Short communication

Preliminary genome-wide association study for wet-dry phenotype in smallholder ovine populations in South Africa

A.H. Molotsi[1#], J.F. Taylor[2], S.W.P. Cloete[1,3], F. Muchadeyi[4], J.E. Decker[2,5], L. Sandenbergh[3] & K. Dzama[1]

[1]Department of Animal Sciences, University of Stellenbosch, Private Bag X1, Matieland, Stellenbosch, 7602, South Africa
[2]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA
[3]Directorate Animal Sciences: Elsenburg, Private Bag X1, Elsenburg, 7607, South Africa
[4]Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa
[5]Informatics Institute, University of Missouri, Columbia, MO 65211, USA

(6 February 2017; Accepted 20 March 2017; First published online 10 April 2017)

Abstract

The aim of this study was to identify single nucleotide polymorphisms (SNPs) associated with genomic region underlying variation in the binomial reproductive trait 'wet-dry' in sheep. The wet-dry phenotype was used to represent the reproductive status of the ewes, divided into two categories, dry (ewes that did not lamb or that lost a lamb) and wet (ewes that had lambed and had at least one suckling lamb). Wet-dry records were obtained from smallholder farmers (n = 176) and Nortier Research Farm (n = 131) for the 2014 breeding season. Ages of the ewes ranged from 1 year to 6+ years. Data from 307 individuals were analysed, of which 172 Dorpers and 4 White Dorpers were from smallholder sheep flocks and 48 Dorpers, 46 Namaqua Afrikaners, 26 South African Mutton Merinos, 4 South African Mutton Merino x Dorper and 7 Dorper x South African Mutton Merino crossbreds were from the research farm. A logistic regression model was fitted to adjust the data for the fixed effects of farm, breed, and age of the ewe and weight at mating as a covariate. Linkage disequilibrium (LD) and inbreeding coefficient were estimated using PLINK. Association analysis was performed using the genome-wide efficient mixed-model association package (GEMMA) to determine whether any significant SNPs were associated with the wet-dry reproductive trait. The wet-dry phenotype differed significantly between the smallholder (0.63 ± 0.04) and research farm flocks (0.79 ± 0.04). Genome-wide LD across all populations was r^2 = 0.36. Dorpers from the smallholder flock exhibited rapid LD decay versus the resource ovine populations. Inbreeding levels were also lower for the smallholder flock (4 ± 0.003%) versus the research flock (13 ± 0.008%). No significant SNPs were identified after correction for false discovery rate. The heritability estimate for wet-dry using SNP information was 0.24. This estimate concurs with the literature and indicates the possibility of using genomic selection to improve reproduction in smallholder sheep flocks.

Keywords: conception rate, Dorper, heritability, Namaqua Afrikaner, reproduction, South African Mutton Merino
Corresponding author: annelind@sun.ac.za

Smallholder sheep farmers in South Africa have been reported to have flocks with low reproductive performance (Marais, 2007; Grobler, 2010). This may be owing to many factors, including nutrition, management and genetics. Genetic selection plays an important role in establishing permanent change that is sustainable for the long term in improving traits related to fitness. Fitness traits are linked to the reproduction of animals and can be measured with various indicator traits. Reproduction can be measured by composite traits such as number of lambs born per ewe lifetime or the number of lambs weaned per ewe lifetime, or by components traits such as litter size, fertility, conception rate and mothering ability (Zishiri *et al*., 2013). Quantifying and measuring fitness traits in smallholder production systems can be challenging owing to a lack of recordkeeping. It is therefore necessary to use easy-to-measure indicator traits that involve minimal recordkeeping and input costs.

URL: http://www.sasas.co.za
ISSN 0375-1589 (print), ISSN 2221-4062 (online)
Publisher: South African Society for Animal Science
http://dx.doi.org/10.4314/sajas.v47i3.9

The wet-dry phenotype (Fourie & Cloete, 1993) is recorded from an udder examination performed during the marking of recently born lambs or at the weaning of lambs. Wet-dry refers to whether a ewe is lactating or not, and can be used as an indicator of reproductive performance of ewes in low-input farming systems. The wet-dry phenotype is a composite trait and includes conception rate and mothering ability. Heritabilities have yet to be estimated for this trait in South Africa. Heritabilities have previously been estimated for Australian Merino flocks and range from 0.09 to 0.17 for wet-dry recorded at weaning and 0.04 to 0.11 when recorded at lamb marking (Lee *et al*., 2010). Heritability estimates for the component conception rate/fertility traits range from 0.01 to 0.30 (Iniguez *et al*., 1986; Tosh *et al*., 2002; Vatankhah *et al*., 2006; Piwczynski & Kowaliszyn, 2013). The underlying genetic regulation of reproduction has not been fully elucidated. However, some genes that influence fecundity and oestrous cycle in sheep have been identified. Genes affecting litter size in sheep include the *BMPR1B/FecB* mutation on chromosome 6, *GDF9* on chromosome 5 and *BMP15* on the X chromosome (Montgomery *et al*., 2001; Souza *et al*., 2001; Davis, 2004; Juengel *et al*., 2004; McNatty *et al*., 2005; Polley *et al*., 2010). Genes influencing the oestrous cycle include *PGFS* (*AKR1B5*/*AKR1C3*), *PGES*, *PGFR,* and *PTGS2* (Kumar *et al*., 2013). Genotyping using SNPs makes it possible to investigate population structure and establish pedigree relationships among animals. Linkage disequilibrium (LD) in a population is important to consider for genome-wide association studies (GWAS). This is because LD is the ability of an allele from one marker to predict the allelic status of another marker (Meadows *et al*., 2008). A preliminary genome-wide association study was conducted to investigate whether the wet-dry phenotype was influenced by variation in any of these genes. If indications of associations were found, then these genes could be further investigated to identify variants that could be used in marker-assisted selected programmes to improve reproduction in smallholder farming operations.

The aim of this study therefore was to identify SNPs associated with quantitative trait loci (QTL) underlying variation in the binomial wet-dry reproductive trait.

The sites of sample collection were Nortier Research Farm and the Ebenheaser smallholder community in Western Cape, South Africa. Ethical clearance was obtained from the Departmental Ethics Committee for Research on Animals (DECRA), namely approval numbers R12/53 for smallholder flocks and R14/100 for the experimental flock from the Western Cape Department of Agriculture. Nortier is located at 32º 5' 32.0833" S, 18º 18' 18.3000" E. Ebenheaser is located at 31º 35' 8.5856" S, 18º 14' 39.2442" E. These farms are located in the West Coast district and are classified under the Succulent Karoo Biome, described by Acocks (1988). The succulent Karoo consists mainly of succulent plants and is a winter rainfall area. The altitude ranges from sea level to 1500 m and the mean annual rainfall is between 20 and 350 mm. A winter lambing season was implemented, and wet-dry data were collected after the lambing season in July–August 2014. The wet-dry phenotype was used to represent the reproductive status of the ewes. The ewes were divided into two categories: dry (ewes that did not lamb, or that lost all their lamb(s) and were thus not lactating) and wet (ewes that had lambed and were suckling one or more lambs). Wet-dry records were obtained from smallholders (n = 176) and the resource flock at Nortier Research Farm (n = 131) for the 2014 breeding season. The ages of the ewes ranged from 1 year to 6+ years. A logistic regression model was fitted in SAS (2012) to adjust the phenotypic data for the fixed effects of farm, breed, age of the ewe and mating weight as a covariate. The sheep sampled from the smallholder farms were identified in a fastSTRUCTURE analysis (Raj *et al*., 2014) as being 172 Dorpers (Dorpersm) and 4 White Dorpers. The sheep sampled from the research farm were identified as 48 Dorpers, 46 Namaqua Afrikaners (Namafr), 26 South African Mutton Merinos (SAMM), 4 South African Mutton Merino x Dorper (SAMMDX) and 7 Dorper x South African Mutton Merino (DSAMMX) crossbreds.

The DNA samples were genotyped using the Ovine 50K SNP bead chip. Genotypic data were analysed using PLINK (Purcell *et al*., 2007). Data for 307 individuals were analysed. Quality control was conducted by setting thresholds for minor allele frequency at 5%, Hardy Weinberg Equilibrium at P <0.001, and genotype call rate per animal at 95%. The data were pruned according to these criteria and 43 500 SNPs were retained and used in the analyses. Linkage disequilibrium and inbreeding coefficient were calculated with PLINK. Association analysis was performed using the genome-wide efficient mixed-model association package (GEMMA) (Zhou & Stephens, 2012), to determine whether any significant large-effect SNPs were associated with the wet-dry reproductive trait.

Breed proportions were fit as covariates, and the farm of origin (smallholder or research farms) was fit as a covariate in the model for the genotypic data. The genetic similarity matrix was estimated using GEMMA. A univariate linear mixed model was fit in GEMMA for testing marker associations with wet-dry accounting for population stratification and pedigree structure. Chip heritability was estimated as the proportion of the phenotypic variance explained by genotypes.

The model used was:

$$y = W\alpha + x\beta + \mu + \varepsilon \qquad \mu \sim MVN_n(0, \lambda_T^{-1}K) \text{ and } \varepsilon \sim MVN_n(0, {}_T^{-1}I_n)$$

Where: y is an n-vector of binary trait values (1 for wet, 0 for dry) for the n individuals
W = (w_1,….., wc) is an n x n matrix of covariate values corresponding to the fixed effects and including a column of 1s
α = is an c-vector of the corresponding fixed effect parameters including the intercept
x = n-vector of marker genotypes
β = effect size of the marker
μ = n-vector of random residual additive genetic effects
ε = is an n-vector of errors
$_T^{-1}$ = is the variance of the residual errors
λ = ratio of additive genetic to residual variance components
K is a known n x n relatedness matrix
I_n = is an n x n identity matrix
MVN_n = the n-dimensional multivariate normal distribution

R was used to plot the $-\log_{10}$(*P*-value) for each SNP genome-wide. Two methods were used to identify significant SNPs: i) using the *P*-value from the likelihood ratio test calculated using GEMMA; and ii) after adjusting the *P*-value to q-values using the q-value package in R (Storey & Tibshirani, 2003) to account for multiple testing. Identification of genes associated with significant SNPs was done using NCBI map viewer for *Ovis Aries* annotation release 102.

The genome-wide LD across all populations was $r^2 = 0.36$. The Dorpersm exhibited the most rapid LD decay with the Namafr breed achieving slower rates of decay (Figure 1). Average inbreeding was also estimated for the various breeds, namely 0.6% for the White Dorper, -7% for the DSAMMX, 1.2% for the SAMMDX, 3% for Dorpersm, 7% for the Dorper, 9% for the SAMM and 24% for the Namafr. Average inbreeding coefficient was 4 ± 0.003% for the smallholder flock and 13 ± 0.008% for the resource flock. Similar inbreeding coefficients and LD levels for LD pruned within each breed have been estimated for the Nortier resource flock, with the Namafr exhibiting the highest level of inbreeding and slow LD decay because of its low effective population size (Sandenbergh *et al*., 2015). The high negative inbreeding coefficient observed for the DSAMMX might be because these crossbred were F_1s and possess higher levels of heterozygosity. The Dorpers and White Dorper obtained from smallholder flocks exhibited low levels of inbreeding in comparison to the purebred Dorper, SAMM and Namafr obtained from the resource flocks. This could be due to the willingness of smallholder farmers to crossbreed. The inbreeding levels for the purebred Dorper and SAMM are in line with literature estimates (Sandenberg *et al*., 2015) and are as expected for close nucleus breeding schemes.

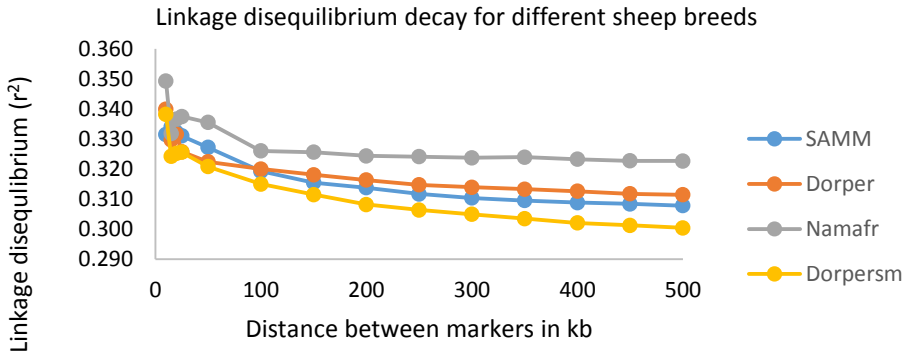

Figure 1 Linkage disequilibrium (r^2) decay with LD pruning within each breed

Farm significantly influenced the wet-dry phenotype. The proportion of wet ewes was significantly lower in the smallholder flocks (0.63 ± 0.04) relative to the research farm flocks (0.79 ± 0.04). These results are consistent with those in the literature, ranging from 0.72 in Merino ewes to 0.80 for SA Mutton Merino ewes in the Bredasdorp region (Fourie & Cloete, 1993). Heritability estimated with the SNP information was $h^2 = 0.24 \pm 0.14$ with the genetic variance component explained by the model as 0.14 and the environmental variance as 0.098. Lee *et al*. (2010) report heritability estimates of 0.09 to 0.17 for wet-dry at weaning and 0.04 to 0.11 for wet-dry at marking. The estimate of the current study falls in the range of heritability

estimates for conception rate (0.01 to 0.30) reported in the literature (Iniguez *et al*., 1986; Lee *et al*., 2009, Piwczynski & Kowaliszyn, 2013). The variation observed in heritability estimates is probably because of different methods used for estimation, and different variance components between breeds and production environments.

The genome-wide association study (Figure 2) suggests that SNP rs428728584 at 14,286,396 bp on chromosome 20 might be associated with wet-dry (P <0.0005, uncorrected for false discovery rate (FDR). However, after adjusting for multiple testing using a q-value at an false discovery rate (FDR) level of 5%, no SNPs were found to be significantly associated with wet-dry phenotype.

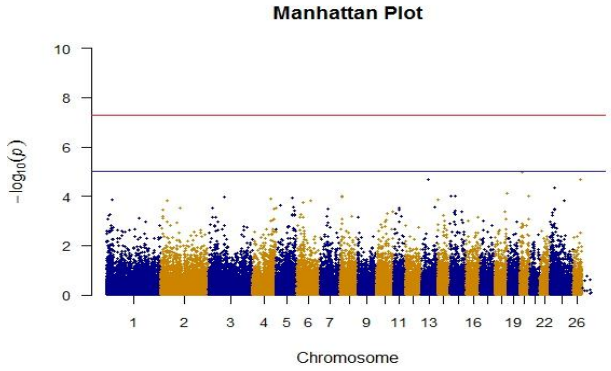

Figure 2 Manhattan plot indicating $-\log_{10}(P)$ values for single nucleotide polymorphisms associated with wet-dry phenotype

The high level of genome-wide LD, however, implies the possibility of identifying SNPs that form haplotype blocks that influence reproduction in sheep. Sandenbergh (2015) accordingly suggested that many loci of small to medium effect might influence the expression of reproduction traits in sheep. The different patterns of LD observed between the breeds implicate the number of markers that can be used for identification of QTL. The smallholder sheep had the shortest LD stretches and thus dense SNP marker panels should be used to identify significant SNPs. The genome-wide association analysis did not yield significant SNPs for wet-dry when a false discovery rate correction was applied to the SNP association *P*-values, which is because of the limited number of records available for this study.

The smallholder sheep flock exhibited low levels of inbreeding, which is promising for future genetic improvement in these populations. It is thus possible to use this approach to estimate additive genetic merit of animals in smallholder populations, allowing an opportunity for selection to improve genetic gain. Identifying causal variants related to the wet-dry phenotype, which is affordable and easy to measure in smallholder farming systems, could aid in improving the intensity of selection for this trait. This is the first study implicating the possible use of SNP data to investigate genetic structure in smallholder sheep populations in South Africa and to estimate a heritability estimate for wet-dry phenotype relevant to smallholder sheep farmers. The next step would be to improve the collection of phenotypic data and increase sample size for further genomic studies. The use of the Ovine SNP 600 K BeadChip or whole genome sequencing could also be of value for future studies.

Acknowledgements

The authors would like to acknowledge the smallholder farmers from the Ebenheaser community for allowing them access to their sheep flocks. They also thank the Western Cape Department of Agriculture for data and genotype information on the Nortier research sheep flock. They acknowledge funding obtained from NUFFIC for genotyping and funding from the National Research Foundation (NRF) Thuthuka Award.

Authors' Contributions

AHM undertook the GWAS analysis and wrote the paper. JFT assisted with the data analysis and editing the paper. SWP and KD were the main supervisors for the project and were responsible for project funding and obtaining phenotypic records from the research sheep flock and smallholder sheep flocks. FM assisted with the GWAS laboratory analysis and data analysis. JED assisted with data analysis and editing the paper. LS contributed genotype data on the research sheep flock and edited the paper. All authors have read and approved the final paper.

Conflict of Interest Declaration

The authors declare that they have no competing interests.

References

Acocks, J.P.H., 1988. Veld types of South Africa. Memoirs of the Botanical Survey of South Africa, 57. SANBI, Pretoria.

Davis, G.H., 2004. Fecundity genes in sheep. Anim. Reprod. Sci. 82-83, 247-253.

Fourie, A.J. & Cloete, S.W.P., 1993. Reproductive performance of commercial Merino, Dohne Merino and SA Mutton Merino flocks in the Southern Cape. S. Afr. J. Anim. Sci. 23, 104-110.

Grobler, H.J.F., 2010. Evaluation survey at extensive farms in the Central Karoo 2009. Agriprobe, June, 6-9.

Iniguez, L.C., Quass, R.L. & Van Vleck, L.D., 1986. J. Anim. Sci. 63, 1769-1778.

Juengel, J.L., Bodensteiner, K.J., Heath, D.A., Hudson, N.L., Moeller, C.L., Smith, P., Galloway, S.M., Davis, G.H., Sawyer, H.R. & McNatty, K.P., 2004. Physiology of GDF9 and BMP15 signalling molecules. Anim. Reprod. Sci. 82-83, 447-460.

Kumar, R., Ramteke, P.W., Nath, A., Pramod, R.K., Singh, S.P., Sharma, S.K. & Kumar, S., 2013. Role of candidate genes regulating uterine prostaglandins biosynthesis for maternal recognition of pregnancy in domestic animals. Int. Sch. Res. Notices Physiol. article ID 854572, http://dx.doi.org/10.1155/2013/854572.

Lee, G.J., Atkins, K.D. & Sladek, M.A., 2009. Genetic parameters for lifetime reproductive performance of merino ewes. Proc. 18th Conf. Assoc. Advanc. Anim. Breed. Genet, Barossa Valley, South Australia, Australia, pp. 382-385 ref.12

Lee, G.J., Sladek, M.A. & Atkins, K.D., 2010. Selection criteria for within flock selection to improve reproduction efficiency. http://www.livestocklibrary.com.au/bitstream/handle/1234/33711/SheepCRC_2_87.pdf?sequence=1 (accessed 21/01/2016)

Marais, P.G., 2007. Evaluation of the genetic potential for growth and wool production of typical Transkei ewes and rams. Research report of the Grootfontein Agricultural Development Institute, 2007. Private Bag x529, Middelburg 5900, South Africa. pp. 33-34.

McNatty, K.P., Smith, P., Moore, L.G, Reader, K., Lun, S., Hanrahan, J.P., Groome, N.P., Laitinen, M., Ritvos, O. & Juengel, J.L., 2005. Oocyte-expressed genes affecting ovulation rate. Mol. Cell. Endocrinol. 234, 57-66.

Meadows, J.R.S., Chan, E.K.F. & Kijas, J.W. 2008. Linkage disequilibrium compared between five populations of domestic sheep. BMC Genet. 9, 61-71.

Montgomery, G.W., Galloway, S.M., Davis, G.H. & McNatty, K.P., 2001. Genes controlling ovulation rate in sheep. Reprod. 121, 843-852.

Piwczynski, D. & Kowaliszyn, B., 2013. Heritability and breeding value of sheep fertility estimated by means of GIBBS sampling method using linear and threshold models. J. Cent. Eur. Agric. 14, 23-32.

Polley, S., De, S., Brahma, B., Mukherjee, A., Vinesh, P.V., Batabyal, S., Arora, J.S., Pan, S., Samanta, A.K., Datta, T.K. & Goswami, S.L., 2010. Polymorphism of BMPR1B, BMP15 and GDF9 fecundity genes in prolific Garole sheep. Trop. Anim. Health Prod. 42, 985-993.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C., 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am. J. Hum. Genet. 81, 559-575.

Raj, A., Stephens, M. & Pritchard, J.K. 2014.fastSTRUCTURE: Variational inference of population structure in large SNP data sets. Genet. 197, 573-589.

Sandenbergh, L., 2015. Identification of SNPs associated with robustness and greater reproductive success in South African Merino using SNP chip technology. PhD thesis. University of Stellenbosch.

Sandenbergh, L., Cloete, S.W.P., Roodt-Wilding, R., Snyman, M.A. & Van der Merwe, A.E., 2015. Genetic diversity and population structure of four South African sheep breeds. Proc. Assoc. Advmt. Anim. Breed. Genet. 21, 294-297.

SAS, 2012. SAS Institute Inc. Cary, NC.

Souza, C.J.H., MacDougall, C., Campbell, B.K. & McNeilly, A.S., 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. J. Endocrinol. 169, R1-R6.

Storey, J.D. & Tibshirani, R., 2003. Statistical significance for genome-wide experiments. Proc. Natl. Acad. Sci. USA. 100, 9440-9445.

Tosh, J.J., Wilton, J.W. & Kennedy, D., 2002. Heritability of fertility in four seasons for ewes under accelerated lambing. Proc. 7th WCGALP, Montpellier, France. Communication No. 08-20.

Vatankhah, W., Poortahmasb, A. & Merzaei, H.R., 2006. Estimation of performance and genetic parameters of reproductive traits in Lori-Bakhtiari sheep using linear and threshold models. EAAP, 57th Annual Meeting, Antyla, 64.

Zhou, X. & Stephens, M., 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44, 821-824.

Zishiri, O.T., Cloete, S.W.P., Olivier, J.J. & Dzama, K., 2013. Genetic parameters for growth, reproduction and fitness traits in the South African Dorper sheep breed. Small Rumin. Res. 112, 39-48