Project No:	ON-00515
Contract No:	PO4500010753
AWI Project Manager:	Bridget Peachey
Contractor Name:	CSIRO Agriculture and Food
Prepared by:	Dr Sonja Dominik
Publication Date:	July 2019

Genotyping of breech flystrike resource – update

Published by Australian Wool Innovation Limited, Level 6, 68 Harrington Street, THE ROCKS, NSW, 2000

Contents

Executive Summary 3
1 Introduction/Hypothesis 5
2 Literature Review 6
3 Project Objectives 8
4 Success in Achieving Objectives 9
5 Methodology 10
5.1 Data 10
5.2 Genotype data and quality control 10
5.3 Phasing and imputation 12
5.4 Phenotype data 12
5.5 Genomic analysis 13
6 Results 15
6.1 Genome-wide association study 15
6.2 Genomic breeding values (GEBV) 20
6.3 Accuracy of the genomic predictions (rGEBV) 22
6.4 SNP effects for breech flystrike and indicator traits 23
7 Discussion 25
8 Impact of Wool Industry – Now & in 5 Years' Time 26
9 Conclusions and Recommendations 27
10 Bibliography 28

Executive Summary

Resistance to breech flystrike has long been a priority area of research for Australian Wool Innovation (AWI). Estimated breeding values for indicator traits of breech flystrike resistance are available through SheepGenetics and provide industry with tools to improve breech flystrike through selection on indicator traits. It has been demonstrated that genetic progress in sheep breeding programs could be fast-tracked with the application of genomic breeding values (van der Werf 2009). This approach requires a large reference population, but it was shown that the major benefit is generated from traits that are difficult or expensive to measure on live animals, and include carcase traits measured after slaughter, reproduction traits measured later in life (van der Werf 2009) and disease traits that cause a compromise in production or welfare. Resistance to breech flystrike falls into the category of such traits that are not directly measured, simply because it is expensive and difficult, and therefore genomic approaches could accelerate genetic gains for this trait.

This project investigated whether genomic approaches based on major genes, such as marker assisted selection, genomic selection based on genomic breeding values or a combination of the two would be the most efficient application of genomic information for selection for breech flystrike resistance.

As part of this project, 576 DNA samples from the WA and NSW Breech Flystrike Resource flocks were genotyped with the 50K Ovine SNP chip. The current data was combined with genotype and phenotype data from WP550 (2014). The previous data set included ~950 genotypes of 600K density. All of the current genotypes were imputed to 600K SNP density and all data analysed as one data set.

The project demonstrated that genomic selection based on genomic breeding values (GEBV) will be the most efficient approach to create benefit to industry within the next 5 years. The current genomic and phenotype resource is unique and an excellent starting point. Genomic breeding values for breechstrike resistance were estimated with an accuracy of 33%. For the GEBV to be a reliable selection tool, this accuracy would need to be improved further. With another 6,000 animals (a total of 7,500 animals) the accuracy of the genomic prediction could be doubled. It is discussed and recommended that a dispersed reference population with opportunistic phenotyping and sample collection in industry flocks would be an efficient way to build the numbers of animals. However, it would need to be demonstrated first that breech flystrike in chemically treated sheep is the same phenotype as collected in the untreated Breech Flystrike Resource flocks.

This study also demonstrated that currently there is no strong evidence that marker assisted selection would be a successful approach to breed for breech flystrike resistance. The pattern of SNP associations found in this study did not point to the existence of a small number of genes with major effects on breech flystrike resistance or the indicator traits. Although some significant SNP associations for breech flystrike and dag were found that would be interesting to explore further. While the results mean that it is unlikely that breech flystrike resistance can be in future predicted by a simple marker test, significant associations were found

through the sheep genome. This means that genomic predictions of breech flystrike resistance are a feasible goal for the application of genomic technology in industry.

Based on the approaches evaluated in this study it is concluded that with the existing genomic breech flystrike data resource as a basis the industry is in a position to create reliable selection tools that provide the fastest genetic progress for industry in five years' time.

It is recommended:

- To establish the relationship between breech flystrike in treated and untreated sheep
- That a dispersed reference population for breech flystrike resistance of at least 7,500 animals (an increase of 6,000 animals) based on opportunistic sample collection and commercial phenotypes is formed over the next 5 years
- To conduct more genomic analyses on the existing data set based on strategic subsets of the data (e.g. analysis for DAG only in the WA data).

1 Introduction/Hypothesis

Breech flystrike is a costly trait to measure. Australian Wool Innovation's investment into genetic solution for breech flystrike has resulted in the availability of estimated breeding values for indicator traits, such as breech wrinkle, breech cover and dag through SheepGenetics to enable genetic and permanent improvement in breech flystrike through selection. However, because selection is still based on indicator traits rather than breech flystrike itself, genetic gains are not at their maximum yet. Genomics provides approaches that could help fast-track genetic gains in the trait of breech flystrike resistance itself.

The data and materials collected on the Breech Flystrike Resource flocks in NSW and WA form an exceptional and unique resource to explore genomic approaches. AWI has already made an investment in genotyping part of the flocks and the work reported here builds on the previous data set and augments the data for more comprehensive analysis. The data set in this study forms the largest reference population for breech flystrike resistance to date.

It was hypothesised that the most efficient pathway of application of the breech flystrike genomic reference population would be the estimation of genomic breeding values (GEBV) for breech flystrike and indicator traits. Genome-wide association studies were conducted for breech flystrike resistance and indicator traits to explore the potential for marker assisted selection approaches.

2 Literature Review

Traits, such as breech flystrike resistance, are difficult to address in breeding programs because it is ethically and economically undesirable to deliberately challenge animals to express the trait. In addition, trait expression is seasonal and highly dependent on environmental conditions. The "gold standard" of phenotype collection has been established by AWI's investment in recording the trait in untreated, intensively monitored research populations. It is unclear, whether a breech flystrike phenotype expressed in a commercial environment that use preventative chemical treatment reflects the same phenotype.

These aspects demonstrate that breech flystrike is a difficult and expensive trait to measure, which would qualify breech flystrike resistance as a potentially good candidate for genomic approaches, such as genomic or marker assisted selection (Meuwissen *et al*., 2001; Dekkers, 2004). Analysis of genomic data forms the basis for recommended pathways of commercial application of genomic technologies (Dekkers, 2004; Goddard, 2009; Daetwyler *et al*. 2010; Zheng *et al*. 2012). The relevant literature was reviewed for AWI Project WP550 (2014). Therefore, we only provide a short overview with the addition of any other applicable literature that has been published since final report delivery for WP550.

Genome-wide association studies will have one of two main possible outcomes, which will indicate the most efficient single or combination approach for application in breeding programs.

1) Large numbers of SNPs with small effect, indicating that we are dealing with a complex trait and commercial application ideally draws on the variation explained by all SNPs, such as genomic breeding values.
2) Small numbers of SNPs of major effect, linked to "major genes", indicating that it is feasible to develop DNA diagnostics.

To date 2,129 major genes for 249 traits are summarised in SheepQTLdp, a database for published major genes (Hu *et al*., 2013). No major genes have been published for flystrike resistance. Only a small number of the reported major genes, such as a test for polled, have been implemented in Australian sheep breeding programs. Some of the reasons being that effect sizes are often too large, like for example Booroola with up to 6 lambs, or the inheritance is complicated (e.g. Callipyge) or the affected traits have undesirable correlations (Dekkers, 2004). Other reasons included an overestimated of the genetic variance explained by these small numbers of significant SNP, and often the association between the markers and the trait effect observed in experimental studies did not carry through to the population level (Boichard *et al.* 2016).

Genomic selection has been adopted with varying speed across the different livestock species. The application in dairy was revolutionary because young bulls could be selected without the need for progeny testing and the generation interval could be reduced, which doubled the annual genetic gain (Boichard *et al.* 2016). The main limiting factor in other species is the cost of phenotyping, in particular for low heritability traits, to achieve a sufficient size of reference population (Boichard *et al.* 2016). Depending on the trait, innovative phenotyping techniques, opportunistic data collection from sensor based

precision farming systems, and the inclusion of commercial data might provide cost-effective and efficient approaches to obtain reference populations of appropriate size (Boichard *et al*. 2016). If the accuracy of a selection index with GEBV would be at least as high as the square root of the heritability of a trait, genetic gains in Merino sheep breeding programs could be increased by up to 40% (van der Werf 2009). For wool traits moderate-to-high GEBV accuracies have been reported in Merino sheep (Daetwyler *et al.* 2010), which means that the rate of genetic gain for these traits will lift significantly with the use of GEBV. Note that the accuracies in the study by Daetwyler *et al.* (2010) are correlations between estimated and genomic breeding values as determined in a validation population, which is different to the accuracy reported in this study.

3 Project Objectives

The project had two main objectives:

1) To create a complete data resource for all drops of the NSW and WA Breech Flystrike Resource flocks by testing 576 sheep from NSW and WA study populations at 50,000 SNP.

2) Identify potential associations between markers and breech strike resistance and indicator trait phenotypes with the full data set.

4 Success in Achieving Objectives

The objectives have been fully achieved.

1) A total of 576 DNA sample from sheep from the NSW and WA Breech Flystrike Resource flocks have been genotyped with 50K Ovine SNP chips by Neogen.

2) The data generated in this project has been merged with the previous HD data after imputation. The full data set now comprises 1,535 HD genotypes. Genome wide association studies have been conducted for breech flystrike, breech wrinkle, breech cover and dags. Genomic breeding values were estimated for all animals.

5 Methodology

5.1 Data

Two genotype and phenotype data sets contributed to the analysis for this study. Data set 1 (600K data) included animals of the drops between 2005 – 2011 of the Breech Flystrike Resource flocks in NSW and WA. Data set 2 (50K data) included animals from the Breech Flystrike Resource flocks in NSW and WA born between 2011 and 2014. In the following both data sets are described independently and also in combination.

5.2 Genotype data and quality control

Data set 1 (600K data) comprised a total of 959 DNA samples from individual sheep of the Breech Flystrike Resource flocks in WA and NSW. Samples were genotyped using the Illumina Ovine HD Beadchip with 606,006 single nucleotide polymorphism (SNP) in CSIRO's Brisbane laboratory using established methods.

Table 1. Description of the quality control criteria for SNP.

SNP criteria	
Genotyping failure	Failure to make a genotype call
GC score	Measure on reliability of genotype call
Level of homozygosity	SNP with the same genotype in 100% of samples are not useful for analysis
Minor allele frequency (MAF)	The occurrence of low frequency of one of the alleles can lead to spurious peaks in genome wide association studies
Heterozygosity	A low level of heterozygous SNPs can indicate a problem with genotyping
Hardy-Weinberg at 1e-15	The Hardy-Weinberg equilibrium describes the relationship between genotype and allele frequencies. In a selected population this might not hold completely, however, a deviation of a large number of SNP would indicate a problem with genotyping.
Sample criteria	
Call rates	% of SNP that could be successfully called
Correlation >0.98	Correlations describes the relationship between samples; a correlation of 1 identifies duplicate samples
Heterozygosity	The level of heterozygosity of the SNPs for a sample; if high a good indication of contamination of the sample
Mapping criteria	These describe the number of SNPs on unknown chromosomes (chromosome 0), and the sex chromosomes (chromosomes X and Y)

The genotypic data were submitted into a quality control (QC) process using snpQC (C. Gondro, https://cgondro2.une.edu.au/CGhomepage). This software is run in an R environment (R Core Team, 2013) and provides the full data pipeline from raw data to a quality controlled data set. The criteria are outlined in Table 1. In summary, quality control excluded 77,188 SNP (12.74%) and 11 animal samples (1.15%), with 52,818 SNPs and 948 samples remaining in the data set. Numbers of excluded SNP and samples are summarised in Table 2. Flagged samples and SNP have been removed from the analysis. That includes SNP on the sex chromosomes X and Y and unmapped chromosomes (chromosome 0).

Table 2. Filtering results of the quality control (note that multiple SNP might not pass multiple quality control criteria).

	600K data (drop 2005-2011)	**50K data (drop 2011-2014)**
SNP criteria	**Number of SNP**	**Number of SNP**
Genotyping fail > 5%	15,907	1497
Median GC scores < 0.5	31,710	2365
All GC scores 0	8,003	441
GC < 0.5 in less than 90% of samples	34,506	3687
100 % homozygous	29,858	198
Minor allele frequency < 0.01	8,507	373
Heterozygosity 3 Standard deviations	4	5
Hardy Weinberg at 1E -15	24,635	1081
Sample criteria	**Number of Samples**	**Number of Samples**
Call rates < 0.9	6	4
Correlation > 0.98	50	0
Heterozygosity 3 Standard deviations	8	8
Mapping criteria	**Number of SNP**	**Number of SNP**
Chromosome 0	1,291	314
Chromosome X	27,314	1449
Chromosome Y	NA	1

The second data set included a total of 576 DNA samples from sheep of the Breech Flystrike Resource flocks in NSW and WA. The samples were genotyped with the Ovine 50K Illumina SNP chip by Neogen in their Gatton laboratory. Nine samples did not pass the filtering criteria (1.56%). From the 54,241 SNPs 6,165 were excluded (11.37%). Out of the total 31,242,816 genotypes, 4,045,497 were excluded (12.95%). Filtering criteria consisted of QC metrics across SNPs, across arrays and on the physical mapping as detailed in the full quality control report. Numbers of excluded SNP and samples are summarised in Table 2. Flagged samples and SNP have been removed from the analysis. That includes SNP on the sex chromosomes X and Y and unmapped chromosomes (chromosome 0).

The total data set comprised 1,535 samples before quality control. The SNP and samples before and after QC are summarized in Table 3.

Table 3. Summary of SNP and sample number before and after quality control for the 600K and 50K and combined data sets.

	600K data	**50K data**
No of samples	959	576
No of SNP	606,006	54,241
Samples excluded	11 (1.15%)	9 (1.56%)
SNP excluded	77,188 (12.74%)	6,165 (11.37%)
Genotypes excluded	80,380,186 (13.83%)	4,045,497 (12.95%)

5.3 Phasing and imputation

The aim was to combine the two data sets. For that purpose, the 50K genotypes needed to be imputed to the 600K density. Data was prepared using PLINK (Purcell *et al.* 2007), phased using the software Eagle v.2.3.5 (Loh *at al*. 2016) and imputation was implemented with Minimac3 (Das *et al.* 2016). Firstly, after phasing missing genotypes in the 600K data were imputed using Minimac3 (Das et al. 2016), followed by the imputation of the 50K genotypes to 600K density. The process of phasing and imputation using Eagle and Minimac3 is described in Al-Mamun *et al.* (2017).

5.4 Phenotype data

The phenotype data set 1 comprised 926 animals with a full set of trait and fixed effect records. Data set 2 contained 567 animals after QC, all of which 554 have a full set of phenotype records. Overall, the phenotype data sets contained 949 from NSW in NSW and 531 from WA in WA, totalling 1,480 animals.

Traits for analysis included breech flystrike (STRIKE), coded as "struck" or "not struck", and breech cover (BCOV), dag score (DAG) and wrinkle score (WRK), all categorised as low, medium or high. Fixed effects that were tested for significance included site (SITE, NSW or WA), sex (SEX, male or female), drop (DROP, 2005-2009 and 2011-2014) and mules status (MULES, mules or unmulesed) (Tables 4). In the Data set 2, the Western Australian site only contributed animals that were born in 2014.

Table 4. Number of records per site across drop

	DROP									**Total**
SITE	'05	'06	'07	'08	'09	'11	'12	'13	'14	
NSW	163	121	92	90	129	133	54	79	83	944
WA	5	43	26	52	126	0	0	0	277	529

On both sites more females than males were included in the data set (Table 5). Mulse status is somewhat confounded with year because both resource flocks were not mulesed after 2011.

Table 5. Number of records per site across sex and mules status

	SEX		**MULES***	
SITE	Male	Female	Mulesed	Unmulesed
NSW	319	625	560	384
WA	57	472	230	299

*from 2011 all animals were unmulesed

Table 6. Number of records for breech flystrike (strike) across site and drop

	DROP									**SITE**	
STRIKE	'05	'06	'07	'08	'09	'11	'12	'13	'14	NSW	WA
Struck	79	69	59	71	125	56	20	27	198	466	238
Not struck	89	95	59	71	13	77	34	52	161	478	290

Numbers of struck vs not struck animals that were included were quite balanced within sites (Table 6). Sufficient numbers of struck and not struck animals were represented in each of

the drops across sites. Also across SEX and mules status representation of stuck and not struck animals was quite balanced (Table 7).

Table 7. Number of records for breech flystrike (strike) across sex and mules status

	SEX		MULES	
STRIKE	Male	Female	Mulesed	Unmulesed
Struck	186	518	372	332
Not struck	189	579	418	350

Table 8. Number of records for breech flystrike (STRIKE) across indicator traits, breechcover (BCOV), dag score (DAG) and wrinkle score (WRK)

	BCOV		
STRIKE	low	medium	high
Not struck	102	355	210
Struck	126	335	278
	WRK		
Not struck	349	131	253
Struck	293	103	343
	DAG		
Not struck	480	67	186
Struck	497	64	178

Struck and not struck animals were balanced across the other indicator traits, namely BCOV, BWRK and DAG. Breech wrinkle and DAG are site specific indicator traits with not much variation in DAG for the NSW site and not much variation in WRK for the WA site. However, across the two sites, high, medium and low categories are well represented in the data set (Table 8).

5.5 Genomic analysis

The software GCTA (Yang *et al.* 2011) was used to estimate a genomic relationship matrix, which was then used in the genome wide association studies (GWAS) and the estimation of GEBV. The GWAS was conducted using GCTA-MLMA which explored associations of SNP with STRIKE, BCOV, BWRK and DAG. The model fitted SITE, DROP, SEX and MULES as fixed effects in the model. The same fixed effects were fitted in the model for the GEBV and these were estimated using GCTA-GREML. Any SNP that exceeded the genome-wide significance threshold were compared against the most recent annotation of the ovine genome from Ensembl Version 3.1.94 (download at ftp://ftp.ensembl.org/pub/release-94/gff3/ovis_aries/Ovis_aries.Oar_v3.1.94.chr.gff3.gz) for genes that are closest to these SNP. For this step closest-feature in bedops (Neph *et al.* 2012) was used.

The accuracy of the genomic predictions (r_{GEBV}) was calculated according to Goddard *et al.* (2011) and Dekkers (2007). It is a function (1) of the accuracy of genomic prediction as predictor of effects captured by markers (rQhat) and the proportion of genetic variance captured by markers (q^2).

$$R_{GEBV} = \sqrt{q^2} \times rQhat \quad (1)$$

Both components are dependent on the effective population size (N_e), the number of independent chromosome segments (M_e, here calculated at in Goddard et al. 2011 and a function of the average chromosome length in Morgan (L), N_e and the number of chromosomes (k)), the numbers of animals in the reference population (N) and the heritability of the trait (h^2) and the number of markers (n_m). All inputs parameters are summarised in Table 9.

Table 9. Parameters for the calculation of the accuracy of the genomic prediction. Effective population size (Ne), the number of independent chromosome segments (Me), average chromosome length in Morgan (L), number of chromosomes (k), the number of animals in the reference population (N) heritability of the trait (h2) the number of markers (nm).

N_e	L	k	h^2	N	n_m
400	1	26	0.1-0.5	1,000-10,000	500,000

The effective population size (Ne) was evaluated by the proportional contribution of the WA and NSW flocks and their respective Ne. Heritabilities were varied from h2=0.1 to 0.5 to capture the trait heritabilities for breech flystrike and indicator traits as estimated for the WA and NSW flocks (Greeff et al. 2014; Smith et al. 2009). The sheep genome has a length of 2,587Mb (https://www.ncbi.nlm.nih.gov/genome?term=ovis%20aries), which translates to the average length of 1 Morgan for the 26 chromosomes of the sheep genome.

6 Results

6.1 Genome-wide association study

Genome-wide association studies (GWAS) were undertaken for breech flystrike and the indicator traits. Results are shown in Figures 1-4. The graphs show the significance level as -log(p-values) on the y-axis. The chromosome-wide significance threshold is drawn at -log10(1e-5). Any SNP that passed this threshold for any of the traits are summarised in Table 10.

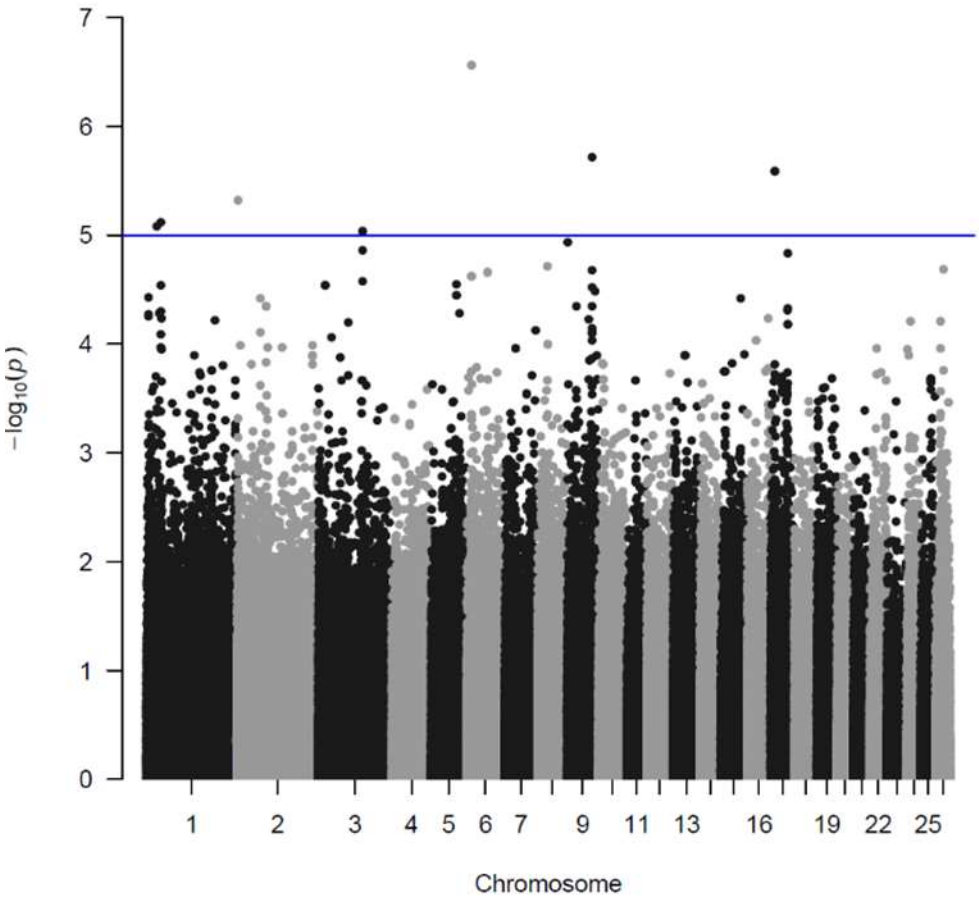

Figure 1: Results of genome-wide association study for breech flystrike. Blue line indicating the chromosome-wide significance threshold.

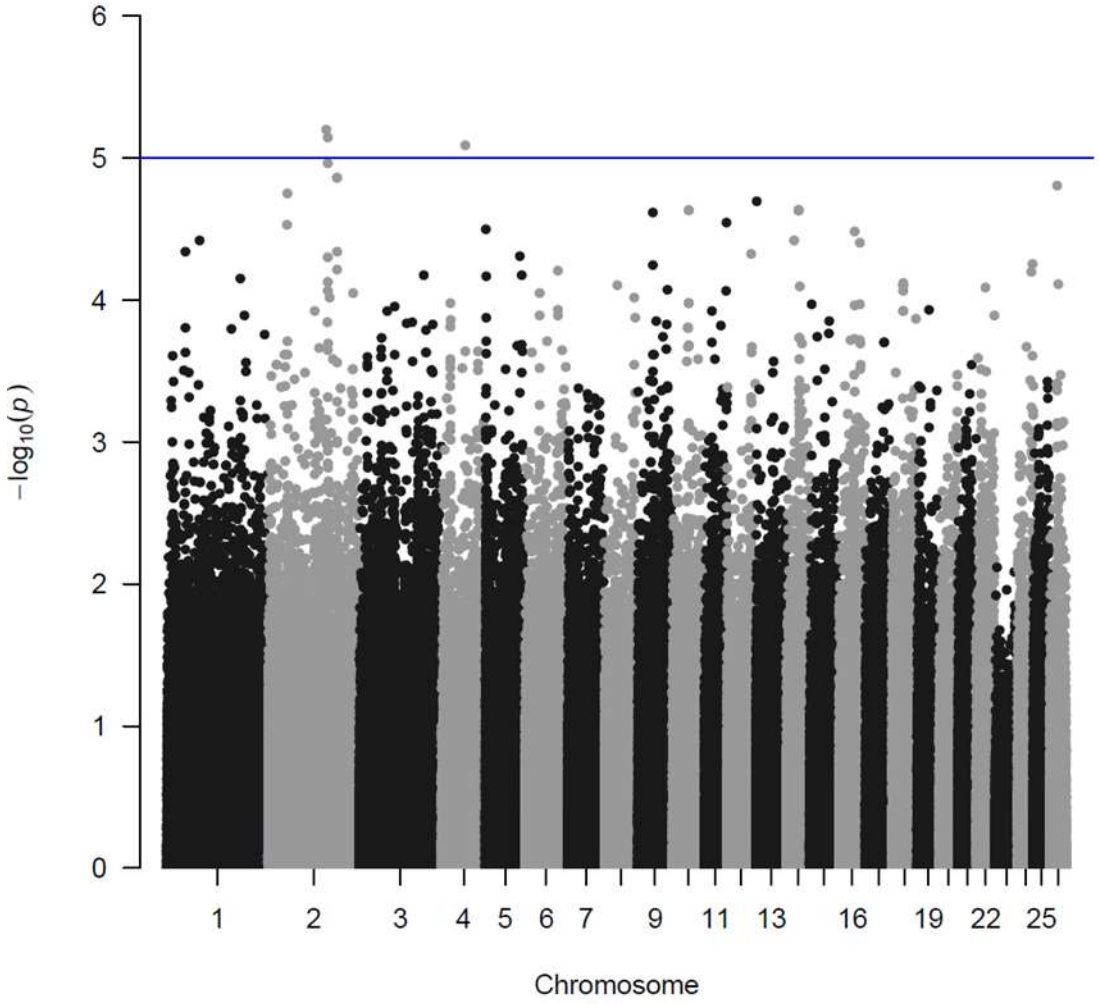

Figure 2: Results of genome-wide association study for breechcover. Blue line indicating the chromosome-wide significance threshold.

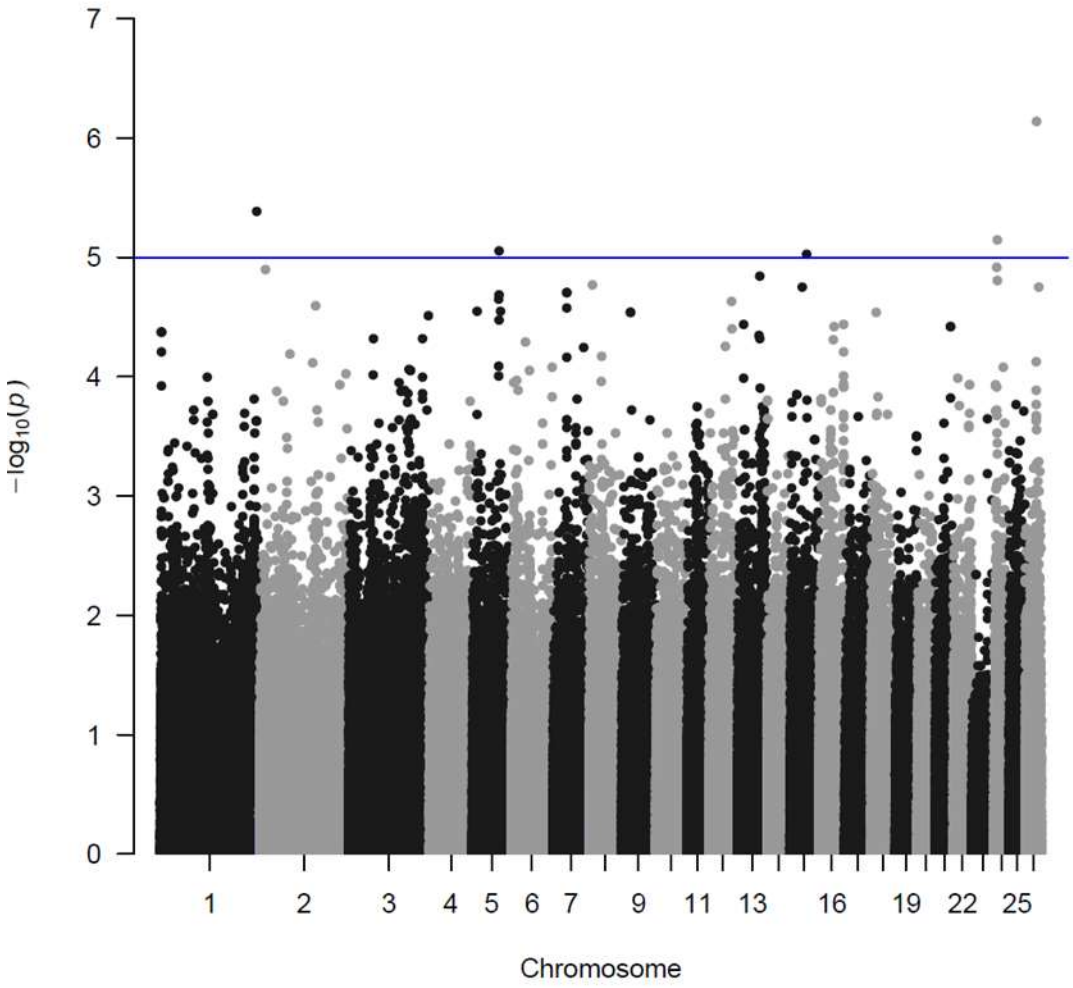

Figure 3: Results of genome-wide association study for breechwrinkle. Blue line indicating the chromosome-wide significance threshold.

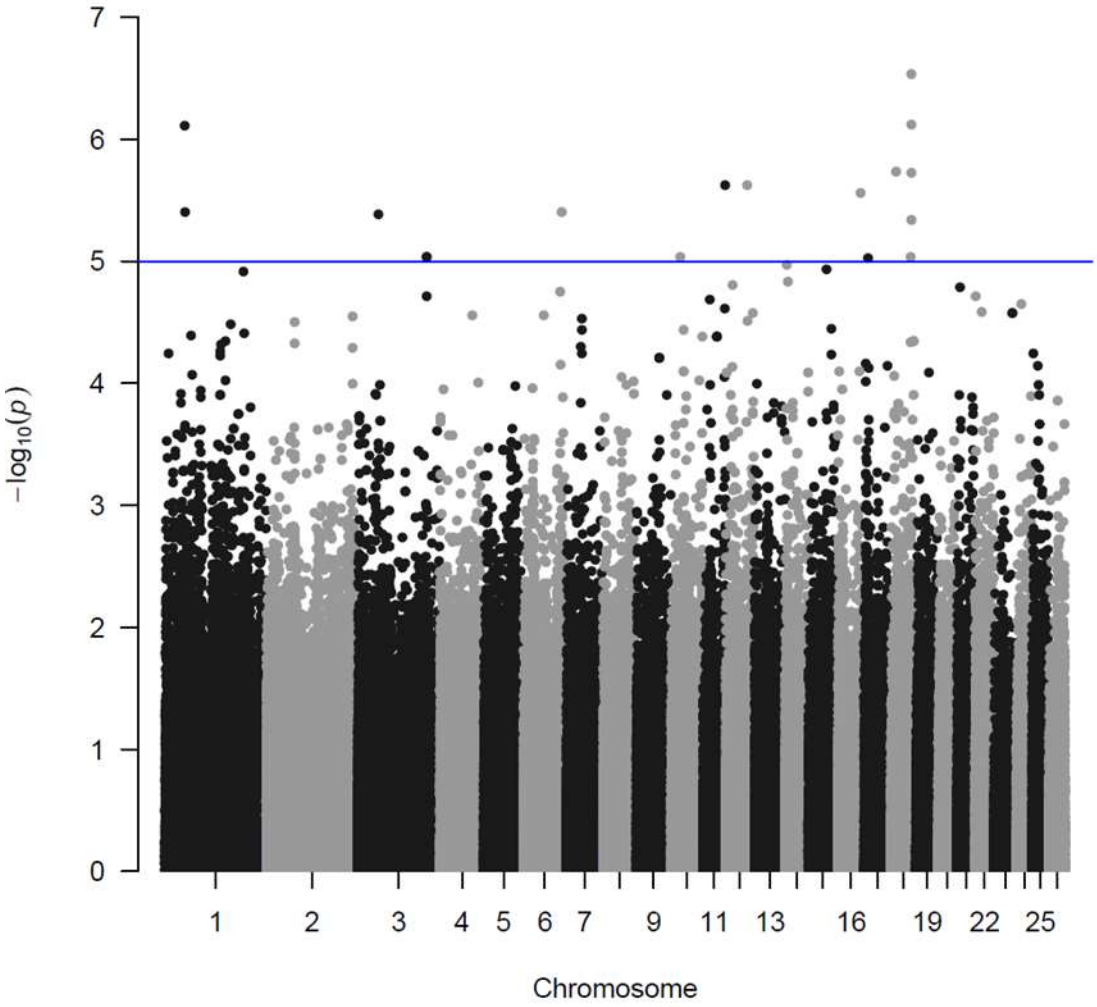

Figure 4: Results of genome-wide association study for dag. Blue line indicating the chromosome-wide significance threshold.

For none of the traits, any outstanding significant SNP peaks formed by multiple SNP were observed, but 32 SNP in total were found to exceed the chromosome-wide significance level across the four traits (Table 10). All of these SNP were checked for the closest genes up or downstream of their position on the chromosome.

For STRIKE, two of the SNP above the chromosome-wide significance threshold on chromosome 3 are in proximity of the gene TFCP2 (Table 10). It can only be hypothesized that it has a functional relationship to breech flystrike, but it is central to various physiological processes because "it encodes a transcription factor that binds the alpha-globin promoter and activates transcription of the alpha-globin gene. The protein regulates erythroid gene expression, plays a role in the transcriptional switch of globin gene promoters, and it activates many other cellular and viral gene promoters" (https://www.genecards.org/Search/Keyword?queryString=TFCP2). In humans, the alpha-globin gene is involved in rare disorders called Alpha-Thalassemia, which is a blood disorder that reduces the production of haemoglobin (https://rarediseases.info.nih.gov/diseases/621/alpha-thalassemia). In 5-29% of people it is associated with abnormalities of the immune system. In Altamurana sheep, a breed of

southern Italy, polymorphisms in the alpha-globin gene have been found, associated with phenotypic variation in red blood cells similar to the human condition (Pieragostini et al. 2003). The authors discuss the link between blood feeding parasites and variation in red blood cells. None of the other genes had a function that appeared of interest to the traits.

Table 10. SNP above chromosome-wide significance level for STRIKE, BCOV, BWRK and DAG and closest gene on the latest Ensemble gene annotation (Ovis_aries.Oar_v3.1.94).

Trait	Chr	Position	SNP name	Closest gene
STRIKE	17	12245506	OAR17_12245506	ZNF827, zinc finger protein 827
STRIKE	17	12247487	OAR17_12247487	ZNF827, zinc finger protein 827
STRIKE	2	3853868	OAR2_3853868	Non-coding RNA
STRIKE	3	134926005	OAR3_134926005	TFCP2, transcription factor CP2
STRIKE	3	134927160	OAR3_134927160	TFCP2, transcription factor CP2
STRIKE	6	15384604	OAR6_15384604	ELOVL6, ELOVL fatty acid elongase 6
STRIKE	9	75913254	OAR9_75913254	YWHAZ, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
BWRK	1	268303449	OAR1_268303449	AK5, Adenylate Kinase
BWRK	5	74469983	OAR5_74469983	ENSOARG00000010398
BWRK	15	46342270	OAR15_46342270	OR52L1, olfactory receptor 52L1
BWRK	24	9373513	OAR24_9373513	TEKT5, tektin 5
BWRK	26	30740352	OAR26_30740352	Non-coding RNA
BCOV	2	158913960	OAR2_158913960	LYPD6, LY6/PLAUR domain containing 6
BCOV	2	162810032	OAR2_162810032	Non-coding RNA
BCOV	4	63902055	OAR4_63902055	KBTBD2, kelch repeat and BTB domain containing 2
DAG	1	53116651	OAR1_53116651	DNAJB4, DnaJ Heat Shock Protein Family (Hsp40) Member B4
DAG	1	53830532	OAR1_53830532	KCNJ6, Potassium voltage-gated channel subfamily J Member
DAG	3	56455584	OAR3_56455584	ABCD2, ATP binding cassette subfamily D member 2
DAG	3	188049305	OAR3_188049305	ITPR2, inositol 1,4,5-trisphosphate receptor type 2
DAG	3	188079614	OAR3_188079614	ITPR2, inositol 1,4,5-trisphosphate receptor type 2
DAG	6	103817831	OAR6_103817831	MSX1, msh homeobox 1
DAG	10	24587722	OAR10_24587722	POSTN Periostin
DAG	11	60209796	OAR11_60209796	ABCA5, ATP-binding cassette subfamily A member 5
DAG	12	58848320	OAR12_58848320	NPHS2, Podocin
DAG	16	61325555	OAR16_61325555	catenin delta 2
DAG	17	10636816	OAR17_10636816	EDNRA, endothelin receptor type A
DAG	18	14160099	OAR18_14160099	SLCO1B1, solute carrier organic anion transporter family member 3A1
DAG	18	53738524	OAR18_53738524	TOGARAM1, TOG array regulator of axonemal microtubules 1
DAG	18	56159080	OAR18_56159080	solute carrier family 24 member 4
DAG	18	56159839	OAR18_56159839	SLC24A4, solute carrier family 24 member 4

DAG	18	56175702	OAR18_56175702	Non-coding RNA
DAG	18	56407757	OAR18_56407757	succinate dehydrogenase complex subunit D

The only cluster of several significant SNP that exceeded the chromosome-wide significance threshold were 4 SNP on chromosome 18 for DAG (56,159,080 – 56,407,757 bp). The analysis of the closest genes did not yield anything of particular interest. However, GOLGA5 has been positioned at 56,405,660 – 56,406,553 bp on chromosome 18, but was not picked up as the closest feature to any of the SNP. GOLGA5 codes for a protein family, the golgins, that are located in the golgi apparatus and apparently are involved in vesicle tethering and docking. The Golgi apparatus is involved in the assembly of peptides, such as antimicrobial peptides that are excreted from the gastric mucosa (Bulet *et al.* 2005), an antimicrobial protein that is involve in the first response to internal parasites in the intestine.

A number of other genes were found to be closest to the other significant SNP, but they did not appear to be of interest to breech flystrike or any of the indicator traits.

6.2 Genomic breeding values (GEBV)

Genomic breeding values were estimated for all four traits using a genomic relationship matrix. The distribution of GEBV are shown in Figures 5 – 8. The distribution for breech flystrike has two peaks, indicating the differentiation of GEBV for high and low breech flystrike resistance for this trait. It was expected to see the same for the indicator traits due to underlying genetic correlations, however, the GEBV for the other three traits display normal distributions. It is hypothesized that the reason is that the samples for genotyping across the two sites were specifically selected to have a good representation of struck vs not struck sheep across sire groups. Some of the indicator traits show very low variation in one of the environments, such as BCOV in WA and DAG in NSW, and variation was therefore confounded with SITE. Some of this variation could have been removed by adjusting for SITE in the model for analysis.

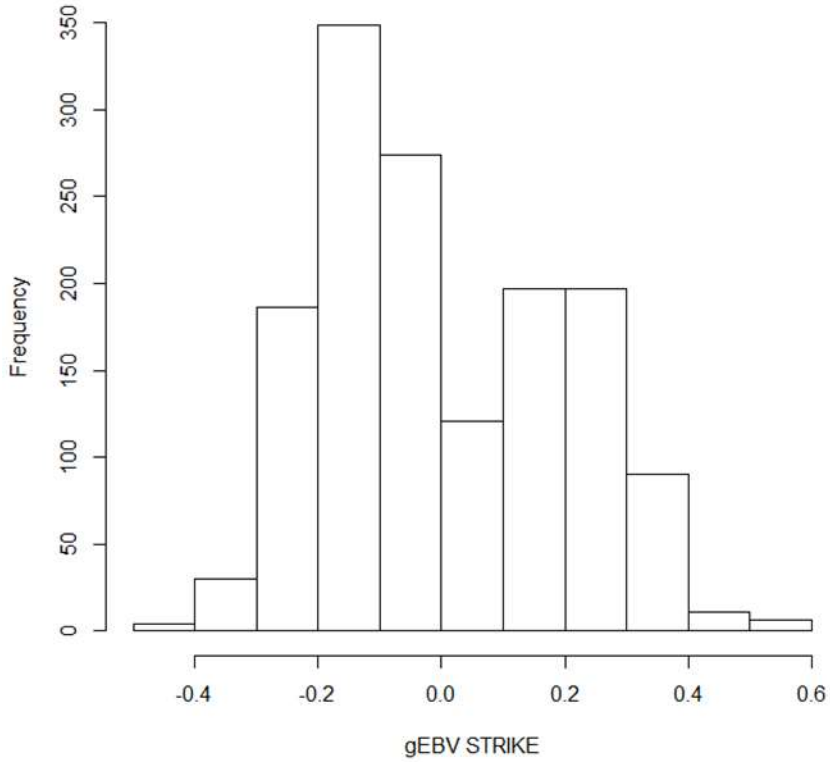

Figure 5: Distribution of genomic breeding values for STRIKE.

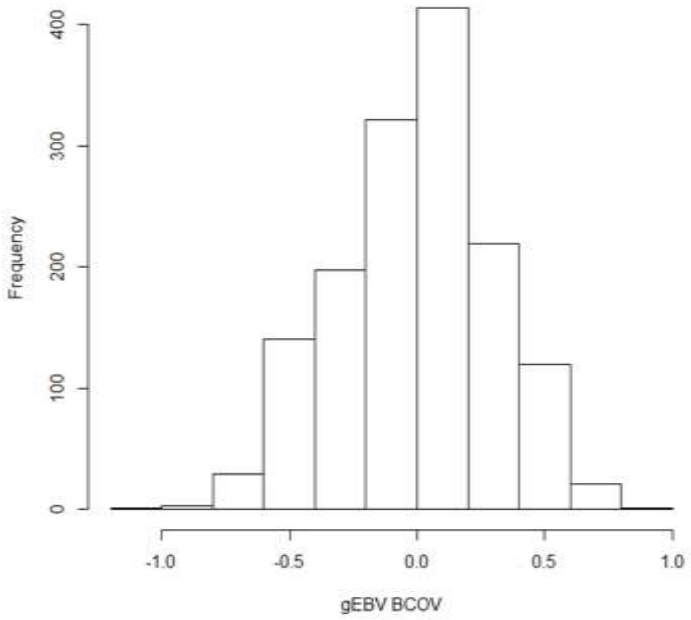

Figure 6: Distribution of genomic breeding values for BCOV.

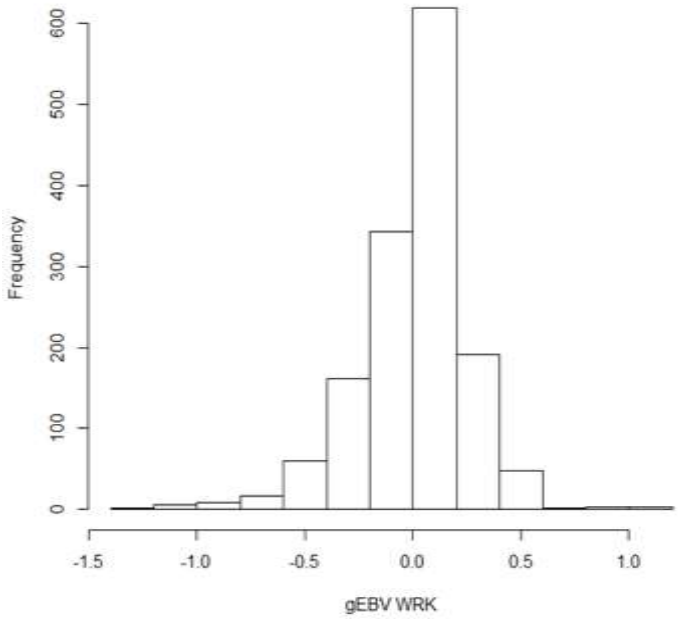

Figure 7: Distribution of genomic breeding values for WRK.

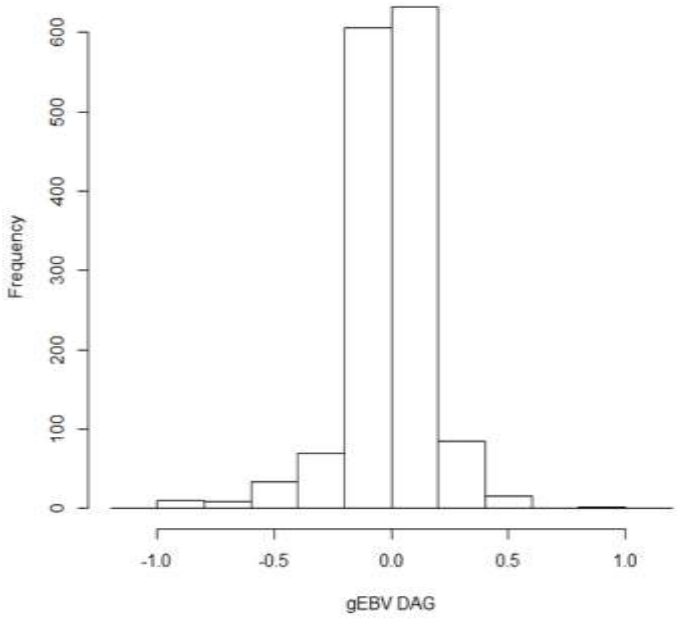

Figure 8: Distribution of genomic breeding values for DAG.

6.3 Accuracy of the genomic predictions (rGEBV)

The accuracy of the genomic prediction has been plotted for different numbers of animals in the reference population for a trait of a heritability of h2=0.3 in Figure 9 and Table 11. With the 1,500 animals that currently form the reference population, an accuracy of rGEBV = 0.33

could be achieved. Figure 9 demonstrates that this accuracy can be double with around 7,500 animals. Increasing the number of animals in the reference population to 10,000 has a diminished effect with only 0.05 higher than with 7,500 animals. However, Table 10 demonstrates the dependency of the accuracy to the heritability. For example, for a trait of heritability of h2=0.1, such as DAG in the NSW flock, the accuracy is rGEBV = 0.20. More animals are required in a reference population if the trait of interest is of low heritability.

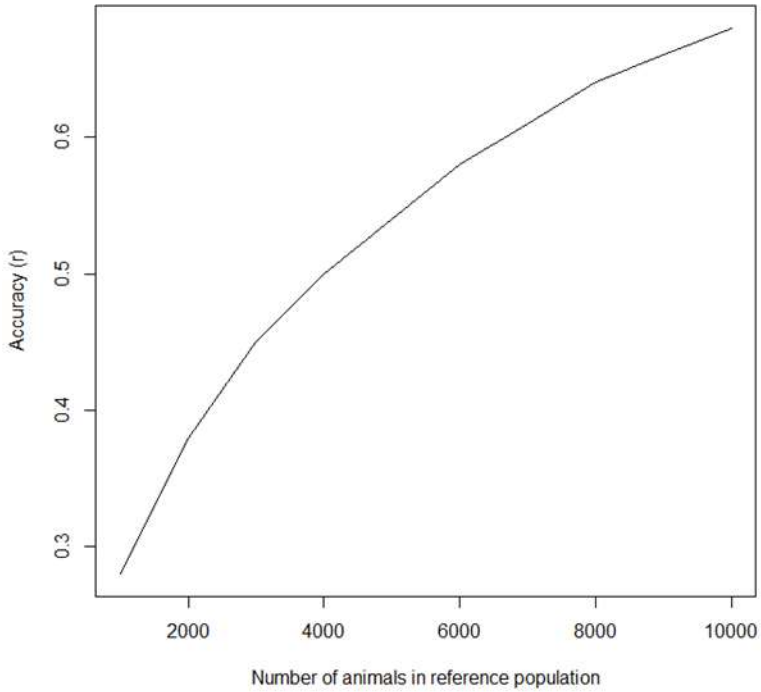

Figure 9: Accuracy (rGEBV) of genomic prediction as a result of number of animals in the reference population for a trait of h2=0.3, such as breech flystrike.

Table 11. Accuracy (rGEBV) for a reference population of 1,500 animals for traits of different heritability

Heritability	Accuracy
0.1	0.20
0.2	0.28
0.3	0.33
0.4	0.38
0.5	0.42

6.4 SNP effects for breech flystrike and indicator traits

Distributions of the solutions for the SNP effects are shown in Figures 10 and 11. Overall, the effects are very small, demonstrating that these effects need to be captured in a genomic breeding value as a tool for industry.

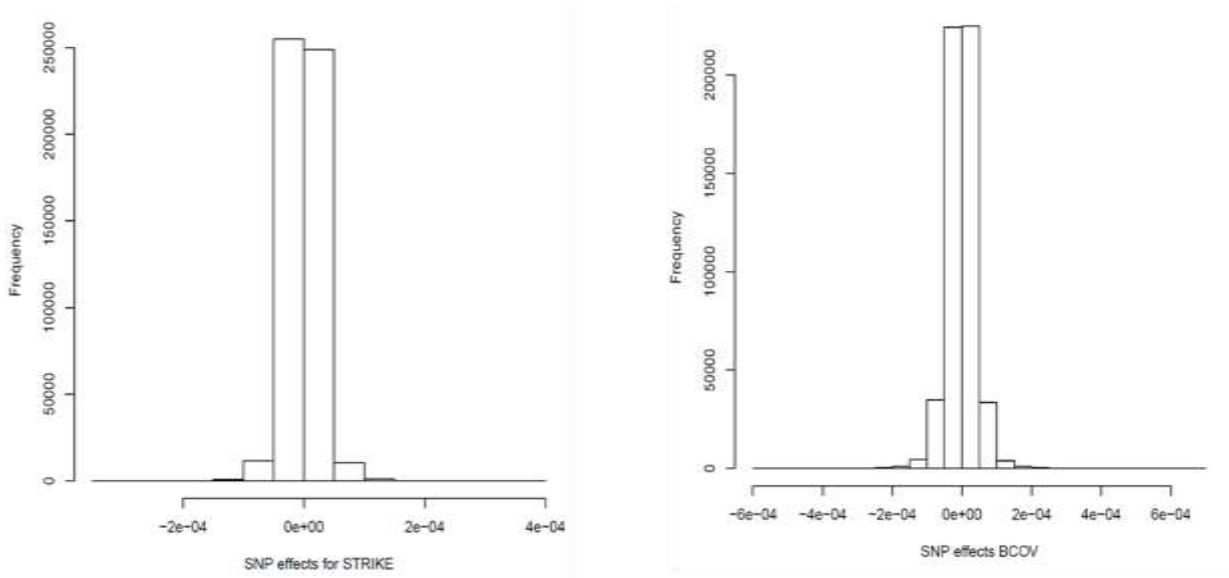

Figure 10: Distribution of SNP effects for STRIKE and BCOV.

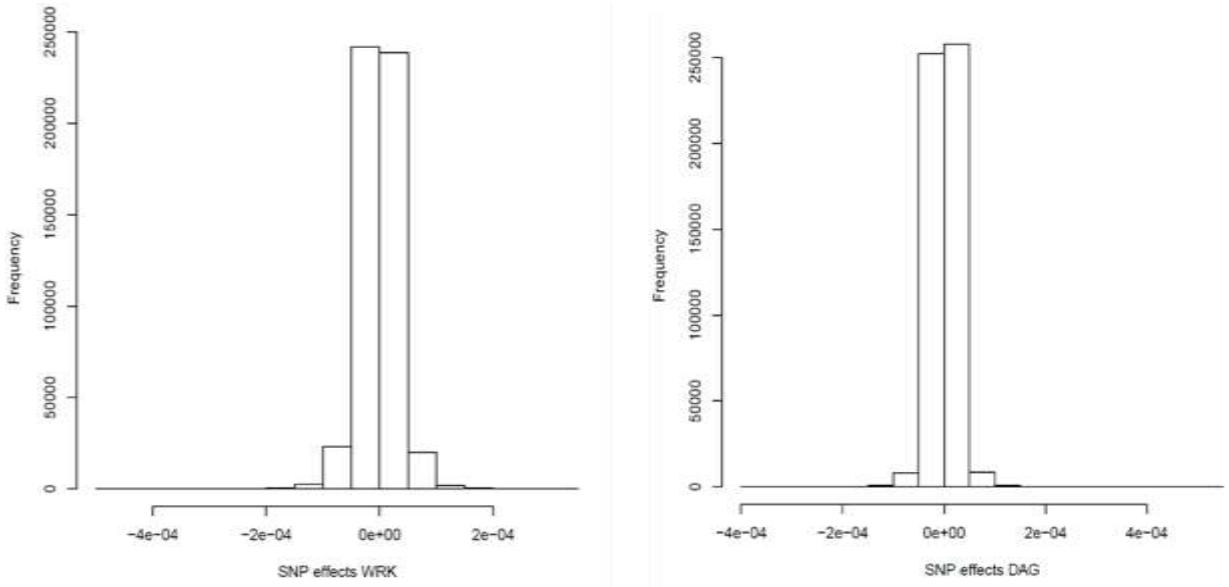

Figure 11: Distribution of SNP effects for WRK and DAG.

7 Discussion

Genomic analyses point at the most efficient approaches for the application of genomic information in breeding programs. These could be based on major genes, such as marker assisted selection, or based on genomic breeding values that draw on the variation explained by all SNP. Here it was hypothesised that a reference population and the estimation of GEBV would be the most efficient genomic approach to fast-track genetic improvement for breech flystrike resistance.

As a first step the existence of SNP that were significantly associated with STRIKE, BCOV, WRK or DAG was explored in a genome-wide association study. Overall, the significance profile for all four traits did not display any outstanding peaks. Although two regions with SNP of significance on chromosomes 3 and 18 have been located in proximity to interesting genes for STRIKE (TFCP2) and DAG (GOLGA5), they do not display peaks formed through a high level of significance of a cluster of several SNP, which would be expected of a region or gene that would harbours a major gene. However, it is possible that the data structure masks some variation and further investigation of these two regions would be warranted. E.g. DAG and BCOV did not display a lot of variation in the NSW and WA environment respectively. Therefore, splitting the data by site or choosing extreme animals (e.g. based on life time strike) are potential further analyses. Although, evidence that TFCP2 or GOLGA5 is weak, it might be worth exploring their role in STRIKE and DAG further.

The current genomic resource of the Breech Flystrike Resource flocks is an excellent start for a reference population. Ideally the accuracy of the GEBV needs to be increased to provide a reliable selection tool for industry and the reference population would also benefit from greater linkage with industry flocks. In order to build on the current reference population other aspects have to be taken into consideration. Breech flystrike remains expensive to measure, therefore it would be costly and impractical to run a specific reference population for this trait, although, the resource could be used for genomic predictions of other traits as well. Therefore, a dispersed reference population would be advantageous based on opportunistic sampling in flocks of key breeders, Merino Lifetime Productivity flocks, sire evaluation flocks - any flock with reliable records. However, in order to include commercial phenotypes in a reference population, it would need to be established first that breech flystrike in sheep that have been chemically treated for flystrike, correlates well with flystrike in untreated sheep. Other options of adding more commercial phenotypes could be explored, such as the collection of “pooled” phenotypes from commercial wool production operations. Based on a sufficient accuracy, it has been recommended that a future reference population should have at least 7,500 phenotypes, which would require the collection of further 6,000 phenotypes in addition to the Breechstrike resource. It is suggested that a time frame of 5 years would a maximum time frame to build a dispersed reference population. It would mean that 1,200 phenotypes need to be collected per year, which includes struck and not struck animals. Assuming 10 struck and 10 not struck records and samples are collected from each contributing property year, then 60 contributing properties would be required. The time frame of 5 years can be shortened with more contributing properties and more phenotypes.

8 Impact of Wool Industry – Now & in 5 Years' Time

The project demonstrated that the genomic resource of the Breech Flystrike Resource Flock from WA and NSW are a great resource that will have impact on the wool industry in the short and long term. In this study GEBV were estimated for breech flystrike resistance but they do not yet have the accuracy to be used as reliable selection tools for industry. The addition of the genomic data of the Breech Flystrike Resource flocks to the current genomic resource for Merino sheep will provide a great increase in accuracy for existing GEBV for wool production traits in the short term.

Within the next five years, a much larger dispersed reference population could be built that increases the accuracy of GEBV to a level that allows accurate direct selection on breech flystrike resistance directly and fast-track genetic improvement for this trait.

9 Conclusions and Recommendations

Genomic selection can provide substantial benefit to wool focussed Merino breeding programs (van der Werf 2009). Industry is in a position to build on the existing genomic breech flystrike resource to create reliable selection tools in a maximum time frame of 5 years that fast-track genetic progress in breech flystrike.

The recommendations arising from this report are:

- To establish the relationship between flystrike in treated and untreated sheep to explore the addition of commercial phenotypes to the reference population
- To establish a dispersed reference population for breech flystrike resistance of at least 7,500 animals (an increase of 6,000 animals) based on opportunistic sample collection and commercial phenotypes is formed over the next 5 years
- To conduct more analyses on the existing data set based on strategic subsets of the data (e.g. analysis for DAG only in the WA data).

10 Bibliography

AL-MAMUN, H.A., BERNARDES, P.A., LIM, D., PARK, B., GONDRO, C. (2017) A guide to imputation of low density single nucleotide polymorphism data up to sequence level. Journal of animal Breeding and Genomics 1(1): 59-68.

BULET, P., STOCKLIN, R., MENIN, L. (2005) Anti-microbial peptides: from invertebrates to vertebrates, Immunological reviews. https://doi.org/10.1111/j.0105-2896.2004.0124.x

BOICHARD, D., DUCROCQ, V., CROISEAU, P., FRITZ, S. (2016) Genomic selection in domestic animals: Principles, applications and perspectives Comptes Rendus Biologies 339 (7-8): 274-277.

DAETWYLER, H.D., PONG-WONG, R., VILLANEUVA, B., WOOLLIAMS, J.A. (2010). The impact of genetic architecture on genome-wide evaluation methods. Genetics 185 (3), 1021-1031.

DAETWYLER, H.D., HICKEY, J.M., HENSHALL, J.M., DOMINIK, S., GREDLER, B., VAN DER WERF, J.H.J., Hayes, B.J. (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population, Animal Production Science 50: 1004–1010.

DAS, S., FORER, L., SCHONHERR, S, SIDORE, C., LOCKE, A.E., KWON, A., VRIEZE, S.I., CHEW, E.Y., LEVY, S., MCGUE, M., SCHLESSINGER, D., STAMBOLIAN, D., LOH, P.R., IACONO, W.G., SWAROOP, A., SCOTT, L.J., CUCCA, F., KRONENBERG, F., BOEHNKE, M., ABECASIS, G.R., FUCHSBERGER, C., (2016) Next-generation genotype imputation service and methods. Nature Genetics 48: 1284-1287.

DEKKERS, J.C.M. (2004). Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. Journal of Animal Science 82: E313-E328.

DEKKERS, J.C.M. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. Journal of Animal Breeding and Genetics 124: 331-341.

GODDARD, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136, 245-257.

GODDARD, M.E., HAYES, B.J., MEUWISSEN, T.H.E. (2011) Using the genomic relationship matrix to predict the accuracy of genomic selection. Journal of Animal Breeding and Genetics 128: 409-421.

GREEFF, J.C., KARLSSON, L.J.E. , SCHLINK, A.C. (2014) Identifying indicator traits for breech strike in Merino sheep in a Mediterranean environment. Animal Production Science 54: 125-140.

HOWIE, J.M., FUCHSBERGER, C., STEPHENS, M., MARCHINI, J., ABECASIS, G.R., (2012) Fast and accurate genotype imputation in genome-wide association studies. Plos Genetics 5, e10005269.

HU, Z.L., PARK, C.A., WU, X.L., REECY, J.M. (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41:D871–9.

LOH, P.R., DANECEK, P., PALAMARA P.F., FUCHSBERGER, C., RESHEF, Y.A., FINUCANE, H.K., SCHOENHER, S., FORER, L., MCCARTHY, S., ABECASIS, G.R., DURBIN, R., PRICE A.L. (2016) Reference-based phasing using the haplotype Reference Consortium panel. Nature Genetics 48: 1443-1448.

PIERAGOSTINI, E., PETAZZI, F., DILUCCIA, A. (2003). The relationship between the presence of extra alpha-globin genes and blood cell traits in Altamurana sheep. Genetic Selection Evolution 35 (Suppl. 1) S121–133. DOI: 10.1051/gse:2003021

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M.A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P.I., DALY, M.J., SHAM, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analysis. American Journal of Human Genetics 81: 559-575.

R CORE TEAM (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ .

SMITH, J.L.; BREWER, H.G.; DYALL, T. (2009) Heritability and phenotypic correlations for breech strike and breech strike resistance indicators in Merinos. In Proceedings of the Association for the Advancement of Animal Breeding and Genetics 18: 334-337

VAN DER WERF, J.H.J. (2009) Potential benefit of genomic selection in sheep. Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics, Adelaide, Australia. pp 38–41.

YANG, J., LEE S.H., GODDARD M.E. and VISSCHER P.M. (2011) GCTA: a tool for Genome-wide Complex Trait Analysis. American Journal of Human Genetics 88(1): 76-82.

ZENG, J., PSZCZOLA, M., WOLC, A., STRABEL, T., FERNANDO, R.L., DORIAN J GARRICK, D.J., DEKKERS, J.C.M. (2012). Genomic breeding value prediction and QTL mapping of QTLMAS2011 data using Bayesian and GBLUP methods. BMC Proceedings 2012, 6 (Suppl 2), S7.