Genome scan for identifying candidate genes for goat lentiviral infections resistance in the Italian Garfagnina goat breed

Francesca Cecchi[a*], Christos Daudosis[b], Riccardo Bozzi[b], Filippo Fratini[a], Claudia Russo[a], Patrizia Bandecchi[a], Carlo Cantile[a], Maurizio Mazzei[a]

[a]Department of Veterinary Sciences, University of Pisa, Viale delle Piagge 2, Pisa, Italy. [b]DISPAA - Sez. Scienze Animali. Università degli Studi di Firenze, Piazzale delle Cascine, 18, Firenze, Italy.

*Corresponding Author: Prof. Francesca Cecchi, email: francesca.cecchi@unipi.it Tel. +39-050-2216879 - Fax +39-050-2210655

Abstract

Small ruminant lentiviruses (SRLVs) are a heterogeneous group of viruses of sheep, goat and wild ruminants, responsible of lifelong persistent infection leading to a multisystem chronic disease. Increased evidences indicate that host genetic factors could influence the individual SRLV resistance. The present study was conducted on the Garfagnina goat breed, an Italian goat population registered on the Tuscan regional repertory of genetic resources at risk of extinction. Forty-eight adult goats belonging to a single flock were studied. SRLV diagnosis was achieved by serological tests and 21 serologically positive animals were identified. All animals were genotyped with the Illumina GoatSNP60 BeadChip and a genome-wide scan was then performed on the individual marker genotypes, in the attempt to identify genomic regions associated with the infection. One SNP was found significant ($P<5\times10^{-5}$) on CHR 18 at 62,360,918bp and it was located within, or nearby, annotated gene. In the region 1Mb upstream-downstream the significant SNP, the *NLRP12* (NLR family pyrin domain containing 12), the *PRKCG* (protein kinase C gamma) and the *CACNG7* (calcium voltage-gated channel auxiliary subunit gamma 7) were found.

Keywords: Garfagnina goat breed; Small Ruminant Lentiviruses; GoatSNP60 BeadChip; GWAS

Maedi/Visna Virus (MVV) and Caprine Arthritis Encephalitis Virus (CAEV) are members of the Retrovidae family, genus Lentivirus. Both viruses are unequivocal closely related lentivirus and thus

designated as small ruminant lentiviruses (SRLV) (Zanoni, 1998). Lentivirus in small ruminants are responsible of a slow progressive infection and the onset of clinical signs involving lungs, mammary glands and joint is insidious and seldom detected in adult animals. Leukoencephalomyelitis in kids (2-4 months) is occasionally observed (Clements and Narayan, 1989). SRLVs infections are endemic in sheep and goat industries in European countries, causing significant economic losses as result of the slow progression and insidious nature of the disease primarily due to premature culling, reduced number of offspring and reduced lamb/kid weights (Petursson et al. 1976). Transmission occurs from infected mothers to offspring by colostrum assumption and between adults mainly through respiratory secretions. In spite of both humoral and cell-mediated immune responses are developed, infected animals fail to clear the virus. Presently, there are no effective vaccines or treatments against lentivirus infections and consequently eradication and prevention of the infection depend mainly on serological diagnosis and early culling strategies. The evidence of some breed differences in susceptibility to SRLV infection (Peterhans et al. 2004) suggests an alternative strategy to reduce the incidence of the infection selecting the most genetic resistant animals. Research carried out in many countries/breeds provided evidence of differential individual and breed susceptibility as well (Larruskain and Begoña 2013). However, selection for resistance to infectious disease is difficult and requires the identification of specific genes or genetic markers linked to resistance either to infection or to disease. Several attempts to identify loci associate with resistance to lentivirus infections have been made during the last 15 years. Some loci were identified: in particular *TMEM154* gene (Transmembrane protein 154) (Clawson et al., 2015; Heaton et al., 2012; Sider et al. 2013; White et al, 2012), *CCR5* (C-C chemokine receptor type 5) (Wang et al., 2014) and other Cytokines (IL2/IL2R, IL4, IL8, IFNy, TGF-β1, MCP-1, GM-CSF) seem to have an important role in the SRLV infection susceptibility (Larruskain and Hugo, 2013). Also the major histocompatibility complex (MHC) region, a polymorphic multi-gene complex located on chromosome 20 in sheep and chromosome 23 in goats, has been also implicated in SRLV infection and SRLV-induced disease (Larruskain and Hugo, 2013).

In a previous paper (Cecchi et al. 2018) an association studies using short tandem repeats (STR markers) was performed in this native Italian goat population and it revealed for the first time that ETH10 locus (on chromosome 5) was weakly associated with the infection.

The aim of the present work is to identify genomic regions associated with SRLV infection in goat using a genome-wide scan analysis. The study was performed on Garfagnina goat individuals. Garfagnina is an Italian native goat breed registered on the Tuscany regional repertory of genetic resources at risk of extinction with about 745 animals belonging to 17 flocks. The study was

performed in a flock located in the Garfagnana district (Media Valle del Serchio, Lucca, Italy) and consisting of 269 females and 20 males. All animals were registered in the herdbook, but genealogical information was not available (Cecchi et al., 2017). All animal procedures used in this study were in agreement with the ethical and animal welfare concerns from the "Committee on the Ethics of Animal Experiments of Minimally Invasive Surgery Centre" and fully complied with recommendations outlined by the Italian laws.

Blood and serum samples were collected from 48 adult (from 2 to 9 years) goats. In order to evaluate the SRLV seroprevalence and characterize the SRLV antibody response sera samples were analyzed by a genotype specific ELISA test (Eradikit SRLV Screening kit, In3diagnostic, Torino, Italy). The genotyping ELISA plates are coated with a mix of gag and env peptides belonging to the three most divergent SRLV viral genotypes: genotype A (Maedi Visna like), B (CAEV like) and E (Roccaverano strain). The serum samples were examined according to the manufacturer's instructions. Specific genotype was assigned when the OD value of a single well was >40% when compared to the OD value of the same sera in others wells coated with different antigens. Sera were considered as SRLV positive when positive on at least one genotype. Serological results indicated twenty-one seropositive goats and twenty-seven serologically negative goats.

All 48 Garfagnina goats were genotyped using the Illumina GoatSNP60 BeadChip (Illumina Inc., San Diego, CA). Single nucleotide polymorphisms (SNP) that fulfilled the following criteria were kept in the analysis: (1) call rate >95%, (2) minor allele frequency >0.005, and (3) no extreme deviation from Hardy-Weinberg proportions (HWP; P-value > 0.000001). After quality control, 48 animals and 51,565 SNP, distributed over 29 autosomes and the X-chromosome, were retained (1,023 SNP were excluded due to low minor allele frequency and 421 SNP due to low call rate).

The association analysis was conducted using the GenABEL package in R (R core team, 2013; GenABEL, 2013). A score test (function "mmscore") was adopted (Amin et al., 2007; Chen and Abecasis, 2007). Firstly, an additive polygenic model was fitted using the genomic relationship matrix:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{a} + \mathbf{e}, \qquad (1)$$

where $\mathbf{y}$ was a vector containing the binary phenotypic records (0=negative and 1=positive); β was a vector with the fixed effects and $\mathbf{X}$ was an incidence matrix that associates each observation to specific levels of factors in β. The random effects in the model consisted of the animal and the residual terms, which were assumed normally distributed as $\boldsymbol{a} \sim N(0, \mathbf{G}\sigma_g^2)$ and $\boldsymbol{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{G}$ was

the genomic relationship matrix, **I** is an identity matrix, and σ_g^2 and σ_e^2 are the additive genomic and residual variances, respectively. The **G** matrix was constructed within the GenABEL R package, where for a given pair of individuals i and j, the identical by state coefficients ($f_{i,j}$) is calculated as:

$$f_{i,j} = \frac{1}{N}\sum_k \frac{(x_{i,k} - p_k) \times (x_{j,k} - p_k)}{p_k \times (1 - p_k)}$$

where N is the number of markers used, $x_{i,k}$ is the genotype of the i^{th} individual at the k^{th} SNP (coded as 0, ½ and 1), p_k is the frequency of the "+" allele and k= 1, …, N.

Then, the residuals obtained in (1) were regressed on the SNP (single marker regression) to test for associations. Associations with P-value $\leq 5 \times 10^{-5}$ were considered significant (Burton et al., 2007). The R package "qqman" was used to produce Manhtattan and quantile-quantile (Q-Q) plots (Turner, 2014). A scan for genes 1Mbp upstream-downstream from the significant SNP was performed using the Ensembl Capra hircus ARS1 database (http://www.ensembl.org/index.html). Annotation of all the significant SNP was performed with the variant effect predictor (https://www.ensembl.org/Tools/VEP) using the Ensembl database and the ARS1 assembly.

Codice campo modificato

Codice campo modificato

Serological analysis revealed that twenty-one out of 48 tested sera (43.75%) resulted positive to SRLV. Four sera reacted specifically to genotype A antigens, five to genotype B and six to genotype E. Genotype characterization was not determined in six sera, although reactive to SRLV.

A principal component analysis revealed a "small" population structure in our dataset (Figure 1). The three identified groups (Group 1: 33 animals; Group 2: 6 animals; Group 3: 9 animals), however, had a similar dispersion of case/control animals. Moreover, the variation explained by the first two principal components was relative small (~19% of the original variance). As reported previously (Cecchi et al., 2017), we postulated that each cluster refers to daughters of related individuals. Animals of SRLV-seropositive group were subdivided in these three groups (Group 1: 13 positive goats; Group 2: 2 positive goats and Group 3: 6 positive goats).

One SNP was found significant ($P < 5\times10^{-5}$) on CHR 18 at 62,360,918bp. Figure 2 represents the Manhattan plot of –log(P-values) for the genome wide association study (GWAS). The red horizontal line indicates a –log10(P-values) of 4.30 (corresponding to P-value = 5×10^{-5}). Chromosome 0 indicates single nucleotide polymorphisms with unknown position on the genome.

The significant SNP was an intron of the zinc finger protein 331 (ZNF331) protein (~62.35-62.37Mb). In the region 1Mb upstream-downstream the significant SNP, the NLRP12 (NLR family pyrin domain containing 12; ~62.48-62.51Mb), the PRKCG (protein kinase C gamma; ~62.32-

63.34Mb) and the CACNG7 (calcium voltage-gated channel auxiliary subunit gamma 7; ~63.34-63.36Mb) were found.

The protein encoded by the gene NLRP12 acts as a suppressor of the inflammatory response by activated macrophages. Mutations of this protein have been reported in patients with primary immunodeficiencies (Borte et al. 2014). The protein encoded by PRKCG gene is one of the PKC family members. This protein kinase is expressed solely in the brain and spinal cord and its localization is restricted to neurons. Alterations of this protein have been related to the development of neurodegenerative diseases in a mouse model (Ji, et al. 2014). The protein encoded by the CACNG7 gene is a type II transmembrane AMPA receptor regulatory protein (TARP) (Chen et al 2007). TARP regulates AMPA receptors (ionotropic non-NMDA post-synaptic receptor for glutamate, important for rapid excitatory synaptic transmission of the central nervous system, SNC).

As previously reported, several attempt to identify loci associate with resistance to lentivirus infections in small ruminant has been made using different markers, including microsatellites (Cecchi et al., 2018; Larruskain et al., 2010) and SNP. In particular, using the first-generation ovine 50kSNP chip, Heaton et al. (2012) conduced a GWAS in a mixed-breed sheep population and two SNPs in the same region of chromosome OAR17 showed the highest and third-highest genome-wide significance. White et al. (2012) in a separate GWAS performed on different sheep breeds confirmed the TMEM154 results of Heaton et al. (2012) and reported 12 additional genomic regions associated with odds of infection and provided 13 regions associated with control of infection.

In conclusion, despite that our study was performed on a limited number of individuals, it describes for the first time the involvement in susceptibility/resistance to goat lentivirus of some genes never identified before confirming that studies on different breeds can lead to different results. Future works may include replication of this study with a larger number of animals and fine mapping of candidate genes.

Acknowledgment This work was supported by grants of the University of Pisa (PRA 2016).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest

Figure 1 - Scatterplot of the first two principal components (PC1 vs. PC2). In brackets the variance explained by each principal component.

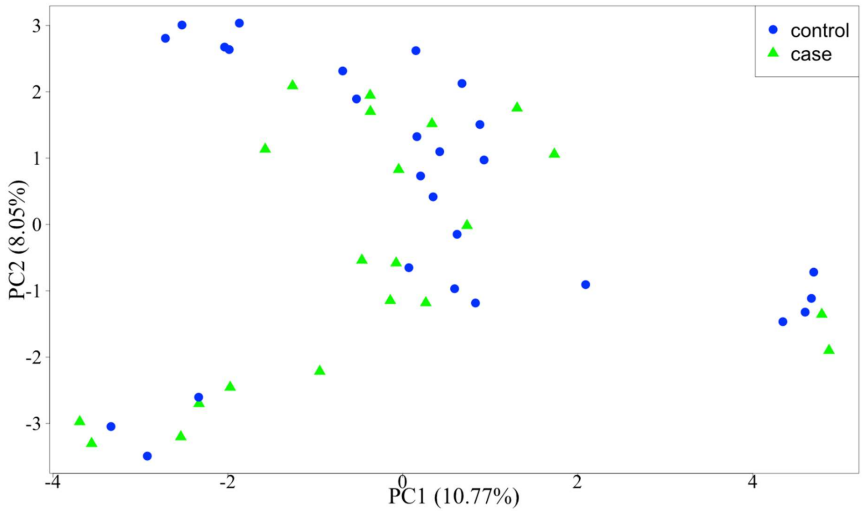
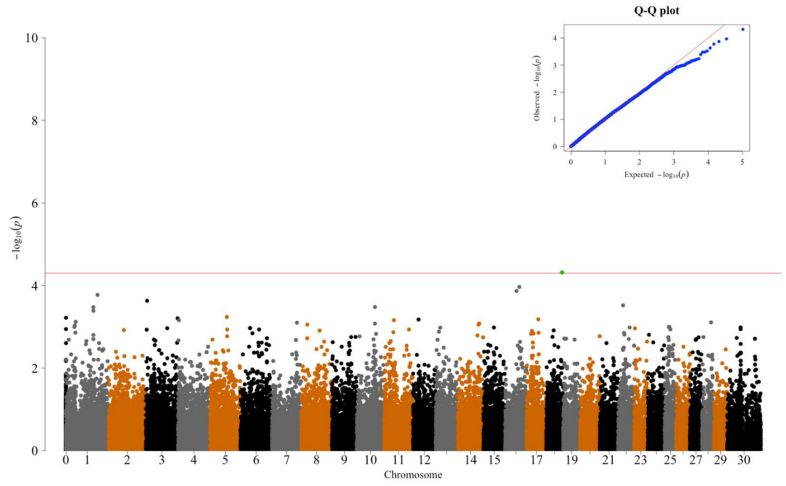


Figure 2 - Manhattan plot of –log(*P*-values) for the genome wide association study (GWAS).



REFERENCES

Amin, N., C.M.V., Duijn, Y.S. and Aulchenko. 2007. A genomic background based method for association analysis in related individuals. PloS One 2:1274.

Borte, S., Celiksoy, M.H., Menzel, V., Ozkaya, O., Ozen, F.Z., Hammarström, L. and Yildiran, A., 2014.Novel NLRP12 mutations associated with intestinal amyloidosis in a patient diagnosed with common variable immunodeficiency. Clinical Immunology, 154, 105-111.

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I.,. Ouwehand, W.H. and Samani, N.J., 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447, 661–678.

Cecchi, F., Russo, C., Iamartino, D., Galiero, A., Turchi, B., Fratini, F., Degl'Innocenti, S., Mazza, R., Biffani, S., Preziuso, G. and Cantile, C., 2017. Identification of candidate genes for paratuberculosis resistance in the native Italian Garfagnina goat breed. Tropical Animal Health and Production, 49, 1135-1142.

Cecchi, F., Fratini, F., Bandecchi, P., Cantile, C. and Mazzei, M., 2018. Investigation on goat lentiviral infections and preliminary association analysis with microsatellites in the native Garfagnina goats breed. Rendiconti Lincei. Scienze Fisiche e Naturali, in press, doi.org/10.1007/s12210-018-0729-0

Chen, W.-M. and Abecasis, G.R,2007. Family-Based Association Tests for Genomewide Association Scans. American Journal of Human Genetics, 81, 913–926.

Chen, R.S., Deng, T.C., Garcia, T., Sellers, Z.M. and Best, P.M., 2007. Calcium channel gamma subunits: a functionally diverse protein family. Cell Biochemistry and Biophysics, 47, 178-186.

191 Clawson, M.L., Redden, R., Schuller, G., Heaton, M.P., Workman, A., Chitko-McKown, C.G.,
192 Smith, T.P. and Leymaster, K.A., 2015. Genetic subgroup of small ruminant lentiviruses that
193 infects sheep homozygous for TMEM154 frameshift deletion mutation A4Δ53. Veterinary
194 Research, 46, 22.
195 Clements, J.E. and Narayan, O., 1989. Biology and pathogenesis of lentiviruses. Journal of General
196 Virology, 70, 1617–39.
197 Heaton, M.P., Clawson, M.L., Chitko-McKown, C.G., Leymaster, K.A., Smith, T.P., et al. 2012.
198 Reduced Lentivirus Susceptibility in Sheep with TMEM154 Mutations. PLoS Genet 8:
199 e1002467.
200 Ji, J. Hassler, M.L., Shimobayashi, E., Paka, N., Streit, R. and Kapfhammer, J.P., 2014. Increased
201 protein kinase C gamma activity induces Purkinje cell pathology in a mouse model of
202 spinocerebellar ataxia 14. Neurobiology Disease, 70, 1-11.
203 Larruskain, A., Minguijón, E., García-Etxebarria, K., Moreno, B., Arostegui, I., Juste, R.A. and Jugo,
204 B.M., 2010. MHC class II DRB1 gene polymorphism in the pathogenesis of Maedi-Visna and
205 pulmonary adenocarcinoma viral diseases in sheep. Immunogenetics, 62(2), 75-83
206 Larruskain, A. and Begoña, J.M., 2013. Retroviral Infections in Sheep and Goats: Small Ruminant
207 Lentiviruses and Host Interaction. Viruses, 8, 2043-2061
208 GenABEL project developers. 2013. GenABEL: Genome-wide SNP association analysis. .R package
209 version 1.8-0.:http://CRAN.R-project.org/package=GenABEL.
210 Peterhans, E., Geenland, T., Badiola, J., Harkiss, G., Bertoni, G., Amorena, B., et al. 2004. Routes of
211 transmission and consequences of small ruminant lentiviruses (SRLVs) Infection and
212 eradication schemes. Veterinary Research, 35, 257–74.
213 Petursson, G., Nathansson, N., Georgsson, G., Panitch, H. and Palsson P.A., 1976. Pathogenesis of
214 Visna. I. Sequential virologic, serologic, and pathologic studies. Laboratory Investigation, 35,
215 402–12.

Sider, LH., Heaton, M.P., Chitko-McKown, C.G., Harhay, G.P., Smith, T.P.L., Leymaster, K.A., Laegreid, W.W. and Clawson, M.L. 2013. Small ruminant lentivirus genetic subgroups associate with sheep TMEM154 genotypes. Veterinary Research, 44, 64.

R Core Team. 2013. R: A language and environment for statistical computing. http://www.R-project.org/.

Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. bioRxiv 005165.

Zanoni, R.G., 1998. Phylogenetic analysis of small ruminant lentiviruses. J. Gen. Journal of General Virology, 79, 1951–1961.

Wang, W., Ye, C., Di Zhang, J.L., Kimata, J.T. and Zhou, P., 2014. CCR5 Gene Disruption via Lentiviral Vectors Expressing Cas9 and Single Guided RNA Renders Cells Resistant to HIV-1 Infection PLOS Published: December 26, 2014 https://doi.org/10.1371/journal.pone.0115987

White, S.N, Mousel, M.R., Herrmann-Hoesing, L.M., Reynolds, J.O., Leymaster, K.A., et al. 2012. Genome-Wide Association Identifies Multiple Genomic Regions Associated with Susceptibility to and Control of Ovine Lentivirus. PLoS ONE 7: e47829.