Genome-wide patterns of copy number variation in case control mastitis study of Polish-HF cows using high-density Illumina BovineSNP50 v3 BeadChip array

Kacper Żukowski[1]*, Piotr Topolski[1], Andrzej Żarnecki[1]

[1]Department of Cattle Breeding and Genetics, National Research Institute of Animal Production, Balice, Poland.
andrzej.zarnecki@izoo.krakow.pl, piotr.topolski@izoo.krakow.pl,
kacper.zukowski@izoo.krakow.pl

*Corresponding author:
Kacper Żukowski, Department of Cattle Breeding and Genetics, National Research Institute of Animal Production, Balice, Poland. Email: kacper.zukowski@izoo.krakow.pl

Kacper Żukowski, Piotr Topolski, Andrzej Żarnecki

Abstract

Background: The mastitis is one of the most important diseases affecting losses of milk yield and functional longevity resulting in economic losses in dairy farms. The cow's inflammation of the mammary gland is a consequence of complex processes in the innate immune system, activation of several cells and accompanying factors that lead to the elimination of microorganism invaders. The study aimed to investigate genetic background by the genome-wide association study (GWAS) for identified copy number variations (CNVs) in one of the most important economically disorders – the mastitis in dairy cattle.

Methods: We investigated 682 Polish-HF cows previously genotyped and analyzed using the following steps: i) identification of CNVs across all the cows based on ARS-UCD1.2 genome map, ii) to perform the GWAS with the utilization of logistic regression models, iii) functional analysis for identified CNVs to indicate putative candidate genes. The cow's genotypes were deposited in the Polish cattle SNP genotypes database (cSNP-db). The phenotype database information covered the veterinary history of mastitis was collected from four experimental farms belonging to the National Institute of Animal Production, Balice, Poland.

Results: The study identified 326 CNVs of which only 28 CNVs were significant with p-value <0.05. The most significant CNVs were identified for BTA17, BTA22, and BTA25, respectively. However, the most interesting from the functional point of view of previously reported CNVs overlapped with genes associated with mastitis were loaded on BTA03 (MSH4), BTA18 (FOXC2, FOXL1, MTHFSD) and BTA27 (TLR3), respectively.
Conclusions: Study clearly concluded that the identification of candidate genes associated with mastitis using of CNVs is feasible. Therefore, one can apply this GWAS approach between the common CNVs and mastitis as a case-control experiment. Our study identified few genes like MSH4, MTHFSD, FOXC2, FOXL1, and TLR3 which were overlapped with tree CNVs to functionally assess and might be considered as the candidate genes for mastitis in Polish HF cattle population. Moreover, obtained results supported by the different genome and transcriptome analysis could lead to a better understood of the genetic background of mastitis.

Keywords: mastitis; cattle; CNVs; GWAS; case-control.

Introduction

The cow's inflammation of mammary gland – mastitis is one of the most important factors determining their production capacity in terms of milk

yield and functional longevity. The udder health is affected by numerous factors related to the udder structure, housing conditions and the way of milking, as well as, other factors related to the functioning of the immune system, among which those of a hereditary nature is the subject of broad breeders' interest, mainly due to the possibility of improving them under the program's breeding. One of the essential elements of the body's response to pathogenic factors, especially microorganisms, appear to be proteins from the β-defensin family, which are expressed in the respiratory system, reproductive system and mammary gland of cows. The activity of these proteins may be affected by genetic polymorphism, which by modifying the sequence of proteins and the activity of their synthesis may play an important role in the functioning of individual elements of the immune response, including elements associated with the action of β-defensins. The β-defensins are amphipathic cationic peptides, acting as antimicrobial peptides (AMP) for gram-negative bacteria, gram-positive bacteria, viruses, fungi and other unicellular parasites [1]. Their antimicrobial activity consists mainly of permeabilisation of the cell membrane or stimulation of hydrolases, leading to degradation of the cell wall [2]. Currently, 58 β-defensin genes are known, located in four clusters in the bovine genome, on chromosomes: 8, 13, 23 and 27, which were named as cluster A, B, C and D, respectively [3–4]. Among the β-defensin genes located within BTA27, there are, among others: LAP, DEFB1, DEFB5, DEFB10 and DEFB4A. A direct relationship between LAP expression and somatic cell count (SCC) has also been demonstrated in which a higher concentration of LAP transcripts was observed in the milk of cattle infected with Streptococcus aureus, Streptococcus bovis, Streptococcus dysgalactiae and E. coli than in healthy cows [5–7]. It has also been found that the DEFB1 gene is inducible expression during mammary gland infection [5], while DEFB5 transcripts are described in coagulase-positive, and coagulase-negative Staphylococcus infected gland [8]. Importantly, β-defensins have also been shown to affect milk performance traits, where β4-defensin polymorphisms (DEFB4A) were significantly associated with protein yield and fat or protein content [9] and milk yield [10] in Polish Holstein-Friesian cattle.

The copy number variation (CNV) is a type of genetic variation of genome sequences up to 5 million base pairs. The CNVs show the high impact on function and gene expression, such as altering gene dosage, disrupting coding sequence, or perturbing long-range gene regulation [11–12]. There were only a few studies using CNVs as genetic markers for genome-wide association study (GWAS) with diseases and economic phenotype: study of Zhou et al. 2016 [13] associates CNVs with growth traits in Bos indicus, studies of Xu et al. 2014 [14] and Zhou et al. 2018 [12] with milk production traits in Holsteins and study of Ben Sassi et al. 2016 [15] with economically important traits in Spanish Holstein. Here, we applied CNV detection and GWAS analysis for mastitis phenotypes reported as case-control of Polish HF dairy cattle. The study aimed to investigate genetic background by the genome-wide association (GWAS) for identified CNV in one of the most important economically disorders – mastitis in dairy cattle.

Materials and Methods

Animals: A total of 682 Polish HF dairy cows with available DNA, identified among high production cows were selected from four experimental farms belonging to the National Institute of Animal Production. The using PLOWET platform collected the observations of mastitis. The user-friendly PLOWET system allows collecting a group of traits associated with health, reproduction, disease prevention and control linked with the history of veterinary treatments, insemination and calving, and postpartum characteristics of calves.

Methodology: The Polish HF Cows were genotyped with the Illumina BovineSNP50 v3 BeadChip (Illumina Inc., San Diego, CA) at the Department of Animal Molecular Biology of the National Institute of Animal Production. Ear tissue samples used to extract DNA were collected in the course of a routine procedure. The raw genotype data were routinely deposited in Polish cattle SNP genotypes database (cSNP) [16]. We per-

formed additional data collection for the log R ratio (LRR) and the B allele frequency (BAF) values. The exclusion criteria for SNPs were minor allele frequency <0.001 and call rate <0.95. The exclusion criteria for animals was call rate <0.95. The position of SNPs was remapped onto the ARS-UCD1.2 cow genome assembly [17]. After quality control and removing SNPs located on Y and MT chromosomes and unassigned chromosome, 52,972 SNPs remained for the copy number variation genome-wide association analysis.

Data were analysed by using the PennCNV software (v1.0.5) with the default parameters and with a dedicated population frequency of the B allele file [18]. According to the hidden types and states of CNV we were able to recognise fill deletion, single copy deletion, single copy duplication, double copy duplication and normal state for homozygote across all autosomes and chromosome X [19]. The identified CNVs with a frequency lower than 5% were removed from the analysis.

Next, we employed a case-control genome-wide association study (GWAS) design with the model assuming correction by the population stratification procedure based on multidimensional scaling analysis [20]. We assume the disease status depends on CNV presence (without distinguishing on type of CNV) through a logistic regression (LR) model:

$$Logit(y) = \beta_0 + \beta_1 CNV$$

Likelihood ratio test (LRT) was used to compare the null model with SNP, tested models, where one predictor variables were incorporated in the model. LRT test the null hypothesis:

$$H_0 : \beta_0 = 0$$

The deviation with the null model (−2 log-likelihood [−2LL] of the null model) was compared to the corresponding term for a given model with explanatory variables (−2 log L of the given model). The differences between the two −2 log L produce the Chi-square statistic (χ^2), with

k degrees of freedom. The associated CNV regions were functionally an-
notated to 27,607 genes deposited in the Ensembl database.

Results

Estimation of population stratification using multidimensional scaling
analysis: To examine the possible population structure in the analysed
Polish HF population, multidimensional scaling analysis was performed
for all 682 Polish HF cows based on the kinship matrix estimated from
SNPs. The four farm sub-group cows were mixing between each other,
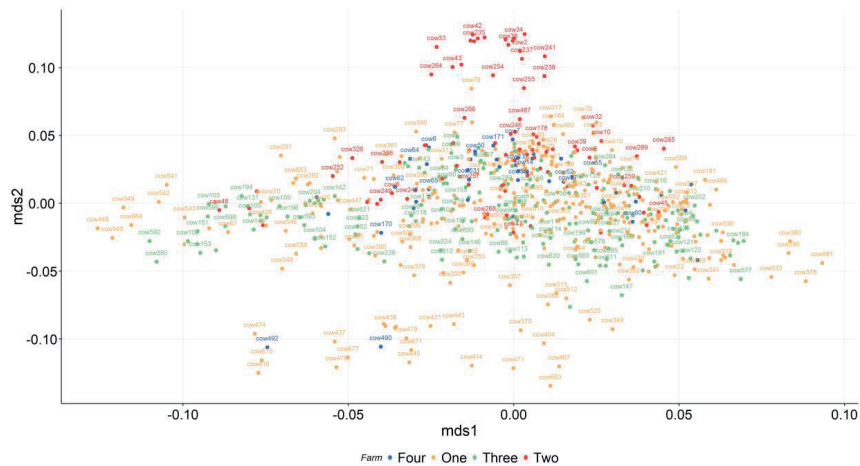and the results were shown in Figure 1.

Figure 1. Multidimensional scaling analysis of the total sample of 682 high produc-
tion Polish HF cows exanimated on four experimental farms

Genome-wide association analysis accounting for copy number
variation: A total of 5627 CNV events was identified using PennCNV
software, and the results were shown in Table 1. Most of them were re-
ported as a single copy deletion (3065). Next, the dataset was limited to
events which were observed with a frequency higher than 5% and based
on this. We were able to identify the highest number of 162 CNV events

for single-copy deletion. Based on these assumptions, the dataset was analysed using a logistic regression CNV model. A comparison of both models with the likelihood-ratio test shows that only 28 significant CNVs (Table 1) with p-value higher than 0.05. However, we were not able to recognise significant CNV region for double copy duplication. The manhattan plot of the obtained p values was presented in Figure 2.

Table 1. The number of CNV events across interpretation state and subsequent analysis steps across all investigated Polish HF cows

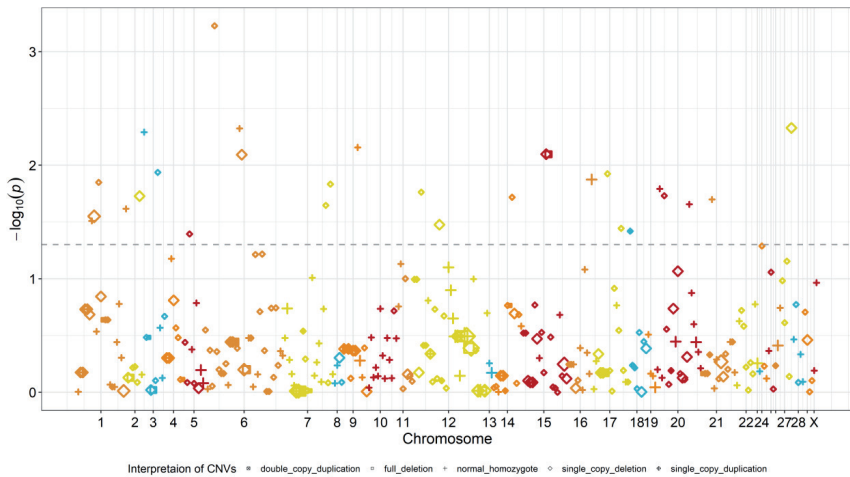| CNV state interpretation	Overall count	CNV count with frequency < 5%	CNV count with p value < 0.05
Double copy duplication	23	6	0
Full deletion	411	11	1
Normal homozygote	1782	117	9
Single copy deletion	3065	162	17
Single copy duplication	346	30	1
Sum	5627	326	28

Figure 2. Manhattan plot showing the LRT values from the test of association of CVN region with mastitis. The dashed line shows a significant level of 0.05

The most significant CNVs were identified for BTA17, BTA22 and BTA25. Most of them were interpreted as single-copy deletion. One of the events, full deletion, was identified on BTA24. The most frequent, represented of 87 cases, were identified on BTA01 and the CADM2 gene was indicated as candidate gene overpaid for region chr1:33655711-33734709. The CNV reported on BTA18 as single-copy duplication overlapped with tree genes: MTHFSD, FOXC2, FOXL1. The longest and covering the largest number of SNPs CNV region was identified on BTA06 and contained 200 kbp and 17 SNPs, respectively. The list of the highly significant CNV events with the assigned candidate genes was presented in Table 2.

Discussion

The analysis of GWAS procedure was preceded by CNV identification. The similar procedure, in consideration of cattle was previously used by Zhou et al. 2016 [12], Zhou et al. 2018 [13] and Xu et al. 2014 [14] respectively, but instead of case-control mastitis design, they used phenotype observations. Moreover, the presented study based on ARS-UCD1.2 instead of UMD3.1 cattle genome assembly previously utilized [12-14] and all SNP positions were remapped to the ARS-UCD1.2 release. Identification of CNVs is highly dependent by the density of SNP map or trivially speaking by used technology or chip [21]. Therefore, we were able to identify only 3845 CNVs events (not including homozygote regions) across 682 Polish HF cows. The shared CNV regions were aligned to barely 80 events, of which 25 constitutes loss of heterozygosity regions. The obtained results fit in summary reported by Upadhyay et al. 2017 [21] and covering studies based on mid (50K) and high-density (777K) microarrays, comparative genomic hybridisation and whole-genome sequencing. There were six CNVs overlapping with 17 protein-coding genes and one lncRNAs ENSBTAG00000050741 with the unknown function and aetiology. However, the high local similarity between ENSBTAG00000050741, human and the mouse reference genome has been shown and this results suggested candidate genes like MYO6 or MPZL2 which played a vital

Table 2. The list of significant CNV regions associated with mastitis with overlapped genes in Polish HF cows

CNV region name	CNV interpret	Chr	CNV start	CNV end	Count of observed region	#SNPs in CNV region	Length of CNV region	p-value	Associated Gene Id
chr1:28434475-33218070	normal homozygote	1	28434475	33218070	10	101	4783596	0.03	
chr1:33655711-33734709	single copy deletion	1	33655711	33734709	87	3	78999	0.02	
chr1:71322090-71380741	single copy deletion	1	71322090	71380741	11	3	58652	0.01	CADM2, NCBP2, PIGZ, SENP5, PAK2, NCBP2AS2
chr1:152960777-156486697	normal homozygote	1	152960777	156486697	10	93	3525921	0.02	
chr2:83650214-83776119	single copy deletion	2	83650214	83776119	37	3	125906	0.01	
chr3:32843393-33110685	normal homozygote	3	32843393	33110685	17	77	267293	<0.01	
chr3:68898933-69057163	single copy deletion	3	68898933	69057163	13	5	158231	0.01	MSH4
chr5:15073278-15127156	single copy deletion	5	15073278	15127156	12	3	53879	0.04	
chr6:8342803-8450316	single copy deletion	6	8342803	8450316	17	4	107514	<0.01	
chr6:51317005-51682000	normal homozygote	6	51317005	51682000	15	22	364996	<0.01	
chr6:51924818-52189115	single copy deletion	6	51924818	52189115	22	17	264298	<0.01	
chr7:76379313-76464440	single copy deletion	7	76379313	76464440	10	3	85128	0.02	
chr7:85944120-85998146	single copy deletion	7	85944120	85998146	11	3	54027	0.01	
chr9:10712212-11341897	normal homozygote	9	10712212	11341897	13	42	629686	<0.01	

Kacper Żukowski, Piotr Topolski, Andrzej Żarnecki

CNV region name	CNV interpret	Chr	CNV start	CNV end	Count of observed region	#SNPs in CNV region	Length of CNV region	p-value	Associated Gene Id
chr12:1606902-1786737	single copy deletion	12	1606902	1786737	17	5	179836	0.01	TDRD3
chr12:56722262-56855497	single copy deletion	12	56722262	56855497	22	5	133236	0.03	
chr14:49908926-50010596	single copy deletion	14	49908926	50010596	10	3	101671	0.01	
chr15:11385265-11451408	full deletion	15	11385265	11451408	24	3	66144	<0.01	
chr16:79862681-80844792	normal homozygote	16	79862681	80844792	41	16	982112	0.01	
chr17:31007492-31204256	single copy deletion	17	31007492	31204256	12	6	196765	<0.01	
chr17:71729438-71789925	single copy deletion	17	71729438	71789925	15	3	60488	0.03	RAB36, BCR, SPEC-C1L, RSPH14, GNAZ, ENSBTAG00000050741
chr18:12358305-12450149	single copy duplication	18	12358305	12450149	19	3	91845	0.03	MTHFSD, FOXC2, FOXL1
chr20:21964055-22216842	normal homozygote	20	21964055	22216842	13	79	252788	0.01	
chr20:42942535-43133525	single copy deletion	20	42942535	43133525	11	5	190991	0.01	
chr20:70694677-71677629	normal homozygote	20	70694677	71677629	10	24	982953	0.02	
chr21:52223103-53177224	normal homozygote	21	52223103	53177224	13	21	954122	0.02	
chr27:16084760-16118966	single copy deletion	27	16084760	16118966	25	3	34207	<0.01	TLR3

role in tumorigenesis of breast cancer [22-23]. Among 17 protein-coding genes, twelve (BCR, CADM2, FOXC2, FOXL1, MSH4, MTHFSD, NCBP2, NCBP2AS2, PAK2, RAB36, TDRD3, TLR3) of them were previously reported as genes directly or indirectly associated with mastitis or breast cancer [21, 24-29] and conversely, one gene was located within a β-defensin region on BTA27 [30].

Conclusions

Our study concluded that the identification of genes associated with mastitis by the using of CNVs is possible. We have successfully applied this GWAS approach between the common CNVs and mastitis as a case-control experiment. We identified few genes like MSH4, MTHFSD, FOXC2, FOXL1, TLR3 overlapped with tree CNVs to functionally assess and therefore they might be considered as the putative candidate genes for mastitis. Furthermore, obtained results supported by the different genome and transcriptome analysis could lead to the better understood of the genetic background of mastitis.

References

[1] Brogden K.A. Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? Nat. Rev. Microbiol 2005;3:238–250.

[2] Bierbaum G., Sahl H.G. Induction of autolysis of staphylococci by the basic peptide antibiotics pep5 and nisin and their influence on the activity of autolytic enzymes. Arch Microbiol 1985;141:249–254.

[3] Meade K.G, Cormican P., Narciandi F., Lloyd A., O'farrelly C. Bovine – defensin gene family: Opportunities to improve animal health? Physiol. Genomics 2013;46:17–28.

[4] Gurao A., Kashyap S.K., Singh R. β-defensins: An innate defense for bovine mastitis. Veterinary World 2017;10:990–998.

[5] Roosen S., Exner K., Paul S., Schroeder J.M., Kalm E., Looft C. Bovine b-defensins: Identification and characterization of novel bovine b-defensin genes

and their expression in mammary gland tissue. Mamm. Genome 2004;15:834–842.

[6] Swanson K., Gorodetsky S., Good L., Davis S., Musgrave D., Stelwagen K., Farr V., Molenaar A. Expression of a beta-defensin mRNA, lingual antimicrobial peptide, in bovine mammary epithelial tissue is induced by mastitis. Infect Immunol 2004;72:7311–7314.

[7] Kazuhiro K., Akamatsu H., Obayashi T., Nagahata H., Higuchi H., Iwano H., Oshida T., Yoshimura Y., Isobe N. Relationship between concentration of lingual antimicrobial peptide and somatic cell count in milk of dairy cows. Vet. Immunol Immunopathol 2013;153:298–301.

[8] Kościuczuk E.M. Lisowski P., Jarczak J., Krzyżewski J., Zwierzchowski L., Bagnicka E. Expression patterns of β-defensin and cathelicidin genes in parenchyma of bovine mammary gland infected with coagulase-positive or coagulase-negative staphylococci. BMC Vet Res 2014;10:1.

[9] Bagnicka E., Strzałkowska N., Szreder T., Prusak B., Jóźwik A., Kosciuczuk E., Krzyzewski J., Zwierzchowski L. A/C polymorphism in the ß-4 defensin gene and its association with phenotypic and breeding values of milk production traits in Polish-Friesian cows. Animal Science Papers and Reports 2008;26:239–250.

[10] Krzyzewski J., Bagnicka E., Strzalkowska N., Jozwik A., Pyzel B., Zwierzchowski L. Association between the polymorphism of bovine beta4-defensin gene and milk traits in Holstein-Friesian cows as computed for standard [305 days] and the whole lactation. Animal Science Papers and Reports 2008;26:191–198.

[11] Zhang F., Gu W., Hurles M.E., Lupski J.R. Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 2009;10:451–481.

[12] Zhou Y., Connor E.E., Wiggans G.R., et al. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. BMC Genomics 2018;19:314.

[13] Zhou Y., Utsunomiya Y.T., Xu L., Hay E.H.A., Bickhart D.M., Alexandre P.A., et al. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genomics 2016;17:419.

[14] Xu L., Cole J.B., Bickhart D.M., Hou Y., Song J., VanRaden P.M., et al. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 2014;15:683.

[15] Ben Sassi N., González-Recio Ó., de Paz-del Río R., Rodríguez-Ramilo S.T., Fernández A.I. Associated effects of copy number variants on economically important traits in Spanish Holstein dairy cattle. J Dairy Sci 2016;99:6371–6380.

[16] Gurgul A., Jasielczuk I., Szmatoła T. et al. Genome-wide characteristics of copy number variation in Polish Holstein and Polish Red cattle using SNP genotyping assay. Genetica 2015;143:145–155.

[17] Schnabel R. ARS-UCD1.2 Cow Genome Assembly: Mapping of all existing variants https://www.animalgenome.org/repository/cattle/UMC_bovine_coordinates/ Accessed 11 Sep 2018.

[18] Wang K., Li M., Hadley D., Liu R., Glessner J., Grant S., Hakonarson H., Bucan M. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research 2007;17:1665-1674.

[19] Colella S., Yau C., Taylor J.M., Mirza G., Butler H., Clouston P., Basset A.S., Seller A., Holmes C., Ragoussis J. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Research 2007;35:2013–2025

[20] Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K., Sale M., Chen W.M. Robust relationship inference in genome-wide association studies. Bioinformatics 2016;26:2867–2873.

[21] Upadhyay M., Da Silva V. H., Megens H.J., Visker M.H.P.W., Ajmone-Marsan P., Bâlteanu V.A., Crooijmans R.P.M.A. Distribution and functionality of copy number variation across European cattle populations. Frontiers in Genetics 2017;8:108.

[22] Guttinger M., Sutti F., Panigada M., Porcellini S., Merati B., Mariani M., Grassi F. Epithelial V-like Antigen (EVA), a Novel Member of the Immunoglobulin Superfamily, Expressed in Embryonic Epithelia with a Potential Role as Homotypic Adhesion Molecule in Thymus Histogenesis. The Journal of Cell Biology 1998;141:1061–1071.

[23] Wang H., Wang B., Zhu W., Yang Z. Lentivirus-Mediated Knockdown of Myosin VI Inhibits Cell Proliferation of Breast Cancer Cell. Cancer Biotherapy and Radiopharmaceuticals 2015;30:330–335.

[24] Alvarez-Garcia I., Miska E.A. MicroRNA functions in animal development and human disease, Development 2015;132:4653–4662.

[25] Olsen M.L., Dilaveri CA Idiopathic granulomatous mastitis: a case report of breast abscess Case Reports 2011;2011:bcr0520114271.

21

[26] Souza, Fernando & Heinemann, Marcos & Gidlund, Magnus & Blagitz, Maiara & Maria, Alice & Paiva, Melville & Della Libera, Alice Maria & Oliveira, Monica & Cerqueira, Pinho & Reis, Luiza & Ramos Sanchez, Eduardo. (2012). The innate immunity in bovine mastitis: The role of pattern-recognition receptors. American Journal of Immunology 2012;166–178. 10.3844/ajisp.2012.166.178.

[27] Werling D., Cheng Z. Combining Genome Wide Association Studies and Differential Gene Expression Data Analyses Identifies Candidate Genes Affecting Mastitis Caused by Two Different Pathogens in the Dairy Cow. Open Journal of Animal Sciences 2015;5:358–393.

[28] Kadri N.K., Harland C., Faux P, et al. Coding and noncoding variants in HFM1, MLH3, MSH4, MSH5, RNF212, and RNF212B affect recombination rate in cattle. Genome Res 2016;26:1323–1332.

[29] Wawrzykowski J., Franczyk M., Hoedemaker M., Pries M., Streuff B., Kankofer M. Preliminary data on possible protein markers of parturition in cows. Reproduction in Domestic Animals 2017; 53.10.1111/rda.13079.

[30] Goldammer T., Kuehn C., Brunner R.M., Weikard R. Evolutionary Break Point Analysis between the Proximal Half of Bovine Chromosome 27 and Conserved Segments of the Human Genome. Cytogenetic and Genome Research 2009;125:33–39.