any other locus (Table 3). Pairwise genetic
342 differentiation estimates were typically higher for control loci (G_{ST}=-0.032–0.123) compared to
343 candidate genes (G_{ST}=-0.015–0.027), even within population groups with similar parasite loads
344 (Fig. 2), highlighting a discrepancy between neutral and adaptive genetic structure. When ac-
345 counting for isolation by distance and differences in anthelmintic medication regime, differences
346 in parasite load were consistently associated with three measures of genetic differentiation for
347 two loci (Table 4), though only locus Lls_CG05 remained significant after correction for multiple
348 testing using the false-discovery-rate method. Finally, Bayesian regression of population-specific
349 local F_{ST} estimates highlighted parasite load as a significant predictor (alongside latitude) for lo-
350 cus Lls_CG06 only. For most loci, no combination of predictors was more likely than a constant
351 model, though medicated grit was a significant predictor for two loci (Table 4).

352 In contrast, individual-level GEE models provided consistent evidence for several positive
353 and negative associations between genotypic variants and parasite load in nine candidate genes,
354 including those highlighted by population genomics analyses. Of 131 SNPs in total, 19 were
355 excluded because they were monomorphic after removing factor levels with frequencies below
356 0.05. All loci were represented by at least two SNPs after exclusion (supplementary materials
357 S2). Parasite load was significantly ($q \leq 0.1$) associated with eight genotypes, seven alleles and
358 five heterozygosity states in candidate gene SNPs, and with four genotypes, two alleles and four
359 heterozygosity states in control locus SNPs (Fig. 3). Taken together, these associations cover
360 eight SNPs in seven out of twelve candidate genes and six SNPs in three out of four control loci
361 (Table 5). Of all SNPs, 70 (53 %) did not significantly deviate from Hardy-Weinberg equilib-
362 rium, 55 (42 %) displayed significant heterozygote deficiency and 6 (5 %) displayed significant
363 heterozygote excess (supplementary materials S2). Haplotype-based analysis was impeded by
364 low sample sizes for most haplotypes (only 86 out of 386 haplotypes were tested), but highlighted
365 significant negative associations for four haplotypes in four candidate genes, of which two genes
366 were not identified in SNP-based analysis (Lls_CG04 and Lls_CG09; supplementary materials
367 S2). Significant models consistently provided a substantially better fit than equivalent models
368 omitting the genetic term ($\Delta\mathrm{QICu} = 4 - 108$), with the exception of one control locus SNP
369 (Table 5). The relative predicted effect sizes of candidate gene SNP variants on parasite load

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

375 The observed associations in candidate genes were predominantly driven by the presence of
376 particular alleles rather than heterozygosity. Although congruent associations across genotypic
377 variants within SNPs did not always allow for separating allele effects from heterozygosity effects
378 (particularly when not all genotypes were sampled; supplementary materials S3), there was no
379 evidence of deviation from Hardy-Weinberg equilibrium in these cases (Table 5). The single
380 exception is Lls_CG05 SNP 15, which displayed heterozygote deficiency and a positive associ-
381 ation between heterozygosity and parasite load, consistent with heterozygote disadvantage. In
382 all cases where an individual allele had a significant effect, the complementary allele either had
383 the opposite but non-significant effect or the model was not run because of low allele frequency
384 (<0.05) (supplementary materials S2). In spite of these highlighted, well supported individual-
385 based associations between alleles and parasite load, these same alleles were not associated with
386 parasite load on a population level. The population-level frequencies of the seven identified
387 candidate gene alleles ranged from 0.00 to 0.44, and no allele was significantly associated with
388 population median parasite load (Fig. 5), though allele C at Lls_CG06 SNP 1 displayed a
389 marginally significant negative association ($\beta_1 = -7.01$; $P = 0.10$), consistent with its negative
390 effect in individual-level GEE models (Table 5).

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

402 (Table 3).

403 Discussion

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

412 Across all analyses, significant associations between parasite load and genotypic variants in
413 at least one SNP or haplotype were detected in nine out of twelve candidate genes. These nine
414 genes are involved in a range of physiological functions that effectively represent a multi-factorial
415 response to long-term parasite insult (Thomas et al., 2010; Poulin, 2013). CYP2K4, GSTK1 and
416 UGT1A1 represent key modification and conjugation enzymes involved in the two main stages
417 of xenobiotics metabolism and detoxification (Guillemette, 2003; Tew & Townsend, 2012; Bock,
418 2014). In infected red grouse, such xenobiotics may originate from primary *T. tenuis* infection
419 or from secondary pathogen infection facilitated through caecal damage and haemorrhaging
420 (Watson et al., 1987). Moreover, CYP2K4 and GSTK1 enzymes are involved in producing
421 and regulating reactive oxygen species, respectively (Lewis, 2002; Symons & King, 2003; Hellou
422 et al., 2012), and may therefore be involved in parasite defence during an immune response or
423 response to oxidative stress caused by other parasite defence mechanisms (Mougeot et al., 2009,
424 2010a). GAL9 is a chicken homologue of β-defensin 9 involved in innate antimicrobial immune
425 responses and interactions with adaptive immune system processes (Mukhopadhyaya et al.,
426 2010; Semple & Dorin, 2012). Further adaptive immune system components are represented
427 by TCB and MFI2, a chicken homologue of the EOS47 eosinophil surface antigen (McNagny
428 et al., 1996; Rahmanto et al., 2012). SUMO3 may be involved in regulation of cell cycle and
429 gene expression, possibly as a response to physiological stress (Yang & Paschen, 2009; Sang
430 et al., 2011). Similarly, CAPRIN1 and CCNL1 are principally involved in cell-cyle control
431 (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; Lochmiller & 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),

464 interferon gamma (Coltman et al., 2001; Stear et al., 2007), Toll-like receptors (Downing et al., 465 2010; Gavan et al., in press) and cytokines (Luikart et al., 2008; Downing et al., 2010; Turner 466 et al., 2012). Landscape-scale application of novel top-down candidate genes for parasite suscep- 467 tibility in *Daphnia* revealed differential patterns of selection at these genes among populations 468 in different stress environments, but no conclusive associations between allele frequencies and 469 parasite prevalence were found (Orsini et al., 2011, 2012). Susceptibility to nematode infection 470 in sheep species has previously been found to be associated with alleles in the MHC and inter- 471 feron gamma genes (Stear et al., 2007), and also with heterozygosity in a T-cell receptor gene 472 (TCRG4) and a cytokine regulator (ADCYAP1) (Luikart et al., 2008), but none of these genes 473 explained nematode abundance in a large-scale study on 960 Soay sheep (Brown et al., 2013). 474 The same study also tested a panel of 123 candidate SNPs derived from genomic and top-down 475 transcriptomic studies on various sheep species, but revealed no more significant associations 476 with nematode abundance than expected by chance (Brown et al., 2013).

477 Identifying novel candidate genes for the genetic basis of complex phenotypes can be ham- 478 pered by relying on transcriptomic assays among different phenotypes to establish functional 479 relevance (e.g., Orsini et al., 2011; Pemberton et al., 2011; Wang et al., 2012; De Wit & Palumbi, 480 2013; Gossner et al., 2013). This is because phenotypic diversity may be mediated by *trans*- 481 regulated variation in gene transcript number rather than sequence polymorphisms in transcripts 482 or tightly linked *cis*-regulatory elements (Stern & Orgogozo, 2008; Duncan et al., 2014). Inte- 483 grating transcriptomic data with genomic assays, as we did for identifying the candidate genes 484 for red grouse (Wenzel et al., 2014), or focussing on genome scans alone (e.g., Manel et al., 2009; 485 Hess & Narum, 2011; Matala et al., 2011; Pespeni & Palumbi, 2013; Milano et al., 2014) is likely 486 to reduce the risk of false positive identification. Great potential also lies in shifting focus from 487 targeting sequence polymorphisms to gene dynamics and epigenetics. For example, Schneider 488 et al. (2014) recently demonstrated a purely transcriptomic application of the candidate gene 489 approach to reveal a transcriptomic basis for diet-induced morphological plasticity in cichlids. 490 The candidate genes were previously discovered using large-scale comparative transcriptomics 491 of divergent phenotypes following diet manipulation (Gunter et al., 2013). This approach avoids 492 the conceptual disjoint between the genome and the transcriptome and is likely to become an 493 important complement to whole-genome population transcriptomics approaches (Ouborg et al., 494 2010; Debes et al., 2012; Matzkin, 2012). In the same vein, genome scans for epigenetic patterns 495 such as cytosine methylation may well facilitate identification of candidate regulatory regions

496 that are associated with phenotypic responses to environmental factors (Duncan et al., 2014;
497 Wenzel & Piertney, 2014).

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

517 In consequence, population genomics signatures may be difficult to detect for complex poly-
518 genic phenotypes unless the phenotype is underpinned by at least some large-effect polymor-
519 phisms under strong selection pressure. Classic systems such as MC1R alleles that are perfectly
520 associated with melanism in mouse populations (Hoekstra et al., 2004; Mullen & Hoekstra, 2008)
521 are unlikely to be the norm for most population genomics scenarios given Fisher's infinitesimal
522 model (Fisher, 1919). Nevertheless, our results show that this does not preclude the detection
523 of fine-scale patterns among individuals. As such, an individual-based landscape genomics ap-
524 proach to link alleles with environmental or phenotypic variables is likely to be more powerful in
525 detecting adaptive genetic discontinuities on continuous landscapes than population genomics
526 approaches (Joost et al., 2007, 2013). Landscape genomics remains a challenging field for the
527 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.

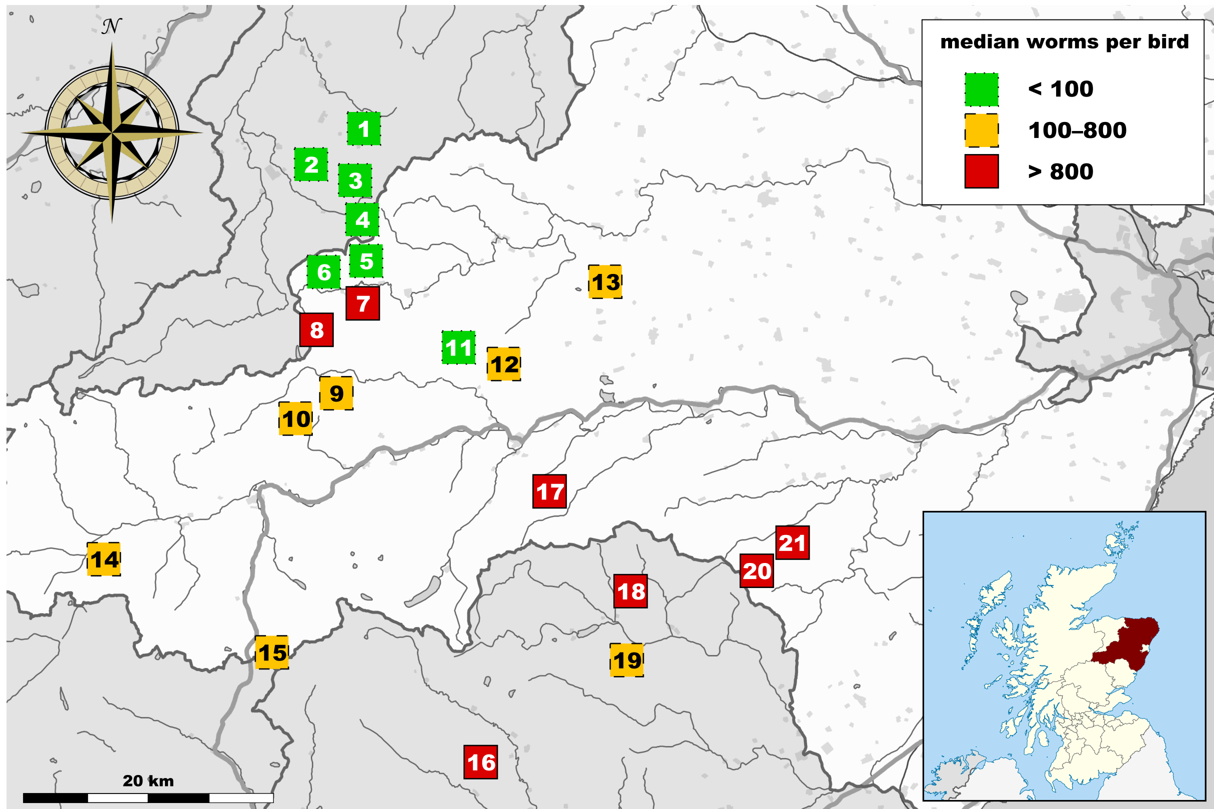

Figure 1: Sites in Aberdeenshire, Angus and Moray that were sampled following grouse sporting shoots in autumn 2012. Median parasite load for each site is indicated by three categories. Detailed locations, sample sizes and parasite loads are presented in table 1.

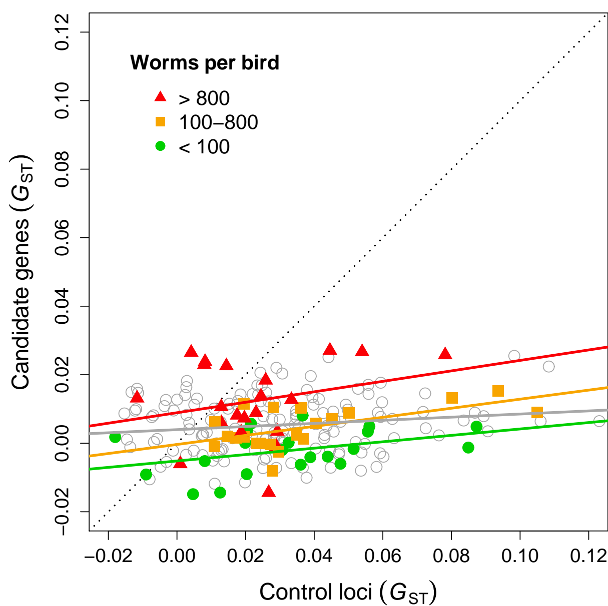

Figure 2: Relationships between genetic differentiation (G_{ST}) among population pairs estimated across candidate genes or control loci. Each data point represents one population pair. Solid lines indicate linear relationships for all data points (grey line) and three subsets representing population 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).

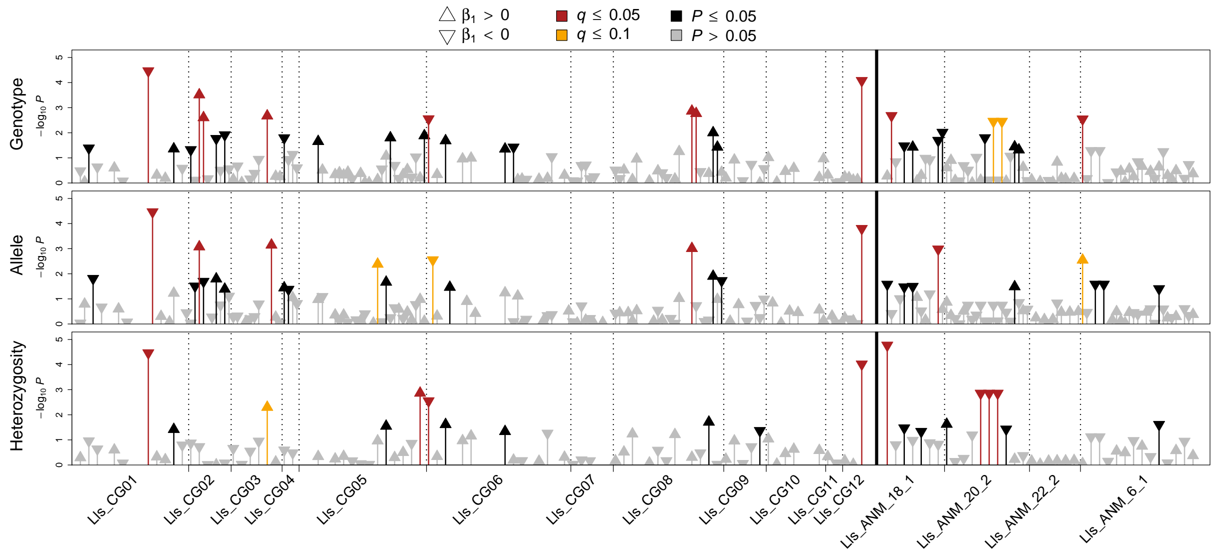

Figure 3: Graphical summary of regression model coefficients and statistical significance of generalised estimating equations (GEE) predicting parasite load by genotypic variation (diploid SNP genotype or presence or absence of an individual allele or heterozygosity) at 131 SNPs in twelve candidate genes 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 ($-\log_{10}$ P-value). Line colours indicates statistical significance levels before and after correction for multiple testing using the false-discovery-rate approach (red: $q \leqq 0.05$, orange: $q \leqq 0.1$ and black: $P \leqq 0.05$). Line symbols indicate the sign of the regression coefficient ("up": positive, "down": negative).

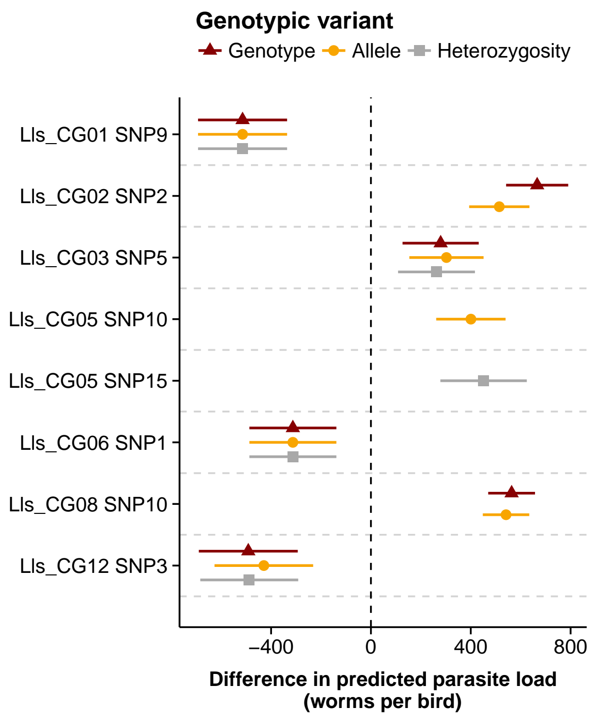

Figure 4: Predicted effect sizes (mean ± SE) of genotypic variants in candidate gene SNPs significantly associated with parasite load. Estimates were derived from differences in GEE-predicted least-square population means between genotypic variants when keeping all other variables constant at mean values. When two genotype effects were significant, only the largest effect is plotted.

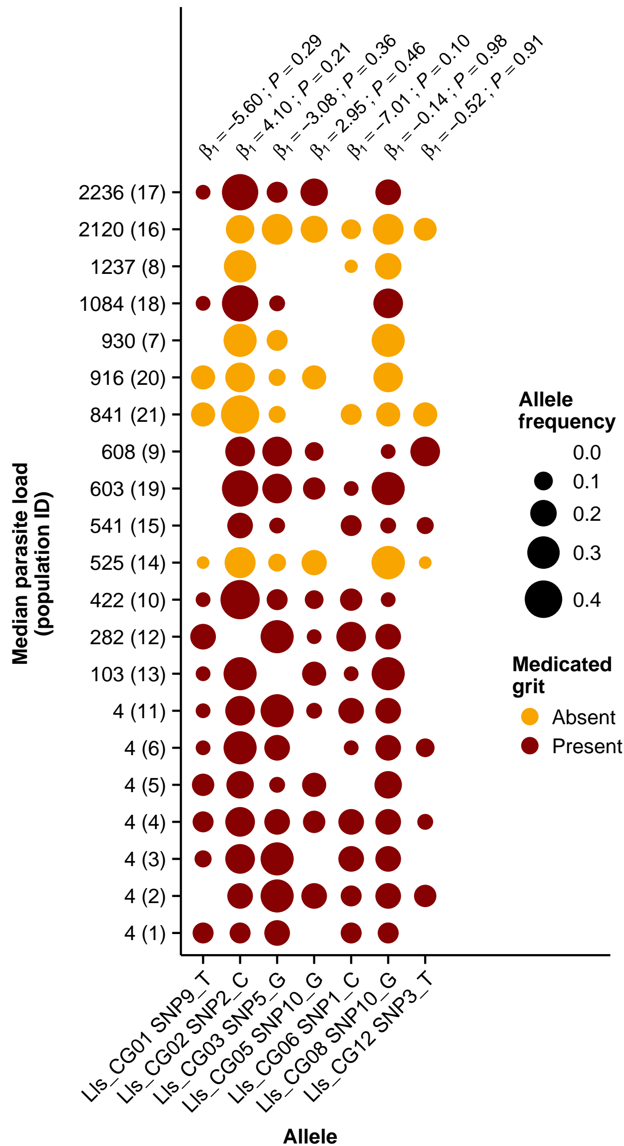

Figure 5: Population-level associations between median parasite load and frequencies of seven candidate gene SNP alleles with significant associations in individual-level GEE models. Populations are ordered by median parasite load. Allele frequencies are represented by circle area and colours indicate 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.

Sampling locations					Sample sizes				Worms per bird		
Site	Estate	Long.	Lat.	Anthelmintic grit	Total	M	F	Y	25 %	Median	75 %
1	Glenlivet	57.29	−3.18	Yes	8	4	4	0	4	4	981
2	Glenlivet	57.25	−3.28	Yes	8	7	1	0	4	4	36
3	Edinglassie	57.24	−3.20	Yes	8	6	2	0	4	4	4
4	Edinglassie	57.21	−3.19	Yes	8	7	1	0	4	4	4
5	Allargue	57.19	−3.29	Yes	7	4	3	0	4	4	41
6	Allargue	57.19	−3.23	Yes	8	6	2	8	4	4	4
7	Delnadamph	57.16	−3.26	No	8	5	3	0	348	930	1616
8	Delnadamph	57.14	−3.30	No	10	9	1	0	937	1237	1837
9	Invercauld	57.10	−3.29	Yes	8	3	5	3	4	608	1826
10	Invercauld	57.08	−3.35	Yes	8	5	3	3	99	422	2856
11	Dinnet	57.12	−3.11	Yes	8	8	0	0	4	4	78
12	Dinnet	57.11	−3.06	Yes	8	6	2	0	100	282	682
13	Tillypronie	57.18	−2.94	Yes	8	3	5	7	4	103	380
14	Mar Lodge	56.95	−3.66	No	11	6	5	4	264	525	1244
15	Invercauld	56.87	−3.40	Yes	8	8	0	0	146	541	1006
16	Airlie	56.81	−3.08	No	13	13	0	0	786	2120	2977
17	Glen Muick	56.99	−3.01	Yes	8	8	0	0	908	2236	3812
18	Invermark	56.94	−2.89	Yes	8	6	2	0	552	1084	1386
19	Invermark	56.89	−2.89	Yes	8	4	4	0	429	603	650
20	Glen Dye	56.95	−2.72	No	6	6	0	3	448	916	1448
21	Glen Dye	56.96	−2.69	No	6	6	0	2	358	841	1509
					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).

ID	Name	Descriptor	Putative function
Lls_CG01	TCB	T-cell receptor beta chain T17T-22	Innate immune response
Lls_CG02	CYP2K4	Cytochrome P450 2K4	Detoxification; oxidative balance
Lls_CG03	GAL9	Gallinacin-9	Innate immune response
Lls_CG04	GSTK1	Glutathione S-transferase kappa 1	Detoxification; oxidative balance
Lls_CG05	CAPRIN1	Caprin-1	Cell cycle; gene expression
Lls_CG06	UGT1A1	UDP-glucuronosyltransferase 1-1	Detoxification
Lls_CG07	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1	Signal transduction
Lls_CG08	MFI2	Melanotransferrin	Immune system
Lls_CG09	CCNL1	Cyclin-L1	Cell cycle; gene expression
Lls_CG10	SPCS2	Signal peptidase complex subunit 2	Peptide translocation
Lls_CG11	MIOS	WD repeat-containing protein mio	Protein modification
Lls_CG12	SUMO3	Small ubiquitin-related modifier 3	Cell cycle; gene expression

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 (D_{T}), Fu & Li's D (D_{FL}) and F (F_{FL}) test statistics for the equilibrium neutral population model. Global population differentiation is presented as G_{ST}, N_{ST} and hierarchical AMOVA statistics (φ_{CT} represents differentiation among the three population groups indicated in Figure 1). Significant statistics are emboldened and significance is indicated with asterisks.

			Polymorphic sites								Haplotypes		Neutral equilibrium tests			Population differentiation				
Locus	n	Size (bp)	P	Ti	Tv	S	NS	NC	k	π	H	Hd	D_{T}	D_{FL}	F_{FL}	G_{ST}	N_{ST}	φ_{CT}	φ_{SC}	φ_{ST}
Lls_CG01	171	312	13	10	3	5	8	0	3.18	0.010	58	0.910	1.325	0.024	0.629	-0.001	-0.006	-0.003	-0.002	-0.005
Lls_CG02	173	191	5	2	3	1	0	4	2.35	0.012	21	0.709	**3.619*****	0.939	**2.228****	-0.012	-0.016	-0.003	-0.012	-0.015
Lls_CG03	168	158	6	4	2	0	0	6	2.08	0.013	13	0.779	**2.310***	1.025	**1.766***	**0.021***	**0.046****	0.004	**0.040****	**0.044****
Lls_CG04	169	206	2	1	1	0	0	2	0.78	0.004	4	0.591	1.876	0.606	1.209	-0.008	-0.005	0.005	-0.007	-0.002
Lls_CG05	140	454	15	9	6	0	0	15	4.17	0.009	49	0.763	1.785	**1.550***	**1.977***	0.017	**0.037***	-0.007	**0.035***	**0.028***
Lls_CG06	166	336	17	10	7	8	9	0	2.49	0.007	63	0.707	-0.163	**1.622***	1.106	**0.082*****	**0.116*****	0.004	**0.109*****	**0.112*****
Lls_CG07	169	182	5	5	0	5	0	0	1.30	0.007	11	0.652	1.198	0.941	1.231	0.013	-0.005	-0.001	-0.002	-0.003
Lls_CG08	172	250	13	11	2	3	2	8	3.43	0.014	42	0.904	1.615	0.735	1.293	0.007	0.014	0.000	0.016	0.016
Lls_CG09	172	281	5	4	1	0	0	5	1.24	0.004	10	0.604	1.060	0.939	1.174	**0.048****	**0.052*****	0.002	**0.045****	**0.047****
Lls_CG10	173	226	7	6	1	5	0	2	1.23	0.005	15	0.677	0.252	1.098	0.957	0.000	-0.016	0.001	-0.016	-0.015
Lls_CG11	151	184	2	1	1	2	0	0	0.76	0.004	4	0.629	1.792	0.612	1.178	0.013	0.005	0.015	-0.006	0.009
Lls_CG12	125	124	3	3	0	3	0	0	0.91	0.007	5	0.666	1.320	0.757	1.117	-0.059	-0.027	-0.005	0.004	-0.001
Lls_ANM_18_1	133	371	7	4	3	0	0	7	2.30	0.006	20	0.877	2.100	1.118	**1.752***	**0.071*****	**0.106*****	-0.004	**0.098*****	**0.095*****
Lls_ANM_20_2	98	401	10	7	3	0	0	10	4.43	0.011	12	0.730	**3.728*****	1.331	**2.632****	**0.149*****	**0.223*****	-0.034	**0.199*****	**0.172*****
Lls_ANM_22_2	98	355	6	6	0	0	0	6	1.28	0.004	8	0.475	0.509	1.066	1.041	0.024	0.038	-0.010	**0.049***	0.040
Lls_ANM_6_1	171	405	15	13	2	0	0	15	4.75	0.012	51	0.878	**2.480***	0.883	**1.818***	0.009	**0.022***	0.006	**0.024***	**0.030****

*: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$

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.

	Isolation by distance (r)			Isolation by stressor (r)			GESTE (P)				
Locus	G_{ST}	G'_{ST}	D	G_{ST}	G'_{ST}	D	Lat.	Long.	Grit	Parasite load	Model
Lls_CG01	0.038	0.042	0.041	-0.065	-0.074	-0.074	0.297	0.394	0.331	0.321	0.205
Lls_CG02	0.001	0.008	0.011	-0.123	-0.120	-0.117	0.457	0.466	0.467	0.472	0.082
Lls_CG03	0.071	0.065	0.067	**0.205***	**0.195***	**0.194***	0.322	0.341	0.449	0.333	0.173
Lls_CG04	0.152	0.107	0.094	-0.002	0.012	0.014	0.113	0.071	**0.997**	0.171	0.684
Lls_CG05	-0.001	-0.037	-0.038	**0.251****	**0.305****	**0.311****	0.256	0.315	0.343	0.191	0.314
Lls_CG06	**0.173***	**0.237***	**0.249***	-0.016	-0.097	-0.092	**0.629**	0.147	0.198	**0.771**	0.395
Lls_CG07	0.095	0.112	0.113	-0.144	-0.159	-0.160	0.361	0.348	0.382	0.389	0.167
Lls_CG08	-0.147	-0.159	-0.160	0.152	0.131	0.130	0.362	0.388	0.355	0.326	0.171
Lls_CG09	0.092	0.102	0.107	-0.043	-0.011	-0.001	0.394	0.378	0.394	0.409	0.136
Lls_CG10	0.092	0.070	0.068	0.046	0.042	0.040	0.279	0.274	0.302	0.330	0.259
Lls_CG11	**0.256****	**0.260****	**0.260****	-0.041	-0.033	-0.030	0.280	0.364	**0.488**	0.268	0.186
Lls_CG12	0.077	0.069	0.068	0.083	0.116	0.131	0.327	0.221	0.249	0.257	0.345
Lls_ANM_18_1	0.016	0.028	0.031	**0.172***	0.099	0.091	0.472	0.249	0.321	0.215	0.247
Lls_ANM_20_2	0.055	0.015	0.008	-0.121	-0.026	-0.052	0.158	0.180	0.305	0.225	0.398
Lls_ANM_22_2	-0.103	-0.064	-0.050	-0.082	-0.060	-0.051	0.447	0.460	0.457	0.465	0.087
Lls_ANM_6_1	0.104	0.088	0.087	-0.096	-0.060	-0.061	0.320	0.218	0.255	0.249	0.348

*: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$

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. ΔQICu indicates the changes of QICu compared to a model without the genetic term (positive ΔQICu indicates improvement).

Locus	SNP	Type	F_{IS}	Comparison	Coefficient ± SE	P-value	Effect	ΔQICu
Lls_CG01	SNP 9	non-synonymous	-0.065	C/T vs. C/C	-0.894 ± 0.216	<0.001	-59%	103
				Allele T	-0.894 ± 0.216	<0.001	-59%	103
				Heterozygosity	-0.894 ± 0.216	<0.001	-59%	103
Lls_CG02	SNP 2	untranslated	-0.157	C/C vs. A/A	1.087 ± 0.301	<0.001	+196%	52
				A/C vs. A/A	0.784 ± 0.259	0.002	+119%	52
				Allele C	0.947 ± 0.283	0.001	+158%	39
Lls_CG03	SNP 5	untranslated	-0.064	A/G vs. A/A	0.354 ± 0.115	0.002	+42%	6
				Allele G	0.382 ± 0.113	0.001	+47%	11
				Heterozygosity	0.331 ± 0.118	0.005	+39%	4
Lls_CG05	SNP 10	untranslated	0.344***	Allele G	0.472 ± 0.165	0.004	+60%	108
	SNP 15	untranslated	0.372***	Heterozygosity	0.532 ± 0.166	0.001	+70%	29
Lls_CG06	SNP 1	synonymous	-0.095	C/T vs. T/T	-0.497 ± 0.166	0.003	-39%	42
				Allele C	-0.497 ± 0.166	0.003	-39%	42
				Heterozygosity	-0.497 ± 0.166	0.003	-39%	42
Lls_CG08	SNP 10	untranslated	0.088	G/G vs. A/A	1.078 ± 0.337	0.001	+194%	64
				A/G vs. A/A	1.022 ± 0.325	0.002	+178%	64
				Allele G	1.048 ± 0.318	0.001	+185%	68
Lls_CG12	SNP 3	synonymous	0.162	C/T vs. C/C	-0.711 ± 0.181	<0.001	-51%	15
				Allele T	-0.586 ± 0.155	<0.001	-44%	16
				Heterozygosity	-0.703 ± 0.180	<0.001	-50%	15
Lls_ANM_18_1	SNP1	untranslated	0.230*	C/G vs. C/C	-0.586 ± 0.190	0.002	-44%	116
				Heterozygosity	-0.669 ± 0.155	<0.001	-49%	102
	SNP7	untranslated	0.575***	Allele G	-0.617 ± 0.188	0.001	-46%	-33
Lls_ANM_20_2	SNP5	untranslated	0.653***	Heterozygosity	-0.786 ± 0.246	0.001	-54%	9
	SNP6	untranslated	0.653***	C/T vs. C/C	-0.836 ± 0.286	0.004	-57%	12
				Heterozygosity	-0.786 ± 0.246	0.001	-54%	9
	SNP7	untranslated	0.653***	A/G vs. A/A	-0.836 ± 0.286	0.004	-57%	12
				Heterozygosity	-0.786 ± 0.246	0.001	-54%	9
Lls_ANM_6_1	SNP1	untranslated	0.214	T/T vs. C/C	-0.588 ± 0.197	0.003	-44%	43
				Allele C	0.505 ± 0.169	0.003	+66%	43

*: $P \leq 0.05$; **: $P \leq 0.01$; ***: $P \leq 0.001$

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.

		SNPs					
		F_{ST}			q		
Locus	Haplotypes	25 %	Median	75 %	Median	$q \leq 0.1$	$q \leq 0.05$
Lls_CG01	**0.008*** **	0.035	0.048	0.112	**0.032**	62.4%	57.4%
Lls_CG02	**0.011*** **	0.033	0.035	0.039	**0.004**	99.8%	98.6%
Lls_CG03	**0.023*** **	0.066	0.109	0.123	0.196	29.2%	9.4%
Lls_CG04	**0.019*** **	0.040	0.042	0.044	**0.012**	99.8%	98.8%
Lls_CG05	**0.016*** **	0.079	0.105	0.136	0.219	21.2%	4.4%
Lls_CG06	0.060	0.121	0.135	0.147	0.411	8.2%	6.2%
Lls_CG07	**0.020*** **	0.035	0.045	0.132	**0.022**	60.0%	59.0%
Lls_CG08	**0.013*** **	0.057	0.077	0.114	0.119	44.8%	27.4%
Lls_CG09	0.063	0.113	0.127	0.136	0.328	2.4%	0.0%
Lls_CG10	**0.015*** **	0.033	0.104	0.128	0.184	42.6%	39.4%
Lls_CG11	**0.039* **	0.050	0.052	0.057	**0.043**	96.0%	61.8%
Lls_CG12	**0.032** **	0.043	0.119	0.131	0.278	34.4%	33.8%
Lls_ANM_18_1	0.078	0.121	0.129	0.142	0.405	0.2%	0.0%
Lls_ANM_20_2	0.094	0.137	0.143	0.150	0.495	0.0%	0.0%
Lls_ANM_22_2	0.067	0.120	0.129	0.138	0.365	1.4%	0.0%
Lls_ANM_6_1	**0.014*** **	0.068	0.113	0.127	0.206	29.8%	22.0%

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

554 # Supplementary data

555 • Document S1: Description of 454 amplicon sequencing run used to supplement Sanger
556 sequence data

557 • Document S2: Complete GEE and Hardy-Weinberg equilibrium test results

558 • Document S3: Summary plots of parasite load in genotypic variants of all SNPs signifi-
559 cantly associated with parasite load

560 # Data Accessibility

561 • Alignment consensus sequences: GENBANK accessions KM236217–KM236228

562 • Full alignments: DataDryad doi:TBC

563 • Haplotype reference panel for PHASE: DataDryad doi:TBC

564 • Phenotypic data: DataDryad doi:TBC

565 # References

566 Allendorf, F.W., Hohenlohe, P.A. & Luikart, G. 2010. Genomics and the future of conservation
567 genetics. *Nat. Rev. Genet.* **11**: 697–709.

568 Amos, W., Driscoll, E. & Hoffman, J. 2011. Candidate genes versus genome-wide associations:
569 which are better for detecting genetic susceptibility to infectious disease? *Proc. R. Soc. Lond.*
570 *B Biol. Sci.* **278**: 1183–1188.

571 Andrew, R.L., Bernatchez, L., Bonin, A., Buerkle, C.A., Carstens, B.C., Emerson, B.C., Garant,
572 D., Giraud, T., Kane, N.C., Rogers, S.M. et al. 2013. A road map for molecular ecology. *Mol.*
573 *Ecol.* **22**: 2605–2626.

574 Bashalkhanov, S., Eckert, A.J. & Rajora, O.P. 2013. Genetic signatures of natural selection in
575 response to air pollution in red spruce (Picea rubens, Pinaceae). *Mol. Ecol.* **22**: 5877–5889.

576 Bedulina, D., Evgen'ev, M., Timofeyev, M., Protopopova, M., Garbuz, D., Pavlichenko, V.,
577 Luckenbach, T., Shatilina, Z., Axenov-Gribanov, D., Gurkov, A. et al. 2013. Expression
578 patterns and organization of the hsp70 genes correlate with thermotolerance in two congener

579 endemic amphipod species (*Eulimnogammarus cyaneus* and *E. verrucosus*) from Lake Baikal.
580 *Mol. Ecol.* **22**: 1416–1430.

581 Benjamini, Y. & Hochberg, Y. 1995. Controlling the False Discovery Rate: a Practical and
582 Powerful Approach to Multiple Testing. *J. R. Statist. Soc. B* **57**: 289–300.

583 Biron, D.G. & Loxdale, H.D. 2013. Host–parasite molecular cross-talk during the manipulative
584 process of a host by its parasite. *J. Exp. Biol.* **216**: 148–160.

585 Bock, K.W. 2014. Homeostatic control of xeno-and endobiotics in the drug-metabolizing enzyme
586 system. *Biochem. Pharmacol.* **90**: 1–6.

587 Bortolotti, G.R., Mougeot, F., Martínez-Padilla, J., Webster, L.M.I. & Piertney, S.B. 2009.
588 Physiological stress mediates the honesty of social signals. *PLoS ONE* **4**: e4983.

589 Brown, E., Pilkington, J., Nussey, D., Watt, K., Hayward, A., Tucker, R., Graham, A., Paterson,
590 S., Beraldi, D., Pemberton, J. et al. 2013. Detecting genes for variation in parasite burden
591 and immunological traits in a wild population: testing the candidate gene approach. *Mol.*
592 *Ecol.* **22**: 757–773.

593 Chamary, J., Parmley, J.L. & Hurst, L.D. 2006. Hearing silence: non-neutral evolution at
594 synonymous sites in mammals. *Nat. Rev. Genet.* **7**: 98–108.

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

598 Cox, R., Newborn, D., Baines, D., Thomas, C.J. & Sherratt, T.N. 2010. No evidence for
599 resistance to fenbendazole in *Trichostrongylus tenuis*, a nematode parasite of the red grouse.
600 *J. Wildl. Manag.* **74**: 1799–1805.

601 De Wit, P. & Palumbi, S.R. 2013. Transcriptome-wide polymorphisms of red abalone (*Haliotis*
602 *rufescens*) reveal patterns of gene flow and local adaptation. *Mol. Ecol.* **22**: 2884–2897.

603 Debes, P.V., Normandeau, E., Fraser, D.J., Bernatchez, L. & Hutchings, J.A. 2012. Differences
604 in transcription levels among wild, domesticated, and hybrid Atlantic salmon (*Salmo salar*)
605 from two environments. *Mol. Ecol.* **21**: 2574–2587.

606 Delahay, R.J. & Moss, R. 1996. Food intake, weight changes and egg production in captive red
607 grouse before and during laying: effects of the parasitic nematode *Trichostrongylus tenuis*.
608 *Condor* **98**: 501–511.

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

612 Dellicour, S. & Mardulyn, P. 2014. SPADS 1.0: a toolbox to perform spatial analyses on DNA
613 sequence data sets. *Mol. Ecol. Res.* **14**: 647–651.

614 Downing, T., Lloyd, A.T., O'Farrelly, C. & Bradley, D.G. 2010. The differential evolutionary
615 dynamics of avian cytokine and TLR gene classes. *J. Immunol.* **184**: 6993–7000.

616 Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled,
617 J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., S, S., Thierer, T. & Wilson, A.
618 2014. Geneious v7.1.2. Available from http://www.geneious.com.

619 Duncan, E.J., Gluckman, P.D. & Dearden, P.K. 2014. Epigenetics, plasticity, and evolution:
620 How do we link epigenetic change to phenotype? *J. Exp. Zool. B Mol. Dev. Evol.* **322**:
621 208–220.

622 Ellegren, H. & Sheldon, B.C. 2008. Genetic basis of fitness differences in natural populations.
623 *Nature* **452**: 169–175.

624 Excoffier, L., Smouse, P.E. & Quattro, J.M. 1992. Analysis of molecular variance inferred
625 from metric distances among DNA haplotypes: application to human mitochondrial DNA
626 restriction data. *Genetics* **131**: 479–491.

627 Fisher, R.A. 1919. The Correlation between Relatives on the Supposition of Mendelian Inheri-
628 tance. *Trans. Roy. Soc. Edin.* **52**: 399–433.

629 Fitzpatrick, M.J., Ben-Shahar, Y., Smid, H.M., Vet, L.E., Robinson, G.E. & Sokolowski, M.B.
630 2005. Candidate genes for behavioural ecology. *Trends Ecol. Evol.* **20**: 96–104.

631 Foll, M. & Gaggiotti, O. 2006. Identifying the environmental factors that determine the genetic
632 structure of populations. *Genetics* **174**: 875–891.

633 Foll, M. & Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for
634 both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977–993.

635 Gaggiotti, O.E. & Foll, M. 2010. Quantifying population structure using the F-model. *Mol.*
636 *Ecol. Res.* **10**: 821–830.

637 Gauly, M. & Erhardt, G. 2001. Genetic resistance to gastrointestinal nematode parasites in
638 Rhön sheep following natural infection. *Vet. Parasitol.* **102**: 253–259.

639 Gavan, M.K., Oliver, M.K., A., D. & Piertney, S.B. in press. Gene dynamics of Toll-like receptor
640 4 through a population bottleneck in an insular population of water voles (*Arvicola amphibius*).
641 *Conserv. Gen.* .

642 Gómez-Díaz, E., Jordà, M., Peinado, M.A. & Rivero, A. 2012. Epigenetics of host–pathogen
643 interactions: the road ahead and the road behind. *PLoS Pathog.* **8**: e1003007.

644 Goslee, S.C. & Urban, D.L. 2007. The ecodist package for dissimilarity-based analysis of eco-
645 logical data. *J. Stat. Softw.* **22**: 1–19.

646 Gossner, A., Wilkie, H., Joshi, A. & Hopkins, J. 2013. Exploring the abomasal lymph node
647 transcriptome for genes associated with resistance to the sheep nematode Teladorsagia cir-
648 cumcincta. *Vet. Res.* **44**: 68.

649 Graham, M.H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* **84**:
650 2809–2815.

651 Griffiths, R., Double, M.C., Orr, K. & Dawson, R.J. 1998. A DNA test to sex most birds. *Mol.*
652 *Ecol. Res.* **7**: 1071–1075.

653 Guillemette, C. 2003. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes.
654 *Pharmacogenomics J.* **3**: 136–158.

655 Gunter, H.M., Fan, S., Xiong, F., Franchini, P., Fruciano, C. & Meyer, A. 2013. Shaping
656 development through mechanical strain: the transcriptional basis of diet-induced phenotypic
657 plasticity in a cichlid fish. *Mol. Ecol.* **22**: 4516–4531.

658 Halekoh, U., Højsgaard, S. & Yan, J. 2006. The R package geepack for generalized estimating
659 equations. *J. Stat. Softw.* **15**: 1–11.

660 Hellou, J., Ross, N.W. & Moon, T.W. 2012. Glutathione, glutathione S-transferase, and glu-
661 tathione conjugates, complementary markers of oxidative stress in aquatic biota. *Environ.*
662 *Sci. Pollut. R.* **19**: 2007–2023.

663 Hemmer-Hansen, J., Nielsen, E.E., Meldrup, D. & Mittelholzer, C. 2011. Identification of single
664 nucleotide polymorphisms in candidate genes for growth and reproduction in a nonmodel
665 organism; the Atlantic cod, *Gadus morhua*. *Mol. Ecol. Res.* **11**: 71–80.

666 Hess, J.E. & Narum, S.R. 2011. Single-nucleotide polymorphism (SNP) loci correlated with run
667 timing in adult Chinook salmon from the Columbia River basin. *Trans. Amer. Fish. Soc.*
668 **140**: 855–864.

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

671 Hoekstra, H. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates.
672 *Heredity* **97**: 222–234.

673 Hoekstra, H.E., Drumm, K.E. & Nachman, M.W. 2004. Ecological genetics of adaptive color
674 polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*
675 **58**: 1329–1341.

676 Hoffmann, A.A. & Willi, Y. 2008. Detecting genetic responses to environmental change. *Nat.*
677 *Rev. Genet.* **9**: 421–432.

678 Hogan, F.E., Cooke, R., Burridge, C.P. & Norman, J.A. 2008. Optimizing the use of shed
679 feathers for genetic analysis. *Mol. Ecol. Res.* **8**: 561–567.

680 Hubbard, A.E., Ahern, J., Fleischer, N.L., Van der Laan, M., Lippman, S.A., Jewell, N., Bruck-
681 ner, T. & Satariano, W.A. 2010. To GEE or not to GEE: comparing population average and
682 mixed models for estimating the associations between neighborhood risk factors and health.
683 *Epidemiology* **21**: 467–474.

684 Hudson, P.J., Dobson, A.P. & Newborn, D. 1992. Do parasites make prey vulnerable to preda-
685 tion? Red grouse and parasites. *J. Anim. Ecol.* **61**: 681–692.

686 Hunt, R.C., Simhadri, V.L., Iandoli, M., Sauna, Z.E. & Kimchi-Sarfaty, C. 2014. Exposing
687 synonymous mutations. *Trends Genet.* **30**: 308–321.

688 Joost, S., Bonin, A., Bruford, M.W., Després, L., Conord, C., Erhardt, G. & Taberlet, P. 2007.
689 A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape
690 genomics approach to adaptation. *Mol. Ecol.* **16**: 3955–3969.

691 Joost, S., Vuilleumier, S., Jensen, J.D., Schoville, S., Leempoel, K., Stucki, S., Widmer, I.,
692 Melodelima, C., Rolland, J. & Manel, S. 2013. Uncovering the genetic basis of adaptive
693 change: on the intersection of landscape genomics and theoretical population genetics. *Mol.*
694 *Ecol.* **22**: 3659–3665.

695 Keenan, K., McGinnity, P., Cross, T.F., Crozier, W.W. & Prodöhl, P.A. 2013. diveRsity: An
696 R package for the estimation and exploration of population genetics parameters and their
697 associated errors. *Methods Ecol. Evol.* **4**: 782–788.

698 Lehtonen, P., Laaksonen, T., Artemyev, A., Belskii, E., Berg, P., Both, C., Buggiotti, L., Bureš,
699 S., Burgess, M., Bushuev, A. et al. 2011. Candidate genes for colour and vision exhibit signals
700 of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity* **108**:
701 431–440.

702 Lewis, D.F.V. 2002. Oxidative stress: the role of cytochromes P450 in oxygen activation. *J.*
703 *Chem. Technol. Biotechnol.* **77**: 1095–1100.

704 Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA poly-
705 morphism data. *Bioinformatics* **25**: 1451–1452.

706 Lim, S. & Kaldis, P. 2013. Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Develop-*
707 *ment* **140**: 3079–3093.

708 Lischer, H. & Excoffier, L. 2012. PGDSpider: an automated data conversion tool for connecting
709 population genetics and genomics programs. *Bioinformatics* **28**: 298–299.

710 Lochmiller, R.L. & Deerenberg, C. 2000. Trade-offs in evolutionary immunology: just what is
711 the cost of immunity? *Oikos* **88**: 87–98.

712 Lotterhos, K.E. & Whitlock, M.C. 2014. Evaluation of demographic history and neutral param-
713 eterization on the performance of FST outlier tests. *Mol. Ecol. Res.* **23**: 2178–2192.

714 Luikart, G., Pilgrim, K., Visty, J., Ezenwa, V.O. & Schwartz, M.K. 2008. Candidate gene
715 microsatellite variation is associated with parasitism in wild bighorn sheep. *Bio. Lett.* **4**:
716 228–231.

717 MacColl, A.D.C., Piertney, S.B., Moss, R. & Lambin, X. 2000. Spatial arrangement of kin
718 affects recruitment success in young male red grouse. *Oikos* **90**: 261–270.

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

722 Manel, S., Joost, S., Epperson, B.K., Holderegger, R., Storfer, A., Rosenberg, M.S., Scribner,
723 K.T., Bonin, A. & Fortin, M.J. 2010. Perspectives on the use of landscape genetics to detect
724 genetic adaptive variation in the field. *Mol. Ecol.* **19**: 3760–3772.

725 Martínez-Padilla, J., Redpath, S.M., Zeineddine, M. & Mougeot, F. 2014. Insights into popu-
726 lation ecology from long-term studies of red grouse *Lagopus lagopus scoticus*. *J. Anim. Ecol.*
727 **83**: 85–98.

728 Martinez-Padilla, J., Vergara, P., Mougeot, F. & Redpath, S.M. 2012. Parasitized mates increase
729 infection risk for partners. *Am. Nat.* **179**: 811–820.

730 Martinez-Padilla, J., Vergara, P., Pérez-Rodríguez, L., Mougeot, F., Casas, F., Ludwig, S.,
731 Haines, J., Zeineddine, M. & Redpath, S. 2011. Condition-and parasite-dependent expression
732 of a male-like trait in a female bird. *Bio. Lett.* **7**: 364–367.

733 Martínez-Padilla, J., Mougeot, F., Webster, L.M.I., Pérez-Rodríguez, L. & Piertney, S.B. 2010.
734 Testing the interactive effects of testosterone and parasites on carotenoid-based ornamentation
735 in a wild bird. *J. Evol. Biol.* **23**: 902–913.

736 Matala, A.P., Hess, J.E. & Narum, S.R. 2011. Resolving adaptive and demographic divergence
737 among Chinook salmon populations in the Columbia River Basin. *Trans. Amer. Fish. Soc.*
738 **140**: 783–807.

739 Matzkin, L.M. 2012. Population transcriptomics of cactus host shifts in *Drosophila mojavensis*.
740 *Mol. Ecol.* **21**: 2428–2439.

741 Mazé-Guilmo, E., Loot, G., Páez, D.J., Lefèvre, T. & Blanchet, S. 2014. Heritable variation in
742 host tolerance and resistance inferred from a wild host–parasite system. *Proc. R. Soc. Lond.*
743 *B Biol. Sci.* **281**: 20132567.

744 McNagny, K.M., Rossi, F., Smith, G. & Graf, T. 1996. The eosinophil-specific cell surface
745 antigen, EOS47, is a chicken homologue of the oncofetal antigen melanotransferrin. *Blood* **87**:
746 1343–1352.

747 Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G.R., Espiñeira,
748 M., Fiorentino, F., Garofalo, G., Geffen, A.J. et al. 2014. Outlier SNP markers reveal fine-
749 scale genetic structuring across European hake populations (*Merluccius merluccius*). *Mol.*
750 *Ecol.* **23**: 118–135.

751 Mitchell-Olds, T., Willis, J.H. & Goldstein, D.B. 2007. Which evolutionary processes influence
752 natural genetic variation for phenotypic traits? *Nat. Rev. Genet.* **8**: 845–856.

753 Mougeot, F. 2008. Ornamental comb colour predicts T-cell-mediated immunity in male red
754 grouse *Lagopus lagopus scoticus*. *Naturwissenschaften* **95**: 125–132.

755 Mougeot, F., Irvine, J.R., Seivwright, L., Redpath, S.M. & Piertney, S. 2004. Testosterone,
756 immunocompetence, and honest sexual signaling in male red grouse. *Behav. Ecol.* **15**: 930–
757 937.

758 Mougeot, F., Martínez-Padilla, J., Blount, J.D., Pérez-Rodríguez, L., Webster, L.M.I. & Piert-
759 ney, S.B. 2010a. Oxidative stress and the effect of parasites on a carotenoid-based ornament.
760 *J. Exp. Biol.* **213**: 400–407.

761 Mougeot, F., Martínez-Padilla, J., Bortolotti, G.R., Webster, L.M.I. & Piertney, S.B. 2010b.
762 Physiological stress links parasites to carotenoid-based colour signals. *J. Evol. Biol.* **23**: 643–
763 650.

764 Mougeot, F., Martínez-Padilla, J., Webster, L.M.I., Blount, J.D., Pérez-Rodríguez, L. & Piert-
765 ney, S.B. 2009. Honest sexual signalling mediated by parasite and testosterone effects on
766 oxidative balance. *Proc. R. Soc. Lond. B Biol. Sci.* **276**: 1093–1100.

767 Mougeot, F. & Redpath, S.M. 2004. Sexual ornamentation relates to immune function in male
768 red grouse *Lagopus lagopus scoticus*. *J. Avian Biol.* **35**: 425–433.

769 Mougeot, F., Redpath, S.M. & Leckie, F. 2005. Ultra-violet reflectance of male and female red
770 grouse, *Lagopus lagopus scoticus*: Sexual ornaments reflect nematode parasite intensity. *J.*
771 *Avian Biol.* **36**: 203–209.

772 Mukhopadhyaya, C., Kumar, R., Brahc, G. et al. 2010. Gallinacin and fowlicidin: two promising
773 antimicrobial peptides in chickens: a review. *Vet. World.* **3**: 297–300.

774 Mullen, L.M. & Hoekstra, H.E. 2008. Natural selection along an environmental gradient: a
775 classic cline in mouse pigmentation. *Evolution* **62**: 1555–1570.

776 Newborn, D. & Foster, R. 2002. Control of parasite burdens in wild red grouse *Lagopus lagopus*
777 *scoticus* through the indirect application of anthelmintics. *J. Appl. Ecol.* **39**: 909–914.

778 Nielsen, R. 2001. Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**:
779 641–647.

780 Oliver, M., Telfer, S. & Piertney, S. 2009. Major histocompatibility complex (MHC) heterozygote
781 superiority to natural multi-parasite infections in the water vole (*Arvicola terrestris*). *Proc.*
782 *R. Soc. Lond. B Biol. Sci.* **276**: 1119–1128.

783 Oppelt, C., Starkloff, A., Rausch, P., Von Holst, D. & Rödel, H.G. 2010. Major histocompatibil-
784 ity complex variation and age-specific endoparasite load in subadult European rabbits. *Mol.*
785 *Ecol.* **19**: 4155–4167.

786 Orsini, L., Jansen, M., Souche, E.L., Geldof, S. & De Meester, L. 2011. Single nucleotide
787 polymorphism discovery from expressed sequence tags in the waterflea *Daphnia magna*. *BMC*
788 *Genomics* **12**: 309.

789 Orsini, L., Spanier, K.I. & De Meester, L. 2012. Genomic signature of natural and anthropogenic
790 stress in wild populations of the waterflea *Daphnia magna*: validation in space, time and
791 experimental evolution. *Mol. Ecol.* **21**: 2160–2175.

792 Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R.K. & Hedrick, P.W. 2010. Conservation
793 genetics in transition to conservation genomics. *Trends Gen.* **26**: 177–187.

794 Pan, W. 2001. Akaike's information criterion in generalized estimating equations. *Biometrics*
795 **57**: 120–125.

796 Pemberton, J., Beraldi, D., Craig, B. & Hopkins, J. 2011. Digital gene expression analysis of
797 gastrointestinal helminth resistance in Scottish blackface lambs. *Mol. Ecol.* **20**: 910–919.

798 Pespeni, M.H. & Palumbi, S.R. 2013. Signals of selection in outlier loci in a widely dispersing
799 species across an environmental mosaic. *Mol. Ecol.* **22**: 3580–3597.

800 Piertney, S. & Oliver, M. 2006. The evolutionary ecology of the major histocompatibility com-
801 plex. *Heredity* **96**: 7–21.

802 Piertney, S.B., Lambin, X., Maccoll, A.D.C., Lock, K., Bacon, P.J., Dallas, J.F., Leckie, F.,
803 Mougeot, F., Racey, P.A., Redpath, S. & Moss, R. 2008. Temporal changes in kin structure
804 through a population cycle in a territorial bird, the red grouse *Lagopus lagopus scoticus*. *Mol.*
805 *Ecol. Res.* **17**: 2544–2551.

806 Piertney, S.B., MacColl, A.D.C., Bacon, P.J. & Dallas, J.F. 1998. Local genetic structure in red
807 grouse (*Lagopus lagopus scoticus*): Evidence from microsatellite DNA markers. *Mol. Ecol.*
808 *Res.* **7**: 1645–1654.

809 Piertney, S.B., Maccoll, A.D.C., Bacon, P.J., Racey, P.A., Lambin, X. & Dallas, J.F. 2000.
810 Matrilineal genetic structure and female-mediated gene flow in red grouse (*Lagopus lagopus*
811 *scoticus*): An analysis using mitochondrial DNA. *Evolution* **54**: 279–289.

812 Piertney, S.B., MacColl, A.D.C., Lambin, X., Moss, R. & Dallas, J.F. 1999. Spatial distribution
813 of genetic relatedness in a moorland population of red grouse (*Lagopus lagopus scoticus*). *Biol.*
814 *J. Linnean Soc.* **68**: 317–331.

815 Piertney, S.B. & Webster, L.M.I. 2010. Characterising functionally important and ecologically
816 meaningful genetic diversity using a candidate gene approach. *Genetica* **138**: 419–432.

817 Poelstra, J.W., Ellegren, H. & Wolf, J. 2013. An extensive candidate gene approach to speciation:
818 diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the
819 European crow hybrid zone. *Heredity* **111**: 467–473.

820 Poulin, R. 2013. Parasite manipulation of host personality and behavioural syndromes. *J. Exp.*
821 *Biol.* **216**: 18–26.

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

824 R Core Team 2014. *R: A Language and Environment for Statistical Computing.* R Foundation
825 for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`.

826 Rahmanto, Y.S., Bal, S., Loh, K.H., Yu, Y. & Richardson, D.R. 2012. Melanotransferrin: Search
827 for a function. *Biochim. Biophys. Acta* **1820**: 237–243.

828 Raymond, M. & Rousset, F. 1995. GENEPOP (version 1.2): population genetics software for
829 exact tests and ecumenicism. *J. Hered.* **86**: 248–249.

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

832 Rousset, F. 2008. Genepop'007: a complete reimplementation of the Genepop software for
833 Windows and Linux. *Mol. Ecol. Res.* **8**: 103–106.

834 Sadd, B.M. & Schmid-Hempel, P. 2009. PERSPECTIVE: principles of ecological immunology.
835 *Evol. Appl.* **2**: 113–121.

836 Sang, J., Yang, K., Sun, Y., Han, Y., Cang, H., Chen, Y., Shi, G., Wang, K., Zhou, J., Wang, X.
837 et al. 2011. SUMO2 and SUMO3 transcription is differentially regulated by oxidative stress
838 in an Sp1-dependent manner. *Biochem. J* **435**: 489–498.

839 Sauna, Z.E. & Kimchi-Sarfaty, C. 2011. Understanding the contribution of synonymous muta-
840 tions to human disease. *Nat. Rev. Genet.* **12**: 683–691.

841 Saunders, L.M., Tompkins, D.M. & Hudson, P.J. 1999. The dynamics of nematode transmission
842 in the red grouse (*Lagopus lagopus scoticus*): Studies on the recovery of *Trichostrongylus*
843 *tenuis* larvae from vegetation. *J. Helminthol.* **73**: 171–175.

844 Schneider, R.F., Li, Y., Meyer, A. & Gunter, H.M. 2014. Regulatory gene networks that shape
845 the development of adaptive phenotypic plasticity in a cichlid fish. *Mol. Ecol.* **23**: 4511–4526.

846 Seivwright, L.J., Redpath, S.M., Mougeot, F., Watt, L. & Hudson, P.J. 2004. Faecal egg
847 counts provide a reliable measure of *Trichostrongylus tenuis* intensities in free-living red grouse
848 *Lagopus lagopus scoticus*. *J. Helminthol.* **78**: 69–76.

849 Semple, F. & Dorin, J. 2012. β-Defensins: Multifunctional modulators of infection, inflammation
850 and more? *J. Innate Immun.* **4**: 337–348.

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

853 Sheldon, B.C. & Verhulst, S. 1996. Ecological immunology: costly parasite defences and trade-
854 offs in evolutionary ecology. *Trends Ecol. Evol.* **11**: 317–321.

855 Shimada, Y., Shikano, T. & Merilä, J. 2011. A high incidence of selection on physiologically
856 important genes in the three-spined stickleback, *Gasterosteus aculeatus*. *Mol. Biol. Evol.* **28**:
857 181–193.

858 Sin, Y.W., Annavi, G., Dugdale, H.L., Newman, C., Burke, T. & Macdonald, D.W. 2014.
859 Pathogen burden, co-infection and major histocompatibility complex variability in the Euro-
860 pean badger (*Meles meles*). *Mol. Ecol.* **23**: 5072–5088. doi:10.1111/mec.12917.

861 Smouse, P.E., Long, J.C. & Sokal, R.R. 1986. Multiple regression and correlation extensions of
862 the Mantel test of matrix correspondence. *Syst. Zool.* pp. 627–632.

863 Stear, M.J., Fitton, L., Innocent, G.T., Murphy, L., Rennie, K. & Matthews, L. 2007. The dy-
864 namic influence of genetic variation on the susceptibility of sheep to gastrointestinal nematode
865 infection. *J. R. Soc. Interface* **4**: 767–776.

866 Stephens, M. & Scheet, P. 2005. Accounting for decay of linkage disequilibrium in haplotype
867 inference and missing-data imputation. *Am. J. Hum. Gen.* **76**: 449–462.

868 Stephens, M., Smith, N.J. & Donnelly, P. 2001. A new statistical method for haplotype recon-
869 struction from population data. *Am. J. Hum. Gen.* **68**: 978–989.

870 Stern, D.L. & Orgogozo, V. 2008. The loci of evolution: how predictable is genetic evolution?
871 *Evolution* **62**: 2155–2177.

872 Symons, A.M. & King, L.J. 2003. Inflammation, reactive oxygen species and cytochrome P450.
873 *Inflammopharmacology* **11**: 75–86.

874 Tew, K.D. & Townsend, D.M. 2012. Glutathione-S-transferases as determinants of cell survival
875 and death. *Antioxid. Redox Signal.* **17**: 1728–1737.

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

878 Turner, A.K., Begon, M., Jackson, J.A. & Paterson, S. 2012. Evidence for selection at cytokine
879 loci in a natural population of field voles (*Microtus agrestis*). *Mol. Ecol.* **21**: 1632–1646.

880 Venables, W.N. & Ripley, B.D. 2002. *Modern Applied Statistics with S*, 4th edn. Springer, New
881 York. URL `http://www.stats.ox.ac.uk/pub/MASS4`. ISBN 0-387-95457-0.

882 Vergara, P., Martínez-Padilla, J., Mougeot, F., Leckie, F. & Redpath, S.M. 2012a. Environmen-
883 tal heterogeneity influences the reliability of secondary sexual traits as condition indicators.
884 *J. Evol. Biol.* **25**: 20–28.

885 Vergara, P., Mougeot, F., Martínez-Padilla, J., Leckie, F. & Redpath, S.M. 2012b. The condition
886 dependence of a secondary sexual trait is stronger under high parasite infection level. *Behav.*
887 *Ecol.* **23**: 502–511.

888 Walsh, N., Dale, J., McGraw, K., Pointer, M. & Mundy, N. 2012. Candidate genes for carotenoid
889 coloration in vertebrates and their expression profiles in the carotenoid-containing plumage
890 and bill of a wild bird. *Proc. R. Soc. Lond. B Biol. Sci.* **279**: 58–66.

891 Wang, B., David, M.D. & Schrader, J.W. 2005. Absence of caprin-1 results in defects in cellular
892 proliferation. *J. Immunol.* **175**: 4274–4282.

893 Wang, Y., Fu, D., Luo, P. & He, X. 2012. Identification of the immune expressed sequence tags
894 of pearl oyster (*Pinctada martensii*, Dunker 1850) responding to *Vibrio alginolyticus* challenge
895 by suppression subtractive hybridization. *Comp. Biochem. Phys. D* **7**: 243–247.

896 Watson, A., Moss, R., Parr, R., Mountford, M. & Rothery, P. 1994. Kin landownership, diff-
897 ferential aggression between kin and non-kin, and population fluctuations in red grouse. *J.*
898 *Anim. Ecol.* **63**: 39–50.

899 Watson, H., Lee, D.L. & Hudson, P.J. 1987. The effect of *Trichostrongylus tenuis* on the caecal
900 mucosa of young, old and anthelmintic-treated wild red grouse, *Lagopus lagopus scoticus.*
901 *Parasitology* **94**: 405–411.

902 Webster, L.M., Johnson, P.C., Adam, A., Mable, B.K. & Keller, L.F. 2008. Absence of three
903 known benzimidazole resistance mutations in *Trichostrongylus tenuis*, a nematode parasite of
904 avian hosts. *Vet. Parasitol.* **158**: 302–310.

905 Webster, L.M.I., Mello, L.V., Mougeot, F., Martínez-Padilla, J., Paterson, S. & Piertney, S.B.
906 2011a. Identification of genes responding to nematode infection in red grouse. *Mol. Ecol. Res.*
907 **11**: 305–313.

908 Webster, L.M.I., Paterson, S., Mougeot, F., Martínez-Padilla, J. & Piertney, S.B. 2011b. Tran-
909 scriptomic response of red grouse to gastro-intestinal nematode parasites and testosterone:
910 Implications for population dynamics. *Mol. Ecol. Res.* **20**: 920–931.

911 Wenzel, M.A. & Piertney, S.B. 2014. Fine-scale population epigenetic structure in relation to
912 gastro-intestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Mol. Ecol. Res.* **23**:
913 4256–4273.

914 Wenzel, M.A. & Piertney, S.B. 2015. In silico identification and characterisation of 17 poly-
915 morphic anonymous non-coding sequence markers (ANMs) for red grouse (*Lagopus lagopus*
916 *scotica*). *Conserv. Gen. Res.* doi:10.1007/s12686-014-0416-0.

917 Wenzel, M.A., Webster, L.M., Blanco, G., Burgess, M.D., Kerbiriou, C., Segelbacher, G., Piert-
918 ney, S.B. & Reid, J.M. 2012. Pronounced genetic structure and low genetic diversity in Euro-
919 pean red-billed chough (*Pyrrhocorax pyrrhocorax*) populations. *Conserv. Gen.* **13**: 1213–1230.

920 Wenzel, M.A., Webster, L.M.I., Paterson, S. & Piertney, S.B. 2014. Identification and char-
921 acterisation of 17 polymorphic candidate genes for response to parasitic nematode (*Tri-*
922 *chostrongylus tenuis*) infection in red grouse (*Lagopus lagopus scotica*). *Conserv. Gen. Res.*
923 doi:10.1007/s12686-014-0323-4.

924 Wilfert, L. & Schmid-Hempel, P. 2008. The genetic architecture of susceptibility to parasites.
925 *BMC Evol. Biol.* **8**: 187.

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

928 Yang, W. & Paschen, W. 2009. Gene expression and cell growth are modified by silencing
929 SUMO2 and SUMO3 expression. *Biochem. Biophys. Res. Commun.* **382**: 215–218.