Acta fytotechn zootechn, 25, 2022(2): 91–96
http://www.acta.fapz.uniag.sk

Original Paper

Genome-wide association study of resistance to mastitis in Czech Holstein cattle

Kristýna Forejt[1], Nina Moravčíková[2], Radovan Kasarda[2], Ludmila Zavadilová[3], Luboš Vostrý[4]*

[1]*Palacký University Olomouc, Faculty of Science, Department of Experimental Biology, Czech Republic*
[2]*Slovak University of Agriculture in Nitra, Slovak Republic*
[3]*Institute of Animal Science, Czech Republic*
[4]*Czech University of Life Sciences Prague, Czech Republic*

Article Details: Received: 2021-12-16 | Accepted: 2021-12-16 | Available online: 2022-06-30

https://doi.org/10.15414/afz.2022.25.02.91-96

Clinical mastitis is an inflammatory disease of the mammary gland that largely impacts dairy farming profitability and welfare. Globally, a massive scientific effort is being made to elucidate the possible link of certain genotypes to the susceptibility to this disease. After data pruning controlling for genotype missingness, minor allele frequency, and population stratification, 51 557 SNPs from 1 042 animals have been analysed using the general linear model (GLM). Two SNPs, BTA-121769-no-rs and BTB-00265951, have demonstrated statistically significant associations ($-\log_{10}(p) > 6.0134$), both located on the chromosome BTA6. The detected SNPs have been annotated within a reference genome. They have been found to lie outside of transcribed regions but within the vicinity of genes essential for the immune response. This finding further supports the case for their significance in the resistance to mastitis. In addition, 14 relatively weaker associations ($-\log_{10}(p) > 4$) have been observed across chromosomes BTA1, 2, 9, 14, 19, 24 and 25.

Keywords: association study, GWAS, SNP, cattle, *Bos taurus*

1 Introduction

Mastitis is one of the most common as well as most costly health concerns in dairy cattle husbandry. It is an umbrella term for any inflammation of the mammary gland, ranging from subclinical infections detected only by a raised somatic cell count in milk to severe inflammation leading to gangrene and sepsis in one or multiple inflamed udder quarters (Hofírek & Haas, 2003). The susceptibility to mastitis is particularly high during the period of 15 to 30 days postpartum and the first two months of lactation (Sodeland et al., 2011; Waller, 2000). It grows with the number of lactations the cow has had (Wolfová et al., 2020).

Affected cows should be temporarily or even permanently removed from milk production, causing significant financial losses through lost milk, culling, medicaments, and additional labour. Recent studies have estimated the average loss at over 400 USD (Kvapilík, 2015; Rollin et al., 2015) per affected cow. Wolfová et al. (2020) have estimated the cost of mastitis in the Czech Republic at 43.63 to 84.84 € per cow and year, with the lactational incidence reaching 0.35, 0.45 and 0.57 for the first, second and third or subsequent lactations, respectively.

Susceptibility to the disease is generally influenced by housing conditions, epizootological situation in regard to the most common causal pathogens (usually bacteria) (Hofírek & Haas, 2003) and the individual genetic predisposition of each animal. The latter has long focused on breeding efforts; however, conventional selection methods have only yielded limited success.

To effectively address this problem, a thorough study of genetic factors influencing immunocompetence in response to the infection and general resilience is required. With the onset of affordable genotyping panels in the dairy industry, genomic approaches are being explored.

***Corresponding Author:** Michal Rolinec, Slovak University of Agriculture in Nitra, Faculty of Agrobiology and Food Resources, Institute of Nutrition and Genomic, Trieda Andreja Hlinku 2, 949 76 Nitra, Slovakia; e-mail: michal.rolinec@uniag.sk ORCID https://orcid.org/0000-0003-4966-7909

Association analysis is best suited for a large number of subjects (thousands), which in turn complicates the effort to eliminate environmental confounding variables. Resistance to mastitis has therefore proven especially exacting to study – previous studies show only limited consistency in the associated regions.

Sodeland et al. (2011) have examined the data of 2589 Red Norwegian sires with 1389776 daughters with records on clinical mastitis and detected associated regions on chromosomes 2, 6 and 20 for the periparturient period and 14 for mastitis in late lactation. Tiezzi et al. (2014) have studied the US population of Holstein bulls and their daughters on their first lactations; they have detected associations on chromosomes 2, 14 and 20. Sahana et al. (2014) ran a crossbreed GWAS on a sample of the Danish cattle population, detecting a strong association on chromosome 6 for Holstein cattle, 13 and 19 for Nordic Red cattle and 16 and 20 for Jersey cattle. Strilacci et al. (2014) have studied the Valdostana Red Pied cattle and detected a total of 171 associated SNPs located across chromosomes 1, 2, 3, 4, 9, 13, 15, 17, 21 and 22. Fang et al. (2017) have detected no association with genome-wide significance. Welderufael et al. (2018) have detected no association with genome-wide significance in the Danish Holstein cattle. Szyda et al. (2019) have combined the study of SNPs with a study of copy number variations (CNVs) in the Polish Holstein-Friesan cattle and concluded that CNVs might be more strongly associated with the mastitis trait than SNPs (191 CNV hits as compared to 17 SNP hits on chromosome 4 and 27).

The present study has examined the association of SNP markers with mastitis in a sample of the Czech population of Holstein cattle to supply further data and enable genomic region prioritisation.

2 Material and methods

2.1 Data

DNA microarray data have been obtained from a total of 1258 Holstein cows via Illumina BovineSNP50 DNA Analysis BeadChip (Illumina Inc., San Diego, USA), testing for 53 218 single-nucleotide polymorphisms (SNPs). Phenotype data have been collected from breeders in a binary format (whether a cow has been diagnosed with mastitis throughout its life or not). The initial dataset consisted of 203 mastitis cases and 1055 controls.

2.2 Data pruning

The initial dataset was pruned using the software PLINK (Purcell et al., 2007): individual call rate >0.9, SNP call rate >0.9 and minor allele frequency >10%. A total of 1661 SNPs and 34 individuals were excluded from further analysis.

The remaining 51557 SNPs and 1224 individuals have been controlled for population stratification, leading to false-positive signals if not accounted for (Hellwege et al., 2018). Four dimensions have been generated via multidimensional scaling (MDS) and plotted against each other in Figure 1a and Figure 1b using the scatter plot function in *R* (R Core Team, 2020). The blue lines represent 2 SD (standard deviation) thresholds beyond which the outliers were excluded from further analysis (n = 216). Individuals located within pale blue areas were included in the final dataset.

2.3 Association analysis

Association analysis was performed in an open-source software TASSEL 5 (Bradbury et al., 2017) using the general linear model (GLM):

$$Y = X\beta + \varepsilon$$

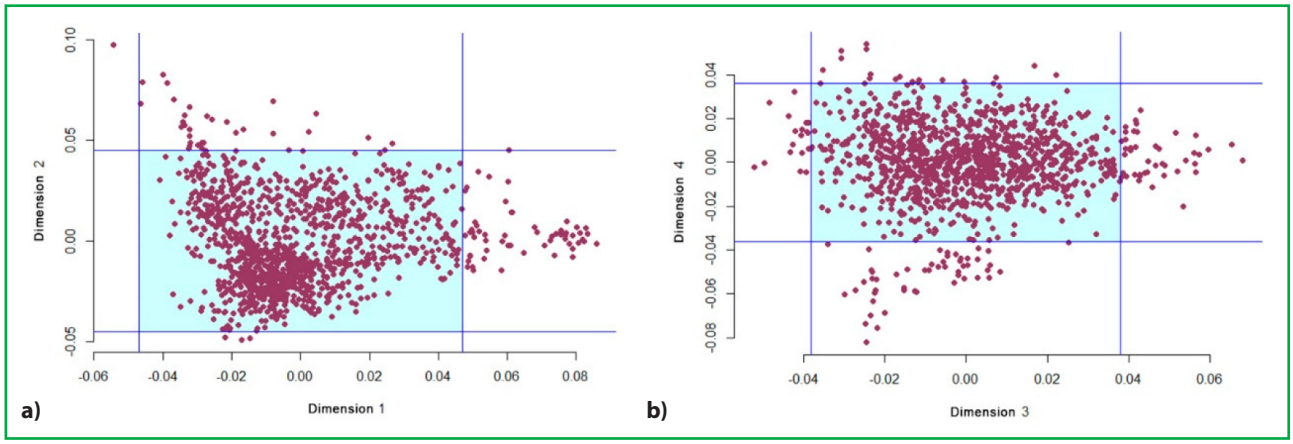

Figure 1 Scatter plots of population stratification using multidimensional scaling (MDS): a) Dimension 1 and 2, b) Dimension 3 and 4

where: Y – a matrix of dependent variables (phenotypes); X – a matrix of independent variables (genotypes); β – a matrix of estimated parameters; ε – a matrix of residual errors

2.4 Post-analysis quality control

Manhattan graph was produced using the $-\log_{10}$ of the *p*-values to depict the strength of observed association across autosomes. Both the Q-Q plot and the Manhattan plot have been generated in R using the qqman package (Turner, 2018).

To account for a multiple testing, the standard Bonferroni correction was applied to the significance threshold:

$$\alpha = \frac{0.05}{51557} = 0.000000097$$

$$-\log_{10}(\alpha) = 6.0134$$

Suggestive value was selected as second threshold:

$$-\log_{10}(p) = 4$$

SNPs exceeding these thresholds were taken for further consideration.

2.5 Annotation and allele effects

SNPs with the strongest detected association have been localised in the ARS-UCD1.2 *Bos taurus* genome assembly using the Basic Local Alignment Search Tool (BLAST) (Yates et al., 2020).

For potential applications in genomic selection, allele effects were estimated for each SNP using TASSEL 5 (Bradbury et al., 2017), and best alleles (with the desirable effect on mastitis resistance) were identified.

3 Results and discussion

Figure 2 shows the Q-Q plot of SNP marker *p*-values. A strong deviation has been observed at its tail while the majority of the plot followed a normal distribution. The results were therefore evaluated as reliable.

SNP markers were sorted by *p*-values; 14 of the most significant ($-\log(p) > 4$) hits are listed in Table 1, which also includes the best allele for each SNP. An animal with the best allele would be less likely to suffer from mastitis.

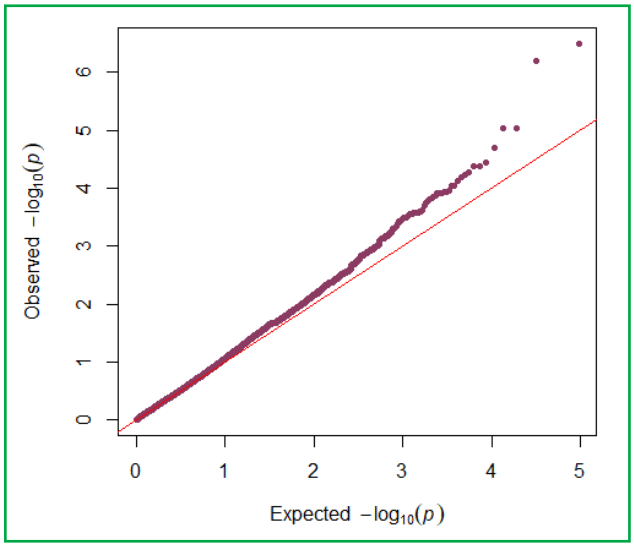

Figure 2 Q-Q plot of the observed null distribution of the *p*-values against the expected null distribution of the *p*-values

Table 1 SNPs manifesting the strongest association to the resistance to mastitis

SNP ID	Chromosome	Position (bp)	*p*-value	$-\log_{10}(p)$	Best allele
BTB-00265951	6	90642682	3.12E-07	6.506	C
BTA-121769-no-rs	6	89912196	6.25E-07	6.204	G
BTB-01283269	14	20196617	9.25E-06	5.034	A
BTA-36062-no-rs	14	20216601	9.25E-06	5.034	T
ARS-BFGL-NGS-109683	25	23825593	1.99E-05	4.701	C
ARS-BFGL-NGS-37189	25	32912638	3.65E-05	4.437	C
ARS-BFGL-NGS-104096	9	7010227	4.11E-05	4.386	C
Hapmap43710-BTA-86183	2	79846105	4.12E-05	4.385	A
BTA-28028-no-rs	1	14551165	5.40E-05	4.268	T
ARS-BFGL-BAC-23887	14	49905143	5.98E-05	4.223	C
ARS-BFGL-NGS-115947	14	5494654	6.50E-05	4.187	C
ARS-BFGL-NGS-1837	19	26556279	7.61E-05	4.119	C
Hapmap50627-BTA-23969	24	13582752	8.84E-05	4.053	G
UA-IFASA-7226	14	36276655	8.97E-05	4.047	C

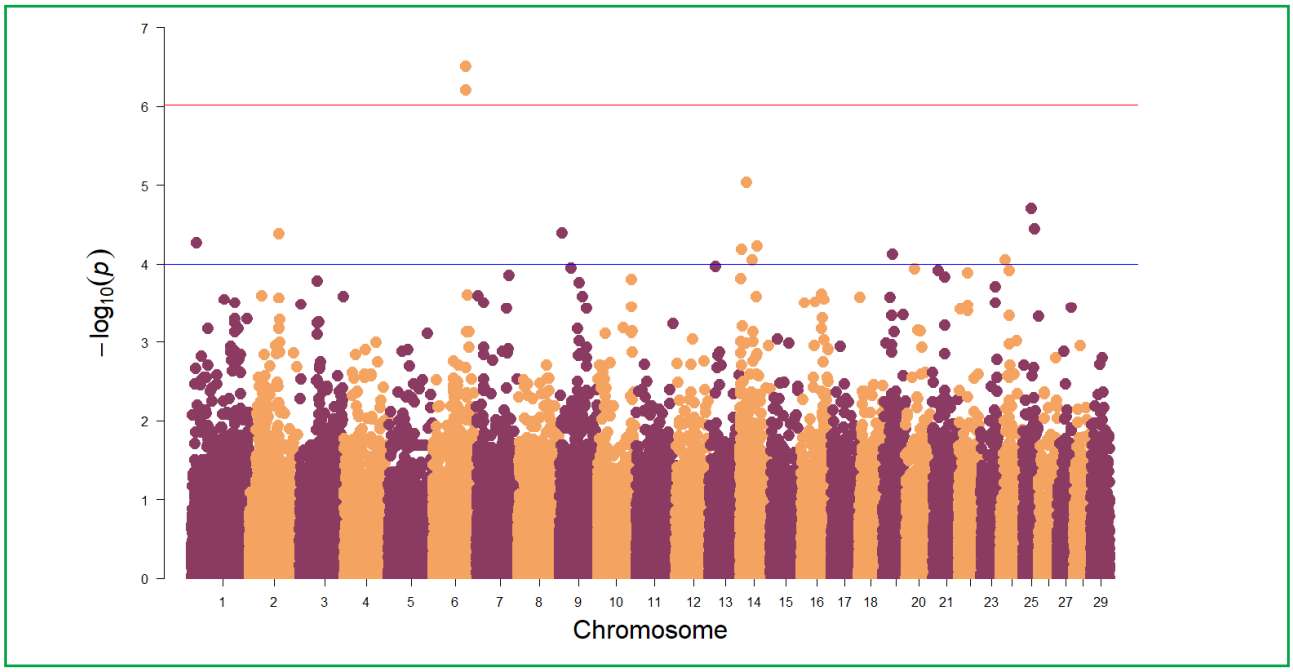

Figure 3 Manhattan plot for each SNP studied for the resistance to mastitis in the Czech Holstein cattle
the red line represents the significance threshold after Bonferroni correction; the blue line represents a suggestive value of $-\log_{10}(p) = 4$

According to Gebreyesus et al. (2019), the incorporation of GWAS results into genomic prediction can be used to improve its reliability.

Figure 3 shows the Manhattan plot of SNP association for the resistance to mastitis. The red line represents the genome-wide significance threshold after Bonferroni correction (Figure 3). The blue line represents a suggestive value of $-\log_{10}(p) = 4$, beyond which SNPs were taken for further consideration.

Two SNPs, BTB-00265951 and BTA-121769-no-rs, have shown a statistically significant association with the mastitis trait (red line). Both are located on chromosome BTA6 within 1 Mb distance from each other. Multiple other SNPs on chromosomes 1, 2, 9, 14, 19, 24 and 25 have surpassed only the suggestive value (blue line).

Both SNPs located on chromosome BTA6 were annotated within the reference genome and scanned for adjacent genes. BTB-00265951 lies outside transcribed regions, but within a cluster of genes including CXCL5, ENSBTAG00000027534, CXCL8, CXCL2 and ENSBTAG00000011961. BTA-121769-no-rs lies outside transcribed regions, but within the immediate vicinity of genes COX18 and ANKRD17.

Gilbert et al. (2013) suggest that two of the closest genes downstream from the identified SNPs may play a role in the immune response: CLX5 codes a C-X-C motif chemokine ligand with heightened expression in bovine culture cells infected with *Staphylococcus aureus*, a common mastitis-causing pathogen. Gene COX18 is coding a mitochondrial cytochrome c oxidase assembly factor (Bourens & Barrientos, 2017).

Chromosome 6, along with chromosome 14, has repeatedly shown significance for various parameters of milk production (milk yield, protein and fat yield) (Jiang et al., 2010; Jiang et al., 2019; Atashi et al., 2020). Some previous GWASs of bovine mastitis have also detected strong associations around the 90 Mb mark on chromosome 6 (Sodeland et al., 2011; Sahana et al., 2014; Abdel-Shafy et al., 2014), although others (Tiezzi et al., 2014; Welderufael et al., 2018; Fang et al., 2017; Strillacci et al., 2014) have not. Out of the latter, Fang et al. (2017) and Welderufael et al. (2018) have not detected any significant associations. Tiezzi et al. (2014) have detected associations located on chromosomes 2, 8, 14, 11, 16, 19, 20, 24 and 29, with chromosome 14 demonstrating the most hits, while Strillacci et al. (2014) have detected associations on chromosomes 1, 2, 3, 4, 9, 13, 15, 17, 21 and 22. Fang et al. (2017) propose that the inconsistency could be partially caused by breed differences.

Sodelant et al. (2011) have observed an association of SNPs between positions 90.67 Mb and 96.19 Mb of chromosome 6 in Norwegian Red cattle. However, they have also detected associations on chromosomes 2, 4, 7, 9, 10, 13, 14, 16, 17, 20 and 29 that the present study has not. Sahana et al. (2014) have observed the strongest association on chromosome 6 across three breeds of dairy cattle (Holstein, Jersey and Nordic Red), but they

have also detected relatively weaker associations on chromosomes 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 18, 19, 20, 22 and 23. Both SNPs detected in the present study lie within the area of interest observed by Sahana et al. (2014). Abdel-Shafy et al. (2014) have also observed significant associations in this area of chromosome 6 in their study of German Holstein cattle, with additional associations on chromosomes 5, 13, 16, 19 and X.

Olsen et al. (2016) have studied the delimited area on chromosome 6 and suggested GC (group-specific component located on 88.69–88.74 Mb of chromosome 6) as the gene most likely responsible for the mastitis associations observed in the region. The protein encoded by this gene is a vitamin D-binding protein with a suggested role in both milk production and immune defence mechanisms.

4 Conclusions

The present study has identified a genomic region on chromosome BTA6 at approximately 90 Mb, which may potentially bear one or more causal mutations relevant to the susceptibility to mastitis. This information can contribute to increase the proportion of genetic variance explained by QTL. The further analysis of the causative mutations would be apropriate to enable direct selection for the deliminate deleterious alleles without the disadvantage of using linked SNP. While more studies are necessary, this research contributes to the knowledge of the mechanisms of resistance to mastitis to be implemented in breeding programs.

Acknowledgements

The authors wish to express their gratitude to the Holstein Cattle Breeders' Association of the Czech Republic for providing genomic data and to concerned breeders for the provision of phenotypic data. The study was supported by the projects QK1810252 and APVV-17-0060.

References

Abdel-Shafy, H. et al. (2014). Short communication: validation of somatic cell score-associated loci identified in a genome-wide association study in German Holstein cattle. *Journal of Dairy Science*, 97(4), 2481–2486. https://doi.org/10.3168/jds.2013-7149

Atashi, H. et al. (2020). Genome-wide association for milk production and lactation curve parameters in Holstein dairy cows. *Journal of Animal Breeding and Genetics*, 137(3), 292–304. https://doi.org/10.1111/jbg.12442

Bourens, M., & Barrientos, A. (2017). Human mitochondrial cytochrome c oxidase assembly factor COX18 acts transiently as a membrane insertase within the subunit 2 maturation module. *Journal of Biological Chemistry*, 292(19), 7774–7783. https://doi.org/10.1074/jbc.M117.778514

Bradbury, P.J. et al. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Fang, L. et al. (2017). Integrating Sequence-based GWAS and RNA-Seq Provides Novel Insights into the Genetic Basis of Mastitis and Milk Production in Dairy Cattle. *Scientific Reports*, 7, 45560. https://doi.org/10.1038/srep45560

Gebreyesus, G. et al. (2019). Reliability of genomic prediction for milk fatty acid composition by using a multi-population reference and incorporating GWAS results. *Genetics Selection Evolution*, 51, 16. https://doi.org/10.1186/s12711-019-0460-z

Gilbert, F.B. et al. (2013). Differential response of bovine mammary epithelial cells to *Staphylococcus aureus* or *Escherichia coli* agonists of the innate immune system. *Veterinary Research*, 44(1), 40. https://doi.org/10.1186/1297-9716-44-40

Hellwege, J.N. et al. (2017). Population Stratification in Genetic Association Studies. *Current Protocols in Human Genetics*, 95, 1.22.1–1.22.23. https://doi.org/10.1002/cphg.48

Hofírek, B. & Haas, D. (2003). Kategorizace zdraví mléčné žlázy, klinické formy mastitid a jejich terapie [Categorization of mammary gland health, clinical forms of mastitis and their therapy]. *Sborník referátů odborného semináře mastitidy skotu ČBS a VFU*, 1, 10–22.

Jiang, J. et al. (2019). A Large-Scale Genome-Wide Association Study in U.S. Holstein Cattle. *Frontiers in Genetics*, 10, 412. https://doi.org/10.3389/fgene.2019.00412

Jiang, L. et al. (2010). Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One*, 5(10), e13661. https://doi.org/10.1371/journal.pone.0013661

Kašná, E. et al. (2018). First national recording of health traits in dairy cows in the Czech Republic. *ICAR Technical Series*, 23, 173–177.

Kvapilík, J. (2014). Mastitidy dojených krav a ekonomické ztráty [Mastitis of dairy cows and economic losses]. *Veterinářství*, 64(12), 946–955.

Olsen, H.G. et al. (2016). Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. *Genetics Selection Evolution*, 48(1), 79. https://doi.org/10.1186/s12711-016-0257-2

Pearson, T.A., & Manolio, T.A. (2008). How to interpret a genome-wide association study. *JAMA*, 299(11), 1335–1344. https://doi.org/10.1001/jama.299.11.1335

Purcell, S. et al. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81. https://doi.org/10.1086/519795

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Available from: https://www.R-project.org/ (accessed February 2021).

Rollin, E. et al. (2015). The cost of clinical mastitis in the first 30 days of lactation: An economic modeling tool. *Preventive Veterinary Medicine*, 122(3), 257–264. https://doi.org/10.1016/j.prevetmed.2015.11.006

Sahana, G. et al. (2014). Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *Journal of Dairy Science*, 97(11), 7258–7275. https://doi.org/10.3168/jds.2014-8141

Snelling, W.M. (2019). A survey of polymorphisms detected from sequences of popular beef breeds. *Journal of Animal Science*, 93(11), 5128–5143. https://doi.org/10.2527/jas.2015-9356

Sodeland, M. et al. (2011). Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Animal Genetics*, 42(5), 457–65. https://doi.org/10.1111/j.1365-2052.2010.02165.x

Strillacci, M.G. et al. (2014). Genome-wide association study for somatic cell score in Valdostana Red Pied cattle breed using pooled DNA. *BMC Genetics*, 15(106). https://doi.org/10.1186/s12863-014-0106-7

Szyda, J. et al. (2019. The genetic background of clinical mastitis in Holstein-Friesian cattle. *Animal*, 13(10), 2156–2163. https://doi.org/10.1017/S1751731119000338

Tiezzi, F. et al. (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS One*, e0114919. https://doi.org/10.1371/journal.pone.0114919

Turner, S. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25), 731. https://doi.org/10.1101/005165

Waller, K.P. (2000). Mammary gland immunology around parturition. Influence of stress, nutrition and genetics. *Advances in Experimental Medicine and Biology*, 480, 231–245. https://doi.org/10.1007/0-306-46832-8_29

Welderufael, B.G. et al. (2018). Genome-Wide Association Study for Susceptibility to and Recoverability From Mastitis in Danish Holstein Cows. *Frontiers in Genetics*, 9,141. https://doi.org/10.3389/fgene.2018.00141

Wolfová, M. et al. (2006). Incidence and economics of clinical mastitis in five Holstein herds in the Czech Republic. *Preventive veterinary medicine*, 77(1–2), 48–64. https://doi.org/10.1016/j.prevetmed.2006.06.002

Yates, A.D. et al. (2020). Ensembl 2020. *Nucleic Acids Research*, 48(D1). D682–D688. https://doi.org/10.1093/nar/gkz966