

Short communication

Evaluation of the OvineSNP50 chip for use in four South African sheep breeds

L. Sandenbergh^{1,2#}, S.W.P. Cloete^{2,3}, R. Roodt-Wilding¹, M.A. Snyman⁴
& A.E. Bester-van der Merwe¹

¹ Department of Genetics, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

² Directorate Animal Sciences: Elsenburg, Private Bag X1, Elsenburg, 7607, South Africa

³ Department of Animal Sciences, Stellenbosch University, Private Bag X1, Matieland, 7602, South Africa

⁴ Grootfontein Agricultural Development Institute, Private Bag X529, Middelburg EC 5900, South Africa

(Received 19 September 2014; Accepted 19 January 2016; First published online 21 March 2016)

Copyright resides with the authors in terms of the Creative Commons Attribution 2.5 South African Licence.
See: <http://creativecommons.org/licenses/by/2.5/za>
Condition of use: The user may copy, distribute, transmit and adapt the work, but must recognize the authors and the South African Journal of Animal Science.

Abstract

Relatively rapid and cost-effective genotyping using the OvineSNP50 chip holds great promise for the South African sheep industry and research partners. However, SNP ascertainment bias may influence inferences from the genotyping results of South African sheep breeds. Therefore, samples from Dorper, Namaqua Afrikaner (NA), South African Merino (SA Merino) and South African Mutton Merino (SAMM) were genotyped to determine the utility of the OvineSNP50 chip for these important South African sheep breeds. After quality control measures had been implemented, 85 SA Merino, 20 Dorper, 20 NA and 19 SAMM samples remained, with an average call rate of 99.72%. A total of 49 517 (91.30%) SNPs on the chip met quality control measures and were included in downstream analyses. The NA had the fewest polymorphic loci, 69.20%, while the SAMM, Dorper and SA Merino had between 81.16% and 86.85% polymorphic loci. Most loci of the SA Merino, Dorper and SAMM had a MAF greater than or equal to 0.3. In contrast, the NA exhibited a large number of rare alleles (MAF < 0.1) and a uniform distribution of other loci across the MAF range (0.1 < MAF ≤ 0.5). The NA exhibited the least genetic diversity and had the greatest inbreeding coefficient among the four breeds. The results of the Dorper, SA Merino, and SAMM compare favourably with those of international breeds and thus demonstrate the utility of the OvineSNP50 chip for these breeds. Effects of SNP ascertainment bias, however, could be seen in the number of non-polymorphic loci and MAF distribution of the three commercial breeds in comparison with those of the NA. The implementation of methods to reduce the effect of SNP ascertainment bias and to ensure unbiased interpretation of genotype results should therefore be considered for future studies using OvineSNP50 chip genotype results.

Keywords: Dorper, genome-wide, Merino, Namaqua Afrikaner, OvineSNP50, SA Mutton Merino, SNP

Corresponding author: LiseS@elsenburg.com

The OvineSNP50 chip was developed by Illumina in collaboration with the International Sheep Genomics Consortium (ISGC) and became commercially available in 2009. This microarray-based system is designed to determine the genotype of approximately 54 000 single nucleotide polymorphisms (SNPs) spaced evenly across the ovine genome. The mean genomic distance between SNPs included in the chip is approximately 51 kb with a median distance equal to 42 kb (International Sheep Genomics Consortium *et al.*, 2010; Kijas *et al.*, 2012). Approximately 500 SNPs discovered through BAC end Sanger sequencing of nine animals, each representing a different breed, are incorporated in the 50K chip. Roche 454 whole-genome shotgun sequencing of six individuals of six different breeds provided approximately 33 000 additional SNPs that are also included in the chip. A total of 15 000 SNPs were identified by reduced representational sequencing (RRS). The RRS was carried out on 60 individuals from 15 breeds (Kijas *et al.*, 2012). The SNP chip has been validated in more than 75 international sheep breeds and has been used in a wide range of applications, which include the elucidation of quantitative trait loci and selection strategies (Miller *et al.*, 2011; Kijas *et al.*, 2012; Demars *et al.*, 2013; Liu *et al.*, 2013; Våge *et al.*, 2013; Gutiérrez-Gil *et al.*, 2014; Phua *et al.*, 2014).

SNP panels developed from a limited number of individuals or breeds are not necessarily representative of the polymorphisms and frequency distribution of alleles in the genome of all members of the species, this being referred to as ascertainment bias (Nielsen & Signorovitch, 2003; Morin *et al.*, 2004). Ascertainment bias may influence parameters estimated from genotyping results and could therefore bias conclusions drawn from such data (Kijas *et al.*, 2009; Albrechtsen *et al.*, 2010). Ascertainment bias could also influence assumptions about a test population as well as the association between certain traits and molecular markers in that population (Heslot *et al.*, 2013). It is therefore vital to recognise the effect of possible ascertainment bias of a marker panel on inferences relating to allele frequency distribution, linkage disequilibrium (LD), population genetic structure, association studies and selection strategies (Clark *et al.*, 2005; Heslot *et al.*, 2013).

The ovine SNP chip could be a valuable genotyping tool for use in South African sheep research and for commercial selection strategies. However, the success of whole-genome SNP studies relies on sufficient numbers of polymorphic SNPs spaced evenly throughout the genome of the test population (Kijas *et al.*, 2009). It is therefore necessary to determine whether these South African sheep breeds are polymorphic for a sufficient number of SNPs included in the OvineSNP50 chip. The potential ascertainment bias of markers included in the OvineSNP50 chip and the potentially unique SNP profile of relatively rare African sheep breeds and subtypes supports evaluating the use of the ovine chip in South African sheep. Here the authors report genotyping results for the dominant wool (SA Merino); dual-purpose (SAMM) and meat (Dorper) breeds in South Africa (Cloete & Olivier, 2010). Genotyping results of the Namaqua Afrikaner (NA) were included as this breed represents the indigenous fat-tailed sheep of South Africa (Qwabe *et al.*, 2013).

The Dorper (n = 20), NA (n = 20) and SAMM (n = 20) samples were obtained from animals maintained as a resource flock on Nortier Research Farm, near Lambert's Bay in the Western Cape, South Africa. The SA Merino samples were from resource flocks maintained at Cradock (n = 50) and Grootfontein (n = 50) in the Eastern Cape (Schoeman *et al.*, 2010). Genotyping was done with the OvineSNP50 chip (Illumina) at GeneSeek Inc. (Lincoln, Nebr, USA) from samples applied to blood cards. GenomeStudio Software v1.0 (Genotyping Module, Illumina) was utilised to call genotypes from SNP intensity data and to ensure the stringency of quality-control parameters. Loci that met these quality-control measures were included for further analyses: >0.25 GenCall score; >0.5 GenTrain score; >0.01 MAF; >0.95 call rate and a sample call rate >0.95. Genotype data that met the quality control criteria were used to determine the number of polymorphic loci and MAF distribution of each breed. The observed heterozygosity and inbreeding coefficient (F_{IS}) were calculated using PLINK v.1.07 (Purcell *et al.*, 2007).

One SAMM sample and 15 SA Merino samples were excluded during quality control owing to low sample call rates. The remaining samples had an average call rate of 99.72%. A total of 49 517 SNPs, 91.30% of SNPs on the chip, met the quality-control measures and had a MAF >0.01 across all samples. These results are similar to other ovine studies, which reported >98% and >90% for sample and SNP loci call rates, respectively (Kijas *et al.*, 2012; Shariflou *et al.*, 2012; Våge *et al.*, 2013; Phua *et al.*, 2014).

The number of polymorphic loci of the SAMM, Dorper and SA Merino ranged between 81.16% and 86.85% (of the total number of SNPs on the chip) and the observed heterozygosity values were between 0.33 and 0.35 (Table 1). These results compare favourably with the number of polymorphic loci and heterozygosity estimates reported in the literature for other commercial sheep breeds (Kijas *et al.*, 2009; 2012).

Table 1 Number of polymorphic loci, mean minor allele frequency (MAF), observed heterozygosity (H_e) and inbreeding coefficient (F_{IS}) of the breeds tested

Sample group	n	Loci with MAF<0.01	Polymorphic loci (%)	Mean MAF	H_e	F_{IS}
NA	20	11921	69.20	0.19	0.28	0.25
SAMM	19	5174	81.16	0.22	0.33	0.12
Dorper	20	4026	83.55	0.23	0.34	0.11
SA Merino	85	1120	86.85	0.26	0.35	0.06

NA: Namaqua Afrikaner; SAMM: South African Mutton Merino; n: number of samples.

Although NA samples from five individuals were included in SNP discovery for the ovine SNP chip (Kijas *et al.*, 2009; 2012), this breed exhibited the fewest polymorphic loci (69.20%), and lowest mean MAF

(0.19) and heterozygosity (0.28). The NA also exhibited the greatest inbreeding coefficient (0.25) among these breeds (Table 1). Similar low levels of genetic diversity have been reported by microsatellite-based studies of locally sourced NA samples (Qwabe *et al.*, 2013), and based on OvineSNP50 genotype information of international NA samples (Kijas *et al.*, 2009; 2012).

The MAF distribution patterns of the SA Merino, Dorper and SAMM were similar to those of most European and American breeds (Kijas *et al.*, 2012) with most loci exhibiting MAFs of at least 30% or greater (Figure 1). In contrast, the NA samples exhibited a large number of rare alleles (MAF < 0.1) and a uniform distribution of polymorphic loci across the MAF range ($0.1 < \text{MAF} \leq 0.5$). Similar results, characterised by a large number of monomorphic loci with many additional loci exhibiting rare alleles, have been observed in BovineSNP50 (Illumina) genotype results of African and indicine cattle breeds (Matukumalli *et al.*, 2009). The limited number of African and indicine individuals and breeds utilised during bovine SNP discovery, resulting in SNP ascertainment bias, has been suggested as underlying these genotype results. Matukumalli *et al.* (2009) therefore report the BovineSNP50 chip's utility in European and other commercial breeds, but warn of reduced power in African and indicine breeds.

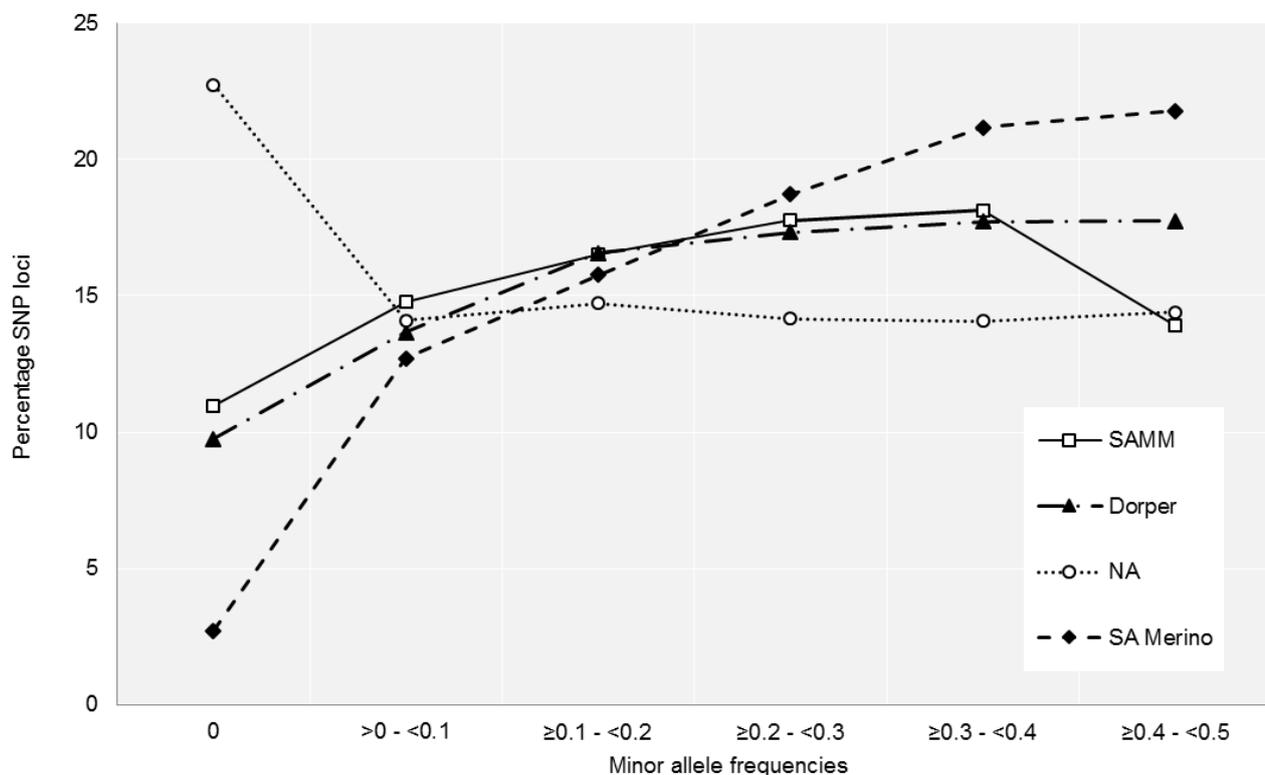


Figure 1 Minor allele frequency distribution of SNP loci included on the OvineSNP50 chip for the South African Mutton Merino (SAMM), Dorper, Namaqua Afrikaner (NA) and South African Merino (SA Merino) breeds.

Whole-genome SNP genotyping results have indicated that the Merino breed is polymorphic for a large number of SNPs included on the OvineSNP50 chip as well as being one of the most genetically diverse livestock breeds (Kijas *et al.*, 2009; 2012; 2014). It is therefore not surprising that the SA Merino samples included in the current study exhibited the highest percentage of polymorphic loci, highest mean MAF and high levels of genetic diversity across the four breeds. Although the Dorper and SAMM were not included during SNP discovery, there were genetic links to other breeds, such as the Merino and Poll Dorset (Milne, 2000), which were included. The percentage of polymorphic markers, mean MAF and heterozygosity of the Dorper and SAMM is somewhat less than that of the SA Merino. These results, however, still compare well with other commercial breeds (Kijas *et al.*, 2012). The indigenous fat-tailed NA exhibited a large number of non-polymorphic loci and a distinct MAF distribution that is most likely the result of SNP ascertainment bias. Nevertheless, the OvineSNP50 chip provides an opportunity to gain a substantial quantity of genomic information in a relatively rapid and inexpensive manner for all four breeds. However, it is important to

recognise the effect ascertainment bias may have on inferences relating to population genetic estimates and genomic selection strategies. A linkage disequilibrium pruning technique, which compensates for SNP ascertainment bias and ensures the unbiased interpretation of genotype data, would therefore be useful to future studies on the local ovine genetic resource (López Herráez *et al.*, 2009; Kijas *et al.*, 2012).

The utility of the OvineSNP50 chip for genotyping the local Dorper, SA Merino and SAMM breeds has been demonstrated. However, studies that utilise the chip to genotype indigenous breeds or breeds that are not included extensively in SNP discovery might be compromised by the effects of SNP ascertainment bias. There is scope for further studies to examine the use of the ovine SNP chip in the other breeds that constitute the South African ovine genetic resource and to characterise the breed diversity of South African sheep using whole-genome SNP data.

Acknowledgements

The authors gratefully acknowledge the financial support of THRIP, Cape Wools SA and the Western Cape Agricultural Research Trust. Research infrastructure was provided by Stellenbosch University and the Western Cape Department of Agriculture.

Authors' contributions

LS conducted the research during her PhD studies and was responsible for drafting and submitting the manuscript. SWPC co-ordinated the research programme, secured funding and identified potential samples. MS provided the Cradock and Grootfontein Merino samples. ABM and RRW, together with the other co-authors, critically analysed and interpreted results, and assisted in the revision of the original and revised manuscript.

References

- Albrechtsen, A., Nielsen, F.C. & Nielsen, R., 2010. Ascertainment biases in SNP chips affect measures of population divergence. *Molec. Biol. Evol.* 27, 2534-2547.
- Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. & Nielsen, R., 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 15, 1496-1502.
- Cloete, S.W.P. & Olivier, J.J., 2010. South African Sheep and Wool Industry. In: *The International Sheep and Wool Handbook*. Ed: Cottle, D.J., Nottingham University Press, Manor Farm, Thrumpton, Nottingham, United Kingdom. pp. 95-112.
- Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., Martyniuk, E. & Bodin, L., 2013. Genome-wide association studies identify two novel *BMP15* mutations responsible for an atypical hyperprolificacy phenotype in sheep. *PLoS Gen.* 9, e1003482.
- Gutiérrez-Gil, B., Arranz, J.J., Pong-Wong, R., García-Gámez, E., Kijas, J. & Wiener, P., 2014. Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS ONE*: e94623.
- Heslot, N., Rutkoski, J., Poland, J., Jannink, J.-L. & Sorrells, M.E., 2013. Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. *PLoS ONE* 8, e74612.
- Kijas, J.W., Townley, D., Dalrymple, B.P., Heaton, M.P., Maddox, J.F., McGrath, A., Wilson, P., Ingersoll, R.G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, H.V., Nicholas, F.W. & Raadsma, H. for the International Sheep Genomics Consortium, 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4, e4668.
- Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B. & other members of the International Sheep Genomics Consortium, 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS One Biol.* 10, e1001258.
- Kijas, J.W., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., Hayes, B.J., Brauning, R., McEwan, J. & The International Sheep Genomics Consortium, 2014. Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Anim. Genet.* 45, 754-757.
- Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X., Zhang, S., Zhou, X., Wei, C., Zhao, F. & Du, L., 2013. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. *BMC Genomics* 14, 229-240.
- López Herráez, D., Bauchet, M., Tang, K., Theunert, C. & Pugach, I., 2009. Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. *PLoS One* 4, e7888.
- Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S. & Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4, e5350.
- Miller, J.M., Poissant, J. & Kijas, J.W., 2011. A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molec. Ecol. Resources* 11, 314-322.
- Milne, C., 2000. The history of the Dorper sheep. *Small Rumin. Res.* 36, 99-102.
- Morin, P.A., Luikart, G., Wayne, R.K. & SNP workshop group, 2004. SNPs in ecology, evolution and conservation. *Trends Ecol. Evol.* 19, 208-216.
- Nielsen, R. & Signorovitch, J., 2003. Correcting for ascertainment bias when analysing SNP data: applications to the estimation of linkage disequilibrium. *Theoret. Pop. Biol.* 63, 245-255.
- Phua, S.H., Brauning, R., Baird, H.J. & Dodds, K.G., 2014. Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. *Anim. Genet.* 45, 240-247.

- 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 tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575.
- Qwabe, S.O., Van Marle-Köster, E., Visser, C., 2013. Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. *Trop. Anim. Health. Pro.* 45, 511-516.
- Schoeman, S.J., Cloete, S.W.P. & Olivier, J.J., 2010. Returns on investment in sheep and goat breeding in South Africa. *Livest. Sci.* 130, 70-82.
- Shariflou, M.R., Wade, C.M., Kijas, J.W., McCulloch, R., Windsor, P.A., Tammen, I. & Nicholas, F.W., 2012. Brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in Merino sheep maps to a 1.1-megabase region on ovine chromosome OAR2. *Anim. Genet.* 44, 231-233.
- The International Sheep Genomics Consortium, Archibald, A.L., Cockett, N.E., Dalrymple, B.P., Faraut, T., Kijas, J.W., Maddox, J.F., McEwan, J.C., Hutton Oddy, V., Raadsma, H.W., Wade, C., Wang, J., Wang, W. & Xun, X., 2010. The sheep genome reference sequence: a work in progress. *Anim. Genet.* 41, 449-453.
- Våge, D.I., Husdal, M., Kent, M.P., Klemetsdal, G. & Boman, I.A., 2013. A missense mutation in growth differentiation factor 9 (*GDF9*) is strongly associated with litter size in sheep. *BMC Genet.* 14, 1-8.