



# How efficiently Genome-Wide Association Studies (GWAS) identify prolificacy-determining genes in sheep

Magda V Benavides, Carlos José H Souza, José Carlos F. Moraes

Embrapa Pecuária Sul, BR 153 Km 603 - Bagé/RS - Brasil. Caixa Postal 242 – Vila Industrial, CEP: 96401-970

Corresponding author: Magda V Benavides

E-mail: magda.benavides@embrapa.br

Genet. Mol. Res. 17 (2): gmr16039909

Received November 08, 2017

Accepted April 23, 2018

Published April 27, 2018

DOI: <http://dx.doi.org/10.4238/gmr16039909>

Copyright © 2018 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Genome wide association studies are performed to narrow down or, better still, pinpoint the region or genes involved in the regulation of a known phenotype. The data concerning litter size usually comes from a restrict number of individuals due to the laborious and costly nature of the phenotyping by lambing data of several breeding seasons. Genome-wide association was performed in a population in which segregates a major gene determinant of prolificacy in sheep. This is the allele of the GDF9 SNP called Vacaria: c943C> T, detected in Ile de France flocks in Southern Brazil. The Vacaria genotype and phenotype data did agree with the GDF9 sheep chromosome 5 regions from genomic analyses. OAR5\_45481559, a marker located within the GDF9 gene (chr5: 41,841,919 bp), had nominal p-value significance but did not reach Bonferroni-corrected significance levels. However, other chromosome 5 SNP markers (s17197, s48166, s25202, and OAR5\_47774570), located at chr5: 43,415,384 - 43,708,878 bp, did show nominal and Bonferroni p-value significances. These first three markers showed to be in linkage disequilibrium with OAR5\_45481559, thus confirming Plink and Emmax (Genome-Wide Association Studies (GWAS) packages) abilities of locating the correct genomic region associated to phenotypes. These results are indicative that GWAS is a useful method for searching candidate genes in prolific animals, since the linkage disequilibrium is verified in a range of 2 Mbp

**Key words:** Vacaria sheep; GDF9; Sheep prolificacy; Genome-wide association studies (GWAS)

## INTRODUCTION

Sheep prolificacy is key to livestock productivity and profit. The increase of born lambs can be induced by hormonal treatment, however, in the last three decades, mutations responsible for sheep prolificacy have been detected (Heaton et al., 2017), most proved to be determined by major genes, and for this reason, some were used for practical applications in animal production.

There have been recent reports of high prolificacy sheep in commercial flocks worldwide (Mullen & Hanrahan, 2014; Souza et al., 2014; Lassoued et al., 2017). Traditionally, the high prolificacy sheep are tested against single nucleotide polymorphism (SNPs) of known mutation, if the phenotype does not fit an earlier studied genotyping system, the search for a new gene of interest needs to be investigated.

Most of the major genes found to date are located at the BMP15 and GDF9 genes (Juengel et al., 2013). However, when prolific sheep with high ovulation rate (OR) are not explained by known genes, the genome-wide association analyses (GWAS) is a tool to search for genetic markers associated with the phenotype of interest, including mutations not linked to major genes.

The objective of this study was to determine the precision of GWAS software packages in finding the most likely gene when phenotyped animals have been genotyped to a causal SNP within the GDF9 gene (Vacaria mutation, Souza et al., 2014). It is important to note that sample size was small, however all animals were half-sisters to strengthen the case-control model used in the analysis.

## MATERIAL AND METHODS

### Searching for prolific ewes

A search was made on the Brazilian Sheep Breeder's Association (ARCO) databank, looking for prolific ewes occurring in commercial stud flocks of eleven woollen and meat sheep breeds raised in South Brazil. Specifically, the Ile de France breed with 76,864 records, from 1990 to 2016, reveals a prevalence of 20% of twins and 0.5% of triplets. Farmers that owned ewes having at least one triplet birth record were contacted to: (1) confirm the fact that hormonal treatment was not used in the farm, (2) survey birth control records, including information of lost lambs, repeatability of twin/triplet lambings, and lambing records in the same ewe (or half-sisters) over the years, and (3) allow blood sampling their animals (high prolificacy sheep and their contemporaries) to detect putative genes associated to the phenotype familiar recurrence. A total of ten flocks were sampled, however, farms with high number of birth rate records, with phenotype repeatability, were selected for the study.

The second step was to genotype all sheep to identify the presence of known mutations that affected ovulation rate (BMPR1B Booroola (Souza et al., 2001; Wilson et al., 2001), GDF9 Embrapa (Silva et al., 2011) and Vacaria (Souza et al., 2014)), and are known to be segregating in Brazilian flocks. Of these, only sheep showing the Vacaria mutation were selected for GWAS analyses.

The Vacaria mutation was genotyped following the methodology (Souza et al. 2014), and the numbers for this population were of 12 heterozygous ewes, 11 non-carrier ewes and one VV homozygous ewes, and the data was analysed by using a case-control model.

### GWAS analysis and identification of genes associated to prolificacy traits in Vacaria sheep

A total of 24 sheep was genotyped using Illumina Ovine 50K and filtered using PLINK (1.07 10May2010 version). Roughly 13% of the 54,241 markers in the Illumina chip were pruned. This data was used for genome-wide association analyses by using Efficient Mixed-Model Association expedited (EMMAX beta 07 Mar 2010 version) to find nominal p-values and these data were ran on PLINK for Bonferroni testing using the --assoc, --perm, and --adjust functions.

PLINK and EMMAX are free GWAS packages that allow case-control analyses and offer permutation tests to identify the most significant markers in the analysis. These packages are normally used in synchrony, plink

allows filtering markers by minor allele frequency (MAF), Emmax analyses phenotype-genotype interaction and Plink is used again for phenotype-genotype interaction, considering Bonferroni correction.

Linkage disequilibrium between SNP markers of interest was calculated using Haploview 4.2 (Feb 16, 2010 version; Barrett et al., 2005).

The objective of the study check if the proposed methodological approach could identify the GDF9 gene region as responsible to the Vacaria phenotype. Data (of 12 non-carrier NN n=12, 11 heterozygous (VN) and, and one homozygous we VV n=1 half-sibs). Phenotypes used were coded as 2, 4 and 1, respectively, for Plink/Emmax/Plink gene search analysis.

An interval of ± 2Mbp of the GDF9 location (chr5: 41,841,919 to 44,735,885 bp) was used to search for genes within the relevant genomic regions in the 3.1 sheep genome International Sheep Genome Consortium (ISGC) version using Ensembl website ([http://www.ensembl.org/Ovis\\_aries/Info/Index](http://www.ensembl.org/Ovis_aries/Info/Index)).

## RESULTS

### Descriptive statistics

Basic statistics for number of born lambs describing each category was described in Table 1.

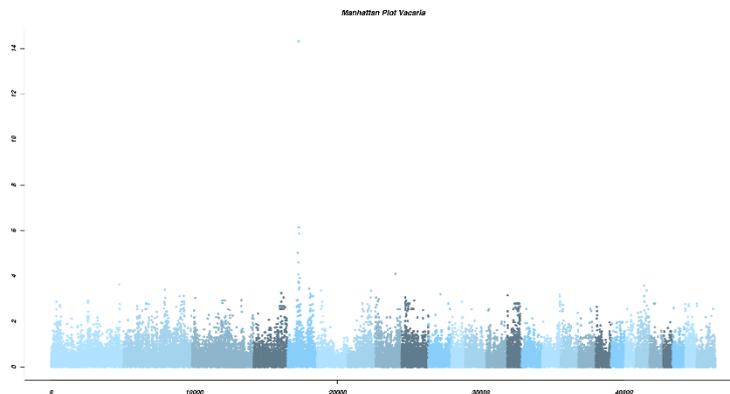
**Table 1.** Born lambs average in the Vacaria genotypes.

Genotypes	n	born lambs average	lower confidence	higher confidence
Vacaria NN	12	1.02	1.00	1.08
Vacaria VN	11	1.75	1.41	2.09
Vacaria VV	1	0	0	0

The VV genotype is sterile due to this mutation, and this animal was included in the analysis as control.

### GDF9 association to Vacaria high- and low- prolificacy ewes

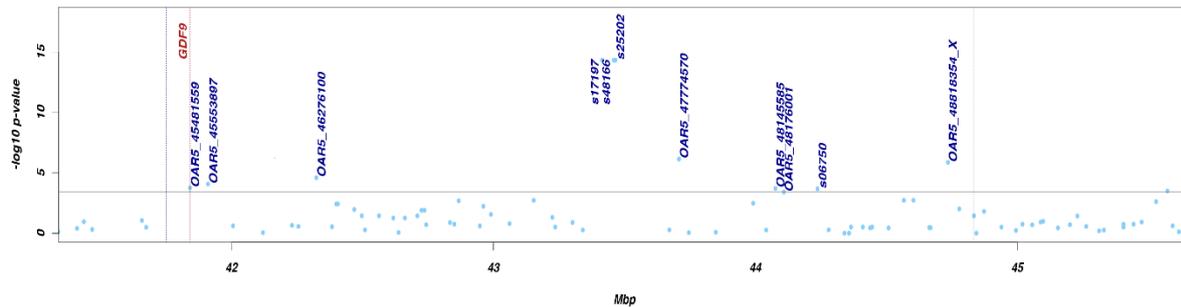
Four ovine chromosome 5 SNP markers showed nominal, Bonferroni (B), and False Discovery Rate (FDR) p-value significances: s17197 ( $p=4.661e^{-15}$ ;  $B=2.162e^{-10}$ ;  $FDR=7.208e^{-11}$ ), s48166 ( $p=4.661e^{-15}$ ;  $B=2.162e^{-10}$ ;  $FDR=7.208e^{-11}$ ), s25202 ( $p=4.661e^{-15}$ ;  $B=2.162e^{-10}$ ;  $FDR=7.208e^{-11}$ ), and OAR5\_47774570 ( $p=7.191e^{-07}$ ;  $B=0.03336$ ;  $FDR=0.0083$ ) (Figure 1).



**Figure 1.** Manhattan plot for Vacaria mutation showing significant markers locations in the ovine chromosome 5.

These markers were located within the ovine chromosome 5: 43,415,384-43,708,878 bp locations.

Other six ovine chromosome 5 SNP markers within the 41,841,919 to 44,735,885 bp range showed nominal significant p-values but did not reach Bonferroni-corrected significance levels (Figure 2 for zoomed image).



**Figure 2.** Ovine chromosome 5 41,841,919-44,735,885 bp SNP significant markers detailed view.

One of them, OAR5\_45481559 (41,841,919 bp), was located within the GDF9 gene (41,841,034-41,843,517 bp). This marker is in linkage disequilibrium with the top three significant markers (s17197, s48166, and s25202), where LOD scores of 3.73 and D' of 1.0 were observed, thus confirming the GDF9 association to the Vacaria sheep.

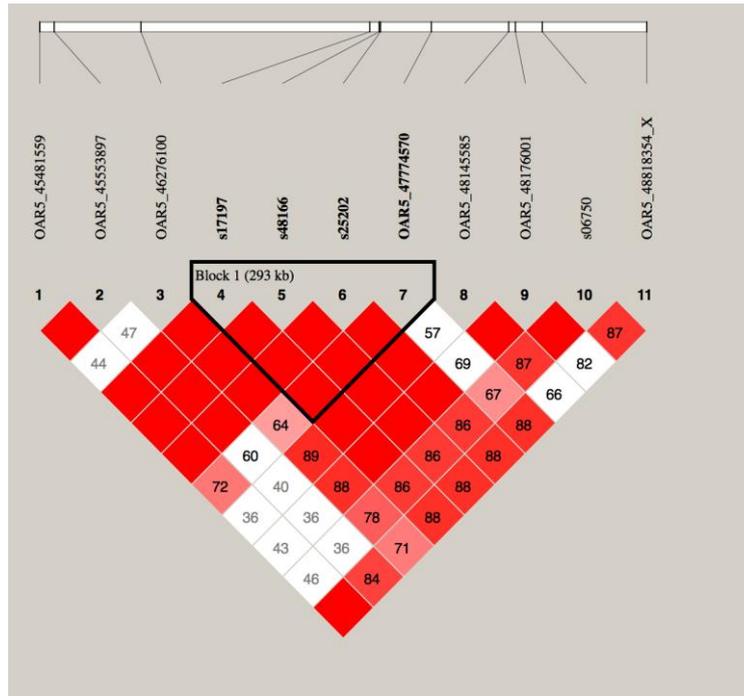
The variants g.43415384G>A (s17197), g.43460085A>G (s48166), and g.43466435A>G (s25202) showed different genotype frequencies between Vacaria ewes, with homozygous A-A-G haplotypes for these three markers defining triplet-, A/G A/G A/G twin-, and homozygous G-G-A single-bearing ewes.

## DISCUSSION

Many aspects may affect the successfulness of finding a genetic association to a phenotype of interest. The number of animals of each phenotype and how accurately phenotypes are assigned to individuals are some of these factors. New mutations generally start in one individual and segregate amongst its offspring, therefore numbers of phenotypes are limited. In such situations, case-control analyses would be advantageous to perform, as opposed to sire-family models, because the former allows the analysis of lower number of individuals.

Lambing single and twins are not unusual in sheep commercial flocks. The frequency of triplet lambs, however, tends to be low and lower still the recurrence within the family, this can indicate the presence of an ovulation rate mutation segregating in a particular flock. Ovulation rate is a trait that varies within an animal across years, therefore correct phenotyping might require many years of recording. As a result, selecting more prolific ewes in a flock can be a long-term undertaking. However, if the mutation is associated to a genomic region, close to a genetic marker, it will be more straightforward to select animals based on their genotypes.

A marker located within the GDF9 gene (OAR5\_45481559, at ovine chromosome 5: 41,841,919 bp) showed to be in linkage disequilibrium with the SNP markers placed the most significant for the analysis (s17197, s48166, and s25202), located at ovine chromosome 5: 43,415,384 - 43,466,435 bp. This indicates Plink/Emmax/Plink were able to pinpoint the correct genomic location of GDF9, since these animals were first genotyped by using a mutation within this gene. It is worth to mention that OAR5\_4774570 is not as closely linked to OAR5\_45481559 as the other three significant markers (Figure 3).



**Figure 3.** Haploview output showing the location of OAR5\_45481559 (close to GDF9) and significant markers found in Plink/Emmax/Plink analyses (s17197, s48166, s25202, and OAR5\_47774570).

This might indicate a possible misplace of some markers on the current version of the ovine genome. Better accuracy of markers locations is expected in coming genome versions.

The fact that the GDF9 gene was located within a 2 Mbp range from the markers selected for Bonferroni correction could also be related to the currently available ovine assembly (OAR 3.1 sheep genome ISGC version), meaning that SNP markers might show slight changes in chromosome location in future assemblies.

Our results showed that Emmax and Plink software packages were feasible tools to search for genetic markers of biological significance in a population with striking phenotypic differences amongst genotypes, even with small number of affected animals.

The identification of novel genes associated to high-prolificacy sheep is helping scientists to uncover the high number of mutations and of phenotypes in sheep, assisting farmers to select for more prolific animals in their flocks. Accurately assigning phenotypes is the base for a successful phenotype-genotype association and case-control software packages such as Plink and Emmax are fit to find genomic regions of interest. This means, in the future, it will be easier to select an animal with the genotype of interest to be part in the flock's next generations.

## CONCLUSION

The Vacaria genotypes showed association to sheep chromosome 5 (43,415,384-43,708,878 bp) regions. The top three significant markers are in linkage disequilibrium with OAR5\_45481559, a marker within the GDF9 gene that showed nominal p-value significance. This linkage confirms the association of GDF9 with Vacaria genotypes. Despite of the reduced number of animals, all of them were half-sibs to strengthen the case-control model used in Plink and Emmax, and the analyses were able to detect the association of the Vacaria allele and the genomic region of 2Mbp comprising the GDF9 gene that is mutated in these animals.

## ACKNOWLEDGEMENTS

The authors thank the Brazilian Sheep Breeder's Association (ARCO) for conceding access to sheep databank permission, and sheep breeders for sharing their flock's reproductive records and allowing blood sampling in ewes of interest. We also thank Embrapa's funding under the 03.13.06.007.00.00 grant.

## REFERENCES

- Barrett JC, Fry B, Maller J and Daly MJ (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21 (2): 263-265. <https://doi.org/10.1093/bioinformatics/bth457>
- Cui XS, Li XY, Jeong YJ, Jun JH, et al. (2006). Gene expression of Cox5a, 5b, or 6b1 and their roles in preimplantation mouse embryos. *Biology of Reproduction* 74 (3): 601-610. <https://doi.org/10.1095/biolreprod.105.045633>
- Gad A, Besenfelder U, Rings F, Ghanem N, et al. (2011). Effect of reproductive tract environment following controlled ovarian hyperstimulation treatment on embryo development and global transcriptome profile of blastocysts: implications for animal breeding and human assisted reproduction. *Human Reproduction* 26 (7): 1693-1707. <https://doi.org/10.1093/humrep/der110>
- Heaton MP, Smith TPL, Freking BA, Workman AM, et al. (2017) Using sheep genomes from diverse US breeds to identify missense variants in genes affecting fecundity. *F1000Research* 6: 1303. <https://doi.org/10.12688/f1000research.12216.1>
- Janowski D, Salilew-Wondim D, Torner H, Tesfaye D, et al. (2012). Incidence of apoptosis and transcript abundance in bovine follicular cells is associated with the quality of the enclosed oocyte. *Theriogenology* 78 (3): 656-669. <https://doi.org/10.1016/j.theriogenology.2012.03.012>
- Juengel JL, Davis GH and McNatty KP (2013). Using sheep lines with mutations in single genes to better understand ovarian function. *Reproduction* 146 (4): R111-R123. <https://doi.org/10.1530/rep-12-0509>
- Lassoued N, Benkhilil Z, Woloszyn F, Rejeb A, et al. (2017). FecX Bar a novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. *BMC Genetics* 18(1): 43-52. <https://doi.org/10.1186/s12863-017-0510-x>
- Mathew SJ, Haubert D, Krönke M, Leptin M (2009). Looking beyond death: a morphogenetic role for the TNF signalling pathway. *J. Cell Sci.* 122 (12): 1939-1946. <https://doi.org/10.1242/jcs.044487>
- Mullen MP and Hanrahan JP (2014). Direct evidence on the contribution of a missense mutation in GDF9 to variation in ovulation rate of Finnsheep. *PLoS One* 9: e95251. <https://doi.org/10.1371/journal.pone.0095251>
- Purcell S, Neale B, Todd-Brown K, Thomas L, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81: 559-575. <https://doi.org/10.1086/519795>
- Ryan J, Artero S, Carrière I, Maller JJ, et al. (2016). GWAS-identified risk variants for major depressive disorder: Preliminary support for an association with late-life depressive symptoms and brain structural alterations. *European Neuropsychopharmacology* 26(1): 113-125. <https://doi.org/10.1016/j.euroneuro.2015.08.022>
- Ryland GL, Doyle MA, Goode D, Boyle SE, et al. (2015). Loss of heterozygosity: what is it good for? *BMC Medical Genomics* 8: 1. [https://doi.org/10.1007/springerreference\\_175071](https://doi.org/10.1007/springerreference_175071)
- Silva BDM, Castro EA, Souza CJH, Paiva SR, et al. (2011). A new polymorphism in the Growth and Differentiation Factor 9 (GDF9) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. *Animal Genetics* 42 (1): 89-92. <https://doi.org/10.1111/j.1365-2052.2010.02078.x>
- Souza CJ, MacDougall C, Campbell BK, McNeilly AS, et al. (2001). The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. *Journal of Endocrinology* 169 (2): R1-R6. <https://doi.org/10.1677/joe.0.169r001>
- Souza CJH, Gonzalez-Bulnes A, Campbell BK, McNeilly AS, et al. (2004). Mechanisms of action of the principal prolific genes and their application to sheep production. *Reproduction, Fertility and Development* 16, 395-401.
- Souza CJH, McNeilly AS, Benavides MV, Melo EO, et al. (2014). Mutation in the protease cleavage site of GDF9 increases ovulation rate and litter size in heterozygous ewes and causes infertility in homozygous ewes. *Animal Genetics* 45 (5): 732-739. <https://doi.org/10.1111/age.12190>
- Wilson T, Wu X-Y, Juengel JL, Ross IK, et al. (2001). Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction* 64 (4): 1225-1235. <https://doi.org/10.1095/biolreprod64.4.1225>
- Ye S, Dhillon S, Ke X, Collins AR, et al. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research* 29 (17): e88-e88. <https://doi.org/10.1093/nar/29.17.e88>