Detection of candidate regions affecting bovine IgM natural antibodies in milk

J.M. Cordero-Solorzano[1,2,4], J.A.J. Arts[3], H.K. Parmentier[3] & H. Bovenhuis[2]

[1] Swedish University of Agricultural Sciences, Animal Breeding and Genetics, Box 7023,750 07, Uppsala, Sweden juan.cordero@slu.se (Corresponding author)
[2] Wageningen University & Research, Animal Breeding and Genomics, P.O. Box 338, 6700 AH, Wageningen, the Netherlands
[3] Wageningen University & Research, Adaptation Physiology Group, P.O. Box 338, 6700 AH, Wageningen, the Netherlands
[4] Animal Health Service of Costa Rica (SENASA), P.O. Box 3-3006, 40104, Heredia, Costa Rica

Summary

Two genomic regions were found to be associated with IgM antibody titers in milk binding lipoteichoic acid (LTA); one on chromosome 17 and the other on chromosome 21. Phenotypes were measured by ELISA and genotypes consisted of imputed 777k SNP. Single SNP analyses were run using an animal model to retrieve significant SNP. In BTA21, a gene for the heavy chains of immunoglobulins, *IGHV* is proposed as a candidate gene as was true for *VPREB3* in BTA17 related to B-cell maturation. These findings provide a further step in better understanding the genetic background of natural antibodies (NAbs), that may be relevant for relevant for the estimation of dairy cattle health.

Keywords: dairy cattle, natural antibodies, genome-wide association

Introduction

Natural antibodies (NAbs) function as a part of the innate immunity, without any antigenic stimulation (Avrameas 1991). They provide a first line of defense against pathogens before specific antibodies (SpAbs) are produced as an adaptive specific immune response. NAbs have a polyreactive nature, low affinity and target both self-antigens and common microbial structures like Pathogen-Associated Molecular Patterns (PAMPs) (Baumgarth *et al.* 2005). Relevant microbial PAMPs include lipoteichoic acid (LTA), lipopolysaccharide (LPS) and peptidoglycan (PGN).

Immunoglobulins are classified into isotypes like IgM, IgG, IgA and others, based on their heavy chain structure. In the early stages of an infection IgM is the first isotype produced, allowing a quick response to a variety of PAMPs or other antigens (Schroeder & Cavacini 2010). Upon antigenic stimulation, B cells may switch isotype and convert into plasma cells that produce more specific IgG antibodies (Liljavirta *et al.* 2014).

LTA is an integral part of the cell wall of gram-positive bacteria, such as *Staphylococcus aureus*, one of the pathogens known to cause mastitis in dairy cattle. This molecule is not synthesized by mammals and upon exposure, it triggers an innate immune response (Schneewind & Missiakas 2014). An important part of this immune response rests on NAbs, making them an interest target to potentially select for disease resistance.

Milk NAb titers are heritable (Wijga *et al.* 2013), but little is known about the

underlying genes. Using genotypic information of thousands of single nucleotide polymorphisms (SNP) from individuals in a population it is possible to perform a genome-wide association study (GWAS). This analysis may detect genomic regions involved in traits potentially leading to the identification of candidate genes (e.g. Hayes & Goddard 2010).

The aim of this study was to identify genomic regions associated with LTA-IgM titers in milk and describe potential candidate genes.

Material and methods

Milk samples and data

Samples and phenotypes were collected as part of the Dutch Milk Genomics Initiative (for details see Stoop *et al.* 2008). Natural IgM antibodies binding LTA were measured in morning milk samples using an indirect ELISA (Ploegaert *et al.* 2010) from 1695 cows from 380 commercial dairy herds in the Netherlands. The average number of days in lactation when milk samples were taken was 166, with a range from 66 to 263 days.

Imputed 777K SNP data from a 50K SNP array were available for 1736 animals from the Milk Genomics project. SNP positions were mapped using BosTau8 assembly of the bovine genome (UMD 3.1.1). Genotyping and imputation of the samples was as described by Duchemin *et al.* (2014).

Genome-wide association study (GWAS)

A GWAS was performed using, the following animal model:

[1]

where y is the observation of the trait of interest; μ is the overall mean of the trait; $dim_{ijklmno}$ is a covariate describing the effect of days in lactation, modeled with a Wilmink curve (Wilmink 1987); $ca_{ijklmno}$ is a covariate accounting for the effect of age at first calving, $season_k$ is the fixed effect with 3 classes of calving season (June–August 2004, September–November 2004, and December 2004–February 2005); $scode_l$ is the fixed effect of sire type in three classes: proven bull, young bull or other proven bull; $animal_o$ is the random additive genetic effect assumed to be distributed as $N(\mathbf{0}, \mathbf{A})$, where **A** is the additive genetic relationships matrix consisting of 12,998 animals (4 generations) provided by CRV (Dutch herdbook), and is the additive genetic variance; SNP_m is the fixed effect of SNP genotype; $herd_n$ is the random herd effect assumed to be distributed as $N(\mathbf{0}, \mathbf{I})$, where **I** is the identity matrix, and is the herd variance; $e_{ijklmno}$ is the random residual effect assumed to be distributed as $N(\mathbf{0}, \mathbf{I})$, where **I** is the identity matrix, and is the residual variance.

The additive genetic variance and the herd variance were fixed at the values calculated from analyses based on model [1] without the inclusion of SNP effects. The additive genetic variance, herd variance and other genetic parameters were previously reported by Wijga *et al.* (2013). All the analyses were performed using ASReml 4.1 (Gilmour *et al.* 2015).

Quality control

To minimize the number of false positives due to multiple testing, a false discovery rate (FDR) threshold was applied using the R package "qvalue" (Storey *et al.* 2015). SNP associations were considered suggestive if 0.05<FDR<0.20 and significant if FDR<0.05. Suggestive and significant associations were grouped into genomic regions; SNPs were considered in the same genomic regions if they were less than 200kb apart. At 200kb the average linkage disequilibrium (r^2) is assumed to decrease approximately to 0.15 (de Roos *et al.* 2008, Khatkar *et al.* 2008).

Markers with genotype classes of 1 to 10 cows were excluded from the GWAS. For significant SNPs, the sensitivity of association was evaluated if they had a genotype class with less than 50 cows. In those cases, that genotype class was removed and the SNP retested to confirm the significance of the association. Phenotypes were regarded as extreme when the sequential increase of the titers within a genotype class showed an interval of at least one titer and there were only one or two values above or below this interval. If extreme NAb titers were present, the SNP was retested without these extreme phenotypes to confirm the association.

Genomic inflation factor (λ) due to residual population stratification, as suggested by QQ plots of $-\log_{10}$ P values (not shown) was accounted for using genomic-control correction.

Results and Discussion

Associations

Our GWAS was performed on 1630 animals, which met the criteria of having both LTA-IgM phenotypes and 777K imputed genotypes. Out of 777962 SNPs, a total of 575806 were used for the analysis after quality control.

We found two genomic regions on two different chromosomes significantly associated with LTA-IgM titers in milk. In total six SNPs were found significant, four on chromosome 21 and two on chromosome 17. There was also one SNP with suggestive association on chromosome 18 (Table 1).

Candidate genes

On chromosome 17 (Figure 1a), several genes are found in the region between 72.9 – 73.3 Mb, but only two have a direct known immune function. One of them is *MIF* (macrophage migration inhibition factor), a cytokine involved in the regulation of macrophage function as well as other innate immune responses (Calandra & Roger 2003). Further analyses should reveal a possible relationship between this gene and natural antibodies.

The other gene is *VPREB3* (Pre-B Lymphocyte Protein 3), which is part of the surrogate light chain (SLC), a heterodimer that is temporarily expressed during B cell development. SLC is composed of two polypeptides, VPREB and IGLL1 (Immunoglobulin lambda-like polypeptide 1) that are homologous to the variable and the constant domain of the immunoglobulin light chain, respectively (Ekman *et al.* 2009). Interestingly, less than 100 kb upstream from the region of interest, we find the genes *VPREB2* and *IGLL1*.

When analyzing the region 71.2 – 71.6 Mb on chromosome 21 (Figure 1b), we find the *IGHV* (Immunoglobulin heavy locus) gene. This locus includes all the segments to produce the heavy chain: V (variable), D (diversity), J (joining), and C (constant). During B cell development, a recombination event rearranges these segments which are then transcribed as

a mu (IgM) heavy chain (Niku *et al.* 2012, Ma *et al.* 2016).

Conclusions

We were able to identify two genomic regions affecting LTA-binding IgM titers in milk. These regions contain genes related to B cell development and immunoglobulins synthesis. Still, further research needs to be made to provide a better understanding of the immunogenetics behind this trait.

Acknowledgements

The first author currently benefits from a joint grant from the European Commission and Swedish farmers' foundation for agricultural research (SLF), within the framework of the Erasmus Mundus joint doctorate “EGS-ABG”.

List of References

Avrameas, S. 1991. Natural autoantibodies: from “horror autotoxicus” to “gnothi seauton”. Immunol. Today 12: 154–9.

Baumgarth, N., J.W. Tung & L.A. Herzenberg. 2005. Inherent specificities in natural antibodies: a key to immune defense against pathogen invasion. Springer Semin. Immunopathol. 26: 347–62.

Calandra, T. & T. Roger. 2003. Macrophage migration inhibitory factor: a regulator of innate immunity. Nat. Rev. Immunol. 3: 791–800.

Duchemin, S.I., M.H.P.W. Visker, J.A.M. Van Arendonk & H. Bovenhuis. 2014. A quantitative trait locus on Bos taurus autosome 17 explains a large proportion of the genetic variation in de novo synthesized milk fatty acids. J. Dairy Sci. 97: 7276–85.

Ekman, A., M. Niku, J. Liljavirta & A. Iivanainen. 2009. Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle. BMC Immunol. 10: 22.

Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham & R. Thompson. 2015. ASReml user guide release 4.1. VSN Int. Ltd.

Hayes, B. & M. Goddard. 2010. Genome-wide association and genomic selection in animal breeding. Genome 53: 876–83.

Khatkar, M.S., F.W. Nicholas, A.R. Collins, K.R. Zenger, J.A.L. Cavanagh, W. Barris, R.D. Schnabel, J.F. Taylor & H.W. Raadsma. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9: 187.

Liljavirta, J., M. Niku, T. Pessa-Morikawa, A. Ekman & A. Iivanainen. 2014. Expansion of the preimmune antibody repertoire by junctional diversity in Bos taurus. PLoS One 9: e99808.

Ma, L., T. Qin, D. Chu, X. Cheng, J. Wang, X. Wang, P. Wang, H. Han, L. Ren, R. Aitken, L. Hammarström, N. Li & Y. Zhao. 2016. Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle. J. Immunol. 196: 4358–66.

Niku, M., J. Liljavirta, K. Durkin, E. Schroderus & A. Iivanainen. 2012. The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. Dev. Comp. Immunol. 37: 457–61.

Ploegaert, T.C.W., S. Wijga, E. Tijhaar, J.J. van der Poel, T.J.G.M. Lam, H.F.J. Savelkoul,

H.K. Parmentier & J. a. M. van Arendonk. 2010. Genetic variation of natural antibodies in milk of Dutch Holstein-Friesian cows. J. Dairy Sci. 93: 5467–5473.
de Roos, A.P.W., B.J. Hayes, R.J. Spelman & M.E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics 179: 1503–12.
Schneewind, O. & D. Missiakas. 2014. Lipoteichoic acids, phosphate-containing polymers in the envelope of gram-positive bacteria. J. Bacteriol. 196: 1133–42.
Schroeder, H.W. & L. Cavacini. 2010. Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125: S41-52.
Stoop, W.M., J.A.M. van Arendonk, J.M.L. Heck, H.J.F. van Valenberg & H. Bovenhuis. 2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch Holstein-Friesians. J. Dairy Sci. 91: 385–94.
Storey, J.D., A.J. Bass, A. Dabney & D. Robinson. 2015. qvalue: Q-value estimation for false discovery rate control.
Wijga, S., H. Bovenhuis, J.W.M. Bastiaansen, J. a. M. van Arendonk, T.C.W. Ploegaert, E. Tijhaar & J.J. van der Poel. 2013. Genetic parameters for natural antibody isotype titers in milk of Dutch Holstein-Friesians. Anim. Genet. 44: 485–492.
Wilmink, J.B.M. 1987. Adjustment of test-day milk, fat and protein yield for age, season and stage of lactation. Livest. Prod. Sci. 16: 335–348.

Table 1. Lead SNP of each chromosome with suggestive or significant associations with LTA-IgM.

SNP	rs ID	chr	Position	MAF [1]	Major/minor allele	Genotype effect (SE)	-log10 P value
BovineHD2100020886	rs135338912	21	71482201	0.30	C/T	CC: -0.17 (0.05) TT: 0.34 (0.07)	6.9
BovineHD1700021382	rs133519711	17	73125915	0.10	A/G	AA: 0.35 (0.06) GG: 0.04 (0.24)	6.5
BovineHD1800014677	rs134833064	18	49839220	0.38	A/G	AA: 0.08 (0.06) GG: -0.28 (0.05)	6.0 [2]

[1] Minor Allele Frequency
[2] Suggestive association (FDR 0.20)

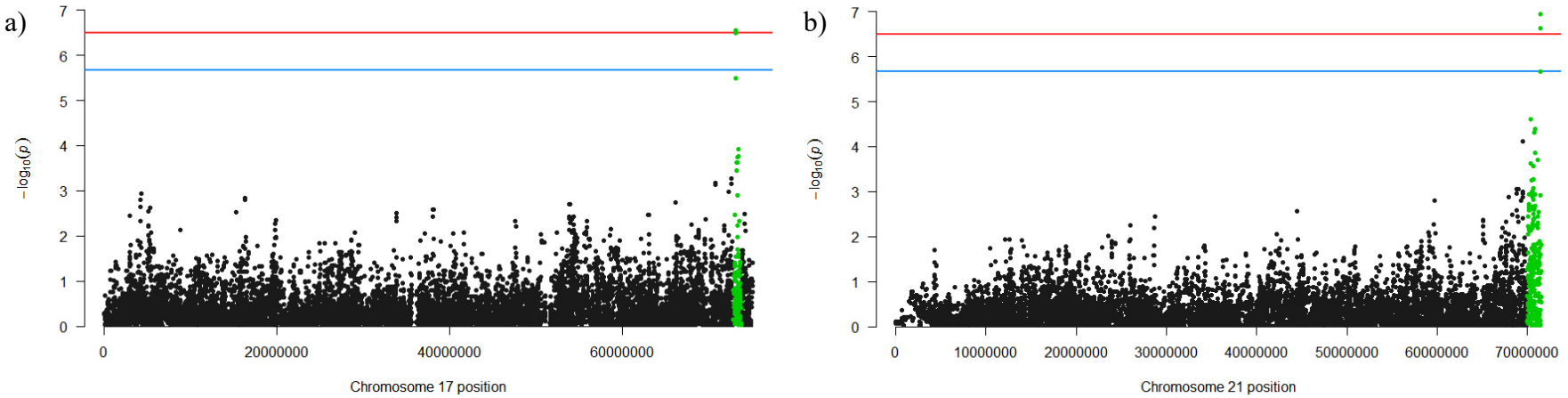

Figure 1. –log10 P values of a) chromosome 17 and b) chromosome 21 from single SNP analysis for LTA-IgM. The green highlighted area is the region of interest. FDR 0.05 threshold is represented by the red line and FDR 0.20 by the blue one.