View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Edinburgh Research Explorer

THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A Comprehensive Genetic Study of Resistance to Nematodes in Sheep using the Ovine SNP Chip

Citation for published version:
Riggio, V, Matika, O, Pong-Wong, R, Moreno, CR, Carta, A & Bishop, SC 2014, 'A Comprehensive Genetic Study of Resistance to Nematodes in Sheep using the Ovine SNP Chip' Paper presented at Proceedings, 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada, 17/08/14 - 22/08/14, pp. 104.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Preprint (usually an early version)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Feb. 2015

A Comprehensive Genetic Study of Resistance to Nematodes in Sheep using the Ovine SNP Chip

V. Riggio[1], O. Matika[1], R. Pong-Wong[1], C.R. Moreno[2], A. Carta[3] and S.C. Bishop[1]
[1]The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK; [2]INRA-GenPhySE, Toulouse, France;
[3]Research Unit: Genetics and Biotechnology, AGRIS Sardinia, Sassari, Italy.

ABSTRACT: Our aims were to: identify loci underlying variation in faecal egg count (FEC) both within and across sheep populations; evaluate the accuracy of genomic EBVs (GEBV) for FEC within and across populations; and explore non-additive genetic variation (i.e. epistasis and heterozygote advantage) for FEC. Data were available on 752 Scottish Blackface lambs, 2,371 Sarda-Lacaune backcross ewes, 1,000 Martinik Blackbelly-Romane backcross lambs and 64 Texel lambs. Phenotypes were FEC for *Nematodirus* and/or *Strongyles* at different ages. Several genomic regions of interest were identified, both within and across populations. Moreover, GEBV had moderate to good within-population predictive accuracy, whereas across-population predictions had accuracies close to zero. Epistasis analysis identified two pairwise SNP interactions significant at the suggestive level for *Strongyles*, and the heterozygote advantage analysis identified some SNPs reaching suggestive significance. Therefore, results suggest the presence a missing heritability undetectable via conventional GWAS, which warrants further exploration.
Keywords: nematode resistance; genetic architecture; sheep

Introduction

Gastrointestinal nematode infections are one of the main health issues in grazing ruminants and have a great impact on sheep industries. Selection for nematode resistance has mainly been based on the use of indicator traits, such as faecal egg count (FEC) (e.g. Bishop and Stear (2001)). However, collecting and quantifying such indicator traits is a costly and time-consuming process and usually requires the animal to undergo parasitic challenge. Therefore, it would be advantageous to select directly for parasite resistance without the requirement for challenge.

To date, several studies have reported quantitative trait loci (QTL) for nematode resistance in sheep (e.g., Crawford et al. (2006); Davies et al. (2006); Sallé et al. (2012)). However, little overall consensus has emerged from these studies, probably because of the genetic complexity of the trait and the fact that these studies are very diverse, involving a variety of sheep breeds, nematode species and experimental approaches. Moreover, it has been shown that standard additive genetic studies generally fail to explain most of the known genetic variation influencing complex diseases (e.g., Manolio et al. (2009); Kemper et al. (2011)), suggesting the presence of a missing heritability undetectable via conventional GWAS.

One of the objectives of the 3SR project (http://www.3srbreeding.eu/) was to develop selectable genetic markers for resistance to nematodes in sheep. This put us in the fortunate position of having direct access to datasets comprising different populations, each of which has nematode resistance phenotypes as well as genotypes from the Illumina® Ovine 50SNP BeadChip (50k SNP chip). In particular, we aimed at: i) identifying loci underlying variation in FEC both within population (Scottish Blackface, SBF) and across populations in a joint-analysis (using the data from SBF, Sarda x Lacaune backcross (SAR) and Martinik Black-Belly x Romane backcross (MBR)); ii) evaluating the potential of genomic selection to predict genomic EBVs (GEBV) for nematode resistance traits both within and across populations; and iii) exploring non-additive genetic variation (i.e., epistasis and heterozygote advantage) in the SBF population.

Materials and Methods

Populations and phenotype data. Different populations were available for the study: 752 lambs from a SBF population, 2,371 ewes from a SAR population, 1,000 lambs from a MBR population, and 64 lambs from a British Texel (BT) population. Phenotypes available were FEC for *Nematodirus* and/or *Strongyles* at different ages, and their average animal effects across repeated measurements, depending on the population. More details on the data structure and on the phenotypes are given in Riggio et al. (2013, 2014a).

Genotype data. All animals from the four populations were genotyped using the 50k SNP chip. The SNP genotypes data were subjected to quality control (QC) measures, specific for each population. After QC, 42,841 SNPs were available for the SBF and BT populations, 44,859 for the SAR, and 42,469 for the MBR. Out of these SNPs, 38,991 were in common among the four populations and used for across-population analyses. For the SBF, a separate QC was carried out for GWAS analyses, leaving 44,388 SNPs available. More details on the genotype data available are given in Riggio et al. (2013, 2014a). Positions of SNPs were obtained from the Sheep Genome browser v2.0 (http://www.livestockgenomics.csiro.au/sheep/).

Heritability estimation. Heritabilities (h^2) were estimated using the genomic relationship matrix (**G**), comprising identity-by-state (IBS) relationships between all animals. For the SBF data, where pedigree was available, h^2 were also estimated using the pedigree-based relationship matrix (**A**), using ASReml (Gilmour et al. (2009)).

Within-population QTL mapping. Two approaches were used to identify loci underlying variation in FEC in SBF. In general, a GWAS analysis was performed with the GenABEL package (Aulchenko et al.

(2007)) in R environment (http://www.r-project.org), using the mmscore function. After Bonferroni correction, significance thresholds were $P < 1.13 \times 10^{-6}$ and $P < 2.25 \times 10^{-5}$ for genome-wide ($P < 0.05$) and suggestive (i.e., one false positive per genome scan) levels, respectively. The second approach was the Regional Heritability Mapping (RHM), in which each chromosome (OAR) is divided into windows of a pre-defined number of SNPs, and the variance attributable to each window estimated (Nagamine et al. (2012)). In this case, significance was evaluated with the likelihood ratio test (LRT), and after Bonferroni correction to account for multiple testing, the LRT thresholds were 13.56 and 9.29 (i.e., $P < 1.15 \times 10^{-4}$ and $P < 2.30 \times 10^{-3}$), respectively. More details about the methodologies, how they were implemented and the models fitted are reported in Riggio et al. (2013).

Across-population QTL mapping. Analyses were performed on the SBF, SAR and MBR populations, using RHM (Nagamine et al. (2012)). Initially, the analyses were performed using only the *Strongyles* data (i.e. average animal effect). Subsequently, a second analysis was performed substituting the *Nematodirus* FEC data for the *Strongyles* data in the SBF population. The **G** matrix was set to be block diagonal by population, to take into account that the three populations are genetically distant (hence linkage phases between marker and causative mutation are likely to differ between populations). Moreover, to account for the population structure (i.e. each population comprising few sire families and hence long stretches of LD), we ran further analyses with **G** matrices created separately for each chromosome under investigation, always excluding the chromosome being interrogated (i.e. 26 different **G** matrices were considered). We termed this the 'n-1' **G** matrix. This avoided QTL effects on individual chromosomes being absorbed by the overall **G** matrix. After Bonferroni correction, the LRT thresholds were 13.38 and 9.11 ($P < 1.27 \times 10^{-4}$ and $P < 2.54 \times 10^{-3}$), for genome-wide ($p < 0.05$) and suggestive significance levels, respectively (Riggio et al. (2014b) for further details).

Genomic prediction within and across populations. In the within-population (SBF) analysis, the predictive ability of the GEBV (i.e., $\hat{g}$) was tested through cross-validation. Validation sets were obtained by setting the phenotype as "unknown" for a defined number of individuals from the training set (TS): five non-overlapping cross-validation sets were created by randomly selecting 150 (152 for the fifth set) lambs at a time, masking each phenotype only once. For each set, the predicted genomic EBVs (PGEBV), i.e. GEBV calculated without phenotypic information on the individual, were estimated.

For the analyses across populations, two combined datasets were used, with SBF, SAR and MBR making the first set (4,123 individuals) and SBF and BT making the other (816 lambs). In the first set, two populations were used as TS to predict the third one (i.e., SAR and MBR to predict SBF; SBF and SAR to predict MBR; and SBF and MBR to predict SAR). Moreover, a few (1 or 10) half-sib family members from SBF were allocated to the TS and used as a connection with the rest of the half-sib family members in the validation set to test for the impact of cross-family links on GEBV.

Genomic prediction accuracies were calculated for each validation set. We estimated the Pearson correlations of PGEBV with the adjusted phenotypes ($r_{\hat{g},\hat{y}}$), and then the accuracy ($r_{\hat{g},g}$) for each validation set was estimated by dividing $r_{\hat{g},\hat{y}}$ by the square root of the h^2 of each trait for that specific validation set. The accuracy for each trait was then obtained by averaging the estimates across groups.

Non-additive genetic variation in SBF data. Average animal effect for *Strongyles* was analysed using BiForce Toolbox (Gyenesei et al. (2012)), to detect epistasis. Pairwise SNP interactions were assessed using contingency tables explained by Gyenesei et al. (2012). Bonferroni corrected genome-wide (and suggestive) significance threshold for the full pairwise genome scan was calculated as P=0.05(1)/(#SNPs*(#SNPs-1)/2), equating to 5.41×10^{-11} (1.08×10^{-9}).

Heterozygote advantage was evaluated on average animal effects for *Nematodirus* and *Strongyles* FEC, by recoding genotypes into two categories, viz. heterozygotes and homozygotes, and a GWAS was then performed accounting for relatedness through the **G** matrix. Significant loci from GWAS were tested for their significance in ASReml, fitting a mixed model.

Results and Discussion

Heritability estimates. Heritabilities for SBF were between 0.04 and 0.27 when estimated with the **G** matrix, and similar results were obtained with the **A** matrix, albeit with slightly higher estimates (Riggio et al. (2013)). When analysed across datasets, h^2 were 0.35 and 0.39 (Riggio et al. (2014b)).

Within-population QTL mapping. GWAS identified two SNPs with genome-wide significant association, whereas several SNPs reached the suggestive level. Not surprisingly, some SNPs were found to be important for more than one trait within the same parasite species, although the significance levels varied with trait. Moreover, there were cases where SNPs were also significant across trait categories, e.g., eight SNPs were associated with both *Strongyles* and *Nematodirus* (Riggio et al. (2013)). RHM identified one region significant at genome-wide level for *Nematodirus* and several other regions significant at the suggestive level. A summary of the identified SNPs/regions can be found in Riggio et al. (2013). Both GWAS and RHM identified genomic regions of interest, with both methods performing well when there was a strong evidence of association. For example, both methods identified similar regions for SNPs associated with *Nematodirus* average animal effect (OAR 14) and *Strongyles* FEC at 16 weeks (OAR 6).

Across-population QTL mapping. When analysing the *Strongyles* data, genome-wide significant

regions were identified on OAR 4, 19 and 20, with the latter being the most significant, whereas several other regions reached the suggestive significance level (Riggio et al. (2014b)). When combining *Nematodirus* and *Strongyles*, a new region on OAR 14 reaching the genome-wide significance was identified (Riggio et al. (2014b)). All our RHM results confirmed a region on OAR 20 encompassing the major histocompatibility complex (MHC) which was previously only identified within SAR population. The OAR 14 region identified in the joint-analysis was the same region previously reported by Riggio et al. (2013). However, this region was not significant in either of the other two populations considered, nor when only the *Strongyles* data was analysed, suggesting therefore that this region is specifically related to the *Nematodirus* infection.

Genomic prediction within and across populations. Correlations between PGEBV and adjusted phenotypes for the cross-validation groups in the SBF data were between -0.027 (*Nematodirus* FEC at 16 weeks) and 0.324 (*Nematodirus* FEC at 20 weeks). Moderate within population accuracies were observed, generally between 0.43 and 0.60, except for *Nematodirus* FEC at 16 weeks with an accuracy of 0.10 and also the lowest h^2. The across population accuracies were very low, sometimes even negative (Riggio et al. (2014a)). However, including a small number of animals in the TS from the population to be predicted was beneficial (Riggio et al. (2014a)). Overall, our results suggest that genomic prediction for nematode resistance may be of value in closely related animals but unlikely to work across breeds with current SNP chip.

Non-additive genetic variation in SBF data. Epistasis analysis identified two pairwise SNP interactions significant at the suggestive level for *Strongyles* (Table 1). However, at present, there appear to be no obvious candidate genes within these regions.

Table 1. Pairwise SNP associations for *Strongyles* average animal effect.

SNP1 (OAR)	SNP2 (OAR)	p-value
Strongyles		
OAR2_158196833 (2)	s27833 (3)	4.93×10^{-10}
OAR2_158196833 (2)	s56018 (3)	6.15×10^{-10}

The heterozygote advantage analysis did not identify any SNPs significant at the genome-wide level. However, some SNPs reached the suggestive significance level (Table 2). All SNPs were significant, when tested in a full mixed model in ASReml. One of the SNPs for *Nematodirus* was in the same region identified by Riggio et al. (2013). When this SNP was fitted as fixed effect in ASReml together with the main SNP from standard GWAS along with their interaction, both SNPs were significant, however the interaction was not significant, suggesting the two effects are independent.

Table 2. Heterozygote advantage analysis for *Nematodirus* and *Strongyles* average animal effects.

SNP (OAR)	Position (bp)	p-value
Nematodirus		
OAR25_25923466 (25)	23694478	2.51×10^{-6}
OAR14_53406640 (14)	50210627	6.21×10^{-6}
Strongyles		
OAR5_94598485 (5)	87049760	5.93×10^{-6}

Conclusion

In conclusion, in this project we have been successful in detecting QTL for nematode resistance both within and across populations. In particular, the joint analysis showed that potentially there are a number of common pathways that are underlying resistance to widely different parasite species. Our results also suggest that genomic prediction for nematode resistance may be of value in closely related animals, but unlikely to work across breeds, with the current SNP chip. Moreover, the non-additive genetic analyses suggest the presence of a missing heritability undetectable via conventional GWAS, which warrants further exploration.

Acknowledgements

These results are obtained through the EC-funded FP7 Project 3SR-245140. BBSRC, Defra and SPASE are also acknowledged.

Literature Cited

Aulchenko, Y. S., Ripke, S., Isaacs, A. et al. (2007). Bioinformatics 23:1294-1296.

Bishop, S. C., and Stear, M. J. (2001). Anim. Sci. 73:389-395.

Crawford, A. M., Paterson, K. A., Dodds, K. G. et al. (2006). BMC Genomics 7:178.

Davies, G., Stear, M. J., Benothman, M. et al. (2006). Heredity 96:252-258.

Gilmour, A., Gogel, B., Cullis, B. et al. (2009). ASReml, Version 3.0.

Gyenesei, A., Moody, J., Laiho, A. et al. (2012). Nucl. Acids Res. 40(W1): W628-W632.

Kemper, K. E., Emery, D. L., Bishop, S. C. et al. (2011). Genet. Res. 93:203-219.

Manolio, T. A., Collins, F. S., Cox, N. J. et al. (2009). Nature 461:747-753.

Nagamine, Y., Pong-Wong, R., Navarro, P. et al. (2012). PLoS ONE 7:e46501.

Riggio, V., Matika, O., Pong-Wong, R. et al. (2013). Heredity 110:420-429.

Riggio, V., Abdel-Aziz, M., Matika, O. et al. (2014a). Animal 8:520-528.

Riggio, V., Pong-Wong, R., Sallé, G. et al. (2014b). J. Anim. Breed. Genet. doi: 10.1111/jbg.12071.

Sallé, G., Jacquiet, P., Gruner, L., et al. (2012). J. Anim. Sci. 90:4690-4705.