

## PREDICTION ACCURACIES FOR POLLED AND HORNED MERINO SHEEP USING DIFFERENT GENETIC MODELS

N. Duijvesteijn<sup>1,2</sup>, S. Weerasinghe<sup>1</sup> and J.H.J. van der Werf<sup>1,2</sup>

<sup>1</sup>School of Environmental and Rural Science, Univ. of New England, Armidale, NSW, 2351

<sup>2</sup>Cooperative Research Centre for Sheep Industry Innovation, Armidale, 2351

### SUMMARY

Previous studies have mapped the responsible locus for the polled phenotype to the 3' region of *RXFP2* at ovine chromosome 10. SNPs to determine whether the insertion is present are neither on the Ovine 50K nor on the OvineHD. In this study we tested different strategies for prediction of the horn phenotype, including single SNP, multiple SNP haplotypes and GBLUP. In total, 4,001 Merino sheep with HD genotype information were used. Prediction accuracies were calculated for each sex separately. Models with the highest prediction accuracies for horned used either single SNPs or 3-SNP haplotypes and also included a polygenic effect estimated based on traditional pedigree relationships. The accuracies of predicting the 'horned' phenotype were 0.338 for females and 0.724 for males. For predicting 'polled' phenotype, the best models were the same but included a genomic relationship matrix. The accuracies were 0.713 for females and 0.618 for males. Results show that prediction accuracy is high using a single SNP, although not unity as the causative mutation is not genotyped, but likely also because females show incomplete penetrance. As long as there is no genotype from a single SNP causative mutation, additional information through pedigree is valuable for the prediction of horned and polled phenotype.

### INTRODUCTION

The genetic background of the polled phenotype has long been studied in horned species such as cattle and sheep (Castle 1940, Georges et al. 1993). The causative mutation in sheep has been mapped to chromosome 10 (Johnston et al. 2011). A 1.78-kb insertion in the 3'-untranslated region of *RXFP2* causes the polled phenotype, as described by Wiedemar and Drogemuller (2015). However, this insertion is not completely explaining the phenotype in different sheep breeds (Lühken et al. 2016). The mode of inheritance is complex as expression differs between sexes and there is not yet a single locus model with complete penetrance. Currently the causative mutation is neither on the Illumina Ovine 50K chip nor on the OvineHD 600K chip. SNPs close to the region of insertion are currently used to predict the phenotype. The aim of this study is to test various strategies for predicting horned or polled phenotypes, including single SNP, multiple SNP haplotypes and SNPBLUP.

### MATERIALS AND METHODS

**Population and phenotypic data.** The data consisted of purebred Merino sheep including Dohne Merino and polled Merino. The phenotype recorded was polled, scurs, knobs or horns, which was analysed as polled / non-polled and horned / non-horned. In total, 4,001 sheep were used. Table 1 shows the distribution of polled and horned status between the two sexes.

**Table 1. Number of observed phenotypes for male and female Merinos**

Sex	Polled	Non-Polled	Horned	Non-Horned
Female	1325	1123	88	2360
Male	1042	511	481	1072

**Genotypes.** Of all 4,001 animals in the dataset, 3,708 were genotyped with the Ovine 50K. The remaining animals were genotyped with the Ovine 12K and imputed up to 50K. All 4,001 animals were further imputed up to 600K. All 600K genotyped animals (~2300) were used for imputation, including 445 animals from the data set used in this study. In total 510,175 SNPs passed quality control and 17,280 SNPs were located at OAR10.

**Statistical analysis.** We applied three methods to predict the phenotype polled or horned status. To select the best single SNP for prediction, we ran a local GWAS for chromosome 10 (OAR10). The single SNP was either fitted solely (base model), or together with a polygenic effect (fitted either by a traditional pedigree or by a genomic relationship matrix).

The second method was using haplotypes. A haplotype was formed using the most significant SNPs from the single SNP GWAS (3, 5 or 10 SNPs). Genotype data was phased using EAGLE. Only haplotypes with a frequency >1% were fitted in the model, and otherwise placed in a bin (sum of all low frequency haplotypes). The number of haplotypes formed from three, five or 10 SNPs, was equal to three, three, and seven, respectively.

The third method was applying a GBLUP analysis using a GRM based on all SNPs from the 600K (Yang et al. 2010) or only those SNPs from OAR10. Additionally, a dominance relationship matrix was added to the model based on the same two sets of SNPs (Zhu et al. 2015). Breeding values from the additive and dominance GRM were summed to get the predicted phenotype.

**Mode of inheritance.** We compared various models where the mode of inheritance was investigated. The model including a sex-dependent effect for the additive and dominance variance resulted in the best predictions (results not shown). Therefore, whenever possible, this mode of inheritance is used for prediction.

**Validation.** A fivefold cross-validation was performed. In each replicate, 20% of the data was randomly blinded and the phenotype was predicted. Prediction accuracy was defined as the correlation of the breeding value with the 0/1 phenotype.

## RESULTS

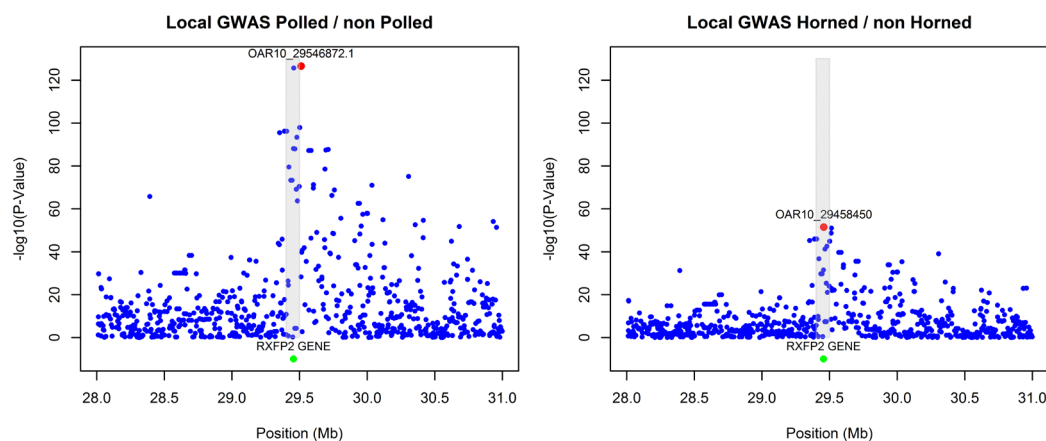
The local GWAS for polled / non polled and horned / non horned clearly indicates the known region with highly significant associations around 29.5 Mb (Figure 1). The most significant SNP for polled was OAR10\_29546872.1 which differed from the most significant SNP for horned which was OAR10\_29458450, although both SNPs are in high LD ( $r^2=0.985$ ). Those SNPs were used to perform the single SNP analyses.

In Table 2 and 3, the frequencies of the genotypes with phenotype polled and horned by sex is shown. Using frequencies to determine the polled or horned status across the validation gives an average prediction accuracy of 0.71 for horned and 0.63 for polled (base model, Table 4). In Table 4, the results of the different genetic models are shown. The highest accuracy for predicting polled was achieved when using a GRM additional to the single SNP, resulting in a correlation of 0.713 for females and 0.618 for males. The highest accuracy for predicting horned was by using pedigree relationships additional to the single SNP, which resulted in a correlation of 0.338 for females and 0.724 for males. Models where haplotypes were used resulted in similar accuracies as the single SNP approach. Haplotypes formed from 3 and 5 SNPs (hap3 and hap5), gave very similar prediction accuracies, where hap10 had a lower prediction accuracy.

## DISCUSSION AND CONCLUSION

The most significant SNP for polled and horned were very close to the causative mutation, (OAR10\_29546872.1: 29512572 and OAR10\_29458450: 29458450) of which OAR10\_29546872.1 has been used by the Sheep CRC (J. van der Werf, pers. comm). Dominik et al. (2012) found a SNP OAR10\_29389966\_X.1 to be most predictive in Merino sheep. This SNP was also in the top10 of most significant SNPs in our GWAS. The reported SNP by Johnston et al. (2011)

OAR10\_29448537.1 did not occur in the top100 SNPs of our GWAS.



**Figure 1. Local GWAS plot for the traits polled and horned of OAR10. The grey rectangle indicates the location of the gene *RXFP2* (29.4-29.5 mb). The most significant SNP is indicated in red**

**Table 2. Frequencies of the SNP OAR10\_29546872.1 per sex for the phenotype polled and probabilities for being polled, and frequencies of the SNP OAR10\_29458450 per sex for the phenotype horned and probabilities for being horned**

Sex	Genotype	Polled			Horned		
		Non Polled	Polled	Probability Polled	Non Horned	Horned	Probability Horned
Female	0	1058	174	<b>0.14</b>	1151	81	<b>0.07</b>
	1	353	811	<b>0.77</b>	1047	6	<b>0.01</b>
	2	25	138	<b>0.84</b>	162	1	<b>0.01</b>
Male	0	675	29	<b>0.04</b>	229	475	<b>0.67</b>
	1	340	385	<b>0.53</b>	719	6	<b>0.01</b>
	2	27	97	<b>0.78</b>	124	0	<b>0.00</b>

A model including pedigree information additional to the single SNP or haplotype had a better prediction accuracy compared to using only a single SNP for the prediction of both horned and polled. When the single SNP was not explicitly fitted (local GRM model), the prediction was reduced. For the trait polled and horned highly predictive SNPS close to the known causative mutation should be modelled explicitly. Applying methods which shrink all SNP effects equally like GBLUP will therefore have a lower prediction accuracy in the presence of a large QTL. Mixture models such as Bayes B or C, should perform better.

Clearly prediction accuracy was not close to one, in spite of highly significant SNPs close to a known causative mutation. This indicates that the most significant SNP is not in complete LD with the causative mutation or it does not confer complete penetrance. This is also indicated by the explained variance from the genotypes. For the trait horned, 85% of the phenotypic variance was explained by the single SNP, and 95% of the phenotypic variance when pedigree was also included. For the trait polled, 67% of the phenotypic variance was explained by the single SNP, and 80% of the phenotypic variance when pedigree was also included.

**Table 4. Prediction accuracies for horned and polled for the different models for the whole dataset or split by sex with or without fitting a polygenic effect (Ped)**

Method	Ped	Correlation Horned			Correlation Polled		
		Average	Female	Male	Average	Female	Male
Single SNP	-	0.711	0.173	0.721	0.630	0.644	0.581
Single SNP	A	0.723	0.338	0.724	0.644	0.665	0.580
Single SNP	GRM	0.719	0.302	0.723	0.686	0.712	0.617
Hap3	A	0.723	0.324	0.726	0.647	0.671	0.579
Hap5	A	0.721	0.324	0.723	0.646	0.673	0.573
Hap10	A	0.673	0.224	0.681	0.632	0.658	0.560
Hap3	GRM	0.722	0.302	0.727	0.687	0.713	0.618
Hap5	GRM	0.721	0.300	0.725	0.676	0.703	0.604
Hap10	GRM	0.696	0.285	0.691	0.670	0.697	0.599
GRM OAR10	GRM	0.391	0.226	0.628	0.617	0.657	0.574
GRM	GRM	0.380 <sup>1</sup>	0.273	0.561	0.580 <sup>2</sup>	0.620	0.526

<sup>1</sup>Four of the five replicates converged. <sup>2</sup>Only two of the five replicates converged

Differences between males and females have been described previously (Dolling 1961, Dominik et al. 2012) in Merino sheep. Possibly incomplete penetrance is causing the sporadic horned phenotype in females, and makes prediction more difficult (prediction accuracy 0.338 vs 0.724 for horned in females and males).

Different approaches to validate the different genetic models (e.g. regress back to 0/1 trait by using a threshold on the predicted phenotypes) could clarify results further, and will be investigated additionally.

To conclude, prediction of polled and horned is already successful using a single SNP (~0.7), although not 1 as the causative mutation is not genotyped (on the new 15K Ovine chip it should be present), but likely also because females show incomplete penetrance. Additional information through pedigree is valuable for the prediction of the horned and polled phenotype as long as the causative mutation is not genotyped.

#### ACKNOWLEDGEMENTS

The authors acknowledge the contributions of people from breeders and many CRC participants that contributed to the Sheep CRC Information Nucleus flocks. Also Klint Gore is acknowledged for his help on retrieving data.

#### REFERENCES

- Castle, W. (1940). *J. Hered.* **31**: 486-487.  
 Dolling, C. (1961). *Crop and Pasture Sci.* **12**: 535-361.  
 Dominik, S., J. Henshall and B. Hayes (2012). *Anim. Genet.* **43**: 468-470.  
 Georges, M., R. Drinkwaterz, T. Kingz, et al. (1993). *Nat. Genet.* **4**.  
 Johnston, S. E., J. McEwan, N. K. Pickering, et al. (2011). *Mol. Ecol.* **20**: 2555-2566.  
 Lühken, G., S. Krebs, S. Rothhammer, et al. (2016). *Genet. Sel. Evol.* **48**: 78.  
 Wiedemar, N. and C. Drögemüller (2015). *Anim. genet.* **46**: 457-461.  
 Yang, J., B. Benyamin, B. P. McEvoy, et al. (2010). *Nat. Genet.* **42**: 565-569.  
 Zhu, Z., A. Bakshi, A. A. Vinkhuyzen, et al. (2015). *Am. J. Hum. Genet.* **96**: 377-385.