GENOME-WIDE ASSOCIATION STUDY USING FIX-LENGTH HAPLOTYPES AND NETWORK ANALYSIS REVEALED NEW CANDIDATE GENES FOR NEMATODE RESISTANCE AND BODY WEIGHT IN BLACKFACE LAMBS

Amir Hossein Khaltabadi Farahani¹*, Hossein Mohammadi², Mohammad Hossein Moradi¹

¹Department of Animal Science, Faculty of Agriculture and Natural Resources, Arak University, Arak, Markazi, Iran

²Department of Animal Science, University of Tabriz, Tabriz, East Azerbaijan, Iran

*Corresponding author: amfarahanikh@gmail.com

Abstract

The objectives of this study were to identify genomic regions by Bayesian methods (BayesA, BayesB, or BayesN) that fit fixed-length haplotypes or SNPs using GenSel. Covariates for haplotype alleles of five lengths (125, 250, 500 kb, 1 or 2 Mb) were generated, and rare haplotypes were removed at three thresholds (1, 5, or 10%). Subsequently, we performed gene network analyses to investigate the biological processes shared by genes that were identified for the same across traits. Genotypes at 41,034 SNPs that were common on OvineSNP50 panel were phased for 751 Scottish Blackface (SBF) lambs. This is the first study to quantify the proportion of genetic variance using haplotypes across the whole genome in an SBF population. The genetic variance explained of haplotype-based GWAS was higher than that of SNP-based GWAS in across traits studied. In this population, fitting 500kb haplotypes with a 1% frequency threshold resulted in the highest proportion of genetic variance explained for nematode resistance and fitting 2Mb haplotypes with a 10% frequency threshold improved genetic variance explained for body weight comparable to fitting SNPs by BayesB. Candidate genes, including *CXCR4*, *STAT4*, *CCL1*, *CCL2*, *CCL3*, *CCL5*, *CCL8*, *CCL16*, *CCL18*, *CARMIL2*, and *HSPA14* were identified for nematode resistance and *ADH5*, *PP-3CA*, and *FABP4* for body weight traits. Network analysis provided annotation results linking to all identified candidate genes. This study supported previous results from GWAS of nematode resistance and body weight and revealed additional regions in the ovine genome associated with these economically important traits. These results suggest that network analysis can provide new information regarding biological mechanisms and genes leading to complex phenotypes, like nematode resistance and body weight of lamb.

Key words: sheep, genome-wide study, gene networks, nematode resistance, body weight

Nematode infections are considered the most critical gastrointestinal disease in sheep production, due to the decreased growth performance of infected individuals

and their subsequent economic effects. In addition, improvement of growth traits is highly emphasized in the sheep industry to improve quantitative traits (Bohlouli et al., 2013; Mohammadi et al., 2013).

Genome-wide association studies (GWAS) aim to identify the genomic regions that are associated with phenotypes and identification of subsets of markers that explain an important portion of the variation in many traits (Zhang et al., 2012, 2013). Recently several studies of GWAS identified genetic variants for disease and productive traits in different sheep breeds (Al-Mamun et al., 2015; Berton et al., 2017; Matika et al., 2016; Riggio et al., 2013; Rovadoscki et al., 2018; Amorim et al., 2018; Wilkie et al., 2017). However, many significant main effects of SNPs were identified via simple linear regression and stepwise regression procedures. However, this approach fits one SNP at a time, and multiple testing correction of results is an issue that limits the power of this approach. Moreover, due to imperfect linkage disequilibrium (LD), the effect of a putative gene/QTL can be only in part captured by a single SNP (Bovo et al., 2019).

The haplotype analysis with data from a GWAS study (Schweer et al., 2018) proved that it substantially improved the amount of the phenotypic variance explained, compared with single SNPs from a particular region of the genome. A haplotype analysis with data from a GWAS study proved that it substantially improved the amount (80%) of the phenotypic variance explained, compared with best single SNPs from a particular region of the genome (Barendse, 2011).

Bayesian methods such as Bayes A and Bayes B assume a heavy tail prior distribution for SNP effects and use Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution (Meuwissen et al., 2001). Although the objective of these methods was to predict the breeding value of selection candidates (genomic breeding values), they do so by estimating the effects of all SNPs. The estimated SNP effect, the proportion of variance explained by a SNP, or the number of times the SNP fits in the model with non-zero effect can be used as criteria to identify locations or genomic regions that affect the trait of interest. Results have shown that these Bayesian methods can effectively detect QTL in simulated (Sun et al., 2011) and real (Wolc et al., 2012; Bovo et al., 2019; Zhao et al., 2019) data. Recently, a new methodology has been developed to address this limitation and allow for a better understanding of the genetic architecture of complex traits through a gene network analysis (Mateescu et al., 2017).

Genomic evaluation of complex traits, such as nematode resistance and body weight, may benefit from the inclusion of biological information, such as gene ontology annotation. Therefore, the specific objective of this study was to evaluate the phenotypic variance explained with Bayesian genome-wide association methods, which fit covariates for fixed-length haplotype alleles compared to SNP alleles. Fixed-length haplotype alleles, ranging from 125 kb to 2 Mb in length, with varying allele frequency thresholds, ranging from 1 to 10%, were fitted using BayesA (Meuwissen et al., 2001), BayesB (Meuwissen et al., 2001), and BayesN (Zeng, 2015). In addition, we performed gene network analyses to investigate the biological processes for the genes, which captured for all the traits associated with weight and parasite resistance.

Material and methods

Phenotype and genotype data

The phenotypic and genotypic data and the statistical models used for the single-marker association analysis were described in detail by Riggio et al. (2013) and are reviewed here. The phenotypic data for nematode resistance and body weight traits in Scottish Blackface lambs were obtained from 751 pedigree-recorded lambs, bred over 3 years (2001–2003) with both sire and dam recorded at birth for all animals. Lambs were born on pasture, over 4 weeks each year, and were continuously exposed to natural mixed nematode infection. Two weeks after commencement of lambing, the ewes who had yet to lamb were moved to a new pasture, creating two management groups. After weaning at 12 weeks of age, the lambs were run as one flock again. Husbandry procedures followed standard commercial practice (Riggio et al., 2013).

Faecal samples were collected from the rectum of the lamb at approximately 16, 20, and 24 weeks of age. FEC was measured on four replicates of each faecal sample, and the average of the four replicates at each time point was used for analyses. At the time of sampling, body weight was recorded for all individuals. In addition, the average animal effect for each trait was estimated by fitting a repeatability model to trait values across the three ages.

Phasing and haplotype construction

The haplotype phase was inferred with the BEAGLE v4.1 software (Browning and Browning, 2009). Afterward, haplotype blocks were constructed based on alternate fixed length of chromosomes, because of its simplicity compared to the other methods of haplotype block construction.

To construct the haplotypes for five different lengths (125 kb, 250 kb, 500 kb, 1Mb, and 2 Mb) the *Oar_v4.0* map of the *Ovis aries* genome was used (Genbank accession: AMGL00000000.2). Rare haplotype alleles were discarded based on their frequency in the dataset at three different frequency thresholds: 1, 5, or 10%. The five choices of haplotype length assessed at each of the three frequency thresholds resulted in 15 scenarios for each haplotype model.

Bayesian GWAS

The analyses were performed using GenSel v4.73R (Garrick and Fernando, 2013) by fitting covariates for either SNPs or haplotype alleles in Bayes A, Bayes B, or Bayes N models. A single Markov chain Monte Carlo (MCMC) chain of length 41,000, including burn-in of 1,000 iterations, was computed for each analysis to obtain posterior estimates of covariate effects, which were used to obtain a direct genetic variance for animals, as described in the following. The prior analysis had shown that correlations and regression coefficients had converged at this chain length.

The SNP model and all 15 scenarios of the haplotype model, comprising five haplotype lengths and three haplotype allele frequency thresholds, were fitted using BayesA for all traits. The BayesA genomic prediction method (Meuwissen et al.,

2001) was used to identify SNPs and genomic regions associated with the measured phenotypes.

The SNP model and one of the fifteen-haplotype scenarios were fitted using BayesB (Meuwissen et al., 2001) and BayesN (Zeng, 2015). The chosen haplotype scenario was the largest (proportion) of total genetic variance explained by each window based on BayesA across all traits and a scenario that fitted a similar number of covariates as the SNP model (Wolc et al., 2012).

A range of values for π (BayesB) and Π (BayesN), collectively referred to as π values, was evaluated to determine the values that led to the highest genetic variance explained. BayesA is identical to BayesB when $\pi=0$. A range of values of π (0.25, 0.50, 0.75, 0.95, 0.99 and 0.995) was compared for all traits to evaluate the sensitivity of BayesB to the assumed π . In addition, ranges of values for π (0.5, 0.8 or 0.95) were assumed to test the sensitivity of BayesN to π .

Selection of haplotype windows, search for candidate genes, and gene network analyses

The top five SNP windows with the highest posterior mean proportion of genetic variance that accounted for the largest proportion of the genetic variance were extended by 0.5-Mb in both directions, and a positional candidate gene approach was conducted using sheep genes. The Map Viewer tool of the ovine genome was used for identification of genes available at National Center for Biotechnology Information (NCBI-<http://www.ncbi.nlm.nih.gov>) in *Oar_v4.0* ovine genome assembly and Ensembl Genome Browser (<http://www.ensembl.org/index.html>) using BioMart Software. The threshold of the top window was chosen based on the literature (Schweer et al., 2018; De Almeida et al., 2016).

In addition, based on human genes with the same description, we carried out four gene network analyses. As needed, human genes were used as background in pathway and gene network investigation.

Pathway enrichment analysis and visualization

Resulting gene lists were analyzed for enrichment of GO annotation terms and pathway analyses using Protein ANalysis THrough Evolutionary Relationships PANTHER v10.0 software (Mi et al., 2016).

An enrichment analysis of gene function was performed using the implementation of a binomial test of overrepresentation. Gene ontology (GO) terms were assessed using the default Ensembl Homo sapiens GO annotation as background for the enrichment analysis. Overrepresentation of GO terms were considered statistically significant at a Benjamini-adjusted P-value ≤ 0.05 .

ClueGO, a Cytoscape plug-in (Bindea et al., 2009), can decipher functionally grouped Gene Ontology (GO) and pathway annotation networks with a hypergeometric test, calculate the kappa coefficient of pathways, and study functional correlations among pathways. CluePedia, another Cytoscape plugin, is a tool used to search new markers potentially related with pathways. Furthermore, ClueGO, combined with CluePedia can derive better results. In the present study, overrepresented gene ontology (GO) terms were identified using ClueGO. The cut-off for considering

a significant over-representation was established by Benjamini-Hochberg multiple-test correction (P -value ≤ 0.05) and kappa coefficient > 0.3 .

The AnimalQTLdb database (Release 38, Apr 24, 2019) was used to verify previous reports of quantitative trait loci (QTL) in the surroundings of significant haploblocks. The main reason to use these tools was to validate our findings.

Results

The number of SNPs in each haploblock varied across the genome (Table 1). The average number of SNPs per haploblock ranged between 2 and 33, and the maximum number ranged from 6 up to 50, along with the haploblock lengths increasing from 125 kb to 2Mb. Haplotype blocks were constructed based on an alternate fixed length of chromosomes, because of its simplicity compared to the other methods of haplotype block construction.

Table 1. Mean and maximum number of SNPs per haploblock length

Haploblock length	Number of haploblocks	Number of SNPs per haploblock ¹	
		mean	maximum
125kb	17,941	2	6
250kb	9,575	4	11
500kb	4,848	8	17
1Mb	2,425	17	28
2Mb	1,210	33	50

¹The minimum number of SNPs per each haploblock was 1 for all haplotype lengths.

BayesA

Optimum haplotype length

In this study, SNP-based GWAS and haplotype-based GWAS were performed for 12 traits related to nematode resistance and body weight traits in Scottish Blackface lambs. In general, the genetic variance haplotype models were more explained than the SNP model.

In this research, the highest genetic variance explained using BayesA was observed at different sizes of haplotype blocks for two different traits. The optimum length of haplotype blocks for which explained the highest proportion of genetic variance was achieved among the four growth traits analyzed generally, in 2Mb and a 10% haplotype allele frequency filter. In addition, our results showed that, across the eight nematode resistance traits, the optimum size of haplotype to obtain the highest genetic variance was 500kb and a 1% haplotype allele frequency.

Overall, for body weight traits, the explained proportion of genetic variance was improved by increasing the length of haplotype blocks from 125 to 2Mb, but in nematode resistance traits, only 500kb explained the highest genetic variance and reduced afterward (data not shown). There are reports indicating that the optimum

haplotype length can vary depending on the type of dataset (i.e., real or simulated data), the distance between SNPs, the studied traits, and also population structure (Ferdosi et al., 2016; Hess et al., 2017).

BayesB and BayesN

Haplotype model choice

Only one scenario from the BayesA analyses was chosen to evaluate whether a BayesB or a BayesN model would further improve the genetic variance explained over the BayesA haplotype model for each trait, i.e. the scenario with 2Mb haploblocks that fitted only the alleles that had a frequency higher than 10% (Hap2Mb-10) for growth traits and the scenario with 500-kb haploblocks that fitted only the alleles that had a frequency higher than 1% in the dataset (Hap500-1) for nematode resistance traits. In this study, the best improvement in the explained genetic variance using BayesB and BayesN was observed at different sizes of haplotype blocks across traits, respectively (data not shown).

In this study, BayesB and BayesN results will be presented for a π value of 0.75 and 0.8 because, in many cases, these values resulted in the highest or close to the highest genetic variance explained. Based on these results, the BayesB Hap2Mb-10 ($\pi=0.75$) model for growth traits and Hap500-1 ($\pi=0.65$) model for nematode resistance traits was chosen as a representative model for the enrichment analysis. Our results from BayesN were consistent with those of Zeng (2015), however, it was surprising that BayesN did not result in the higher explained genetic variance comparing BayesB for haplotype models (data not shown).

Moreover, the results of our study showed that the proportion of genetic variance for nematode resistance traits was the highest for shorter haploblocks, i.e. 500kb with haplotype allele frequency 1%. With smaller genetic variation, the haplotype-based model is expected to yield less improvement in terms of explained genetic variance, since there is less variation in the QTLs for the haplotype alleles to capture.

Gene ontology enrichment analysis

Top windows were identified in the GWAS by considering all 751 animals and 41,034 SNPs and by applying the BayesB method ($\pi = 0.75$ for growth traits and $\pi = 0.65$ for nematode resistance) (Table 2 and 3). The chosen haplotype scenario was the largest (proportion) of total genetic variance explained by each window based on BayesB across all traits.

For growth traits, seven relevant genomic regions (single and overlapping) were found on OAR1 (OAR for *Ovis aries* chromosome), 2, 3, 6, 8, 20, and 23 (Table 2). For nematode resistance, 17 relevant genomic regions were located on OAR1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 22 and 23 (Table 3).

For growth traits, 149 genes located in the relevant genomic regions were identified, among which three genes were described to be related to body weights in the literature (Table 2). For nematode resistance, 935 genes were found in the detected relevant regions, of which twelve genes were related to parasite resistance based on literature (Table 3).

Table 2. Top windows explaining the largest percentages of body weight genetic variance obtained by BayesB analyses ($\pi=0.75$) along with associated candidate genes and previously reported QTL

Window ¹	Position of first SNP ²	Position of last SNP ³	chr ⁴	N SNP ⁵	Candidate Gene ⁶	QTL ⁷
486	242549454	242989342	1	41	—	Carcass fat percentage (20846385) ⁸
1037	242529435	242973674		36		Hot carcass weight and Body weight (20846385)
1040	244061250	244473057	2	35	—	Hot carcass weight and Body weight (20846385)
1038	243002057	243488052	2	32	—	Hot carcass weight and Body weight (20846385)
1039	243534446	243905042	2	31	—	Hot carcass weight and Body weight (20846385)
356	184182468	185928568	2	28	—	Internal fat amount (20846385)
1069	9504432	9959967	3	37	—	—
1160	55059556	55430870	3	23	—	Body weight (19389264)
501	22076908	23991456		43		Body weight (20846385) Average daily gain (19389264)
502	24002187	25991750	6	27	<i>PPP3CA</i>	Body weight (20846385)
503	26035930	27952676	6	36	<i>ADH5</i>	Body weight (20846385)
606	14076382	15997390	6	48	—	Internal fat amount (20846385)
1073	22025253	23967105	8	44	<i>FABP4</i>	Metacarpal length (17542849)
1074	24046661	25993444	20	27		Body weight (19076734) Maximum daily gain (19076734)
1167	56195700	57815619	20	—	—	—
			23	26	—	—

¹window coded number by GenSel according to physical map order.²start physical position the window.³end physical position the window.⁴chromosome number.⁵Number of SNPs within the window.⁶best candidate gene(s) in the region.⁷Quantitative trait loci.⁸Pub-Med id.

Table 3. Top windows explaining the largest percentages of nematode resistance genetic variance obtained by BayesB analyses ($\pi=0.65$) along with associated candidate genes and previously reported QTL

Window ¹	Position of first SNP ²	Position of last SNP ³	chr ⁴	N SNP ⁵	Candidate Gene ⁶	QTL ⁷
1	2	3	4	5	6	7
49	24070623	24415725	1	7	–	–
775	111987606	111987606		1	–	Somatic cell score (21693024) ⁸
			2			<i>Trichostrongylus</i> adult and larva count (16846521)
914	181085259	181456739	2	9	<i>CXCR4</i> , <i>STAT4</i>	–
1012	230015703	230482669	2	10	–	<i>Haemonchus contortus</i> FEC (23051556)
1283	116540180	116979911		8	–	<i>Trichostrongylus colubriformis</i> FEC (12047222)
			3			
1069	9504432	9959967			–	<i>Haemonchus contortus</i> FEC (19291139)
			3			
1416	183016939	183494868		10	–	<i>Nematodirus</i> FEC (22440166)
			3			
1408	179355428	179462570	3	3	–	–
1512	7049362	7458340	4	8	–	–
1598	50067372	50477036	4	11	–	–
1555	28518332	28900286	4	8	–	–
2011	30025516	30480469		11	–	Fecal egg count (22221026)
			6			
2075	62051221	62434378		7	–	Fecal egg count (22221026)
			6			
2009	29050589	29468607		11	–	Fecal egg count (22221026)
			6			
1988	18516843	18930178	6	7	–	–
1989	19106019	19411405	6	7	–	–
2439	27011783	27499284	8	10	–	<i>Trichostrongylus</i> adult and larva count (16846521)
2738	86084891	86467285	9	9	–	–
2907	75528230	75954556		11	–	Fecal egg count (19863786)
			10			
2814	29003338	29413536		12	–	Somatic cell score (21693024)
						Fecal egg count (19863786)
			10			
2793	18514582	18943787	10	9	–	–
2958	14527896	14942800	11	10	<i>CCL1</i> , <i>CCL2</i> , <i>CCL3</i> , <i>CCL5</i> , <i>CCL8</i> , <i>CCL11</i> , <i>CCL16</i> , <i>CCL18</i>	<i>Trichostrongylus</i> adult and larva count (16846521)
3202	74530063	74980433	12	8	–	Immunoglobulin A level (26791855)

Table 3 – contd.

1	2	3	4	5	6	7
3311	50100212	50209237	13	3	–	–
3445	34084625	34471958	13	8	–	<i>Nematodirus</i> FEC (16391549)
3268	28511594	28976864	13	10	<i>HSPA14</i>	<i>Salmonella abortusovis</i> susceptibility (22497440)
3462	42517080	42931938	14	11	CARMIL2	<i>Nematodirus</i> FEC (16391549)
3775	55009115	55495234	16	10	–	–
3804	69500806	69939000	16	9	–	–
3865	28053147	28468074	17	8	–	Immunoglobulin A level (26791855) Fecal egg count (22221026)
4078	62010518	62473576	18	12	–	–
4050	48015086	48480610	18	11	–	<i>Haemonchus contortus</i> FEC (19291139)
4549	16573236	16985452	23	8	–	<i>Haemonchus contortus</i> FEC (23051556) Immunoglobulin E level (16846521)
4524	4044507	4340330	23	8	–	–

¹window coded number by GenSel according to physical map order.²start physical position the window.³end physical position the window.⁴chromosome number.⁵Number of SNPs within the window.⁶best candidate gene(s) in the region.⁷Quantitative trait loci.⁸Pub-Med id.

Table 4. List of gene ontology terms that were significantly over-represented in the set of genes located in significant windows from the GenSel analysis for nematode resistance and body weight

Category	GO term	No. of genes in list	Fold enrichment	P-value	FDR P-value
1	2	3	4	5	6
Parasite					
Biological process slim	positive regulation of GTPase activity	10	14.03	4.53E–09	8.10E–06
	cellular response to tumor necrosis factor	10	12.01	1.91E–08	1.70E–05
	granulocyte chemotaxis	10	8.25	5.87E–07	2.62E–04
	protein secretion	8	5.52	1.30E–04	1.55E–02
	positive regulation of ERK1 and ERK2 cascade	10	4.43	1.18E–04	1.62E–02
	inflammatory response	10	4.01	2.60E–04	2.73E–02
	cytokine-mediated signaling pathway	10	3.93	3.01E–04	2.99E–02

Table 4 – contd.

1	2	3	4	5	6
Molecular function slim	protein transporter activity	10	6.79	3.29E-06	1.68E-03
Cellular Component slim	endoplasmic reticulum exit site	8	19.80	1.17E-08	5.26E-06
PANTHER Protein Class	apolipoprotein	5	15.03	2.53E-05	2.72E-03
	chemokine	10	8.42	4.92E-07	1.06E-04
Body weight					
Biological process slim	diterpenoid metabolic process	3	80.34	8.19E-06	1.46E-02
	small molecule catabolic process	4	17.85	8.31E-05	4.95E-02
Molecular function slim	zinc ion binding	4	20.40	4.98E-05	2.54E-02
	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	4	17.49	9.00E-05	2.30E-02
PANTHER Pathways	T cell activation	6	13.97	5.28E-06	8.65E-04
PANTHER Protein Class	major histocompatibility complex antigen	4	37.26	4.79E-06	1.03E-03

The proximity of the QTLs with the association regions and the involvement of QTL in across traits (e.g., body weight, average daily gain, hot carcass weight, fecal egg count, trichostrongylus adult, haemonchus contortus FEC, and immunoglobulin A level) support the results obtained for the examined genes (Table 2 and 3).

It is noteworthy that the present study has detected an association in regions different from that reported by a previous research on the Scottish Blackface breed (Riggio et al., 2013). This can be because of herd particularities, such as the extent of LD, allelic frequencies, and statistical approaches, in addition to the use of haplotypes in the current study.

The PANTHER classifications are presented according to biological process, molecular function, and cellular component in Figures 1–3.

The most enriched GO terms for genes near SNP associated with nematode resistance and body weight are presented in Table 4. Consistently, enriched Biological Process Slim (BPS) GO terms were cellular response to tumor necrosis factor, granulocyte chemotaxis, inflammatory response, and cytokine-mediated signaling pathway; the PANTHER Protein Class that was enriched in parasite resistance gene lists was binding (Table 4). An additional 3 BPS and 2 MFS GO terms and PANTHER Pathways were enriched in the growth gene list. The number of genes within each of these GO terms varied from 3 up to 10. Diterpenoid metabolic process term consisted of a minimum of 3 genes followed by cellular response to tumor necrosis factor, granulocyte chemotaxis, inflammatory response, and cytokine-mediated signaling pathways with ten genes in each. After Benjamini-Hochberg correction, none of the Cellular component Slim (CCS) GO terms were significantly enriched in the growth traits gene list.

Figure 1. Distribution of the GO “Biological Process” (BP) categories for genes represented by the significantly associated haplotypes in GWAS for nematode resistance and body weight. A) The PANTHER overrepresentation test grouped 935 annotated genes into 16 biological processes for nematode resistance. B) The PANTHER overrepresentation test grouped 149 annotated genes into 11 biological processes for body weight

Figure 2. Distribution of the GO “Molecular function” (MF) categories for genes represented by the significantly associated haplotypes in GWAS for nematode resistance and body weight. A) The PANTHER overrepresentation test grouped 935 annotated genes into 7 molecular function classes for nematode resistance. B) The PANTHER overrepresentation test grouped 149 annotated genes into 8 molecular function classes for body weight

Figure 3. Distribution of the GO “Cellular component” (CC) categories for genes represented by the significantly associated haplotypes in GWAS for nematode resistance and body weight. A) The PANTHER overrepresentation test grouped 935 annotated genes into 8 cellular components classes for nematode resistance. B) The PANTHER overrepresentation test grouped 149 annotated genes into 5 cellular components classes for body weight

Figure 4. Gene network of biological processes for nematode resistance (A) and function annotation for body weight (B). Node size is negatively correlated with the adjusted P-value of the pathway. Edge between nodes reflects shared or common genes: the edge width is positively correlated with the number of common genes. Different node colors represent different functional groups. The most significant pathway of each group has a highlighted label

Discussion

The Bayesian method used in this study (i.e., BayesA vs. BayesB vs. BayesN) had much impact on the explained genetic variance for both SNP and haplotype models. Haplotype models explained more genetic variance than the SNP model for

all the traits of interest. The top of windows that explained the largest fraction of genetic variation in common between haplotype models ranged from 18% to 24% and from 20% to 32% for nematode resistance and body weight, respectively.

Fitting covariates for haplotype alleles rather than SNPs could increase explained genetic variance through improved ability to detect ancestral relationships among individuals (i.e. identity-by-descent), higher LD between causal mutations and haplotype alleles, or greater ability to capture short-range epistatic effects (i.e., between loci that are present within the same haploblock) (Karimi et al., 2018).

We evaluated the BayesB and BayesN models to determine which model would be more suitable for haplotype analyses and if either model could outperform BayesA. When a large proportion of the variation in a trait is explained by a few large QTL, BayesA, which estimates non-zero effects for all SNPs or haplotype alleles, is less efficient than models such as BayesB, which estimate non-zero effects for a proportion of the SNPs or haplotype alleles (Meuwissen et al., 2001).

In our study, two Bayesian mixture models were evaluated in addition to BayesA: BayesB, which samples each haplotype allele regardless of the genomic region, and BayesN, which samples haplotype alleles within a genomic region jointly. As implemented in our study, the BayesN haploblock model can be considered as analogous to a BayesB model where the haploblock, rather than the haplotype allele, is sampled as being associated with the trait or not. There are previous reports indicating that the gain obtained by haplotype-based models vary among traits in pigs (Meuwissen et al., 2014), and cattle (Ferdosi et al., 2016; Hess et al., 2017). In agreement with the results of the present study, Hess et al. (2017) showed that employing haplotype-based approaches instead of individual SNP could reduce bias and improved genomic prediction accuracy in a cattle population. In general, these findings may suggest that in order to find the optimum length of haplotype block, each dataset or population should be evaluated individually.

An investigation of genes overrepresented in the 29 pathways revealed ten genes common to at least eight pathways. Eight genes, namely *CCL2*, *CCL3*, *CCL4*, *CCL5*, *CCL6*, *CCL8*, *CCL11*, and *CCL18* were found in common among at least seven pathways. It is interesting to note that a majority of the common genes that we detected are involved in response to tumor necrosis factor: response to cytokine, granulocyte chemotaxis, positive regulation of MAPK cascade, inflammatory response, and cytokine-mediated signaling pathway. This is not surprising given the role of chemokines and cytokines in nematode resistance through their involvement in the response to immune system responsible for parasite infections.

Chemokines have a relatively small molecular weight (8–14 kD) and are a specific class of cytokines. The various roles of chemokines make them interesting candidate genes.

CCL1 is produced by activated monocytes/macrophages and T-lymphocytes, and acts as a potent attractant for Th2 cells and a subset of T-regulatory cells.

CCL2 is the key chemokines for recruitment of monocytes, which compose the second line of cellular immune defense. It is important for innate immunity and helps initiate the humoral immune response (Schukken et al., 2011). The present study

may support the evidence for suggesting the association of this gene with nematode resistance in Blackface lambs.

Mainly multiple myeloma cells and acts produce *CCL3* as chemoattractant for monocyte-derived cells which, differentiate into macrophages or osteoclasts. These cells in turn promote inflammation, angiogenesis, osteolytic lesions, and immune-response impairment (Botta et al., 2016). Lindholm-Perry et al. (2017) reported *CCL3* expression of cytokine genes and receptors in white blood cells associated with divergent body weight.

CCL5 (*RANTES*) attracts lymphocytes, eosinophils, and basophils. Recently, Gossner et al. (2013) reported chemokine *CCL5* was up-regulation (3.16 fold) in Blackface lambs to resistant nematode *Teladorsagia circumcincta*.

CCL11, commonly known as eotaxin-1, exert their functions on a variety of cell types, including lymphocytes and monocytes.

CCL16 is known under various names, including the liver expressed chemokine (LEC) and Monotactin-1 (MTN-1). *CCL16* is expressed in the thymus, liver, spleen, and stimulates the migration of monocytes and lymphocytes.

CCL18, a chemokine predominantly produced by monocyte-derived cells with M2 phenotype, was highly expressed in serum of EOC patients and is a circulating EOC biomarker (Stranger et al., 2011). Our findings may reflect the functional effect of these genes related to nematode resistance, but further studies are required to confirm it.

STAT4 gene located on sheep chromosome 2 encodes a transcription factor, namely signal transducer and activator of transcription 4, involved in the signaling pathways of several cytokines, including interleukin (IL)-12, the type I interferons (IFN) and IL-23. It has been suggested that *STAT4* plays a key role in the regulation of Th1 and Th17 cell responses.

CXCR4 gene encodes a CXC chemokine receptor specific for stromal cell-derived factor-1, while the encoded protein has seven transmembrane regions and is located on the cell surface. Berton et al. (2017) reported *CXCR4* gene to be associated with red blood cells (RBC) trait indicating resistance to gastrointestinal, particularly *Haemonchus contortus*, in Santa Inês breed.

Protein phosphatase 3, catalytic subunit, alpha isozyme (*PPP3CA*) activates myogenin gene transcription. Barbero et al. (2017) have reported the association of this gene with early growth-related functions in Nellore heifers. Moradi et al. (2012) have also reported the *PPP2CA* gene has been selected in thin and fat tail sheep breed divergently. Therefore, this observation may suggest that this gene family could be associated with fat deposition in sheep breeds.

The *FABP4* gene is a member of the FABP family, which are fatty acid-binding proteins. Polymorphisms in the *FABP4* gene are associated with lipid metabolism (lipolysis and lipogenesis), marbling, backfat deposition, and carcass weight (Ardicli et al., 2017; Maharani et al., 2012). Thus, the obtained results indicate that the *FABP4* gene can be used as a marker of meat production in lamb.

A functionally grouped annotation network was developed based on 935 and 149 unique and annotated genes for nematode resistance and growth traits, respectively, and the network was visualized using the ClueGO plug-in for Cytoscape (Figure 4).

Only 588 genes were recognized by ClueGO, 429 (75.53%) were functionally annotated in the “Biological Process” ontology, and 315 (55.46%) were associated with representative terms and pathways after applying the selection criteria. Thirty-five GO terms were significantly represented in this network. The biological process with the highest percentage of the genes per GO term (approximately 35%) was observed for the “lymphocyte chemotaxis.” Of interest, the number of associated genes was found in biological processes related to the host immune, inflammatory and defense responses, response to interferon- γ as well as regulation of natural killer cell chemotaxis (Figure 4). Higher connectivity between GO terms with similar biological processes was found as expected, but a high priority in terms of future research will be placed on genes common among several different GO terms as these might point toward key regulator genes with a higher impact on the nematode resistance complex. The type of analyses used in this study and aimed at dissecting and understanding the gene networks and their contribution to the phenotypic expression of complex traits is highly dependent on the level of annotation of the respective genome but will further our general knowledge of gene function.

The gene network technique employed in this work advances the genomic analysis of complex traits beyond the simple marker association analysis by allowing the inclusion of markers, which initially are not able to reach a very stringent genome-wide significance status. These markers and the genomic regions they represent could be legitimate markers and regions explaining a small portion of the variation in these complex traits, but they do not have large enough effects to reach significance. The danger of a false positive is overcome through the gene/network enrichment analysis where a true false positive gene would most likely be eliminated while genes with a real but small effect on the trait will be validated through their biological role in a specific pathway contributing to a trait of interest. However, it is important that these results be validated through additional functional analyses at the gene expression or proteomics level.

Conclusions

Moving towards the use of haplotype models for genomic association studies can define the QTL locations more precisely. The traditional method of fitting SNP genotypes as covariates relies on LD between the marker and QTL. The corresponding location is that of the marker instead of the actual QTL of interest. Fitting covariates for fixed-length haplotype alleles rather than SNPs can increase the explained genetic variance.

Haplotype length and filtering based on haplotype allele frequency have a large impact on explained genetic variance and are therefore important parameters to optimize for the population and the analysis that is performed because non-optimized applications may decrease efficiency. In this dataset, for nematode resistance traits shorter haploblocks (500kb) and body weight longer haploblocks (2Mb) resulted in the highest genetic variance explained.

The Bayesian model that was used for haplotype models (BayesA, BayesB, or BayesN) had much impact on explained genetic variance, as long as the π values were less than 0.75 and 0.8 for the BayesB and BayesN models. The BayesB model

that consistently gave the highest genetic variance explained was the model that fitted 500-kb and 2Mb haploblocks with a 1 and 10% haplotype allele frequency filter for nematode resistance and body weight, respectively.

Several candidate regions related to immune system development and activation, inflammatory response, regulation of lymphocytes, and leukocytes proliferation were found in this study. These candidate regions may be useful in the selection of animals with higher resistance to the nematode, and consequently, reduce the anthelmintic drugs costs, also the production losses linked to the use of it. In addition, several genomic regions associated with body weight have been found using Bayesian GWAS method in this study.

Acknowledgments

The authors would like to thank Dr. Riggio for sample collection and genotyping.

References

- Al-Mamun H.A., Kwan P., Clark S.A., Ferdosi M.H., Tellam R., Gondro C. (2015). Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet. Sel. Evol.*, 47: 66.
- Amorim S.T., Eler J.P., Mattos E., Grigoletto L., Lemos M.V.A., Baldi F., Ferraz J.B. (2018). Genome wide association study (GWAS) for body weight traits in Santa Inês sheep breed. *Proc. 10th World Congress on Genetics Applied to Livestock Production*, Vancouver, Canada, 17–22.08.2014, p. 708.
- Ardicli S., Dincel D., Samli H., Balci F. (2017). Effects of polymorphisms at LEP, CAST, CAPN1, GHR, FABP4 and DGAT1 genes on fattening performance and carcass traits in Simmental bulls. *Arch. Anim. Breed.*, 60: 61–70.
- Barbero M.M.D., Santos D.J.A., Takada L., de Camargo G.M.F., Freitas A.C., Diaz I., de Souza F.R.P., Tonhati H., Albuquerque L.G., Oliveira H.N. (2017). Prospecting polymorphisms in the PPP3CA and FABP4 genes and their association with early pregnancy probability in Nellore heifers. *Livest. Sci.*, 203: 76–81.
- Barendse W. (2011). Haplotype analysis improved evidence for candidate genes for intramuscular fat percentage from a genome wide association study of cattle. *PLoS One*, 6: e29601.
- Berton M.P., Silva R.M.O., Peripolli E., Stafuzza N.B., Martin J.F., Alvarez M.S., Gavinã B.V., Toro M.A., Banchemo G., Oliveira P.S. (2017). Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. *J. Anim. Sci., Biotechnol.*, 8: 73.
- Bindea G., Mlecnik B., Hackl H., Charoentong P., Tosolini M., Kirilovsky A., Fridman W.H., Pagès F., Trajanoski Z., Galon J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25: 1091–1093.
- Bohloul M., Mohammadi H., Alijani S. (2013). Genetic evaluation and genetic trend of growth traits of Zandi sheep in semi-arid Iran using random regression models. *Small Rum. Res.*, 114: 195–201.
- Botta C., Di Martino M.T., Ciliberto D., Cucè M., Correale P., Rossi M., Tagliaferri P., Tassone P. (2016). A gene expression inflammatory signature specifically predicts multiple myeloma evolution and patients survival. *Blood Cancer J.*, 6: e511.
- Bovo S., Mazzoni G., Bertolini F., Schiavo G., Galimberti G., Gallo M., Dall'Olio S., Fontanesi L. (2019). Genome-wide association studies for 30 haematological and

- blood clinical-biochemical traits in Large White pigs reveal genomic regions affecting intermediate phenotypes. *Sci Rep.*, 9: 7003.
- Browning B.L., Browning S.R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.*, 84: 210–223.
- De Almeida Santana M.H., Junior G.A., Cesar A.S., Freua M.C., da Costa Gomes R., da Luz E Silva S., Leme P.R., Fukumasu H., Carvalho M.E., Ventura R.V., Coutinho L.L., Kadarmideen H.N., Ferraz J.B. (2016). Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle. *J. Appl. Genet.*, 57: 495–504.
- Ferdosi M.H., Henshall J., Tier B. (2016). Study of the optimum haplotype length to build genomic relationship matrices. *Genet. Sel. Evol.*, 48: 75.
- Garrick D.J., Fernando R.L. (2013). Implementing a QTL detection study (GWAS) using genomic prediction methodology. In: *Genome-Wide Association Studies and Genomic Prediction. Methods Mol. Biol.*, 1019: 275–298.
- Gossner A., Wilkie H., Joshi A., Hopkins J. (2013). Exploring the abomasal lymph node transcriptome for genes associated with resistance to the sheep nematode *Teladorsagia circumcincta*. *Vet. Res.*, 44: 68.
- Hess M., Druet T., Hess A., Garrick D. (2017). Fixed-length haplotypes can improve genomic prediction accuracy in an admixed dairy cattle population. *Genet. Sel. Evol.*, 49: 54.
- Karimi Z., Sargolzaei M., Robinson J.A.B., Schenkel F.S. (2018). Assessing haplotype-based models for genomic evaluation in Holstein cattle. *Can. J. Anim. Sci.*, 98: 750–759.
- Lindholm-Perry A.K., Artegoitia V.M., Miles J.R., Foote A.P. (2017). Expression of cytokine genes and receptors in white blood cells associated with divergent body weight gain in beef steers. *Agri Gene.*, 6: 37–39.
- Maharani D., Jung Y., Jung W.Y., Jo C., Ryoo S.H., Lee S.H., Yeon S.H., Lee J.H. (2012). Association of five candidate genes with fatty acid composition in Korean cattle. *Mol. Biol. Rep.*, 39: 6113–6121.
- Mateescu R.G., Garrick D.J., Reecy J.M. (2017). Network analysis reveals putative genes affecting meat quality in Angus cattle. *Front. Genet.*, 8: 171.
- Matika O., Riggio V., Anselme-Moizan M., Law A.S., Pong-Wong R., Archibald A.L., Bishop S.C. (2016). Genome-wide association reveals QTL for growth, bone and *in vivo* carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genet. Sel. Evol.*, 48: 11.
- Meuwissen T.H., Hayes B.J., Goddard M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819–1829.
- Meuwissen T.H., Odegard J., Andersen-Ranberg I., Grindflek E. (2014). On the distance of genetic relationships and the accuracy of genomic prediction in pig breeding. *Genet. Sel. Evol.*, 46: 49.
- Mi H., Huang X., Muruganujan A., Tang H., Mills C., Kang D., Thomas P.D. (2016). PANTHER: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res.*, 45: 183–189.
- Mohammadi H., Moradi Shahrehabak M., Moradi Shahrehabak H. (2013). Analysis of genetic relationship between reproductive vs. lamb growth traits in Makoei ewes. *J. Agr. Sci. Tech.*, 15: 45–53.
- Moradi M.H., Nejati-Javaremi A., Moradi-Shahrehabak M., Dodds K.G., McEwan J.C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.*, 13: 10.
- Riggio V., Matika O., Pong-Wong R., Stear M.J., Bishop S.C. (2013). Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. *Heredity*, 110: 420–429.
- Rovadoscki G.A., Pertile S.F.N., Alvarenga A.B., Cesar A.S.M., Pértile F., Petri ni J., Franzo V., Soares W.V.B., Morota G., Spangler M.L. (2018). Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Inês sheep. *BMC Genomics*, 19: 375.

- Schukken Y.H., Günther J., Fitzpatrick J., Fontaine M.C., Goetze L., Holst O., Leigh J., Petzl W., Schuberth H.J., Sipka A. (2011). Host-response patterns of intramammary infections in dairy cows. *Vet. Immunol. Immunopathol.*, 144: 270–289.
- Schweer K.R., Kachman S.D., Kuehn L.A., Freetly H.C., Pollak J.E., Spangler M.L. (2018). Genome-wide association study for feed efficiency traits using SNP and haplotype models. *J. Anim. Sci.*, 96: 2086–2098.
- Stranger B.E., Stahl E.A., Raj T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 187: 367–383.
- Sun X., Habier D., Fernando R.L., Garrick D.J., Dekkers J.C. (2011). Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. *BMC Proc.*, 3: S13.
- Wilkie H., Riggio V., Matika O., Nicol L., Watt K.A., Sinclair R., Sparks A.M., Nussey D.H., Pemberton J.M., Houston R.D. (2017). A candidate gene approach to study nematode resistance traits in naturally infected sheep. *Vet. Parasitol.*, 243: 71–74.
- Wolc A., Arango J., Settar P., Fulton J.E., O'Sullivan N.P., Preisinger R., Habier D., Fernando R., Garrick D.J., Hill W.G. (2012). Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. *Anim. Genet.*, 43: 87–96.
- Zeng J. (2015). Whole genome analyses accounting for structures in genotype data. Ames: Iowa State University.
- Zhang H., Wang Z., Wang S., Li H. (2012). Progress of genome wide association study in domestic animals. *J. Anim. Sci. Biotechnol.*, 3: 26.
- Zhang L., Liu J., Zhao F., Ren H., Xu L., Lu J., Zhang S., Zhang X., Wei C., Lu G. (2013). Genome-wide association studies for growth and meat production traits in sheep. *PLoS One*, 8: e66569.
- Zhao Y., Zhu H., Lu Z., Knickmeyer R.C., Zou F. (2019). Structured genome-wide association studies with Bayesian hierarchical variable selection. *Genetics*, 212: 397–415.

Received: 29 IX 2019

Accepted: 28 II 2020