Short communication

Validation of the 50k Illumina goat SNP chip in the South African Angora goat

S.F. Lashmar, C. Visser[#] & E. van Marle-Köster
Department of Animal and Wildlife Sciences, University of Pretoria, Private Bag X20,
Hatfield, 0028, South Africa

Abstract

Tools for the genomic evaluation of goats have generally lagged behind those for other species. However, the recent availability of the goat SNP50 consortium bead chip has marked a positive change for this small ruminant species. Polymorphic loci can differ greatly between breeds of the same species. Exclusion of fibre-producing breeds, such as the Angora goat, during the development of this genotyping array necessitates the validation of SNPs included on the chip to allow for genomic applications that would accelerate genetic progress in mohair yield and quality. Forty eight unrelated Angora goats, displaying phenotypic variation in two important price-determining traits, namely fibre diameter and fleece weight, were genotyped with the goat SNP50 consortium bead chip. Results revealed that 46 983 SNP (88.1%) of the 53 347 called SNPs were polymorphic (MAF>0.05). After quality control, 3 960 SNP were filtered from further analysis for violating Hardy-Weinberg and call-rate parameter thresholds, leaving 43 759 (82%) of the 53 347 SNPs to be validated for downstream analysis. Observed and expected heterozygosity values of 0.365 and 0.370, respectively, were obtained for polymorphic SNPs. A total of 30 357 SNPs in linkage disequilibrium (LD) were removed to obtain a set of independent markers, resulting in a final SNP density of 1 SNP/ ~226 kb. Results indicate that the goat SNP50 bead chip was informative in the Angora goats that were studied, and should be useful in examining the underlying genetic variation.

Genetic research on the South African Angora goat has been extensive, and has included quantitative studies (Snyman & Olivier, 1996; Visser & Van Marle-Köster, 2009) pertaining to the estimation of genetic parameters for traits of economic importance and molecular research using microsatellites (Visser & Van Marle-Köster, 2009; Visser *et al.*, 2010; 2011a; b; 2013). Although microsatellite markers have greater polymorphism on a per-marker basis, they are not as abundant as Single Nucleotide Polymorphisms (SNPs), lack sufficient coverage across the genome, and have limited automation options (Vignal *et al.*, 2002). The recent publication of a radiation hybrid map (Du *et al.*, 2012), a reference genome sequence (Dong *et al.*, 2013), and the consequent development of the 50K SNP chip (Illumina Inc., San Diego, Calif) (Tosser-Klopp *et al.*, 2014) for the domestic goat marked a positive change for this species.

A key initiative in establishing the International Goat Genome Consortium (IGGC) was the development of a moderate- to high-density genotyping tool for the domestic goat (Tosser-Klopp *et al.*, 2014). Due to the fact that three separate SNP discovery projects were underway before their consolidation within the IGGC, the detection of variants for the goat species was done through two pipelines – one for dairy and mixed breeds, and the other for meat breeds. Fibre-producing breeds were not included in any of these pipelines, and thus not in the development of this chip (Tosser-Klopp *et al.*, 2014). Research has shown that polymorphic loci differ greatly between breeds of the same species (Garrick, 2011), and high levels of polymorphism have been observed in breeds that were not included in the development of the goat SNP chip (Kijas *et al.*, 2013; Tosser-Klopp *et al.*, 2014). This suggests that high levels could be observed across goat breeds that were not sampled for SNP discovery, and thus the developed chip could be useful for Angora goats. South Africa is a major producer of mohair and the use of genomic technology for genetic

improvement needs due consideration (Visser & Van Marle-Köster, 2014). The aim of this study was to validate the use of the goat SNP50 bead chip in the South African Angora population.

Whole blood samples of 48 unrelated Angora goat kids were collected from the DNA biobank for small stock research and conservation (Grootfontein Agricultural Development Institute, National Department of Agriculture). Animals with varying performance in two important price-determining phenotypic traits, namely fibre diameter and fleece weight, were chosen to ensure a diverse group of animals, and thus encapsulate a broad spectrum of genetic variation. For fleece weight, phenotypic values ranged from 0.5 kg for inferior animals to 2.4 kg for superior animals, and for fibre diameter phenotypic values ranged from 35.7 micron for inferior animals to 19.5 micron for superior animals. Blood samples were transported on ice to the University of Pretoria (UP). DNA was extracted from the whole blood samples at the Animal Breeding and Genetics Laboratory (UP), using Qiagen DNeasy Blood and Tissue kit® (Whitehead Scientific (Pty) Ltd, Cape Town, South Africa) according to the manufacturer's protocol. Genomic DNA quantity and quality for all samples were estimated with the NanoDrop spectrophotometer (NanoDrop ND-1000) and plate reader (SpectraMax® Microplate Reader) at the Genetics Department of the University of Pretoria, and the Qubit® 2.0 fluorometer (Life Technologies (Pty) Ltd, Carlsbad, Calif, USA) at the ARC Biotechnology Platform. The average DNA concentration obtained over all the samples was 74 ng/μL with an average 260 : 280 ratio of 2.03.

Genotyping was conducted at the ARC Biotechnology Platform with the Illumina goat SNP50 bead chip, which features 53 347 SNP probes distributed across the whole goat genome, and provides an average inter-SNP spacing of ~40 kb (Tosser-Klopp *et al.*, 2012). Genotyping was conducted over three days, and included overnight whole-genome amplification, followed by fragmentation, precipitation and re-suspension of the samples in a hybridization buffer. Hybridization of the DNA to the bead chips occurred overnight for 20 hours in a hybridization oven at 48 ºC. After hybridization, the bead chips were washed, stained and dried. Processed bead chips were imaged with the Illumina iScan Reader, after which data were transferred to Illumina GenomeStudio 1.9.0 software for analysis.

GenomeStudio was utilized to generate PLINK (Purcell *et al.*, 2007) input files, including the MAP (SNP panel) files and PED (genotypes per individual) files, which were used to perform sample-based and marker-based quality control measures to filter non-informative individuals and SNPs. Sample-based quality control was based on rates of missing genotypes, and SNP-based quality control was based on call rate, minor allele frequency and Hardy-Weinberg equilibrium (HWE). SNPs that had a call rate below 98%, MAF below 5% or violated HWE (P <0.001) were removed from further analysis. After quality control, PLINK was used to calculate observed and expected heterozygosity across all autosomes. The --indep-pairwise (50 5 0.2) option was then used to remove SNPs that were in linkage disequilibrium in order to attain an independent set of SNPs for downstream analysis.

Sample-based quality control was performed first. No individuals were removed for having a call rate below 98% with the average call rate across the 48 samples being 99.6%. A total of 53 347 SNPs were considered for analysis before marker-based quality control filtering, including 1417 unmapped SNPs and 1986 SNPs located on the X-chromosome. Results following marker-based quality control are indicated in Table 1.

Table 1 Results following marker-based quality control

Number of SNPs included	SNPs<98% Call Rate	SNPs<5% MAF	SNPs<HWE (P <0.001)	Total SNPs removed	SNPs validated for downstream analysis
53 347	2279	6364	1681	9588	43 759

After the procedures for MAF filtering, it was found that the remaining 46 983 (88.1%) SNP were polymorphic for the sampled individuals. An average MAF of 0.25 was found across the 48 samples. After quality control, the observed and expected rates of heterozygosity of the polymorphic autosomal SNPs were 0.365 and 0.370, respectively. Only autosomal SNPs and SNPs with known genomic location, that is, mapped SNPs, were considered for linkage-based filtering, and therefore only 41 016 of the 43 759 SNPs were used. A total of 30 357 SNPs were removed after LD pruning, resulting in a set of 10 659 independent SNPs. SNP densities were estimated before and after LD pruning and are summarized in Table 2.

Quality control, including sample-based quality control and marker-based quality control, is an important step in utilizing SNP data, and is performed to remove potential biases introduced during study

Table 2 SNP densities before and after linkage disequilibrium filtering

	Minimum gap (bp)	Maximum gap (kb)	Median gap (kb)	SNP density (average gap)
Before LD pruning	2	971	46	1 SNP/~59 kb
After LD pruning	139	1993	142	1 SNP/~226 kb

LD: linkage disequilibrium.

design, as well as errors from genotyping laboratory procedures and subsequent genotype calling (Anderson *et al.*, 2010). Sample-based filtering relates to the sample call rate, and can be influenced by factors such as DNA quality and concentration, as well as genotype calling algorithms (Anderson *et al.*, 2010). The average call rate of 99.6% obtained in this study is comparable with call rates of above 99.5% (Kijas *et al.*, 2013) and 99.9% (Tosser-Klopp *et al.*, 2014) reported in international studies in which numbers of Angora goats were genotyped using the 50K goat SNP chip. These results compare favourably with the average call rate of approximately 99.9% obtained across 10 goat breeds during the validation phase of the development of the goat SNP chip (Tosser-Klopp *et al.*, 2014). In comparison with other validation studies performed for livestock in South Africa, the average call rate obtained can be compared with a range of 96.7% (Holstein) to 99.7% (Angus) obtained during an evaluation of the performance of the bovine SNP50 bead chip in South African cattle breeds (Qwabe *et al.*, 2013).

The successful application of SNP arrays depends largely on the degree of polymorphism in the various breeds within each species (Fan *et al.*, 2010). An average MAF of 0.25 obtained across all 53 347 loci compares favourably with average MAF ranges of 0.18 to 0.23 in horses using the equine SNP50 bead chip (McCue *et al.*, 2013), and a value of 0.24 for African N'Dama and Sheko breeds using the bovine SNP50 bead chip (Matukumalli *et al.*, 2009). During an evaluation of the bovine SNP50 bead chip in four South African cattle populations, Qwabe *et al.* (2013) reported MAF ranges between 0.17 for Nguni cattle and 0.22 for Holstein cattle.

After filtering out SNPs with low MAF, 46 983 SNPs (88.1%) proved to be polymorphic for the South African Angora goat, which is comparable with the number of polymorphic loci found in other goat breeds (Jinlan: 45 648, Skopelos: 50 908) that were not included in the SNP discovery process (Tosser-Klopp *et al.*, 2014). However, differences in sample sizes among studies should be considered in comparing the results. The number of polymorphic loci compares favourably with values of 28 869 and 35 084 SNPs obtained for African N'Dama and Sheko breeds, respectively, during Illumina's bovine SNP50 content validation study (Matukumalli *et al.*, 2009). The level of polymorphism compares favourably with the range of 35 843 for Nguni to 41 078 for Holstein when the bovine SNP50 chip was validated for cattle breeds in South Africa (Qwabe *et al.*, 2013). All these comparisons were based on similar quality control parameters in previous studies and in the current study.

The average observed heterozygosity for the 48 animals was calculated as 0.365, which was similar to the expected value of 0.370. The observed heterozygosity was lower than the value of 0.442 obtained for Angora goats sampled by Kijas *et al.* (2013).

To perform certain population genetic analyses such as principle component analysis (PCA), it is ideal to include only independent SNPs (Anderson *et al.*, 2010). Markers that are in LD were inactivated in order to remove bias caused by linkage and to obtain a set of SNPs with the maximum number of independent markers (Davis *et al.*, 2011). Following LD pruning, there should be no SNP pair within a given window, usually 50 kb, that are correlated, taken as r^2 >0.2 (Anderson *et al.*, 2010). The SNP density obtained before LD-based filtering (1 SNP/~59 kb) was lower than the densities of 1 SNP/~53 SNP obtained for Alpine and Saanen goats, but higher than the densities of 1 SNP/~67kb and 1 SNP/~71kb for SA Boer and Toggenburg goats, respectively (Brito *et al.*, 2014). SNP densities obtained after LD-based filtering (1 SNP/~226kb), however, were considerably lower than the original average inter-SNP distance range of 30 - 90 kb, as stated by Tosser-Klopp *et al.* (2014).

It can be concluded from the results that the breed-based application of the goat SNP50 bead chip for the South African Angora goat is possible. Even though no fibre-producing breed was included in developing this genotyping tool, the high level of polymorphism observed in this study suggests that the goat SNP50 bead chip will allow for applications such as genome-wide association studies, diversity studies, parentage verification, selection signatures and eventually genomic selection.

References

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat. Protoc. 5, 1564-1573.

Brito, L.F., Jafarikia, M., Grossi1, D.A., Maignel, L., Sargolzaei1, M. & Schenkel, F.S., 2014. Characterization of linkage disequilibrium and consistency of gametic phase in Canadian goats. In: Proc 10th World Cong. Genet. App. Livest. Prod. Vancouver, BC, Canada, 2014. Poster presentation, 882.

Davis, N.A., Pandey, A. & McKinney, B.A., 2011. Real-world comparison of CPU and GPU implementations of SNPrank: a network analysis tool for GWAS. Bioinformatics 27, 284-285.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X.Y., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang. J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J. & Wang, W., 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat *(Capra hircus)*. Nat. Biotechnol. 31, 135-141.

Du, X.Y., Womack, J.E., Owens, K.E., Elliott, J.S., Sayre, B., Bottcher, P.J., Milan, D., Garcia Podesta, M., Zhao, S.H. & Malek, M., 2012. A whole-genome radiation hybrid panel for goat. Small Rumin. Res. 105, 114-116.

Fan, B., Zhi-Qiang, D., Danielle, M., Gorbach, M. & Rothschild, F., 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Austral. J. Anim. Sci. 23, 833-847.

Garrick, D.J., 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Gen. Sel. Evol. 43, 1-11.

Kijas, J.W., Ortiz, J.S., McCulloch, R., James, A., Brice, B., Swain, B., Tosser-Klopp, G. & International Goat Genome Consortium, 2013. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. Anim. Genet. 44, 325-335.

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., Sonstegard, T.S. & Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4, e5350-5063.

McCue, M. & Mickelson, J., 2013. Genomic tools and resources: Development and applications of an equine SNP genotyping array. In: Equine Genomics. (1st ed.). Ed. Chowdhary, B.P., John Wilkey & Sons Inc., 113-124.

Purcell, S., Neale, B. & Todd-Brown, K., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Gen. 81, 559-575.

Qwabe, S.O., Van Marle-Köster, E., Maiwashe, A. & Muchadeyi, F.C., 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. S. Afr. J. Anim. Sci. 43, 1, 64-67.

Tosser-Klopp, G., Bardou, P., Cabau, C. & the International Goat Genome Consortium, 2012. Goat genome assembly, availability of an international 50 K SNP chip and RH panel: an update of the International Goat Genome Consortium projects. In: International Plant & Animal Genome Conference, 14 - 18 January 2012, San Diego, CA, USA.

Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H.C.M., Jamli, S., Jiken, A.J., Klopp, C., Lawley, C.T., McEwan, J., Martin, P., Moreno, C.R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhiere, E., Rupp, R., Sarry, J., Sayre, B.L., Tircazes, A., Wang, J., Wang, W., Zhang, W. & the International Goat Genome Consortium, 2014. Design and Characterization of a 52K SNP Chip for Goats. PLoS ONE 9, e86227.

Vignal, A., Milan, D., SanCristobal, M. & Eggen, A., 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Gen. Sel. Evol. 34, 275-305.

Visser, C. & Van Marle-Köster, E., 2009. Genetic variation of the reference population for quantitative loci research in South African Angora goats. AGR Information 45, 113-119.

Visser, C. & Van Marle-Köster, E., 2014. Strategies for the genetic improvement of South African Angora goats. Small Rumin. Res. 121, 89-95.

Visser, C., Crooijmans, R.P.M.A. & Van Marle-Köster, E., 2010. A genetic linkage map for the South African Angora goat. Small Rumin. Res. 93, 171-179.

Visser, C., Crooijmans, R.P.M.A., Bovenhuis, H. & Van Marle-Köster, E., 2011a. QTL for mohair traits in South African Angora goats. Small Rumin. Res. 100, 8-14.

Visser, C., Van Marle-Köster, E. & Friedrich, H., 2011b. Parentage verification of South African Angora goats, using microsatellite markers. S. Afr. J. Anim. Sci. 41, 250-255.

Visser, C., Van Marle-Köster, E., Snyman, M.A., Bovenhuis, H. & Crooijmans, R.P.M.A., 2013. Quantitative trait loci associated with pre-weaning growth in South African Angora goats. Small Rumin. Res. 112, 15-20.