UNIVERSITY OF LJUBLJANA
BIOTECHNICAL FACULTY

Marija ŠPEHAR

GENOMIC EVALUATION AND ASSOCIATION STUDIES OF CORRELATED TRAITS IN DAIRY AND DUAL PURPOSE CATTLE BREEDS

DOCTORAL DISSERTATION

GENOMSKO VREDNOTENJE IN ASOCIACIJSKE ŠTUDIJE KORELIRANIH LASTNOSTI PRI GOVEDU MLEČNIH IN KOMBINIRANIH PASEM

DOKTORSKA DISERTACIJA

Ljubljana, 2015

On the basis of the Statute of the University of Ljubljana and according to the decisions of the Senate of the Biotechnical Faculty and the University Senate dating 11th of Febuary 2014, it has been confirmed that the candidate fulfils all the conditions for acquiring a PhD in Doctorate Postgraduate Study of Biotechnical Sciences, Animal Science. Assist. prof. Gregor Gorjanc PhD has been appointed as candidate's supervisor and prof. Ino Čurik PhD as co-supervisor.

Commission on doctoral dissertation assessment and defense:

Chairman: Prof. Dr. Peter DOVČ
University of Ljubljana, Biotechnical Faculty, Department of Animal Science
Member: Assist. Prof. Dr. Klemen POTOČNIK
University of Ljubljana, Biotechnical Faculty, Department of Animal Science
Member: Prof. Dr. Johann SÖELKNER
University of Natural Resources and Life Sciences, Vienna

Date of dissertation defense: 24.03.2015

I, the undersigned doctoral candidate declare that this doctoral dissertation is a result of my own research work and that the electronic and printed versions are identical. I am hereby non-paidly, non-exclusively, and spatially and timelessly unlimitedly transferring to University the right to store this authorial work in electronic version and to reproduce it, and the right to enable it publicly accessible on the web pages of Digital Library of Biotechnical faculty.

Marija ŠPEHAR

Na podlagi Statuta Univerze v Ljubljani ter po sklepu Senata Biotehniške fakultete in sklepa Komisije za doktorski študij Univerze v Ljubljani z dne 11. 2. 2014 je bilo potrjeno, da kandidatka izpolnjuje pogoje za opravljanje doktorata znanosti na Interdisciplinarnem doktorskem študijskem programu Bioznanosti, znanstveno področje Znanost o živalih. Za mentorja je bil imenovan doc. dr. Gregor Gorjanc ter za somentorja prof. dr. Ino Čurik.

Komisija za oceno in zagovor:

Predsednik:	Prof. dr. Peter DOVČ
	Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko
Član:	Doc. dr. Klemen POTOČNIK
	Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko
Član:	Prof. dr. Johann SÖELKNER
	Universität für Bodenkultur, Wien

Datum zagovora: 24.03.2015

Podpisana izjavljam, da je disertacija rezultat lastnega raziskovalnega dela. Izjavljam, da je elektronski izvod identičen tiskanemu. Na univerzo neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravici shranitve avtorskega dela v elektronski obliki in reproduciranja ter pravico omogočanja javnega dostopa do avtorskega dela na svetovnem spletu preko Digitalne knjižnice Biotehniške fakultete.

Marija ŠPEHAR

KEY WORDS DOCUMENTATION

DN Dd
DC UDC 636.2:575(043.3)=111
CX cattle/genomic selection/GWAS
CC AGRIS L10/5214
AU ŠPEHAR, Marija
AA GORJANC, Gregor (supervisor)/ČURIK, Ino (co-supervisor)
PP SI-1000 Ljubljana, Jamnikarjeva 101
PB University of Ljubljana, Biotechnical Faculty, Interdisciplinary Doctoral Programme in Biosciences, scientific field Animal Sciences
PY 2015
TI GENOMIC EVALUATION AND ASSOCIATION STUDIES OF CORRELATED TRAITS IN DAIRY AND DUAL PURPOSE CATTLE
DT Doctoral Dissertation
NO XX, 122 p., 14 tab., 24 fig., 11 ann., 202 ref.
LA en
AL en/sl
AB Selection of domestic animals is commonly based on phenotypic and pedigree information that are used to estimate breeding value (EBV). Development of techniques of molecular genetics and inclusion of thousands of single-nucleotide polymorphisms (SNPs) as an additional information into the system of genetic evaluation leads to so called genomic selection. The advantage of this approach in cattle selection is that breeding values can be evaluated at the birth, which can lead up to 100% greater genetic gain per year. In the first part of this thesis, the accuracy of progeny and genomic evaluations of Slovenian BSW bulls was evaluated utilizing all the available data. The accuracy of evaluations was assessed using univariate national evaluation, international DGV, and bivariate national evaluation incorporating international DGV as a correlated trait. Results indicate that integration of DGV from a large consortium into the national evaluation as a correlated trait enabled combination of all the available data and avoiding the need to publish several estimates of breeding values per animal. Bivariate approach is an important for bulls with a small number of daughters per tested bull. The issue of double counting was neglected due to small number of Slovenian BSW bulls used for the development of DGV equation in a consortium. In the second part, SNP marker panels were used in genome wide association studies (GWAS) to determine associations between SNPs and dairy traits (milk, fat, and protein yield) in Slovenian BSW bulls. The Bonferroni correction resulted in 52 significant SNPs using single SNP analysis. Correction for population stratification based on admixture components was shown to have great impact on the analysis and then no significant associations among SNPs and dairy traits have been detected. Results from GWAS will facilitate the understanding of the genetic architecture of complex traits and could be used to improve breeding programmes. In the last part, GWAS was performed in order to characterize genomic regions with associations having different magnitude and direction of effect on dairy (milk and fat yield) and beef traits (net daily gain and carcass quality) in dual purpose populations (German Fleckvieh and Italian Pezzata Rossa) providing the possibility to dissect genetic correlations between traits. Genome partitioning of genetic covariance between dairy and beef traits showed that estimated allele substitution effects from GWAS could be used as an input for the estimation of contribution partition of genetic covariances. The genetic architecture of the covariation contribution is mostly polygenic with the presence of a number of genes with large effects. The presented approach will provide additional knowledge to enhance understanding of genetic basis of covariation and will contribute to the identification of genes and pathways associated with dairy and beef traits in dual purpose populations.

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD Dd
DK UDK 636.2:575(043.3)=111
KG govedo/genomska selekcija/GWAS
KK AGRIS L10/5214
AV ŠPEHAR, Marija, mag., dipl.ing. agr.
SA GORJANC, Gregor (mentor)/ČURIK, Ino (somentor)
KZ SI-1000 Ljubljana, Jamnikarjeva 101
ZA Univerza v Ljubljani, Biotehniška fakulteta, Interdisciplinarni doktorski študij Bioznanosti, področje znanost o živalih
LI 2015
IN GENOMSKO VREDNOTENJE IN ASOCIACIJSKE ŠTUDIJE KORELIRANIH LASTNOSTI PRI GOVEDU MLEČNIH IN KOMBINIRANIH PASEM
TD Doktorska disertacija
OP XX, 122 str., 14 pregl., 24 sl., 11 pril., 202 vir.
IJ en
AL en/sl
AL Selekcija domačih živali običajno temelji na fenotipskih podatkih in poreklu, ki se uporabljajo za oceno plemenskih vrednosti (EBV). Razvoj tehnik molekularne genetike in vključitev več tisoč SNP označevalcev kot vir dodatne informacije v sistem ocenjevanja plemenskih vrednosti vodi do t.i. genomske selekcije. Prednost tega načina pri selekciji goveda je možnost ocene plemenskih vrednosti že ob rojstvu, kar lahko poveča letni genetski napredek tudi za 100%. V prvem delu disertacije je bila ocenjena točnost progenega in genomskega vrednotenja bikov rjave pasme v Sloveniji z uporabo vseh razpoložljivih podatkov. Točnost vrednotenja je bila narejena z enolastnostnim nacionalnim vrednotenjem, mednarodnimi direktnimi genomskimi vrednostmi (DGV) in dvolastnostnim nacionalnim vrednotenjem, ki je vključevalo mednarodno DGV kot korelirano lastnost. Rezultati kažejo, da je vključevanje DGV iz velikega konzorcija v nacionalno vrednotenje kot korelirano lastnost, omogočilo kombinacijo vseh razpoložljivih podatkov, kar izključuje potrebo po objavi več plemenskih vrednosti za žival. Dvolastnostni način je pomemben za bike, z majhnim številom hčera na testiranega bika. Prispevek dvojnega štetja podatkov je bil zanemarljiv zaradi majhnega števila rjavih bikov, ki so sodelovali pri razvoju DGV enačbe v konzorciju. V drugem delu so bili SNP označevalci vključeni v genomske asociacijske študije (GWAS) za določanje povezav med SNP-i in lastnostmi mlečnosti (količina mleka, maščobe in beljakovin) pri slovenskih rjavih bikih. Bonferroni korekcija je zaznala 52 značilnih SNP označevalcev, na osnovi posameznih analiz na vsakem označevalcu. Korekcija na razdeljenost populacije na podpopulacije na podlagi rezultatov admixture programa je imela pomemben vpiv na analizo v kateri ni bilo značilnih povezav med SNP označevalci in lastnostmi mlečnosti. Rezultati GWAS bodo olajšali razumevanje genetske strukture kompleksnih lastnosti in se uporabijo za izboljšanje rejskih programov. V zadnjem delu je bila GWAS narejena za karakterizacijo genomskih regij s povezavami med lastnostmi mlečnosti (količina mleka in maščobe) in mesnatosti (dnevni neto prirast in klavna kakovost) z različno jakostjo in smerjo pri kombiniranih pasmah (nemška in italijanska lisasta pasma), kar zagotavlja možnost razčlenitve genetskih korelacij med lastnostmi. Porazdelitev genetskih kovarianc na ravni genoma med lastnostmi mlečnosti in mesnatosti je pokazala, da lahko vpliv zamenjave alelov iz GWAS uporabimo za oceno porazdelitve prispevka genetskih kovarianc. Genetska arhitektura prispevka kovariacije je v glavnem poligena s prisotnostjo številnih genov z velikimi vplivi. Opisani način bo zagotovil nova znanja za boljše razumevanje genetskega ozadja kovaricij, kar bo pripomoglo k identifikaciji genov in poti, ki so povezane z lastnostmi mlečnosti in mesnatosti pri kombiniranih pasmah govedi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ANNEXES

Annex A11: Genomic location of SNP markers (red triangle for negative covariance and green triangle for positive covariance), QTL (blue line), and known genes (blue triangle) affecting dairy and beef traits by chromosome

ABBREVIATIONS AND SYMBOLS

QTL	Quantitative trait locus
EBV	Estimated breeding value
MAS	Marker assisted selection
SNP	Single nucleotide polymorphism
PA	Parent average
EBV	Estimated breeding value
MACE	Multiple across country evaluation
DGV	Direct genomic value
GEBV	Genomically enhanced breeding value
LD	Linkage disequilibrium
GWAS	Genome-wide association studies
BSW	Brown Swiss
GFV	German Fleckvieh
IPR	Italian Pezzata Rossa
MY	Milk yield
FY	Fat yield
PY	Protein yield
NG	Net daily gain
CQ	Carcass quality
CR	Call rate
MAF	Minor allele frequency
HWE	Hardy–Weinberg equilibrium

GLOSSARY

Alleles	Different forms of a gene, which produce variations in a genetically inherited trait
Assembly	Putting sequenced fragments of DNA into their correct chromosomal positions
Autosome	A chromosome not involved in sex determination. The diploid bovine genome consists of 60 chromosomes: 29 pairs of autosomes (BTA), and 1 pair of sex chromosomes (the X and Y chromosomes)
Base	One of the molecules that form DNA and RNA molecules
Base pair (bp)	Two nitrogenous bases (adenine and thymine or guanine and cytosine) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs
Bioinformatics	The science of managing and analysing biological data using advanced computing techniques. Especially is important in analysing genomic and transciptomics research data
Candidate gene	A gene located in a chromosome region suspected of being involved in a disease
Case – control study	Compares cases (that is, a selected group of individuals: for example, those diagnosed with a disorder) with controls (that is, a comparison group of individuals: for example, those who are not diagnosed with the disorder). Genome-wide association case–control studies test whether genetic marker allele frequencies differ between cases and controls
Cell	Basic unit of any living organism that carries on the biochemical processes of life
Chromosome	Complex DNA chain that contains genetic information. Chromosomes occur as paired sets throughout the genome
Centimorgan (cM)	Unit of measure of recombination frequency. One cM is equal to 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing-over in a single generation
Complex traits	Traits controlled by a combination of many genes and environmental factors
Crossing over	The breaking during meiosis of one maternal and one paternal chromosome, the exchange of corresponding

	sections of DNA, and the re-joining of the chromosomes. This process can result in an exchange of alleles between chromosomes
Deoxyribonucleic acid (DNA)	The molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T)
DNA marker	A DNA sequence that exists in two or more forms that can be used to genotype individual organisms
Estimated breeding value	An estimate of the additive genetic merit for a particular trait that an individual will pass on to its descendant's
Fine mapping	Extensively genotyping or sequencing a region of the genome that was identified in genome-wide association studies to identify the causal variant
Gene	The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule)
Gene locus (pl. loci)	Gene's position on a chromosome or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean expressed DNA regions
Gene mapping	Determination of the relative positions of gene s on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them
Genetic architecture	Genetic model (the number of single-nucleotide polymorphisms, effect sizes, allele frequency, etc.) underlying a phenotypic trait
Genomic breeding value	An estimate of an animal's genetic merit, including genomic information
Genetic improvement	Deliberate genetic change in a population of domestic animals or plants brought about by human control of their selection and breeding that makes them more suitable for the purpose for which they are kept
Genetic marker	Gene or DNA sequence at a known location on a chromosome that can be used to identify individuals or species. A genetic marker may be a short or long DNA sequence. Common types of genetic markers are RFLPs

	(Restriction Fragment Length Polymorphism), SSLPs (Simple Sequence Length Polymorphism), SNPs (Single Nucleotide Polymorphism), and SSRs (Simple Sequence Repeat DNA marker). Genetic markers play a role in genetic engineering because they allow breeders to locate and utilize genes of interest for genetic gains
Genetic polymorphism	Differences between DNA sequences
Gene pool	All the variations of genes in a species
Genome	All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs
Genomics	The discipline of genetics that encompasses gene mapping, gene sequencing, and determining gene function
Genome mapping	Determining a set of landmarks or genetic markers – genes or short DNA sequences – in the genome that will enable researchers to find new genes. A genome map is one - dimensional–a straight line with landmarks that stand for genes or DNA sequences. They guide a researcher toward a gene that is suspected to be involved in some process that is of interest
Genomic selection	Selection of animals for breeding based on estimated breeding values calculated from the joint effects of genetic markers covering the whole genome
Genome-wide association studies (GWAS)	Studies in which hundreds of thousands (or millions) of genetic markers are tested for association with a phenotypic trait
Genome-wide-significant	A term describing the statistical significance threshold that accounts for multiple testing in GWAS
Genotype	The genetic make-up of a cell or organism
Genotype imputation	Inference of missing genotypes or untyped single-nucleotide polymorphisms using statistical techniques
Haplotype	The particular combination of alleles on the same chromosome typically inherited together
Hardy-Weinberg equilibrium (HWE)	Holds at a locus when the two alleles within an individual are statistically independent
Heritability	The proportion of phenotypic variance attributed to genetic differences among individuals in a population
Heterozygote	The presence of different alleles at one or more loci on homologous chromosomes
Homozygote	An organism that has two identical alleles of a gene
Linkage analysis	Genotyping using a collection of genetic markers across

	the genome, and examining how those genetic markers segregate with the disease across multiple families
Kilobase (kb)	A unit of length for DNA fragments equal to 1000 nucleotides
Linkage disequilibrium (LD)	The absence of linkage equilibrium so that the allele at one locus is correlated with the allele at another locus
Locus (pl. loci)	The position on a chromosome of a gene or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean expressed DNA regions
Marker-assisted selection (MAS)	Use of genetic/DNA markers to guide in choosing specific plants or lines with targeted traits for new variety development
Megabase (Mb)	Unit of length for DNA fragments equal to 1 million nucleotides and roughly equal to 1 cM (centimorgan)
Minor allele frequency	The frequency of the less frequent allele in a two-allele polymorphism
Multivariate analyses	The simultaneous inclusion of two or more phenotypes in one analysis when testing the association with a genetic variant
Nucleotide	Subunit of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA). Thousands of nucleotides are linked to form a DNA or RNA molecule
Phenotype	The observable or measurable characteristics of an individual
Pleiotropy	A gene or genetic variant that affects more than one phenotypic trait
Polygenic	Controlled by many genes
Population genetics	The study of variation in genes among a group of individuals
Population stratification	Also termed population heterogeneity or confounding by ethnicity. Population stratification refers to a situation where subgroups of individuals are on average more related to each other than to other members of the wider population.
Qualitative trait	Traits that are easily classified into distinct phenotypic categories. These phenotypes are under the genetic control of only one or a few genes with little or no environmental

	modification to obscure the genetic effects
Quantitative trait	Refers to phenotypes or characteristics of an organism that vary in degree and can be attributed to multi-gene effects (as opposed to single or a few genes) and their environment. These traits do not fit into discrete phenotypic classes, but instead form a continuum of phenotypes that blend to form a continuous variability
Quantitative trait	A measurable trait that depends on the cumulative action of many genes and the environment, and that can vary among individuals over a given range to produce a continuous distribution of phenotypes
Quantitative trait locus (QTL)	Describes a chromosomal region containing one or more genes involved in the expression of a polygenic trait. QTL are identified by the association between a set of genetic markers
Segregation	The normal biological process whereby the two pieces of a chromosome pair are separated during meiosis and randomly distributed to the germ cells
SNPs	Single base pair positions in DNA. This is the most common form of DNA variation, alterations to a single base. Most SNPs are associated with DNA variation and are used as genetic markers
Univariate analysis	Test of association between one phenotype and a genetic variant

1 GENERAL INTRODUCTION

Genome research in farm animals differs from those in humans. The identification of single disease loci is less important since animals with inherited disorders are pulled from breeding. Most of economically important traits in livestock have a quantitative expression. These traits are influenced by many genes of predominantly small effect and environment which results in continuous variation of phenotypes (Falconer and Mackay, 1996). A chromosomal region that contains one or more genes involved in the expression of a quantitative trait is called quantitative trait locus (QTL) (Fischer et al., 2003). The presence of QTL is detected by mapping studies that shows differences in phenotypic expression between individuals having different QTL alleles (Anderson, 2001). The purpose of QTL mapping in farm animals is to apply the genomic information in practical breeding programmes in order to enhance selection programs (Anderson, 2001; Dekkers and Hospital, 2002) and to understand the genetic architecture of the economicaly important traits.

1.1 IMPROVEMENT OF GENETIC EVALUATIONS USING MOLECULAR INFORMATION

Genetic improvement of quantitative traits is commonly based on phenotypic and pedigree information that are used to estimate additive genetic effect of all the genes of an individual or the so called estimated breeding value (EBV). Statistically the estimation is carried out using the mixed model methodology formulated by Henderson (1984). In dairy cattle, phenotypic data are collected through various recording schemes (milk and fertility recording, type classification, etc.) on daughters of progeny tested bulls. The latter are four to six years old when the first crop of data on their daughters is collected. At that time the accuracy of bull EBV is 0.90 or more. Without progeny testing which is a source of phenotypic information that pertains to bulls, the accuracy of EBV using only pedigree information (i.e., average parental breeding values) is around 0.60 (e.g., Schefers and Weigel, 2012).

1.1.1 Mapping genomic regions underlying quantitative traits

Considerable technological developments and discoveries in the field of molecular genetics during 1990's allowed the identification of genomic regions with QTL for economically important traits. These findings were the result of linkage map developments and genotyping based on molecular markers (Kappes et al., 1997). Improvements in animal breeding programmes through the use of molecular markers linked to QTL to predict the performance of animals is called marker assisted selection (MAS). The inclusion of

molecular information in traditional selection schemes through MAS became a valuable tool in selecting traits of interest (Meuwissen and Goddard, 1996; Spelman and Garrick, 1997). According to traditional MAS, relatively small number of genetic markers was included in genetic evaluation (Fernando and Grossman, 1989). At that time considerable effort has been put in to the construction of genetic maps containing 746 (Barendse et al., 1997) and 1250 microsatellite markers (Kappes et al., 1997) distributed across the whole genome. Genetic markers such as restriction fragment length polymorphisms (RFLP), variable number tandem repeats (VNTR), microsatellites and single nucleotide polymorphisms (SNPs) were commonly used in these studies to track the inheritance of specific regions of chromosomes. Even though MAS has improved the efficiency of breeding programs (Dekkers, 2004), its implementation has been very limited (Boichard et al., 2010). Only a limited number of genes have been identified. The confidence interval of the marker location was large and was impossible to test for two linked QTL emphasizing the need for rendering the fine-mapping (Visscher and Goddard, 2004; Boichard et al., 2003). Therefore an alternative approach has to be applied to utilize the information from a large number of molecular markers.

1.1.2 High-throughput genotyping

Sequencing the genome (bovine or human or rice) simply means the determination of the linear order of four bases in the genome: adenosine (A), thymidine (T), guanine (G) and cytosine (C). The cattle genome sequencing and assembly was completed in 2009 (The Bovine Genome Sequencing and Analysis Consortium, 2009). The DNA of the whole genome shotgun sequences belong to the Hereford cow Dominette and her sire Domino. The sequencing was assembled by the Baylor College of Medicine Human Genome Sequencing Center (HGSC, 2104). The genome assembly went through several improvements, from assemblies Btau_1.0, Btau_3.1 to Btau_4.2 in 2009. The sequences were also assembled by the Center for Bioinformatics and Computational Biology at University of Maryland and generated an alternate assembly UMD2 (Zimin et al., 2009). Since then, both the Btau and UMD assemblies have been upgraded to the current versions, Btau_4.6.1 and UMD3.1. The total length of the cattle genome is approximately 2.8 bilion base pairs (bp) DNA (Zimin et. al., 2009).

For the purposes of genetic studies, single nucleotide polymorphisms (SNPs, pronounced 'snips') are typically used as markers of a genomic region. Each SNP represents a difference in a nucleotide. For example, a SNP may replace the nucleotide cytosine (C) with the nucleotide thymine (T) in a certain stretch of DNA. SNPs are located at approximately 700 base pairs in the *Bos taurus* and every 300 base pairs in Bos indicus genome (The Bovine HapMap Consortium, 2009). There are nearly four million SNPs in

the *Bos taurus* genome which were identified during sequencing (Seidel, 2010). The genome sequence assembly from the Hereford cow and sequences sampled from six additional breeds were used to interrogate 37,470 SNPs in 497 animals from 19 geographically and biologically diverse breeds (The Bovine HapMap Consortium, 2009). This set was the basis for further expansion of SNP discoveries. Van Tassel et al. (2008) added an additional 23,357 SNPs to the bovine collection, studying 66 animals from dairy breed (Holstein), and six beef breeds (Angus, Red Angus, Gelbvieh, Hereford, Limousin, and Simmental). The availability of thousands of SNPs has led to the development of SNP chips capable for assaying large number of SNPs simultaneously.

Two most common SNP chip for cattle are coming from Illumina (Illumina Inc., San Diego, CA, USA) and Affymetrix (Santa Clara, CA). These two competing technologies have been recently reviewed (Distefano and Taverna, 2011) and offer different approaches to measure SNP variation. For example, the Affymetrix platform prints short DNA sequences as a spot on the chip that recognizes a specific SNP allele. Alleles (i.e. nucleotides) are detected by differential hybridization of the sample DNA. Illumina on the other hand uses a bead-based technology with slightly longer DNA sequences to detect alleles. The Illumina chips are more expensive to make but provide better specificity. The most used SNP chip contains approximately 54,000 SNPs and is called the Bovine50K SNP chip (Illumina. Inc., 2011). The densest is so called BovineHD beadchip (Illumina. Inc., 2010) with more than 700,000 SNPs that covers the entire bovine genome. The high-density chip enables to capture genetic effects at a better resolution and might result in substantially higher accuracies of genomic breeding values (Pausch, 2013).

The availability of affordable high-density panels of thousands of SNPs has led to the abundant use of this information in livestock breeding programs. This is commonly referred to as genome-wide or genomic selection (Meuwissen et al., 2001). Genomic selection is based on the inference of breeding values based on the sum of SNP association effects with phenotype across the whole genome (Meuwissen et al., 2001; Solberg et al., 2008). Statistically the SNP association effect is a regression coefficient of phenotype on SNP genotype, while genetically the SNP association effect is an average allele substitution effect of a particular SNP. Both representations describe the additive genetic effect of a SNP association between phenotypic values and genomic region of a SNP. The sum of all SNP associations represents an estimate of total additive genetic (breeding) value of an individual. In dairy cattle this estimate is often called direct genomic value (DGV) and is based solely on genotypic information of an individual animal and population based SNP associations.

For the implementation of genomic selection, phenotyped and genotyped reference population is needed to estimate the SNP associations and assemble them in the prediction equation of a DGV which is in turn used to estimate DGV for non-phenotyped individuals (Meuwissen et al., 2001; Goddard and Hayes, 2007). Different types of breeding value estimates (EBV from pedigree prediction or progeny test and DGV from genomic information) can be blended in one value often called genomically enhanced breeding value (GEBV) using various approaches (VanRaden 2008; Kachman, 2008; Aguilar et al., 2010). The accuracy of genomic evaluation mainly depends on the number of genotyped and phenotyped individuals in the reference population (Goddard, 2009; Schefers and Weigel, 2012). Use of genomic data generally provides less information than phenotypic data from a progeny test and consequently lower accuracy of EBV. However, the advantage of using genomic data is to increase the accuracy of EBV of young or non-phenotyped animals from about 0.60 (parent average) to 0.80 (Hayes et al., 2009a). Even though these accuracies are lower than with progeny test the early use of young bulls in artificial insemination shortens generation interval considerably and therefore increases genetic gain per unit of time (Schaeffer, 2006).

The highest gains of using genomic selection could be expected in populations with large a number of genotyped and phenotyped animals, especially progeny tested bulls. An example of such large reference population is the European consortium in Holstein breed (EuroGenomics) which in the year 2011 included more than 17,000 progeny tested bulls from France, Germany, Netherlands and Scandinavian countries (Denmark, Sweden, and Finland) (Liu et al., 2011). Since then the number of progeny tested bulls increased by inclusion of Spain and Poland and new routinely tested bulls. The links between these populations are created by extensive world-wide use of semen of several popular bull families in Holstein breed. Additionally, all the data can be easily interchanged, due to the international across country evaluation performed by the Interbull in Sweden. Similar development of a large international reference population had been done in the Brown Swiss breed, where the reference population is consisted of almost 8,000 bulls from 7 countries (Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States of America) via the Intergenomics project operated by the Interbull (Zumbach et al., 2010; Jorjani et al., 2012). The need for large reference populations has therefore steered a lot of international cooperation, which is essential for introduction of genomic selection in small populations.

Reference populations in beef cattle are usually much smaller than in dairy cattle due to more disconnected (sub)populations in one or several countries. In these populations SNP prediction equations could be developed on experimental populations of reasonable size and later used to compute the DGV of other animals in whole national population. In such cases DGVs can be blended with the national EBV from pedigree and/or phenotype

inference via the use of bivariate analysis (MacNeil et al., 2010; Johnston et al., 2010; Johnston et al., 2012). Mäntysaari and Strandén (2010) proposed to use the same approach to blend genomic and conventional EBV. However, this approach can lead to a potential double counting of information, which could be avoided by using method of Vandenplas and Gengler (2012).

1.2 QUANTITATIVE TRAIT LOCI MAPPING METHODOLOGIES

Detection of QTL relies on linkage disequilibrium (LD) between the genetic markers and QTL. The existence of LD implies that small segments of chromosomes exists in the current population and are descended from the same common ancestor. Therefore, chromosome segments which are identical by descent will not carry only the identical marker haplotypes but also will carry identical QTL alleles. In the case when two animals carry chromosomes which are identical by descent then their phenotypes will be correlated (Hayes et al., 2005).

Two main approaches are used in dissecting genetic loci or genes: linkage and association analysis. A key difference between linkage and association mapping is in the precision with which they map the location of quantitative trait loci (QTLs).

1.2.1 Linkage analysis

This statistical method tests for the co-segregation of linked alleles (i.e. alleles that are inherited together) within family members. It is most effective in identifying rare alleles with large effect sizes (Concannon et al., 2009). It is based on Mendelian inheritance and uses affected relatives to identify regions on chromosomes that are shared more frequently than expected by chance. The main idea is that markers which are found in vicinity on the QTL have a tendency to be inherited together when passed on to offspring. Thus, if some disease is often passed to offspring along with specific markers, then it can be concluded that the gene which is responsible for the disease are located close on the chromosome to these markers.

Linkage studies are performed when the pedigree of related individuals is known and phenotype is present in some of the family members. Since linkage analysis uses recombination events only within the recorded pedigree the confidence interval for the position of the QTL is typically large (Darvasi et al., 1993). Two most common mapping designs proposed by Weller et al. (1990), the daughter design (DD) and granddaughter design (GDD) were used for detection of QTL in dairy cattle. QTL mapping studies resulted in identification only a few genes. For example in Holstein population, a QTL for

milk yield and fat content was detected on chromosome 14 (Coppieters et al., 1998; Looft et al., 2001; Thomsen et al., 2000). Further studies revealed the underlying gene DGAT1 in Dutch and New Zealand Holstein populations (Grisart et al., 2002) as well as in German Fleckvieh population (Winter et al., 2002). A *Myostatin* gene responsible for double muscling was mapped in beef cattle (Charlier et al., 1995). There are several disadvantages of linkage studies. A large number of families with several affected generations have to be identified. Linkage analysis usually mapped the QTLs to a large interval of 20 centimorgans (cM) or more (Meuwissen et al., 2001) which made it difficult to identify the underlying mutation and to use the marker information in animal breeding programmes. This kind of studies is less helpful for complex traits, such as all economically important traits (Pearson and Manolio, 2008).

1.2.2 Association analysis

It is based on association between marker alleles and trait of interest (or disease). Association is statistical term about the co-occurrence of alleles and phenotypes. The method is suitable for common alleles with modest and small effect sizes (Concannon et al., 2009). Allelic association is occured when there is a significant change in the frequency of a marker allele for a disease or trait (phenotype) compared to the control population. This method tests for association of genetic markers (typically SNPs) with the trait of interest in population-based (case/control or quantitative trait model) or family-based sample in the form of parent-offspring trios (transmission disequilibrium test). Case-control study compares two groups of individuals: cases (individuals with disease) and controls (individuals without disease). The assumption is higher prevalence of susceptibility alleles for disease or trait of interest in cases than in control group. Susceptibility alleles could be detected through direct comparison of allele frequency between cases and controls (McCarthy et al., 2008). Family-based design approach examines the transmission of alleles from heterozygous parents to affected offspring that is observed more frequently than expected by chance (Smith and Newton-Chech, 2009).

1.2.3 Genome-wide association studies (GWAS)

Development of SNP marker panels opened the possibility to perform genome-wide association studies (GWAS) to hunt for associations between phenotypic values and genomic regions in order to prioritize regions for further gene or regulatory element discovery (Hirschhorn and Daly, 2005). From 2005, GWAS took the complete dominance in association studies. In GWAS, association analysis is used to detect association of common variants (allelic variants present in more than 5% of the population) with the modest to large effects (Manolio et al., 2009) and to identify chromosome regions that

harbour the genes or regulatory elements related to the traits of interest (e.g., Goddard and Hayes, 2009; Bush and Moore, 2012; Eggen, 2012; Montaldo et al., 2012).

These studies were first used for the analysis of human disease data in order to reveal the genetic architecture of diseases (e.g., Hirschhorn and Daly, 2005; Massey and Parkes, 2007). Results were encouraging and lead to the identification of genomic regions associated with Crohn's disease (Barrett et al. 2008), diabetes (Frayling, 2007), cancer (Easton and Eeles, 2008), etc. Association studies in the field of livestock species are used to understand the genetic control of economically important traits such as milk (e.g., Grisart et al., 2002; Pryce et al., 2010; Mai et al., 2010), meat (e.g., Bolormaa et al., 2011), and health (e.g., Murdoch et al., 2011). Such knowledge can be used to enhance biological understanding as well as to enhance the methods of genomic selection. The use of genome-wide data enhances such hunting simply due to greater resolution of genome. For example in dairy cattle, significant associations using genome-wide data were detected for milk yield on BTA8, BTA9, BTA10, BTA11, BTA 13, BTA25, and BTA29 (Hayes et al., 2009b; Bolormaa et al., 2010). Milk fat and protein content was associated with SNPs on BTA5, BTA6, BTA11, BTA14, BTA19, and BTA26 (Pryce et al., 2010; Schopen et al., 2011).

Beside catlle, GWA studies were also performed in other domestic animals including pigs, sheep, chickens, and dogs. GWAS in filed of pig was performed using commercial Illumina PorcineSNP60 BeadChip and revealed significant associations for androstenone (Duijvesteijn et al., 2010) on chromosomes 1 and 6 and for skatole (Ramos et al., 2011) on chromosome 6. Grindflek et al. (2011) found that 28 chromosome was related to boar taint in commercial Landrace and Duroc. Illumina OvineSNP50 BeadChip and GWAS methodology was used to identify significant SNPs associated with growth and meat production traits (birth weight, weaning weight, 6-month weight, eye muscle area, fat thickness, preweaning gain, post-weaning gain, daily weight gain, height at withers, chest girth, and shin circumference) and to explore and forecast the major candidate genes in sheep (Zhang et al., 2013). In the study, 36 significant SNPs were identified for seven traits, and 10 of them reached genome-wise significance level for post-weaning gain. In Spanish Churra sheep, Garcia-Gomez et al. (2012) performed GWAS for milk production traits and reported significant association for fat and protein content located on chromosome 3 (OAR3). GWAS in chichen was based on use Illumina ChickenSNP60 BeadChip. Studies were focused on chicken body weight. Gu et al. (2011) reported significant SNPs in a region of the chicken chromosome 4. GWAS about chicken growth was reported by Xie et al (2012) where significant SNPs and genes were detected for 18 traits mostly on chicken chromosome 1. In layer chickens the egg production and quality traits are important traits. GWAS on chicken egg production and quality traits using two populations including White Leghorn and Brown-Egg Dwarf Layers revealed significant

SNPs associated with egg production and quality traits (Liu et al., 2011). In dogs, GWAS was reported for canine atopic dermatitis (cAD; Wood et al., 2009) using the Illumina Canine SNP20 array showing that GWAS could be useful tool for identify genetic risk factors for cAD. GWAS for Arrhythmogenic right ventricular cardiomyopathy (ARVC) was carried out by Meurs et al. (2010) using the canine 50k SNP array in adult Boxer dogs and identified several regions significantly associated with ARVC expecially on chromosome 17. Genetic variants associated with intervertebral disc calcification in Dachshunds were identified by Mogensen et al. (2011). They reported signifficant association in a region of chromosome 12.

1.2.3.1 Linkage disequilibrium

GWAS can be successfully implemented due to the existing linkage disequilibrium (LD) between dense markers and QTL for the trait of interest (Hayes et al., 2009a; de Roos et al., 2008). LD is a property of SNPs on a contiguous stretch of genomic sequence that describes the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP within a population (Bush et al., 2012). The term LD was formed by population geneticists in order to mathematically describe changes in genetic variation within a population over time. It is related to the concept of chromosomal linkage, where two markers on a chromosome remain physically joined on a chromosome through generations of a family. Recombination events within a family from generation to generation breaks chromosomal segments. This effect is magnified through generations, and in a population of fixed size undergoing random mating, repeated random recombination will break part segments of contiguous chromosome until all alleles in the population are independent. The rate of LD decay is dependent on the population size, the number of founding chromosomes in the population, and the number of generations for which the population has existed (Bush et al., 2012).

Many measures of LD have been proposed (Devlin and Risch, 1995) and all are related to the difference between the observed frequency of co-occurrence for two alleles (i.e. a two-marker haplotype) and the frequency expected if the two markers are independent. The two most commonly used measures of linkage disequilibrium are D' and r^2. D' is a population genetics measure that is related to recombination events between markers and is scaled between 0 and 1. A D' value of 0 indicates complete linkage equilibrium, which implies frequent recombination between the two markers and statistical independence under principles of Hardy-Weinberg equilibrium. A D' of 1 indicates complete LD, indicating no recombination between the two markers within the population. For the purposes of genetic analysis, LD is generally reported in terms of r^2 which is statistical measure of correlation. High r^2 values indicate that two SNPs carry similar information. Therefore, only one of the two SNPs needs to be genotyped to capture the allelic variation. There are dependencies

between these two statistics: r^2 is sensitive to the allele frequencies of the tow markers, and can only be high in regions of high D'.

1.2.3.2 Principles of GWAS

The basic design of a GWAS in livestock species is that a sample of animals are recorded for a trait of interest and assayed for a genome-wide panel of markers to detect statistical associations between the trait and any of the markers (Goddard and Hayes, 2009). Data from a GWAS are analysed by testing the significance of one SNP at a time using a linear model Single-SNP and multiple-SNP (multipoint-SNP) analyses are common approaches of SNP-based analyses.

Single-SNP analysis is a series of single-locus statistic tests, examining each SNP independently for association to the phenotype. Quantitative traits are generally analyzed using generalized linear model (GLM) approach that includes fixed effects (sex, age, cohort or group to which the animal belongs), the polygenic breeding value of each animal (due to all other genes affecting the trait), and the effect of a SNP. Only the additive effect of the SNP is tested by coding the SNP genotype as 0 for one homozygote, 1 for the heterozygote, and 2 for the other homozygote. One of the most important covariates considered in genetic analysis is a population structure (or admixure) which mostly consists of a mixure of breeds in the sample of animals used in analysis. The correction for population structure can be done by including breed in the statistical model used for analysis. However, cattle is bred in half-sibling families. Relationships among the animals in the sample cause LD between loci even if they are unlinked. For instance, if the sire of a family carries rare alleles at two unlinked loci, his offspring will be more likely than other animals to carry both rare alleles. This problem can be solved by including a polygenic term in the statistical model with the appropriate numerator relationship matrix. It is also possible to estimate the relationship matrix from the markers instead of from the pedigree, then to fit this matrix directly or its principal components (Patterson et al., 2006). The assumption here is that the principal components will capture population stratification as these are likely to be the main axes of variation in the genomic relationship matrix. Pausch et al. (2011) applied this approach to a GWAS in Fleckvieh cattle for calving ease and growth traits. Although the inclusion of a polygenic term in the statistical model eliminates associations between a QTL and markers that are not linked to it, it does not eliminate associations between QTL and linked markers. If there are relationships among the animals used for the GWAS, the associations discovered represent a mixture of associations caused by LD and associations caused by linkage.

When testing for a statistically significant difference between means, each SNP is tested separately and a *p-value* is generated for each SNP. If the *p-value* is set to 0.05, there is a 5% error margin for each single SNP to pass the test. There are several strategies to adjust the *p-values*: Bonferroni adjustment, false discovery rate, and permutation testing. Bonferroni correction is the simplest approach to correct for multiple testing. This approach adjusts the α value from α = 0.05 to α =(0.05/k) where k is the number of statistical tests conducted. For a typical GWAS using 500,000 SNPs, statistical significance of a SNP association would be set at 10^{-7}. This correction is the most conservative, and assumes that each association test of the 500,000 is independent of all other tests. An alternative approach is to compute the false discovery rate (FDR) (Benjamini and Hochberg, 1995). The false discovery rate is an estimate of the proportion of significant results or SNPs (α= 0.05) that are declared significant but acctually they are false positives. If the FDR is 0.05, it means that only 5% of the significant SNPs are false discoveries. Permutation testing is another approach for establishing significance in GWAS. The idea of permutation tests was first described by Fisher (1935), with the assumption that all possible permutations would be enumerated. A key concept of this approach is the idea that the distribution of an estimator or test statistic can be estimated by an empirical re-sampling distribution. In the multiple testing much more precision is needed. A useful strategy is to start with 1,000 permutations and continue to larger numbers only if p is small enough to be interesting.

An alternative to the single SNP analysis is to fit all SNPs simultaneously and treating them as random effects. The main obstacle for using multiple regression appears because the number of predictors is greater than the number of observations (Tibshirani, 1996; Logsdon et al., 2012). Since p >> n standard maximum likelihood estimation cannot be performed. Regularisation based on penalized maximum likelihood method was considered in order to avoid model overfitting. Therefore, variable selection implemented by penalized linear regression could identify predictors (or SNPs) with nonzero effects. Variable selection approaches with continuous penalties include ridge regression (Hoerl and Kennard, 1970), lasso (Tibshirani, 1996), and elastic net (Zou and Hastie, 2005). Ridge regression is known to shrink the coefficients of correlated predictors towards each other, allowing them to borrow strength from each other. Ridge regression shrinks coefficients with smaller variances more than it does coefficients with larger variances. Under the assumption that less variable predictors are less associated with the outcome, then ridge shrinkage is more pronounced for unimportant than for important predictors. The lasso tends to pick only one or a few of highly correlated variables included in the model and shrinks the rest to 0. Elastic net is a linear combination of the L1 penalty and L2 penalty. It was designed as a form of the lasso that would be more stable when used on data sets characterized by highly correlated variables (Zou and Hastie, 2005).

1.2.3.3 Limitations of GWAS

One of the most pronounced limitations of GWAS is that selected SNPs capture only a fraction of total genomic variance, so called 'missing heritability' phenomenon (Maher, 2008; Manolio et al., 2009). For example, human height is highly heritable trait (80 to 90%). Using a single SNP analysis over 30,000 people more than 40 loci were found to be associated with the height. Only 5% of the height's heritability is explained by the genetic variaton at those 40 loci. Analyzing all SNPs jointly for their influence on height, 45% of the height's heritability is explained by SNPs. It is not clear why the common variants fail to explain the majority of the genetic heritability of most human diseases. One possible explanation to the missing heritability could be some interaction between different gene (epistasis). These interactions could be hard to detect when analyzing one SNP at the time, as the marginal effect of a single SNP will be small. Another explanation is that part of the increased risk can be explained by many rare variants, which are present among less than 1 % of the population. This suggests that there could be heterogeneity, where different genetic profiles can cause diseases that are diagnostically the same (Manolio et al., 2009). Furthermore, GWA studies are focused on common variation only, and many associated variants are not causal (Manolio et al., 2009). Results need replication in independent samples in different population. In association studies potential causes of inconsistent results are due to population stratification, genetic heterogeneity, random error such as false positive or false negative results, and study design problems (Silverman and Palmer, 2000; Cardon and Bell, 2001).

Despite the limitations outlined above, GWAS has proven itself well-suited to the identification of common SNP-based variants with modest to large effects on phenotype. The studies give possibility to define the molecular mechanisms of the disease or phenotypic expression through functional analysis of a gene and its associated variants.

1.3 ADDITIONAL KNOWLEDGE FOR BREEDING PROGRAMS

One of the practical problems in breeding programs is selection on multiple traits with variable genetic correlations between traits. In particular the limitation is posed by (strongly) negative genetic correlations limiting the space for improvement in the future. Such unfavourable correlations can be due to closely linked genes or pleiotropy. Combination of linked genes with unfavourable (opposite) effects can arise due to selection for a certain trait proliferating also the remaining genes in the same genomic regions. In some cases this can also lead to unfavourable genetic changes in other traits, due to unfavourable genetic correlations. Another explanation for this phenomenon can be due to pleiotropy. Pleiotropic genes are genes influencing more than one trait. Increasing the frequency of such genes via selection will result in the same unfavourable genetic

changes in other traits. Classical example of antagonistic traits in dairy cattle is milk yield and fertility. For example in Norwegian Red cattle Olsen et al. (2011) reported QTL on BTA12 having opposite effect on non-return rate and milk traits. Another example with antagonistic effect has been reported for calving ease and postnatal growth traits where two QTLs were identified on BTA14 and BTA21 in the German Simmental breed (Pausch et al., 2011). Due to polygenic architecture for economically important traits there can be genomic regions with favourable, unfavourable, and no genetic correlations that all jointly result in a global genetic correlation between traits. Genome-wide data provides an opportunity to dissect the genome in such regions to facilitate further gene discovery as well as to enhance breeding. Most studies were focused on milk production and fertility traits in dairy cattle. However, these aspects have not yet been studied in dual purpose breeds, where the antagonism is not only present between milk production and fertility, but also between milk and beef production.

1.4 HYPOTHESES

On the basis of the issues described above, the present thesis will investigate the following scientific problems.

1.4.1 Implementation of genomic evaluation in small populations

Implementation of genomic evaluation in small populations is challenging due to the requirement of large numbers of genotyped and phenotyped animals to build reference population of sufficient size. Alternative to own reference population is to use reference population combined from populations from several countries and to integrate this information into national evaluations. Hypotheses for this part are:

- Integration of direct genomic values based on a combined reference population from several countries provides significantly more accurate estimates of breeding values for routine national genomic evaluation than parent average prediction.
- Appropriate statistical methods can correct for the double counting of national data in the process of integration international information (MACE EBV, DGV, and GEBV) based on a combined reference population into national genomic evaluation.

1.4.2 Genome-wide association studies for dairy traits in Brown-Swiss breed

- Address the aspects of the GWAS for implementation of genomic selection for dairy traits: milk yield (MY), fat yield (FY), and protein yield (PY) in Slovenian Brown Swiss breed based on single SNP analysis and to SNP association signals across the whole genome.

1.4.3 Decomposition of genetic correlations due to linkage or pleiotropy using genomic data

- Genome-wide association can be used to characterize genomic regions with associations having different magnitude and direction of effect on different traits providing the possibility to dissect genetic correlations between traits.
- Genetic correlations between dairy and beef traits can be mapped combining single trait genome-wide association studies.
- Multivariate modelling can be used to map genetic correlations between dairy and beef traits at the whole genome, on chromosomal, and (sub)chromosomal levels, and to identify the candidate genes affecting these traits.

2 SCIENTIFIC WORKS

2.1 ACCURACY OF GENOMIC PREDICTION FOR MILK PRODUCTION TRAITS WITH DIFFERENT APPROACHES IN A SMALL POPULATION OF SLOVENIAN BROWN BULLS

2.1.1 Introduction

Genetic improvement of quantitative traits in dairy cattle is commonly based on phenotypic and pedigree information that are used to infer (estimate) breeding values (EBV). The process of providing reliable breeding values based on polygenic model is time consuming. Phenotypic data in dairy cattle are collected through various recording schemes (milk and fertility recording, type classification, etc.) on daughters of progeny tested bulls. The later are four to six years old when the accuracy of their EBV is 0.90 or more. Before progeny testing, the accuracy of EBV (parent average) is around 0.60 (e.g., Schefers and Weigel, 2012).

In recent years, the availability of affordable high-density panels of single-nucleotide polymorphisms (SNP) has led to abundant use of this information in selection decisions commonly called genome-wide or genomic selection (Meuwissen et al., 2001). Genomic selection is based on the inference of breeding value based on the sum of SNP or haplotypes effects across whole genome (Meuwissen et al., 2001; Solberg et al., 2008). Due to the direct use of marker data obtained marker based breeding value is often called direct genomic value (DGV). For the implementation of genomic selection, phenotyped and genotyped reference population is used to derive the prediction equation of DGV. This equation is then used to estimate DGV for non-phenotyped individuals (Meuwissen et al., 2001; Goddard and Hayes, 2007). Both types of breeding values (EBV and DGV) can be blended in one value - genomically enhanced breeding value (GEBV) using various approaches (VanRaden 2008; Kachman, 2008; Aguilar et al., 2010). The advantage of using genome-wide data is to increase the accuracy of EBV from about 0.60 to 0.80 for young or non-phenotyped animals (Hayes et al., 2009). VanRaden et al. (2009) reported the accuracies of GBV equal to 0.70 averaged across traits in Holstein-Friesian dairy cattle in USA. If the accuracy of GEBV is high enough, early use of young bulls in artificial insemination will shorten generation interval and increase genetic gain per unit of time (Schaeffer, 2006).

The highest increase in accuracy can be expected in populations with a large number of genotyped and phenotyped animals (Goddard, 2009). When populations are small prediction equations for EBV using SNP data can be developed on the experimental populations of reasonable size and later blended with the national evaluation system as

proposed in beef cattle populations (MacNeil et al., 2010; Johnston et al., 2012). An alternative is to develop prediction equations on a larger combined international data as in Brown breed, where an international reference population had been setup with almost 8,000 bulls from Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States of America via the InterGenomics consortium (Jorjani et al., 2012). Within the scope of the InterGenomics consortium specific prediction equations for each country are developed based on all available SNP data and multiple across country evaluation (MACE) EBV on a country's specific scale (Jorjani et al., 2012). Each bull has therefore several evaluations at the national level (parent average for young bulls and progeny test EBV for proven bulls) and at the international level (SNP genotype based EBV called direct genomic value (DGV) and a blended genomically enhanced breeding value of DGV and parent average EBV for young bulls and progeny test MACE EBV for proven bulls). All these evaluations complicate the publication of results and inhibit the use of all the available data.

The objective of this study was to evaluate the accuracy of progeny and genomic evaluations in a small population of Slovenian Brown bulls utilizing all the available data. Specifically, the accuracy of evaluations was assessed using: univariate national evaluation based on phenotype and pedigree data, international DGV, and bivariate national evaluation incorporating international DGV as a correlated trait.

2.1.2 Material and methods

2.1.2.1 Phenotype and genotype data

Phenotypic data in the analysis consisted of 1,342,134 test-day records for milk, fat, and protein yields from 56,670 cows recorded between years 1997 and 2011 and used in the routine national genetic evaluation. The total number of animals in the pedigree was 79,573. The cows were progeny of 736 bulls. Among these, 184 bulls born between 1990 and 2007 were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1) comprising of 54,001 SNP markers. These bulls were of predominantly Slovenian origin with nine bulls originating from other countries (Austria and Germany), but were imported to Slovenia as live animals and used only in Slovenia. In addition, international direct genomic value (DGV) was available from the InterGenomics consortium for the analysed traits (Jorjani et al., 2012). This information was considered for 399 bulls that had daughters in the national data, namely for 184 Slovenian and 215 foreign bulls.

Preliminary edits of national SNP data were carried as follows. SNPs were considered if a call rate was higher than 90% by SNP and 80% by bull. All together, 6,889 SNPs did not

meet these criteria and were therefore excluded. SNPs with minor allele frequencies lower than 0.05 were also excluded (13,340 cases). The departure from the Hardy-Weinberg equilibrium at a threshold of $P < 0.0001$ was the next edit with 7,490 SNPs failing this criterion. Finally, SNPs that could not be mapped or that were on the X chromosome were excluded, leaving a final set of 34,450 SNP for 184 bulls. Missing genotypes were imputed using the gpig program (Strandén, 2010), which implements the imputation method based on linear best unbiased prediction (Gengler et al., 2007).

2.1.2.2 Evaluation methods

Based on the available data the following approaches were used for the evaluation of EBVs. The first approach was univariate repeatability test-day model (U) based on the national phenotypic and pedigree data:

$$y = \boldsymbol{Xb} + \boldsymbol{Z}_a\boldsymbol{a} + \boldsymbol{Z}_c\boldsymbol{c} + \boldsymbol{Z}_p\boldsymbol{p} + \boldsymbol{e} \quad \text{... (1)}$$

where **y** is a vector of phenotypic observations, $\boldsymbol{b}$ is a vector of parameters for fixed effects, $\boldsymbol{a} \sim N(\boldsymbol{0}, \boldsymbol{A}\sigma_a^2)$ is a vector of parameters for breeding values with pedigree relationship matrix $\boldsymbol{A}$, $\boldsymbol{c} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_c^2)$ is a vector of parameters for herd effect, $\boldsymbol{p} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_p^2)$ is a vector of parameters for permanent environment effect, and $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2)$ is a vector of residuals, while $\boldsymbol{X}$, $\boldsymbol{Z}_a$, $\boldsymbol{Z}_c$, and $\boldsymbol{Z}_p$ are incidence matrices linking $\boldsymbol{y}$ and $\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{c}$, and $\boldsymbol{p}$.

The second approach was based on the bivariate model (B) combining the U approach and DGV as a correlated trait into a national estimate (Kachman, 2008; Mäntysaari and Strandén, 2010):

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{dgv} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_y & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_{dgv} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \mu \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_{a_y} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_{a_{dgv}} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_y \\ \boldsymbol{a}_{dgv} \end{bmatrix} + \boldsymbol{Z}_{c_y}\boldsymbol{c}_y + \boldsymbol{Z}_{p_y}\boldsymbol{p}_y + \begin{bmatrix} \boldsymbol{e}_y \\ \boldsymbol{e}_{dgv} \end{bmatrix} \quad \text{... (2)}$$

where DGV is incorporated via covariance $\sigma_{a_y, a_{dgv}}$:

$$\begin{bmatrix} \boldsymbol{a}_y \\ \boldsymbol{a}_{dgv} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \sigma_{a_y}^2 & \sigma_{a_y, a_{dgv}} \\ \sigma_{a_y, a_{dgv}} & \sigma_{a_{dgv}}^2 \end{bmatrix} \otimes \boldsymbol{A} \right) \quad \text{... (3)}$$

Estimation of required (co)variance components was performed with the residual maximum likelihood method. Conditional on these estimates, breeding values were evaluated. All computations were performed with VCE (Kovač et al., 2002). In addition to the U and B approach, we have used the international DGV as a direct predictor of a national proof.

2.1.2.3 Validation of evaluations

For the purpose of validation, the complete data was divided into training and validation subset by removing daughter phenotypes in the validation subset (Table 1). Training subset was comprised of 701 bulls born before the year 2004. Validation subset was comprised of 35 genotyped bulls born in the years between 2004 and 2007. Evaluations using the U and B approaches and different data sets were named: U_t – the U approach using the training subset, U_c – the U approach using the complete data set; B_t – the B approach using the training subset, and B_c – the B approach using the complete data set. EBVs evaluated with a certain approach (e.g., U_c) were always obtained for all bulls and were denoted accordingly (e.g., $EBV_{U,c}$).

Table 1: Description of data sets used in evaluation
Tabela 1: Opis setov podatkov uporabljenih za izvrednotenje

Data set	No. of bulls	Birth year	Origin of bulls with direct genomic value	
			Slovenian	Foreign
Complete	736	1990 - 2007	184	215
Training	701	1990 - 2003	149	215
Validation	35	2004 - 2007	35	/

Validation of predictions was assessed via theoretical and empirical accuracies separately for proven and validation bulls and separately for different evaluations ($EBV_{U,t}$, $EBV_{U,c}$, $EBV_{B,t}$, $EBV_{B,c}$). Theoretical accuracy was evaluated from the variance of prediction errors and additive genetic variance in base generation, while empirical correlations were computed as a correlation between EBVs from different evaluations.

2.1.3 Results and discussion

Heritability estimates were 0.28 for milk, 0.21 for fat, and 0.25 for protein yield and were the same for both the U and the B approach (Table 2). In the B approach, heritabilities for DGV were equal to 1.00 (after rounding) since DGV is in principle a fully heritable trait with a complete penetrance due to the same SNP equation being used for all animals. Genetic correlation between the phenotype based EBV and DGV was 0.86 for milk, 0.80 for fat, and 0.79 for protein yield.

These results suggest that DGV is a useful early predictor. Applications of such a bivariate approach are not present in the literature for dairy traits. For beef populations, lower genetic correlations between phenotypic based EBV and DGV are found. The average of reported values was 0.48 in Angus (Saatchi et al., 2011), 0.50 in Simmental (Saatchi et al., 2012), and 0.55 in Limousine breed (Saatchi et al., 2012). These correlations are lower

than in our case which can be attributed to several factors. Training set used for the development of DGV equation was larger in the InterGenomics consortium. In addition, dairy populations tend to have a smaller effective population size with stronger linkage disequilibrium leading to tighter linkage between SNP markers and potential QTLs (Goddard, 2009).

Part of the data used for the development of DGV equation was also used in this study, which leads to double counting of data in the presented analysis. One of the possibilities to avoid double counting of shared data at both the national and international levels is based on Bayesian methods. Quaas and Zang (2006) and Legarra et al. (2007) proposed two different Bayesian procedures where external or international information (EBV and associated accuracies) were integrated into the internal or national evaluation as prior information. Improvements of these methods were given by Vandenplas and Gengler (2012) in order to take into account the double counting among related external animals. In this case, the use of different sources of information was performed via the modification of a prior distribution for the additive genetic effect. In this way, correction for double counting can be applied directly opening opportunity to improve current analyses (Vandenplas and Gengler, 2012).

In the current study, the number of such bulls used for the development of DGV equation was small compared to the total number of bulls in the InterGenomics consortium leading to the negligible amount of double counting. Mäntysaari and Strandén (2010) proposed a method for the correction of double counting for the B approach and found that required corrections tend to be small. Therefore, the issue of double counting was neglected in this analysis.

Table 2: Estimates of (co)variance component ratios and correlations (± standard errors) for breeding values based on the national phenotype information $(\mathbf{h^2_y})$, breeding values based on the direct genomic value $(\mathbf{h^2_{dgv}})$, herd $(\mathbf{c})$, permanent environment $(\mathbf{p})$ effects, and residual $(\mathbf{e})$ using two approaches and complete data set

Tabela 2: Ocene deležev komponent (ko)varianc in korelacij (± standardnih napak) za plemenske vrednosti na osnovi nacionalnih fenotipskih informacij $(\mathbf{h^2_y})$, plemenske vrednosti na osnovi direktnih genomskih vrednosti $(\mathbf{h^2_{dgv}})$, čredo $(\mathbf{c})$, vpliv stalnega okolja $(\mathbf{p})$ in ostanek $(\mathbf{e})$ z uporabo dveh pristopov in popolnega seta podatkov

Item	Trait	Milk yield		Fat yield		Protein yield	
	Approach[1]	U_c	B_c	U_c	B_c	U_c	B_c
h^2_y		0.28 ± 0.003	0.28 ± 0.003	0.21 ± 0.001	0.21 ± 0.002	0.25 ± 0.003	0.25 ± 0.003
h^2_{dgv}		/	1.00 ± 0.000	/	1.00 ± 0.000	/	1.00 ± 0.000
c^2_y		0.22 ± 0.005	0.23 ± 0.005	0.21 ± 0.001	0.21 ± 0.005	0.28 ± 0.005	0.28 ± 0.006
p^2_y		0.18 ± 0.002	0.18 ± 0.002	0.15 ± 0.001	0.15 ± 0.001	0.15 ± 0.002	0.15 ± 0.001
e^2_y		0.32 ± 0.002	0.32 ± 0.002	0.43 ± 0.001	0.43 ± 0.003	0.32 ± 0.003	0.32 ± 0.003
$\mathrm{Cor}(a_y, a_{dgv})$		/	0.86 ± 0.030	/	0.80 ± 0.030	/	0.79 ± 0.003
$\mathrm{Cor}(e_y, e_{dgv})$		/	0.00 ± 0.000	/	0.00 ± 0.000	/	0.00 ± 0.000

[1]U_c – phenotypic and pedigree data used in the national genetic evaluation; B_c – bivariate analysis based on national genetic evaluation (U_c) and DGV for all genotyped bulls in the national pedigree

The average theoretical accuracy of proven bulls was 0.98 for all milk traits using the Uc approach and the inclusion of DGV via the Bc approach did not lead to the significant increase in the accuracy (Table 3). This can be attributed to the sizeable number of daughters per bull (231 on average). However, for a small number of bulls with a low number of daughters, a significant increase in accuracies was observed. The average theoretical accuracy of parent average information for bulls in the validation set (U_t) was 0.58 for milk and protein yield, and 0.52 for fat yield (Table 3). The inclusion of DGV in the B_t analysis led to an increase in the accuracy up to 0.89. Further increases in the theoretical accuracy were achieved using both, the U_c and B_c analyses, when these validation bulls were progeny tested.

Table 3: Average theoretical accuracies (minimum and maximum) for milk traits using different approach and data for a group of proven and validation bulls
Tabela 3: Povprečne teoretične natančnosti (najmanjše in največje) za lastnosti mlečnosti z uporabo različnih pristopov in podatkov za skupino progeno testiranih in mladih bikov

Bulls	No. of daughters	Data set	Accuracy		
			Milk yield	Fat yield	Protein yield
Proven (n=149)	231 (40 - 1067)	U_c	0.98 (0.52 - 0.99)	0.98 (0.43 - 0.99)	0.98 (0.51 - 0.99)
		B_c	0.99 (0.87 - 0.99)	0.98 (0.84 - 0.99)	0.99 (0.83 - 0.99)
Validation (n=35)	57 (1 - 213)	U_t	0.58 (0.30 - 0.68)	0.52 (0.38 - 0.64)	0.58 (0.29 - 0.68)
		B_t	0.89 (0.86 - 0.91)	0.86 (0.82 - 0.88)	0.85 (0.81 - 0.88)
		U_c	0.92 (0.62 - 0.99)	0.90 (0.56 - 0.99)	0.91 (0.61 - 0.99)
		B_c	0.96 (0.89 - 0.99)	0.95 (0.86 - 0.99)	0.95 (0.85 - 0.99)

[1]U_c – phenotypic and pedigree data used in the national genetic evaluation on complete data set; B_c – bivariate analysis based on Uc and DGV for all genotyped bulls in the national pedigree on complete data set; U_t – phenotypic and pedigree data used in the national genetic evaluation on training subset; B_t – bivariate analysis based on Ut and DGV for all genotyped bulls in the national pedigree on training subset

In line with the theoretical accuracies in proven bulls there were also empirical correlations between EBVs from different approaches (Figure 1). For proven bulls high correlations were observed between the national phenotype based evaluation ($EBV_{U,c}$) and genomic prediction using DGV (0.98) for milk yield. With the B approach correlations were similar. Results for fat and protein yield were similar to those of milk yield (Figure 2 and Figure 3). Since proven bulls had a sufficient number of daughters, blending of DGV with phenotype evaluation did not change their EBV significantly.

Empirical correlations in a set of validation bulls were lower. Correlation between parent average ($EBV_{U,t}$) and progeny evaluation ($EBV_{U,c}$) was 0.49 in milk yield (Figure 1). Using DGV as a sole predictor had better correlation (0.56) with progeny evaluation than pedigree prediction via parent average. Inclusion of DGV in the B_t approach led to a small increase in correlation between the progeny evaluation and prediction via the genomically

enhanced B_t approach (cor($EBV_{U,c}$, $EBV_{B,t}$)=0.61). Such a low correlation can be attributed to the low number of daughters per bull in the validation set, decreasing the correlation between these two evaluations due to the initial low accuracy of the U_c evaluation (Table 2). Finally, when combining both progeny and DGV data correlation between the $EBV_{U,c}$ and $EBV_{B,c}$ was 0.92. Similar trends in empirical correlations in the validation set of bulls were observed also for fat and protein yield, with higher values for fat yield and lower values for protein yield (Figure 2 and Figure 3).

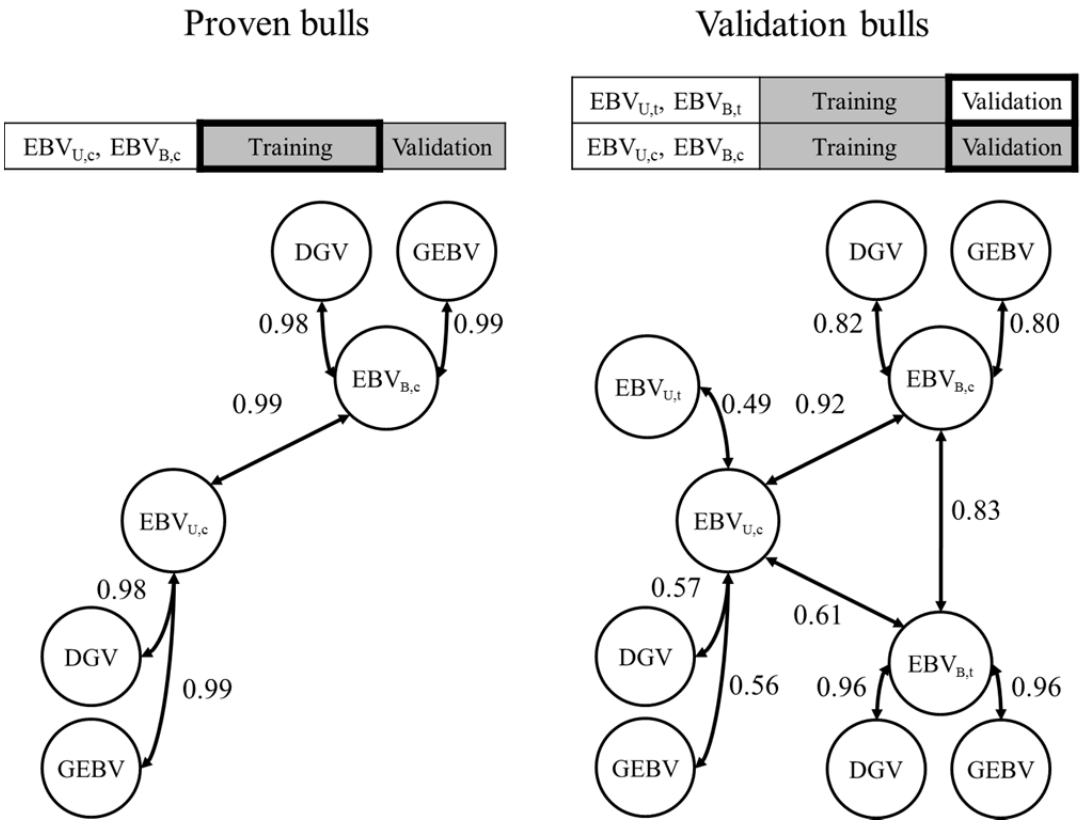


Figure 1: Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for milk yield
Slika 1: Empirične korelacije med ocenjenimi plemenskimi vrednostmi iz različnih pristopov in setov podatkov za progeno testirane in mlade bike za količino mleka

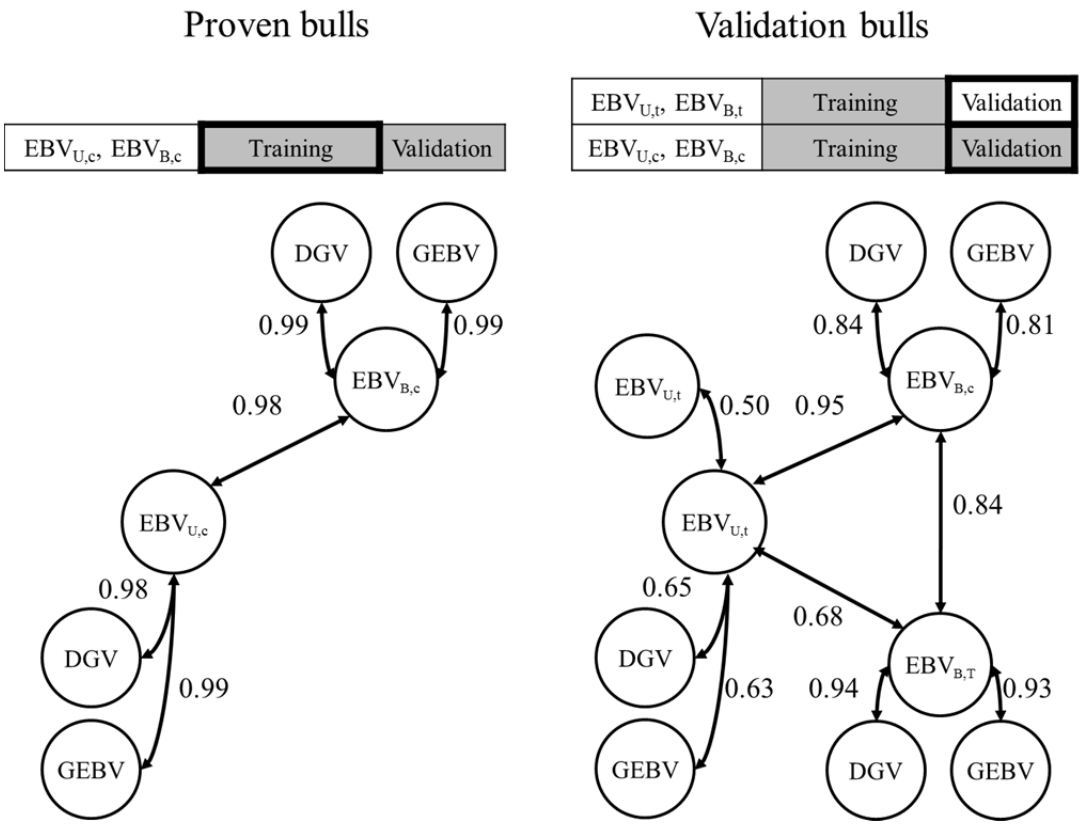


Figure 2: Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for fat yield

Slika 2: Empirične korelacije med ocenjenimi plemenskimi vrednostmi iz različnih pristopov in setov podatkov za progeno testirane in mlade bike za količino maščobe

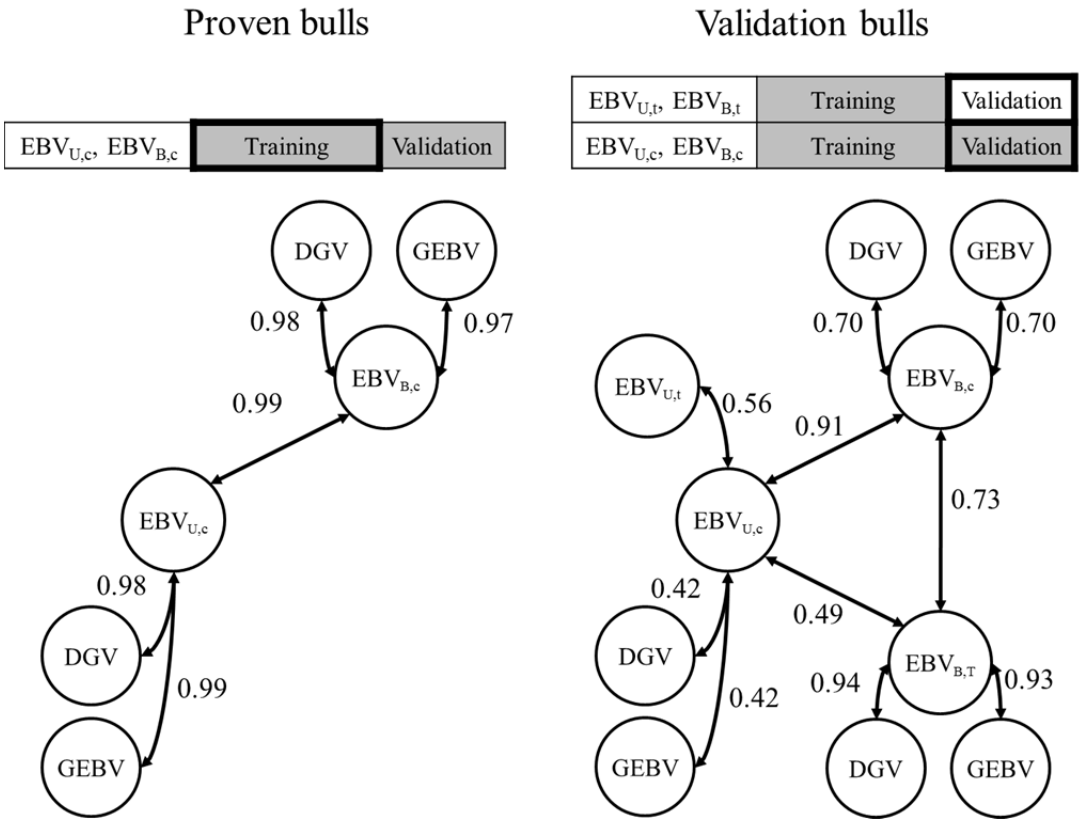


Figure 3: Empirical correlations between EBVs from different approaches and dataset for proven and validation bulls for protein yield
Slika 3: Empirične korelacije med ocenjenimi plemenskimi vrednostmi iz različnih pristopov in setov podatkov za progeno testirane in mlade bike za količino beljakovin

Empirical correlations show the benefit of combining all the available information via the B approach in the analysed population characterized by a small population size and a small number of daughters per tested bull. The availability of genomic information from a larger consortium provided the opportunity to implement genomic prediction in this population even though the size of own reference population is not sufficiently large. This paves the way of integrating the external genomic information into other small populations of dairy cattle with the national evaluation system. Additional benefit of the B approach is the unified system integrating all the available data in a single EBV instead of reporting several EBV from different data sources for the same animal.

2.1.4 Conclusion

Different approaches of integrating genomic information into a national evaluation system for small population of BSW breed were evaluated in comparison to the conventional evaluation based on phenotype and pedigree data. Results indicate that integration of DGV

from a large consortium into the national evaluation as a correlated trait enabled combination of all the available data. In addition, this approach provides a way to automatically blend all the results in a single value avoiding the need to publish several estimates of breeding values per animal.

2.2 AN INTEGRATION OF EXTERNAL INFORMATION FOR BROWN SWISS BULLS INTO THE SLOVENIAN GENETIC EVALUATION FOR MILK YIELD

2.2.1 Introduction

The decision which animals should be selected as parents for the next generation is based on the inferred (estimated) breeding value (EBV) of individual animals. Phenotypic data collected through various recording schemes (milk and fertility recording, type classification, etc.) and pedigree information are used to infer EBV in so called 'conventional' genetic evaluation. From early 1980s, exchange of frozen bulls' semen became widespread and the same bulls have been used extensively in many countries. However, genetic evaluation is based only on internal phenotypic data since external data used to select foreign bulls are not available on the national level. In such case genetic evaluation could be biased and less accurate (Bonaiti and Boichard, 1995). In order to compare the EBV of bulls evaluated in different countries, the international genetic evaluation was performed at Interbull. The evaluation is performed using a Multiple Across Country Evaluation method (MACE, Schaeffer, 1994) where EBVs of bulls from different countries were treated as different traits. Each country participating at Interbull received international EBV for bulls with greater accuracy of estimates calculated on population scale.

Genomic selection is based on the inference of EBV based on the sum of SNP markers across whole genome (Meuwissen et al., 2001; Solberg et al., 2008). Marker based BV is often called direct genomic value (DGV) since it is estimated from the direct use of marker data. Both types of BV (EBV and DGV) can be blended in genomically enhanced breeding value (GEBV) using various approaches (VanRaden, 2008; Kachman, 2008; Aguilar et al., 2010). The highest gain with genomic selection could be expected in the populations with large number of genotyped animals, especially progeny tested bulls. In BSW breed, reference population consisted of almost 8,000 genotyped bulls from seven countries (Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States) via Intergenomics project operated at the Interbull (Jorjani et al., 2012). Within the scope of the InterGenomics consortium specific prediction equations for each country are developed based on all available SNP data and multiple across country evaluation (MACE) EBV on a country's specific scale (Jorjani et al., 2012). The part of the data used for calculation of MACE EBV and for development of DGV equation comes from Slovenian population (191 bulls) leading to potential bias or double counting of information. Since EBV of an animal takes into account information from its own records and from records of all relatives (parents and progeny), double counting of contributions can be due to own records and contributions due to relationships (Misztal and Wiggans, 1988; VanRaden, 2001).

One of the possibilities to avoid double counting of shared data at both, internal (national) and external (international) levels are based on Bayesian methods. Vandenplas and Gengler (2012) pointed that Bayesian approach means that a prior distribution of EBV is changed to what is known from an external source of information. Quaas and Zang (2006) and Legarra et al. (2007) proposed two different Bayesian procedures where external information (EBV and associated accuracies) were integrated into the internal evaluation as prior information. Improvements of these methods were given by Vandenplas and Gengler (2012) and Vandenplas et al. (2014) in order to take into account the double counting among related external animals. The integration of the external information (EBV or GEBV and associated reliabilities) into internal evaluation leads to improved ranking and more reliable EBV for animals with external evaluation which is more similar to the ranking of a hypothetical joint evaluation of internal and external animals (Vandenplas and Gengler, 2012).

The objective of this study was to apply Bayesian approach developed by Vandenplas et al. (2014) to integrate and blend several external sources of information (MACE EBV and GEBV) into an internal genetic evaluation for BSW breed in Slovenia. In order to achieve the objective, aforementioned approach was applied to avoid double counting of contributions due to records and due to relationships generated by the integration of several sources of information.

2.2.2 Material and methods

Two datasets, named internal and external, were used in the analysis. Internal dataset was referred to phenotypic data and pedigree information. Phenotypic data consisted of 1,286,699 test-day records for daily milk yield from 56,383 cows recorded between years 2000 and 2014. The total number of animals in the pedigree was 82,739. Internal information included EBVs and associated reliabilities (EBV_S, REL_S) for daily milk yield from the routine national genetic evaluation ($EVAL_S$) in April 2014 (Potočnik, 1999) provided by Biotechnical Faculty, Animal Science Department. A total of 82,739 animals were associated with the available EBV_S.

The statistical model used in the internal national evaluation system is a univariate repeatability test-day model:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z}_a\boldsymbol{a} + \boldsymbol{Z}_c\boldsymbol{c} + \boldsymbol{Z}_p\boldsymbol{p} + \boldsymbol{e} \quad \text{... (4)}$$

where $\boldsymbol{y}$ is a vector of phenotypic observations for daily milk yield, $\boldsymbol{b}$ is a vector of unknown parameters for fixed effects, $\mathbf{a} \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2)$ is a vector of unknown parameters for additive genetic effect (breeding values) with covariance matrix equal to a pedigree based numerator relationship matrix ($\mathbf{A}$), $\mathbf{c} \sim N(\mathbf{0}, \mathbf{I}\sigma_c^2)$ is a vector of unknown parameters for herd effect, $\mathbf{p} \sim N(\mathbf{0}, \mathbf{I}\sigma_p^2)$ is a vector of unknown parameters for permanent

environment effect, and **e** ~ N(**0**, $\mathbf{I}\sigma_e^2$) is a vector of residuals, while $\boldsymbol{X}$, $\boldsymbol{Z}_a$, $\boldsymbol{Z}_c$, and $\boldsymbol{Z}_p$, are incidence matrices linking **y** and **b**, **a**, **c**, and **p**.

External dataset was related to information obtained from the international genetic evaluations via Interbull. For that purpose, EBV_{MACE} and REL_{MACE} from the MACE evaluation ($EVAL_{MACE}$) as well as EBV_{GEBV} and REL_{GEBV} from Intergenomics project ($EVAL_{GEBV}$) were parts of external data for 5,760 bulls. Internal and external information were further harmonized between the Slovenian and Interbull evaluations (MACE and GEBV) by adjusting scales and mean differences towards the original expression of the daily milk yield in the Slovenian national genetic evaluations ($EVAL_S$).

In order to take double counting into account, N sources of external information have been intergrated into the internal evaluation E_0. External animals associated with the i^{th} source of external information have EBV_{Ei} summarized by the vector of external EBV $\widehat{\boldsymbol{u}}_{E_i(A_i)}$, and by prediction error (co)variance matrix $\boldsymbol{D}_{E_i(A_i)}$.
The following set of multi-trait mixed model equations (5) proposed by Vandenplas et al. (2014) were used to integrate several sources of external information into the internal evaluation:

$$\begin{bmatrix} \boldsymbol{X}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{X}_{E_0} & \boldsymbol{X}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{Z}_{E_0} \\ \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{X}_{E_0} & \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{Z}_{E_0} + \boldsymbol{G}_{E_0}^{-1} + \sum_{i=1}^{N}\left(\boldsymbol{D}_{E_i}^{-1} - \boldsymbol{G}_{E_i}^{-1}\right) \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{E_0} \\ \widehat{\boldsymbol{u}}_{E_0} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{X}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{y}_{E_0} \\ \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{y}_{E_0} + \sum_{i=1}^{N}(\boldsymbol{D}_{E_i}^{-1}\widehat{\boldsymbol{u}}_{E_i}) \end{bmatrix} \quad \text{... (5)}$$

where $\boldsymbol{X}_{E_0}$ and $\boldsymbol{Z}_{E_0}$ are incidence matrices relating records in $\boldsymbol{y}_{E_0}$ to the vector of fixed effects $\widehat{\boldsymbol{\beta}}_{E_0}$ and the vector of random additive genetic effects $\widehat{\boldsymbol{u}}_{E_0}$. $\boldsymbol{G}_{E_i}^{-1}$ is the inverse of the internal additive genetic (co)variance matrix associated with the internal genetic evaluation E_0 that includes all internal and external animals and $\boldsymbol{R}_{E_0}^{-1}$ is the inverse of the residual (co)variance matrix.
Prediction error (co)variance matrix $\boldsymbol{D}_{E_i}^{-1}$ is approximated by $\boldsymbol{D}_{E_i}^{-1} = \boldsymbol{G}_{E_i}^{-1} + \boldsymbol{\Lambda}_{E_i}$ where $\boldsymbol{\Lambda}_{E_i}$ is a block diagonal variance matrix with one block per animal (Quaas and Zhang, 2006; Vandenplas et al., 2012) and $\boldsymbol{\Lambda}_{E_i} \sim \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{Z}_{E_i}$.
Each block diagonal variance matrix $\boldsymbol{\Lambda}_{E_i}$ is equal to $\boldsymbol{\Delta}_{E_{i(j)}}\boldsymbol{R}_0^{-1}\boldsymbol{\Delta}_{E_{i(j)}}$ for j^{th} animal ($j = 1, 2, ...$), where the matrix $\boldsymbol{R}_0$ is a matrix of residual (co)variance among traits. The j^{th} matrix $\boldsymbol{\Delta}_{E_{i(j)}}$ is a diagonal matrix with elements $\sqrt{R_{ijk}}$ for k^{th} trait ($k = 1, 2, ...$).

R_{ijk} is the effective number of records i.e. records equivalents, for the j^{th} animal for the k^{th} trait associated with the i^{th} source (Vandenplas et al., 2012; Vandenplas et al., 2013) and express the contributions due to relationships and/or due to records considered for the evaluation of an animal. It is equal to 0 for internal animals, because all contributions are only due to the relationships among external and internal animals (Vandenplas et al., 2014).

For external animals, $RE_{ijk} = \frac{1-h_k^2}{h_k^2} * \frac{REL_{ijk}}{1-REL_{ijk}}$ for the j^{th} animal for the k^{th} trait associated with the i^{th} source, where h_k^2 is the heritability of the k^{th} trait when double counting of contributions due to relationships among external animals is not taken into account (Misztal and Wiggans, 1988). If double-counting of contributions due to relationships among external animals is taken into account, RE_{ijk} only expresses the amount of contributions due to records and can be estimated through a two-step algorithm (Vandenplas et al., 2012; Vandenplas et al., 2014). In the first step external animals associated with only contributions due to relationships were determined (Vandenplas et al., 2014). In the second step the amount of contributions due to records (expressed as RE) for external animals associated with contributions due to relationships and own records were estimated.

A set of multi-trait mixed model equations (6) proposed by Vandenplas et al. (2014) that integrate several sources of external information and take double-counting of contributions due to records between external and internal information into account, can be written as follows:

$$\begin{bmatrix} \boldsymbol{X}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{X}_{E_0} & \boldsymbol{X}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{Z}_{E_0} \\ \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{X}_{E_0} & \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{Z}_{E_0} + \boldsymbol{G}_{E_0}^{-1} + \\ & \sum_{i=1}^{N}\left(\boldsymbol{D}_{E_i}^{-1} - \boldsymbol{G}_{E_i}^{-1}\right) - \sum_{i=1}^{N}\left(\boldsymbol{D}_{I_i}^{-1} - \boldsymbol{G}_{I_i}^{-1}\right) \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}}_{E_0} \\ \widehat{\boldsymbol{u}}_{E_0} \end{bmatrix} = \begin{bmatrix} X'_{E_0}R_{E_0}^{-1}y_{E_0} \\ \boldsymbol{Z}'_{E_0}\boldsymbol{R}_{E_0}^{-1}\boldsymbol{y}_{E_0} + \sum_{i=1}^{N}\left(\boldsymbol{D}_{E_i}^{-1}\widehat{\boldsymbol{u}}_{E_i}\right) - \sum_{i=1}^{N}\left(\boldsymbol{D}_{I_i}^{-1}\widehat{\boldsymbol{u}}_{E_i}\right) \end{bmatrix} \quad \text{... (6)}$$

where $\boldsymbol{G}_{I_i}$ is the genetic (co)variance matrix for all animals for the internal information included into i^{th} source of external information, $\widehat{\boldsymbol{u}}_{E_0}\begin{bmatrix} \widehat{\boldsymbol{u}}_{I_i(A_i^0)} \\ \widehat{\boldsymbol{u}}_{I_i(A_i)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_{I_i(A_i^0A_i)} & G^{-1}_{I_i(A_i^0A_i)}\widehat{\boldsymbol{u}}_{I_i(A_i)} \\ & \widehat{\boldsymbol{u}}_{I_i(A_i)} \end{bmatrix}$ is the vector of internal EBV associated with the i^{th} source of external information that includes internal information and $\boldsymbol{D}_{I_i}^{-1}$ is the inverse of the prediction error (co)variance matrix associated with $\widehat{\boldsymbol{u}}_{I_i}$.

Blending several sources of external information by avoiding double counting of contributions due to records is based on the equation that can be written as follows:

$$\left(\boldsymbol{G}_{E_0}^{-1} + \sum_{i=1}^{N}\left(\boldsymbol{D}_{E_i}^{-1} - \boldsymbol{G}_{E_i}^{-1}\right) - \sum_{i=1}^{N}\left(\boldsymbol{D}_{I_i}^{-1} - \boldsymbol{G}_{I_i}^{-1}\right)\right)\widehat{\boldsymbol{u}}_{E_0}$$

$$= \sum_{i=1}^{N}\left(\boldsymbol{D}_{E_i}^{-1} - \widehat{\boldsymbol{u}}_{E_i}\right) - \sum_{i=1}^{N}\left(\boldsymbol{D}_{I_i}^{-1} - \widehat{\boldsymbol{u}}_{I_i}\right) \quad \text{... (7)}$$

Slovenian and international (MACE EBV or GEBV) information were blended using equation (7) for the following four cases: with or without consideration of double counting of contributions due to relationships and with or without consideration of double counting of contributions due to records. Double counting of contributions due to relationships was possible since all animals associated with Slovenian and/or international information was related. Double counting of contributions due to records was also possible because international information (MACE EBV or GEBV) for 5,760 bulls included contribution provided by the internal genetic evaluation ($EVAL_S$).

To test the importance of double counting, the following evaluations were considered:
1) Slovenian BLUP ($EVAL_S$, EBV_S, REL_S),
2) Slovenian BLUP and MACE information were blended without considering double counting of contributions due to records and due to relationships ($EVAL_{NDC}$, EBV_{NDC}, REL_{NDC}),
3) Slovenian BLUP and MACE information were blended by considering only double counting of contributions due to records ($EVAL_{DCR}$, EBV_{DCR}, REL_{DCR}),
4) Slovenian BLUP and MACE information were blended by only considering double-counting of contributions due to relationships among all animals ($EVAL_{DCP}$, EBV_{DCP}, REL_{DCP}),
5) Slovenian BLUP and MACE information were blended by considering both double counting of contributions due to records and relationships among all animals ($EVAL_{DCRP}$, EBV_{DCRP}, REL_{DCRP}).
$EVAL_{MACE}$ was used as the reference evaluation for bulls evaluated through $EVAL_{MACE}$.

To test the importance of double counting based on $EVAL_{GEBV}$, evaluation was performed for the following cases:
1) National BLUP and GEBV information were blended without considering double counting of contributions due to records and due to relationships ($EVAL_{NGNDC}$, EBV_{NGNDC}, REL_{NGNDC}),
2) National BLUP and GEBV information were blended by considering only double counting of contributions due to records ($EVAL_{NGDCR}$, EBV_{NGDCR}, REL_{NGDCR}),
3) National BLUP and GEBV information were blended by considering only double counting of contributions due to relationships among all animals ($EVAL_{NGDCP}$, EBV_{NGDCP}, REL_{NGDCP}) and,

4) National BLUP and GEBV information were blended by considering both double counting of contributions due to records and relationships among all animals ($EVAL_{NGDCPR}$, EBV_{NGDCRP}, REL_{NGDCRP}).
$EVAL_{GEBV}$ was considered as the reference evaluation for bulls evaluated through $EVAL_{GEBV}$.

All evaluated cases are presented in Table 4. Reliabilities for evaluations from $EVAL_{NDC}$ to $EVAL_{NGDCPR}$ were computed using the equation $REL = 1 - PEV/\sigma_g^2$ where σ_g^2 is the genetic variance for the corresponding trait and PEV is the prediction error variance obtained from the diagonal element of the inverted left hand- side of the equation (7).

Table 4: Bayesian evaluations for testing the importance of double-counting issues
Tabela 4: Bayesovska vrednotenja za testiranje pomembnosti dvojnega štetja

Bayesian evaluations	Double-counting accounted	
	Records	Relationships
$EVAL_{NDC}$		
$EVAL_{DCR}$	X	
$EVAL_{DCP}$		X
$EVAL_{DCRP}$	X	X
$EVAL_{NGNDC}$		
$EVAL_{NGDCR}$	X	
$EVAL_{NGDCP}$		X
$EVAL_{NGDCRP}$	X	X

As already mentioned $EVAL_{MACE}$ was considered as the reference evaluation for bulls evaluated through $EVAL_{MACE}$. Comparisons were performed between $EVAL_{MACE}$ and $EVAL_{S,}$ $EVAL_{NDC,}$ $EVAL_{DCR}$, $EVAL_{DCP}$, and $EVAL_{DCRP}$. Furthermore, $EVAL_{GEBV}$ was considered as the reference evaluation for bulls evaluated through $EVAL_{GEBV}$ and was compared with Bayesian evaluations from $EVAL_{NGNDC}$ to $EVAL_{NGDCRP}$. All comparisons were performed based on:
a) Spearman's rank correlation coefficients (r) of EBV_{MACE} or EBV_{GEBV} with corresponding evaluations,
b) MSE of evaluations (i.e. mean squared errors expressed as a percentage of average MSE of EBV_S or EBV_{NGNDC}),
c) regression coefficients (a),
d) R^2 of the regressions of $EVAL_{MACE}$ or $EVAL_{GEBV}$ on the four other evaluations,
e) RE_{tot}, and
f) average REL for the corresponding evaluations.

Comparisons were performed for two groups of bulls: internally used and internally unused bulls. Internally used bulls were associated with either Slovenian and MACE information (400 bulls) or Slovenian and GEBV information (320 bulls) and had daughters with records in the Slovenian internal dataset. A group of internally unused bulls included bulls that have either Slovenian and MACE information (5,360 bulls) or Slovenian and GEBV information (5,344 bulls) but had no daughters with records in the Slovenian internal dataset. All computations were performed using the BLUPF90 program (Misztal et al., 2002) modified to implement the equations (5), (6), and (7). SAS software (SAS Institute, 2011) was used for comparison among evaluations.

2.2.3 Results and discussion

Blending information (Slovenian BLUP and MACE) without considering double counting of contributions due to records and due to relationships ($EVAL_{NDC}$) was considered as well as double counting of contributions due to records ($EVAL_{DCR}$), and double counting of contributions due to relationships among all animals ($EVAL_{DCP}$) to study their influences on prediction of $EVAL_{MACE}$ for internally used bulls (Table 5). Internally used bulls associated with Slovenian and MACE information (400 bulls) had on average 112.5 internal daughters with records. The average REL_S was 0.87 while the average REL_{MACE} was 0.90.

Results for r, MSE, a, and R^2 for prediction of EBV_{MACE} by evaluations from $EVAL_S$ to $EVAL_{DCRP}$ are shown in Table 5 for daily milk yield. Blending of Slovenian and MACE information by taking double counting of contributions due to records and due to relationships into account (i.e. $EVAL_{DCRP}$) led to a ranking that was more similar to the MACE ranking (r = 0.99) than to the internal ranking (i.e. $EVAL_S$; r = 0.86), despite the fact that internally used bulls had a large number of daughters with records in the Slovenia. Rank correlations between EBV_{MACE} and EBV_S increased by 0.13 points to achieve a rank correlation between EBV_{MACE} and EBv_{DCRP} that was 0.99 (Table 5). The MSE, a, and R^2 showed that accuracy of predictions of EBV_{MACE} by EBV_S or by EBV_{DCRP} increased when external information was integrated. Integration of MACE information also increased the average REL by 0.03 points compared to REL_S (Table 5). Also, the average REL_{DCRP} for the 400 bulls was 0.06 points higher than the average REL_{MACE} (Table 5). The difference in average REL, as well as the differences between EBV_{MACE} and EBV_{DCRP} based on MSE, a, and R^2 can be explained by the fact that MACE did not include all information available for animals in the Slovenia. According to Interbull requirements, EBV_S of bulls was included into MACE if it had at least ten daughters with records within ten herds at the internal level.

Table 5: Parameters obtained between $EVAL_{MACE}$ and evaluations based on National BLUP and MACE EBV ($EVAL_{S,}$ $EVAL_{NDC,}$ $EVAL_{DCR,}$ $EVAL_{DCP,}$ $EVAL_{DCRP}$) for 400 internally used bulls
Tabela 5: Parametri ocenjeni med $EVAL_{MACE}$ in vrednotenja ki izhajajo iz nacionalnega BLUP in MACE EBV ($EVAL_S$, $EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$, $EVAL_{DCRP}$) za 400 bikov v interni uporabi

Genetic evaluations	r[1]	MSE[2]	a[3]	R^2 [4]	RE_{tot} [5]	REL[6]
$EVAL_S$	0.86	100.00	0.88±0.03	0.73	/	0.87±0.20
$EVAL_{NDC}$	0.99	8.21	0.97±0.01	0.98	26.31	0.95±0.06
$EVAL_{DCR}$	0.99	5.50	0.99±0.01	0.98	2.21	0.94±0.06
$EVAL_{DCP}$	0.98	11.28	0.96±0.01	0.97	24.87	0.94±0.06
$EVAL_{DCRP}$	0.99	8.51	0.99±0.01	0.98	2.56	0.93±0.07

[1]r = rank correlation between $EVAL_{MACE}$ and $EVAL_{S,}$ $EVAL_{NDC,}$ $EVAL_{DCR,}$ $EVAL_{DCP}$ or $EVAL_{DCRP}$; [2]MSE = mean squared error expressed as a percentage of the average internal mean squared error; [3]a = regression coefficient ± standard error; [4]R^2 = coefficient of determination of the regression of MACE EBV on EBV estimated by $EVAL_{S,}$ $EVAL_{NDC,}$ $EVAL_{DCR,}$ $EVAL_{DCP}$ or $EVAL_{DCRP}$; [5]RE_{tot} = total amount of record equivalents; [6]REL = average reliability ± SD

Parameters r, a, and R^2 associated with $EVAL_{NDC,}$ $EVAL_{DCR}$ and $EVAL_{DCP}$ for the 400 bulls were similar to the r, a, and R^2 of $EVAL_{DCRP}$, although a slight advantage was observed for $EVAL_{DCRP}$. Therefore, the four blending evaluations led to similar rankings as MACE for the 400 internally used bulls (i.e. rank correlations equal to 0.99 on average). The result indicate that $EVAL_{DCRP}$, i.e. a Bayesian evaluation that blended internal and external information and avoided most double counting of contributions due to records and due to relationships, was successful in integrating MACE information for internally used bulls.

Double counting could be also observed based on RE_{tot} (Table 5). With regard to double counting of contributions due to relationship for the 400 internally used bulls, RE that was free of contributions due to records (i.e. RE that included only contributions due to relationship) for EBV_{MACE} was equal to 1.44. These amounts of RE free of contributions due to relationships represented 94.5% of the RE that contributed to MACE information. While double counting of contributions due to relationship was present for $EVAL_{DCR}$ (i.e. the blending evaluation that considered only double counting of contributions due to records), the contributions due to relationship were small and their double counting had little effect on the prediction of EBV_{MACE} for the internally used bulls. RE that was free of contributions due to records (i.e. RE that included only contributions due to relationship) for EBV_{MACE} was 2.21. If double counting of contributions due to relationships was considered only, RE_{tot} associated with $EVAL_{DCP}$ was 24.87, while RE_{tot} associated with $EVAL_{DCPR}$ was 2.56. Thus, 22.31 RE were considered twice (i.e. records were double counted) by $EVAL_{DCRP}$.

Double counting of contributions due to records ($EVAL_{DCR}$) affected the prediction of EBV_{MACE} for internally used bulls slightly according to all parameters. $EVAL_{DCR}$ had

small advantage compared to $EVAL_{DCP}$ based on r, MSE, and R^2 for analysed trait. Furthermore, $EVAL_{DCR}$ was more reliable based on MSE than $EVAL_{DCRP}$, but parameter a indicated that $EVAL_{DCR}$ was less reliable compared to $EVAL_{DCRP}$. These evaluations also showed that EBV_{MACE} for the 400 internally used bulls were slightly better predicted when contributions due to records (EBV_{DCR}) were considered. Based on these results, it can be stated that double counting of contributions due to relationships and due to records had little effect on EBV for internally used bulls.

The Bayesian analysis was also performed for Holstein breed in Walloon region in Belgium (Vandenplas et al., 2014). Their results indicate that evaluations which blended internal and external information and avoided double counting of contributions due to records and due to relationships was successful in integrating MACE information and had little effect on EBV for internally used bulls. For Walloon internally used bulls, the average national REL (REL_W) was slightly lower, from 0.74 to 0.76 as well as the average REL_{MACE} (0.88) than is current study. Similar rankings of internally used bulls were gained based on four evaluations as in Slovenian example. Rank correlations also increased (by 0.104 points) from EBV_{MACE} to $EVAL_{DCRP}$. Results for r, MSE, a, and R^2 showed that accuracy of EBV_{MACE} prediction increased when external information was integrated by four evaluations. No preference was observed between $EVAL_{DCR}$ and EVA_{DCP} based on r, MSE, a, and R^2. Indeed, r and R^2 were similar for these two evaluations, while $EVAL_{DCR}$ was more reliable based on MSE, but parameter a indicated that $EVAL_{DCP}$ was more reliable. Different approach was proposed by VanRaden and Tooker (2012) althouh they found similar correlations between EBV_{MACE} and combined EBV for bulls with only external daughters for yield traits (between 0.991 and 0.994). Their strategy included compution of external deregressed proofs (DRP) from EBV_{MACE} and inclusion of one extra record based on these DRP, weighted by the associated daughter equivalents (DE) for the bull. Internal contributions in MACE information for bulls with internal and external daughters were considered by subtracting the number of internal DE from the total and by using internal EBV to compute external DRP. However, the main advantage of the approach used in the current study was to avoid a pre-processing deregression step or computation of external EBV (Vandenplas et al., 2014).

Of the internally used bulls, 320 had Slovenian and GEBV information and internal daughters with records. Results for r, MSE, a, and R^2 for prediction of EBV_{GEBV} by evaluations from $EVAL_{NGNDC}$ to $EVAL_{NGDCRP}$ are shown in Table 6 for daily milk yield. Blending of national BLUP and GEBV information by taking into account double counting of contributions due to records and contributions due to relationships (i.e. $EVAL_{NGDCRP}$) lead to the same ranking (r = 0.99) as observed for other evaluations based on national BLUP and GEBV (Table 6). Parameters r, a, and R^2 associated with $EVAL_{NGNDC}$, $EVAL_{NGDCR}$ and $EVAL_{NGDCP}$ for the 320 internally used bulls were similar to the r, a, and

R^2 of $EVAL_{NGDCRP}$. Therefore, the four blending evaluations led to similar rankings as GEBV for the 320 internally used bulls (i.e. rank correlations equal to 0.99 on average).

Table 6: Parameters obtained between $EVAL_{GEBV}$ and evaluations based National BLUP and GEBV ($EVAL_{NGNDC}$, $EVAL_{NGDCP}$, $EVAL_{NGDCR}$, $EVAL_{NGDCRP}$) for 320 internally used bulls
Tabela 6: Parametri ocenjeni med $EVAL_{GEBV}$ in vrednotenja ki izhajajo iz nacionalnega BLUP in GEBV EBV ($EVAL_{NGNDC}$, $EVAL_{NGDCP}$, $EVAL_{NGDCR}$, $EVAL_{NGDCRP}$) za 320 bikov v interni uporabi

Genetic evaluations	r[1]	MSE[2]	a[3]	R^2 [4]	RE_{tot} [5]	REL[6]
$EVAL_{NGNDC}$	0.99	2.29	1.00±0.01	0.99	60.57	0.98±0.02
$EVAL_{NGDCR}$	0.99	1.23	1.00±0.01	0.99	35.71	0.97±0.02
$EVAL_{NGDCP}$	0.99	2.58	1.00±0.01	0.99	59.03	0.98±0.02
$EVAL_{NGDCRP}$	0.99	1.28	1.00±0.01	0.99	31.88	0.97±0.02

[1]r = rank correlation between $EVAL_{GEBV}$ and $EVAL_{NGNDC}$, $EVAL_{NGDCR}$, $EVAL_{NGDCP}$, or $EVAL_{NGDCRP}$; [2]MSE = mean squared error expressed as a percentage of the average internal mean squared error; [3]a = regression coefficient ± standard error; [4]R^2 = coefficient of determination of the regression of GEBV on EBV estimated by $EVAL_{NGNDC}$, $EVAL_{NGDCR}$, $EVAL_{NGDCP}$ or $EVAL_{NGDCPR}$; [5]RE_{tot} = total amount of record equivalents; [6]REL = average reliability ± SD

Double counting of contributions due to relationship for the 320 internally used sires based on RE_{tot} (Table 6) showed that RE free of contributions due to records (i.e. RE that included only contributions due to relationship) for EBV_{GEBV} was equal to 1.54. These amounts of RE free of contributions due to relationships represented 97.5% of the RE that contributed to GEBV information. While double counting of contributions due to relationship was present for $EVAL_{NGDCR}$ (i.e. the blending evaluation that considered only double counting of contributions due to records), the contributions due to relationship were small and their double counting had little effect on the prediction of EBV_{GEBV} for the internally used bulls. RE that was free of contributions due to records (i.e. RE that included only contributions due to relationship) for EBV_{GEBV} was 35.71. If double counting of contributions due to relationships was considered only, RE_{tot} associated with $EVAL_{NGDCP}$ was 59.03, while RE_{tot} associated with $EVAL_{NGDCPR}$ was 31.88. Thus, 27.15 RE were considered twice (i.e. records were double counted) by $EVAL_{NGDCRP}$.

These evaluations showed that EBV_{GEBV} for the 320 internally used bulls were slightly better predicted when contributions due to records (EBV_{NGDCR}) were considered. Double counting of contributions due to records ($EVAL_{NGDCR}$) affected the prediction of EBV_{GEBV} for internally used bulls slightly since $EVAL_{NGDCR}$ had small advantage compared to $EVAL_{NGDCP}$ based on MSE for analysed trait. Furthermore, $EVAL_{NGDCR}$ was more reliable based on REL than $EVAL_{NGDCRP}$. Based on these results, it can be stated that double counting of contributions due to relationships and due to records had little effect on EBV_{GEBV} for internally used bulls.

Group of internally unused bulls that had only external daughters with records included 5,360 bulls with Slovenian and MACE information. The average REL_S of these bulls was 0.17 (Table 7) and the average REL_{MACE} was 0.72. Integration of MACE information increased the average REL_S by 0.58 points, resulting in average REL_{DCRP} equal to 0.72. Since they had only external daughters, Slovenian contributions only included contributions due to relationships and no contributions due to records. The nearly correct estimation of contributions due to relationships led to same average REL_{MACE} and average REL_{DSRP} (Table 7).

Table 7: Parameters obtained between $EVAL_{MACE}$ and evaluations based on National BLUP and MACE EBV ($EVAL_S$, $EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$, $EVAL_{DCRP}$) for 5,360 internally unused bulls
Tabela 7: Parametri ocenjeni med $EVAL_{MACE}$ in vrednotenja ki izhajajo iz nacionalnega BLUP in MACE EBV ($EVAL_S$, $EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$, $EVAL_{DCRP}$) za 5,360 bikov ki niso v interni uporabi

Genetic evaluations	r[1]	MSE[2]	a[3]	R^2 [4]	RE_{tot} [5]	REL[6]
$EVAL_S$	0.51	100.00	0.94±0.02	0.26	/	0.17±0.09
$EVAL_{NDC}$	0.99	0.09	1.00±0.00	0.99	3.15	0.75±0.05
$EVAL_{DCR}$	0.99	0.06	1.00±0.00	0.99	3.15	0.75±0.05
$EVAL_{DCP}$	0.99	0.17	0.99±0.01	0.99	2.51	0.72±0.05
$EVAL_{DCRP}$	0.99	0.12	1.00±0.00	0.99	2.51	0.72±0.05

[1]r = rank correlation between $EVAL_{MACE}$ and $EVAL_S$, $EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$ or $EVAL_{DCRP}$; [2]MSE = mean squared error expressed as a percentage of the average internal mean squared error; [3]a = regression coefficient ± standard error; [4]R^2 = coefficient of determination of the regression of MACE EBV on EBV estimated by $EVAL_S$, $EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$ or $EVAL_{DCRP}$; [5]RE_{tot} = total amount of record equivalents; [6]REL = average reliability ± SD

Results for r, MSE, a, and R^2 for prediction of EBV_{MACE} by $EVAL_{NDC}$ and $EVAL_{DC}$ are shown in Table 7 for the 5,360 bulls for milk yield. Blending of Slovenian and MACE information led to similar rankings of these bulls for all blending evaluations ($EVAL_{NDC}$, $EVAL_{DCR}$, $EVAL_{DCP}$, and $EVAL_{DCRP}$). Rank correlations between EBV_{MACE} and EBV for the four blending evaluations increased from 0.26 to 0.99. These rank correlations indicated that the blending method was also successful for bulls with only external information. These results were confirmed by a decrease of MSE and by regression coefficients (a) close or equal to 1.0, with a R^2 equal to 0.99 for all blending evaluations (Table 7). Because double counting can be only attributed to contributions due to relationships for these bulls, $EVAL_{NDC}$ and $EVAL_{DCR}$ had almost the same values for parameters. Small difference was observed only for MSE between these two groups of evaluations. This was also observed for $EVAL_{DCP}$ and $EVAL_{DCRP}$ (Table 7).

As already mentioned double counting is only based on contributions due to relationships (EBV_{DCP}) and no contributions due to records. This could be also observed based on RE_{tot} (Table 7). With regard to double counting of contributions due to relationship for the 5,360 internally unused bulls, RE that was free of contributions due to records ($EVAL_{DCR}$, i.e. RE that included only contributions due to relationship) for EBV_{MACE} was very small and

equal to 0.64. These two evaluations showed that EBV_{MACE} were slightly better predicted when contributions due to relationships (EBV_{DCP}) were considered. However, all these results showed that contributions due to relationships had little effect on the prediction of $EBV_{MACE.}$ In Holstein breed in Walloon region in Belgium, Vandenplas et al. (2014) also showed that contributions due to relationships had little effect on the prediction of EBV_{MACE} for internally unused bulls. For the mentioned data set, $EVAL_{NDC}$ and $EVAL_{DCR}$ had similar values for analysed parameters (r, MSE, a, and R^2). Therefore, ranking of these bulls was similar among evaluations.

Among internally unused bulls that had only external daughters with records, 5,344 of them had Slovenian and GEBV information. Results for r, MSE, a, and R^2 for prediction of EBV_{GEBV} by evaluations from $EVAL_{NGNDC}$ to $EVAL_{NGDCRP}$ are shown in Table 8 for milk yield. Blending of national BLUP and GEBV information by taking double counting of contributions due to records and due to relationships into account (i.e. $EVAL_{NGDCRP}$) lead to the same ranking (r = 0.99) and indicated that the blending method was also successful for bulls with external genomic information. The four blending evaluations led to similar rankings as GEBV for the 5,344 internally unused bulls (i.e. rank correlations equal to 0.99 on average). These results were also confirmed by the almost same values of parameters r, a, and R^2 for $EVAL_{NGNDC}$ and $EVAL_{NGDCR}$ (Table 8) because double counting can be only attributed to contributions due to relationships.

Table 8: Parameters obtained between $EVAL_{GEBV}$ and evaluations based either on National BLUP and GEBV ($EVAL_{NGNDC,}$ $EVAL_{NGDCR,}$ $EVAL_{NGDCP,}$ $EVAL_{NGDCRP}$) for 5,344 internally unused bulls
Tabela 8: Parametri ocenjeni med $EVAL_{GEBV}$ in vrednotenja ki izhajajo iz nacionalnega BLUP in GEBV ($EVAL_{NGNDC,}$ $EVAL_{NGDCR,}$ $EVAL_{NGDCP,}$ $EVAL_{NGDCRP}$) za 5,344 bikov ki niso v interni uporabi

Genetic evaluations	r[1]	MSE[2]	a[3]	R^2 [4]	RE_{tot}[5]	REL[6]
$EVAL_{NGNDC}$	0.99	0.01	1.00±0.00	0.99	11.61	0.91±0.03
$EVAL_{NGDCR}$	0.99	0.01	1.00±0.01	0.99	11.09	0.91±0.03
$EVAL_{NGDCP}$	0.99	0.01	1.00±0.01	0.99	11.09	0.91±0.03
$EVAL_{NGDCRP}$	0.99	0.01	1.00±0.01	0.99	10.75	0.91±0.03

[1]r = rank correlation between $EVAL_{GEBV}$ and $EVAL_{NGNDC,}$ $EVAL_{NGDCR,}$ $EVAL_{NGDCP,}$ or $EVAL_{NGDCRP}$; [2]MSE = mean squared error expressed as a percentage of the average internal mean squared error; [3]a = regression coefficient ± standard error; [4]R^2 = coefficient of determination of the regression of GEBV on EBV estimated by $EVAL_{NGNDC,}$ $EVAL_{NGDCR,}$ $EVAL_{NGDCP}$ or $EVAL_{NGDCPR}$; [5]RE_{tot} = total amount of record equivalents; [6]REL = average reliability ± SD

Double counting based on RE_{tot} (Table 8) showed that RE that was free of contributions due to records ($EVAL_{DCR}$, i.e. RE that included only contributions due to relationship) for EBV_{GEBV} was equal to 0.52 and represented 95.5% of the RE that contributed to GEBV information. If double counting of contributions due to relationships was considered only, RE_{tot} associated with $EVAL_{NGDCP}$ was 11.09, while RE_{tot} associated with $EVAL_{NGDCPR}$ was 10.75. Thus, only 0.34 RE were considered twice (i.e. records were double counted) by $EVAL_{NGDCRP}$. Since double counting was contributed only by contributions due to

relationship, $EVAL_{NGDC}$ and $_{NGDCRP}$ had little effect on the prediction of EBV_{GEBV}. These evaluations showed that double counting of contributions due to relationships and due to records had little effect on GEBV for internally unused bulls.

2.2.4 Conclusion

The proposed Bayesian method blended and integrated several sources of external information (MACE EBV and GEBV) into an internal genetic evaluation for BSW bulls in Slovenia. The results showed that the method was able to avoid double counting of contributions due to records and due to relationships. Therefore, double counting of contributions due to relationships and due to records had little effect on MACE EBV and GEBV for internally used and unused bulls. Furthermore, because all available external sources of information were correctly integrated, relatives of external animals benefited and received more reliable EBV.

2.3 GENOME-WIDE ASSOCIATION STUDY FOR DAIRY TRAITS IN SLOVENIAN BROWN SWISS BREED

2.3.1 Introduction

In dairy cattle production, all economically important traits are influenced by many genes that appear as continuous variation (Falconer and Mackey, 1996). The identification of genomic regions with quantitative trait loci (QTL) for economically important traits was result of linkage maps development containing polymorphic genetic markers (Kappes et al., 1997). The inclusion of molecular information in traditional selection scheme using marker assisted selection (MAS) became a voluble tool in selecting traits of interest (Meuwissen and Goddard, 1996; Spelman and Garrick, 1997). The availability of dense maps thousands of single-nucleotide polymorphisms (SNPs) made feasible use of this information in selection decisions, commonly called genome-wide or genomic selection (Meuwissen et al., 2001). These SNPs can be used to derive direct genomic values (DGV) via the estimation of SNP effects across the whole genome (Meuwissen et al., 2001; Solberg et al., 2008). Genomic selection allows achieving an extra genetic gain compared to the traditional EBV, due to the higher accuracy of EBV for young non-phenotyped animals (Hayes et al., 2009a). Furthermore, the ability to reduce the generation interval due to an earlier estimation of the genetic merit is another advantage of using genomic selection (Schaeffer, 2006).

Development of SNP marker panels also enabled genome wide association studies (GWAS) with a purpose to determine associations between DNA polymorphisms and phenotypes (Hirschhorn and Daly, 2005). In the GWAS studies, association analysis was used to identify chromosome regions that harbour the genes or regulatory elements related to the traits of interest (e.g., Goddard and Hayes, 2009; Bush and Moore, 2012; Eggen, 2012; Montaldo et al., 2012). The first use of GWAS was in the analysis of human diseases in order to reveal the genetic basis of disease (Massey et al., 2007). These studies gave encouraging results in identification of genomic locations associated with Crohn's disease (Barrett et al., 2008), diabetes (Frayling, 2007) and cancer (Easton and Eeles, 2008). GWAS was extended to the field of livestock species (like cattle) to discover superior variants of all genes contributing to the phenotype of economically important milk (Bolormaa et al., 2010; Pryce et al., 2010; Mai et al., 2010), meat (Bolormaa et al., 2011) and health traits (Murdoch et al., 2011). The results from GWAS could be used to identify markers that will improve the accuracy of EBV for these traits and to increase the rate of genetic improvement (Meuwissen et al. 2001).

Methods used in GWAS are based on assumption that SNPs have the feasibility of QTL detection and mapping using historical population linkage disequilibrium (LD), which

requires that a SNP allele have to be in LD with the QTL allele across the entire population. The pattern of LD observed in a population depends on the history of the population, especially the history of its effective population size (Ne) (Hayes et al., 2003), domestication, breed formation and selection influencing their recent histories (Kemper et al., 2012). For example, N_e for *Bos taurus* was >50,000 prior to domestication, from 1,000 to 2,000 after domestication, and declined to about 100 in recent times in many breeds due to intensive artificial insemination and assortative mating scheme in milk production (Goddard and Hayes, 2012). Therefore, GWAS can be successfully implemented in dairy cattle, due to the existing LD between high dense markers and QTL for the trait of interest (Hayes et al., 2009a; de Roos et al., 2008).

The Brown Swiss (BSW) breed is a cosmopolitan breed selected for dairy characteristics and ability to produce in less favoured areas mostly in Alpine regions of Austria, Germany, Italy, Switzerland, and Slovenia. Original Brown breed (Braunvieh) from Switzerland was exported to the United States in 1869 (Yoder and Lush, 1937), and after strong selection of this breed, previously classified as a dual-purpose breed, it was established as a dairy type and received a name Brown Swiss breed. Development of Slovenian BSW breed started at the beginning of 20th century when Original Brown breed bulls were imported from Austria and Switzerland with the manner to improve cows of local grey cattle. Lately also breeding heifers were imported. From the 1960s, the American BSW breeding bulls were used for improving cows of Brown breed in order to increase milk production (Ferčej et al., 1980). Bulls with 50% of American BSW genotype were imported from neighbouring countries. Moreover, five bulls with 100% American BSW genotype were also used for insemination. From 1976 to 1980 bulls were no longer imported but have been bred in Slovenia. The breed is mainly spread in the western part of Alpine area, as well as southern and northern part of Slovenia with harsh environmental conditions. According to the annual report (Sadar et al., 2013), 6.95% of all active cattle (or 31,087 heads) belong to the BSW breed. It is the second most important breed for milk production (Sadar et al., 2013). The average production is 5,554 kg of milk in the standard lactation with 4.06% of fat and 3.39% of protein content.

A potential challenge in GWAS is the presence of population stratification (family substructure) that can result in false-positive or negative associations and may lead to confounding results (Devlin and Roeder, 1995; Pritchard et al., 2000). In dairy cattle, population stratification could be result of genetic selection that caused allele frequency changes (Sonstegard et al., 2010) and artificial insemination which increased the presence of related animals in a randomly selected samples and the presence of large half-sib families (Ma et al., 2012). In order to take into account population stratification linear models have been suggested as a method of choice to reduce false-positive or negative associations (Kang et al., 2008).

The objective of this study was to address the aspects of the GWAS for implementation of genomic selection for dairy traits: milk yield (MY), fat yield (FY), and protein yield (PY) in BSW. Single SNP analysis was performed to detect significant associations among SNPs and dairy traits across the genome using two models: 1) linear regression model considering one marker at a time, and 2) admixture components were included in the 1) model for accounting the population stratification.

2.3.2 Material and methods

2.3.2.1 Animals and phenotypic data

Phenotypic, pedigree, and genotypic data were obtained for 191 BSW bulls born between the years 1990 and 2007 provided by the by the Slovenian Agricultural Institute (phenotypic and pedigree data) and Biotechnical Faculty, Animal Science Department (genotypic data). Phenotypes considered were milk yield (MY), fat yield (FY), and protein yield (PY). These phenotypes were obtained from the routine national genetic evaluations (Potočnik, 1999) and were available in the form of estimated breeding values (EBVs) with the corresponding reliabilities for the BSW bulls. Altogether, daughter yield deviations (DYDs) of 182 progeny tested and genotyped BSW bulls born between 1990 and 2007 were included in the analysis (Table 9). These bulls predominantly originated from Slovenia. The exceptions were nine bulls imported to Slovenia as live animals from Austria and Germany. These bulls were used only in Slovenia.

Table 9: Descriptive statistics of the phenotypic values (estimated breeding values - EBVs) for dairy traits with the corresponding reliabilities (R^2) in the Brown Swiss breed
Tabela 9: Opisna statistika fenotipskih vrednosti (ocenjene plemenske vrednosti - OPV) za lastnosti mlečnosti z ustreznimi zanesljivostmi (R^2) pri rjavi pasmi

Trait	Bulls	Mean		Std		Minimum		Maximum	
		DYD	R^2 (%)	DYD	R^2 (%)	DYD	R^2 (%)	DYD	R^2 (%)
MY	182	0.32	94	0.76	4	-1.75	56	2.51	99
FY	182	0.01	93	0.03	4	-0.07	52	0.09	99
PY	182	0.01	93	0.03	5	-0.07	55	0.07	99

2.3.2.2 Marker genotypes and quality control

The DNA was extracted from the semen samples and 191 BSW bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1) comprising of 54,001 SNP markers. Genotypes of the bulls were scored as 0 (BB), 1 (AB) and 2 (AA). Several quality controls of genotypic data were applied to investigate the SNPs integrity and usefulness. These controls included genotyping call rate (CR) by SNP and animal, minor allele

frequency (MAF), departure from Hardy–Weinberg equilibrium (HWE), and parent–progeny conflicts (PPC). The number of SNPs per chromosome was also checked.

The CR was the first step in data quality checking. CR is the proportion of genotypes that are successfully identified. It is calculated across all SNPs in the DNA sample of genotyped bulls. Allele frequency was calculated for each SNP to detect MAF as the next criterion in data filtering. It is related to allele with the lowest frequency in the particular locus. Many of the genotyped SNPs could be monomorphic, i.e. they do not have two genotype variants but just one. The filter criterion is related to sample size and varies from 1 to 5%. The general statement of the HWE is that the genotype and allele frequencies in a large, randomly mating population with no selection, mutation or migration remain constant over generations (Falconer and Mackay, 1996). Deviation from the HWE has been widely used for detecting genotype errors (Hosking et al. 2004). However, besides genotype error, HWE could also be caused by selection, inbreeding, assortative mating, and stratification/admixture. Deviation from HWE was tested for every SNP using the chi-square statistic with two degrees of freedom (8).

$$\chi 2 = \sum_{i=1}^{n} \frac{(\text{observed frequency} - \text{expected frequency}_i)^2}{\text{expected frequency}_i} \quad \ldots (8)$$

The SNP genotypes for each bull were compared with the SNP genotypes of its sire (12 pairs) to determine the number of opposite homozygous genotypes between the parent and progeny (e.g., AA vs. BB). The mean of PPC was low (<0.0001%) and below the threshold (<200 discrepancies) set for the data exclusion. The number of SNP markers or bulls excluded due to quality control is shown in Table 10.

Table 10: The number of SNP markers or bulls excluded due to quality control criteria
Tabela 10: Število označevalcev SNP ali bikov izključenih zaradi kontrole kakovosti

Criteria	No of SNPs
Call rate ≥ 90% by SNP marker	5,975
Call rate ≥ 80% by bull	8
Minor allele frequency < 5%	10,619
Departure from the HWE ($P < 10^{-5}$)	538
Unmapped and Chr X SNP	2,419
In the further analysis	34,450

The SAS statistical package (SAS Institute, 2011) was used to make edits leaving a final set of 34,450 SNPs placed on 29 *Bos taurus* autosomes (BTA) for 182 bulls that are progeny tested. After the quality control any missing genotypes were imputed using the

gpig program (Strandén, 2010) using the linear approximation method (Gengler et al., 2007).

2.3.2.3 Statistical analyses

Two types of analyses were conducted to detect SNP association signals. Initially, a single trait genome-wide association study was performed to estimate genome-wide associations between phenotypes and SNPs for each of the dairy traits. Then, admixture analysis was performed on the data to adjust for the effect of population stratification. Admixture software (Alexander et al., 2009), was used to determine the proportion of genotypes originated from different clusters in the population (*K*). Bulls in BSW data set were divided in four clusters coming from different ancestral populations. Separation was determined using admixture cross-validation procedure, which gives standard error of the cross validation estimate.

GWAS between phenotypes and SNP markers for each dairy trait were estimated using the following univariate linear mixed model (9) to estimate association for one marker at a time:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e} \quad \text{... (9)}$$

where $\boldsymbol{y}$ is a vector of phenotype values (DYDs), $\boldsymbol{b}$ is the vector that holds the intercept and allele substitution effect for the SNP marker of interest, while $\boldsymbol{X}$ is the incidence matrix linking $\boldsymbol{y}$ respectively with $\boldsymbol{b}$, and $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma_e^2)$ is a vector of residuals. Then, admixture components adjusting for population stratification were included in the model as covariates. R software package (R Development Core Team, 2011) was used to test associations between SNPs and dairy traits.

In the GWAS, significance threshold was calculated using Bonferroni correction for multiple testing. Correction based on *p-value* of 0.05 resulted in Bonferroni threshold of 1.45×10^{-6}. SNPs with the smaller *p-values* than the Bonferroni threshold are considered as significant. The results of GWAS were plotted using Manhattan plot based on the obtained *p*-value of each SNP. The *p-values* of the association test were transformed to –log10 (*p*-values) for each SNP versus its chromosomal locations (Ding et al., 2009). It shows locations of the statistically significant SNPs across the chromosomes (horizontal axis) associated with their effect (vertical axis) on the traits of interest. The higher position of the dots, the stronger the genetic association.

2.3.3 Results and discussion

Genotyped bulls with progeny were born between 1990 and 2007 (Figure 4), with four to 17 bulls per year. The average number of phenotyped daughters per bull ranged between 200 and 400 for bulls born before the year 2000. In the later years, the average number of phenotyped daughters decreased.

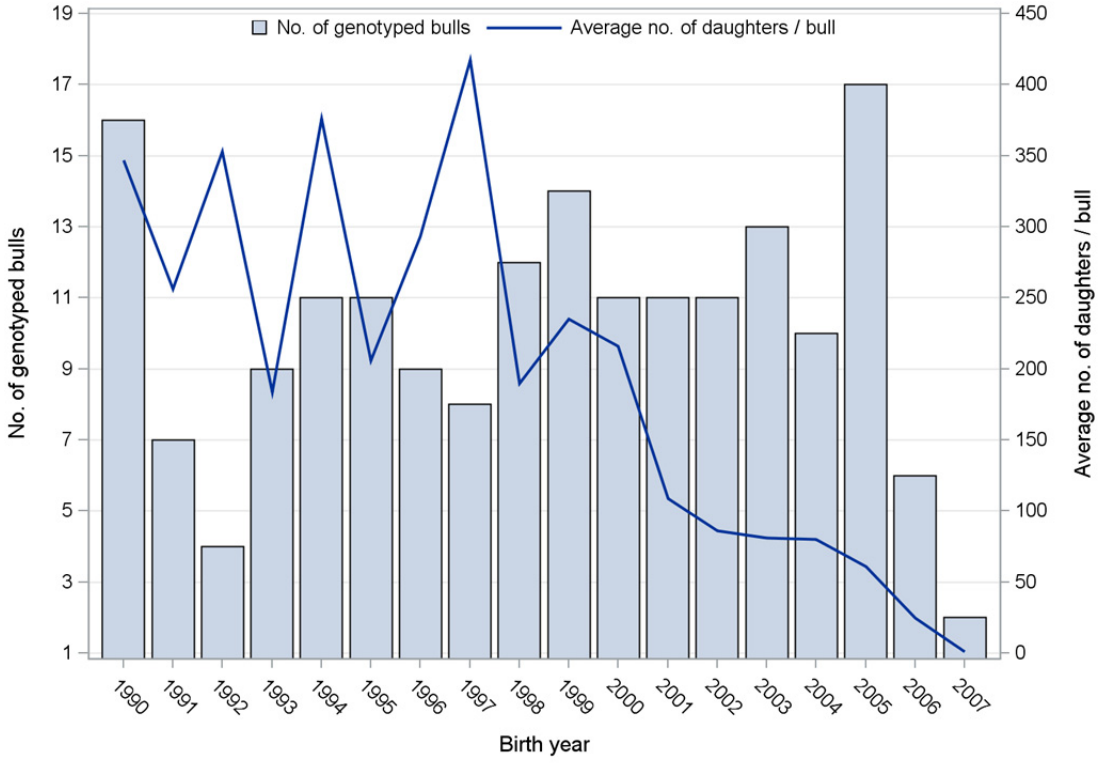


Figure 4: Number of genotyped bulls born per year (bars) and average number of daughters per genotyped bull
Slika 4: Število genotipiziranih bikov rojenih po letih (stolpci) in povprečno število hčera na genotipiziranega bika

2.3.3.1 Results of quality criteria

Number of SNPs excluded from analysis differed among filtering criteria. SNPs were included in analysis if CR was higher than 90% by SNP. Altogether, 5,975 SNPs failed this criterion and were excluded from the analysis (Figure 5). Eight bulls were excluded from the analysis because more than 20% of SNPs per animal missing (Figure 6). CR criteria differ among species and populations. The SNPs with CR lower than 95% and individuals with CR lower than 85% were excluded from the analysis of Danish and Swedish Holstein

bulls (Sahana et al., 2011). The threshold of SNPs CR exclusion was 90% in Australian Holstein bulls (Hayes et al., 2010) and four breeds of pigs (Large White, Duroc, Pietrain and Landrace; Harlizius et al., 2011). Lower SNPs CR criterion (70%) was used for quality control in Holstein bulls in Israel (Weller et al., 2011). However, this criterion is higher (97%) in the studies of human genome (The Wellcome trust case control consortium, 2007).

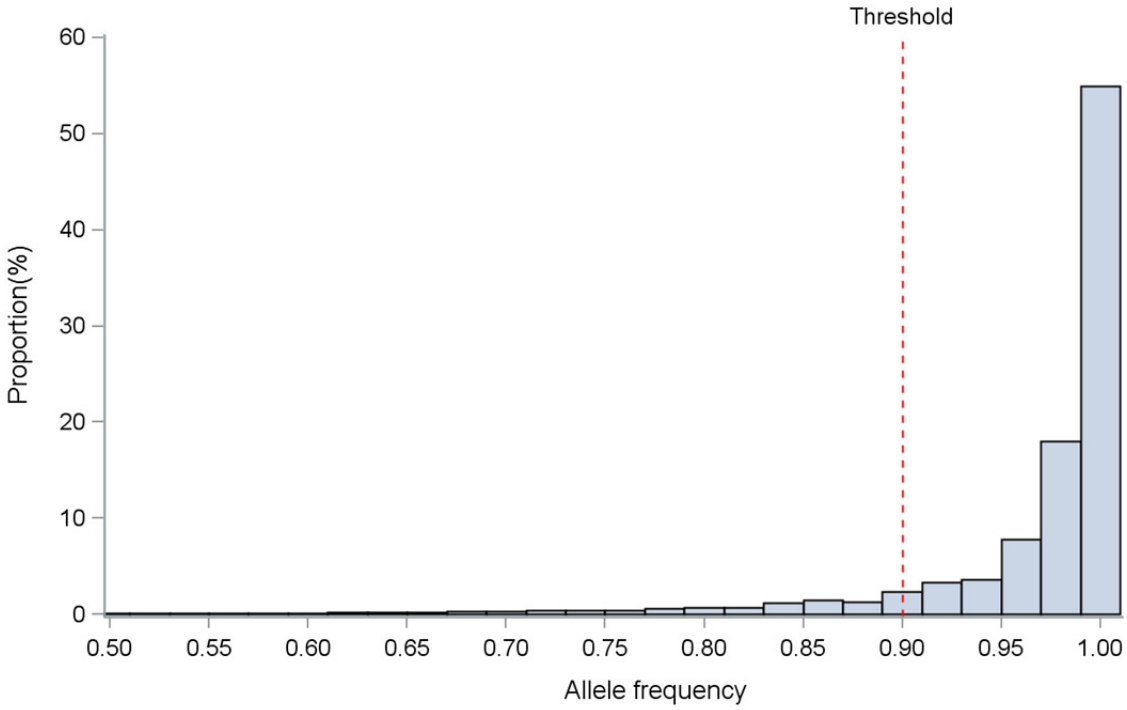


Figure 5: Genotype frequency per SNP
Slika 5: Frekvenca genotipa po SNP označevalcu

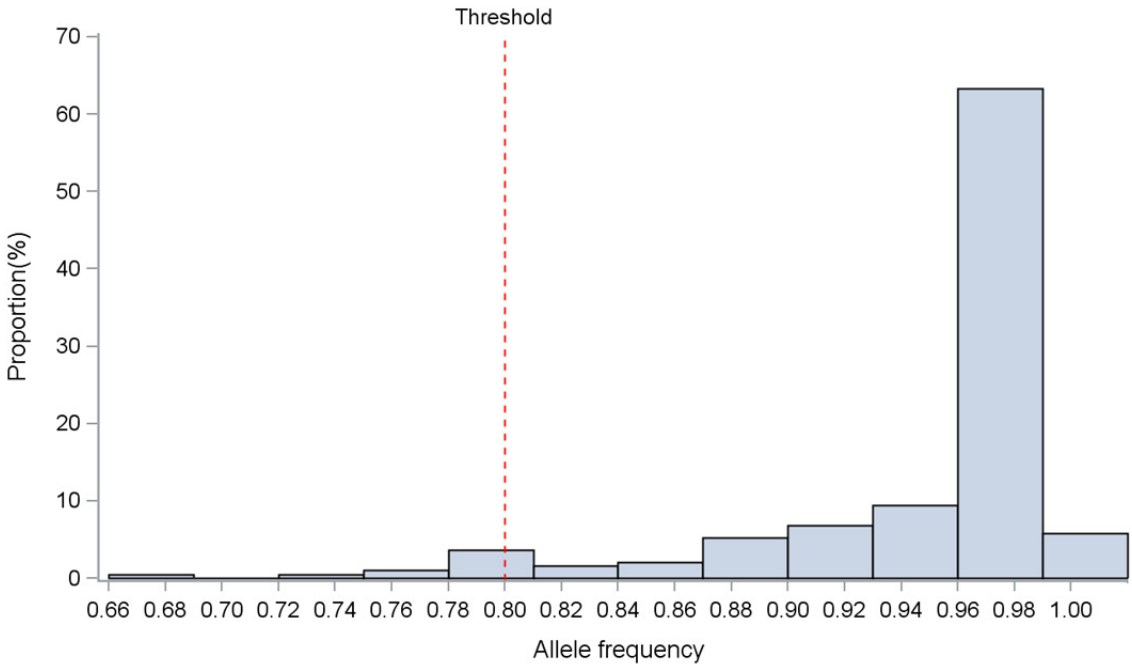


Figure 6: Genotype frequency per animal
Slika 6: Frekvenca genotipa po živali

SNPs with MAF lower than 5% (Figure 7) were excluded (10,619 cases or 25%) from the analysis. The same MAF was used in the study of Wiggans et al. (2009). In Australian data set analysed by Hayes et al. (2011), 2% of the SNPs were monomorphic. Even lower frequency (0.3%) for the minor allele was used by Harlizius et al. (2011).

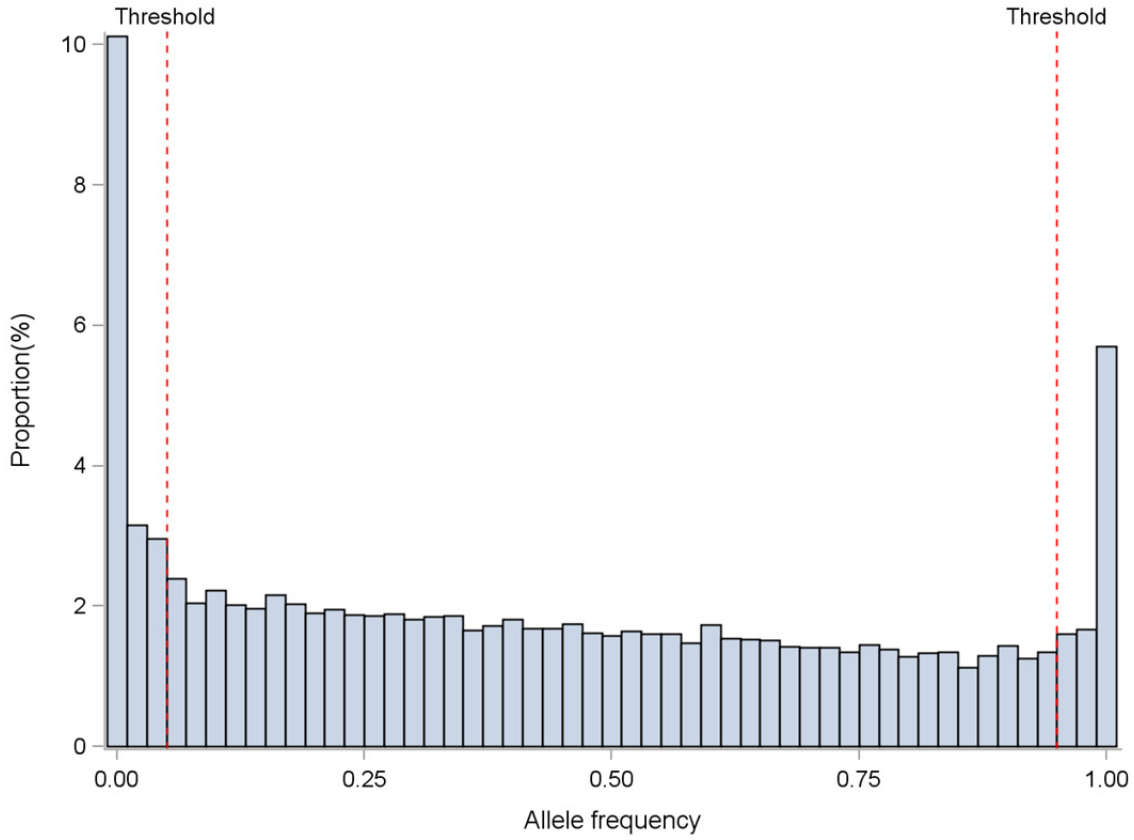


Figure 7: Allele frequency
Slika 7: Frekvenca alel

For each SNP, genotype frequency was calculated and denoted as P (AA) and Q (BB) for homozygous and H (AB) for heterozygous genotype. Difference between observed genotype frequencies obtained from the data and expected frequency from the HWE was normally distributed (Figure 8, 9, and 10). The departure from the HWE was set at the threshold of $P < 10^{-5}$, and 538 SNPs were excluded from the analysis. The same threshold was considered for testing HWE in Australian Holstein bulls (Hayes et al, 2011). In the studies with the human genome (The Wellcome Trust Case Control Consortium, 2007), $P < 10^{-4}$ was used as threshold. Wiggans et al. (2009) applied the departure of the heterozygotes frequency from that expected under HWE. The SNPs were excluded if the difference between observed and expected genotype was > 0.15.

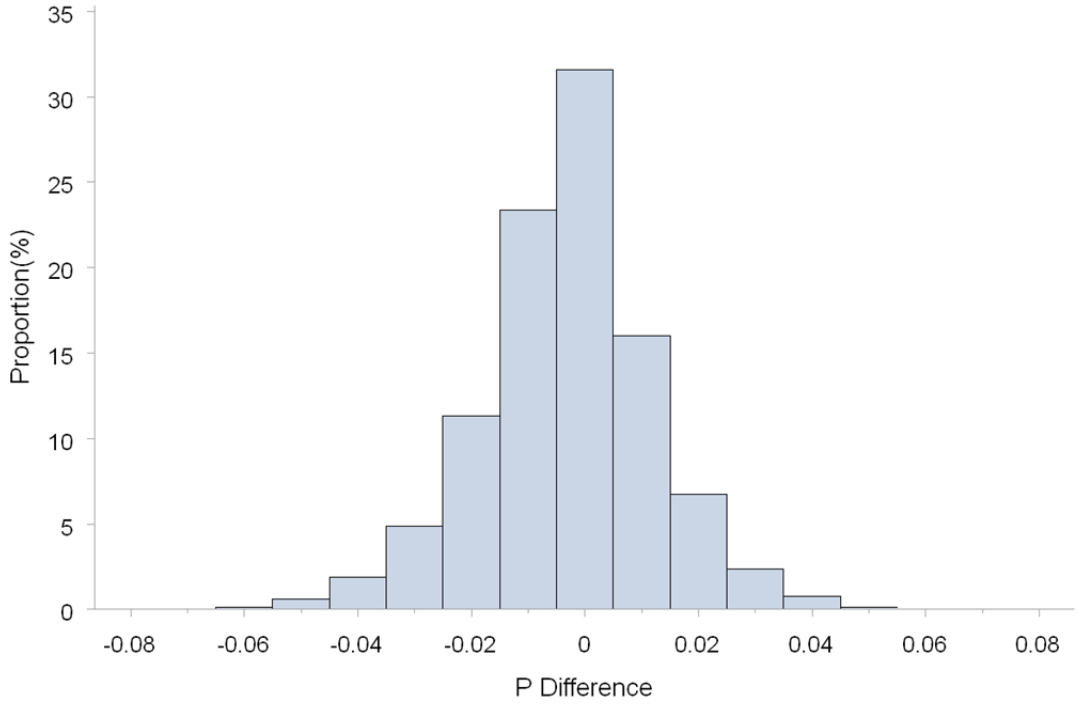


Figure 8: Differences between observed and expected genotype frequency for homozygous genotype AA (P difference)
Slika 8: Razlike med dejansko in pričakovano frekvenco genotipa za homozigotni genotip AA (P razlika)

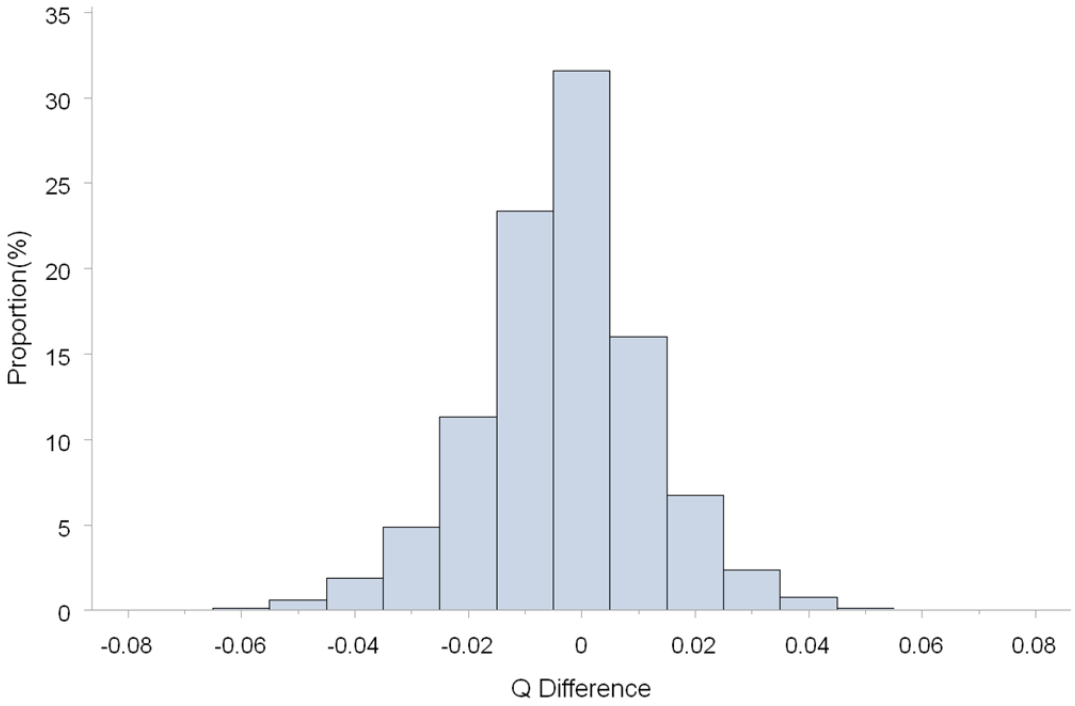


Figure 9: Differences between observed and expected genotype frequency for homozygous genotype BB (Q difference)
Slika 9: Razlike med dejansko in pričakovano frekvenco genotipa za homozigotni genotip BB (Q razlika)

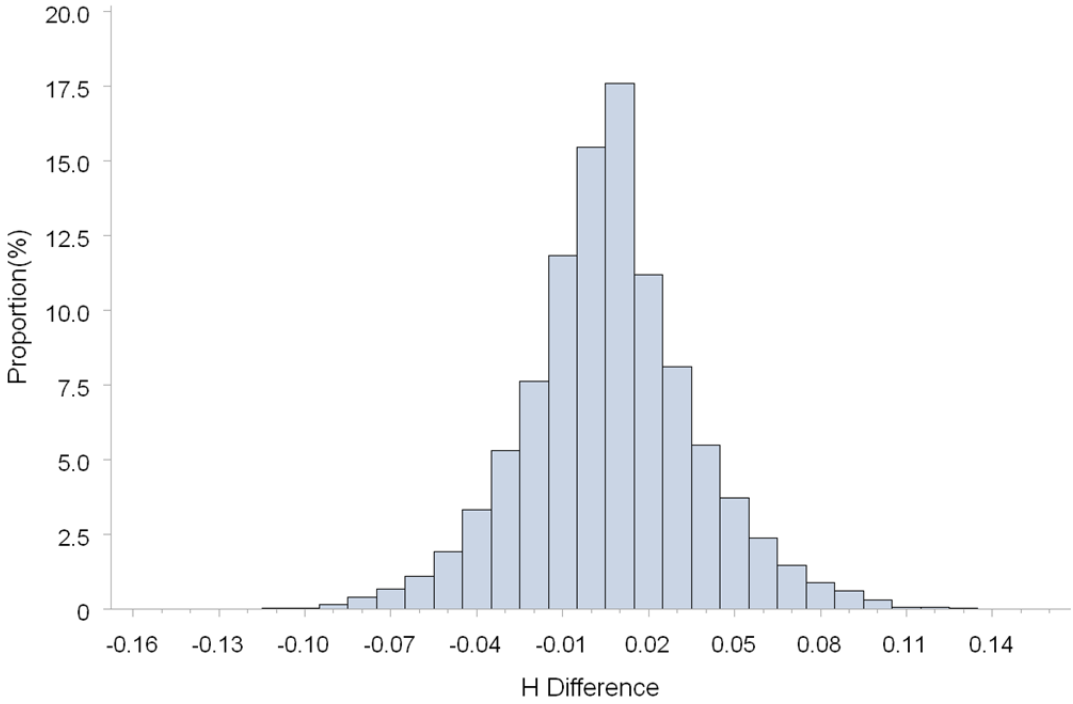


Figure 10: Differences between observed and expected genotype frequency for heterozygous genotype AB (H difference)
Slika 10: Razlike med dejansko in pričakovano frekvenco genotipa za heterozigotni genotip AB (H razlika)

Number of SNPs differed among BTA (Figure 11). The largest number of SNPs was mapped to BTA1 and their number gradually decreased towards BTA29. SNPs that could not be mapped to BTA (1,672) or SNPs mapped to X chromosome (747) were excluded from the analysis.

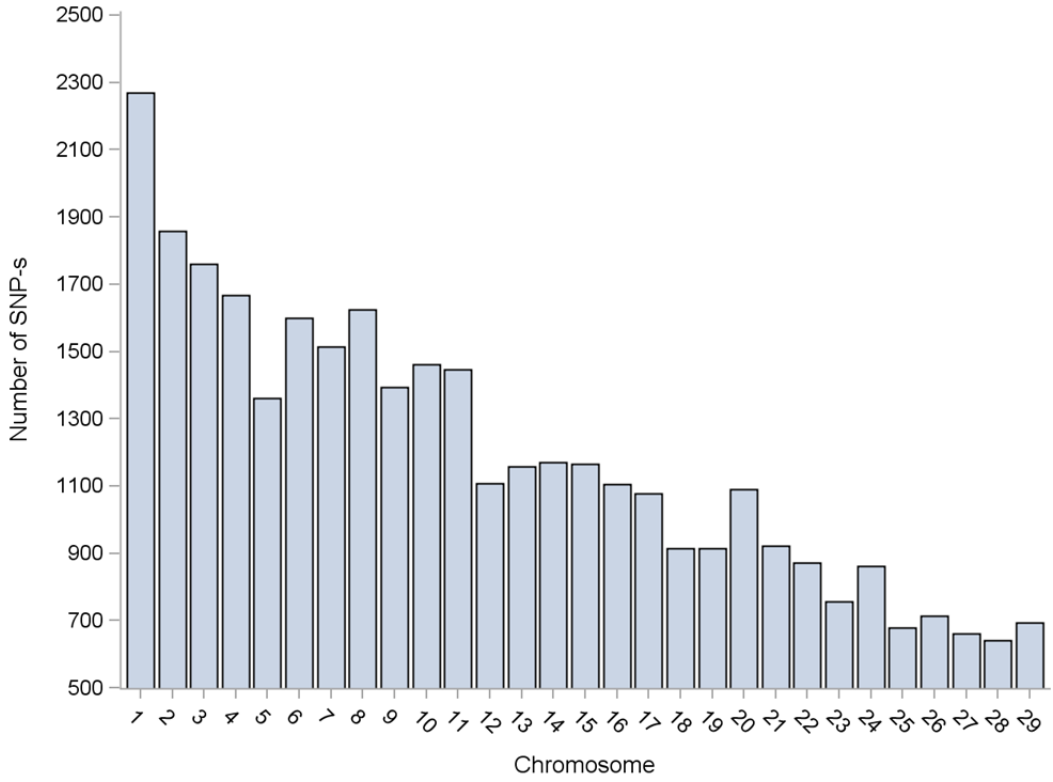


Figure 11: Number of SNPs per chromosome
Slika 11: Število SNP označevalcev po kromosomu

2.3.3.2 Correction for population stratification

The proportion of each subpopulation was estimated as unsupervised analysis in admixture program assuming from two to ten populations (K). Four ancestral populations were accepted as the best match describing population stratification due to lowest cross-validation error in K=4. Different genotypes proportions per animal divided in four clusters (Figure 12) were added as a fixed effect affecting SNP influence in GWAS analysis.

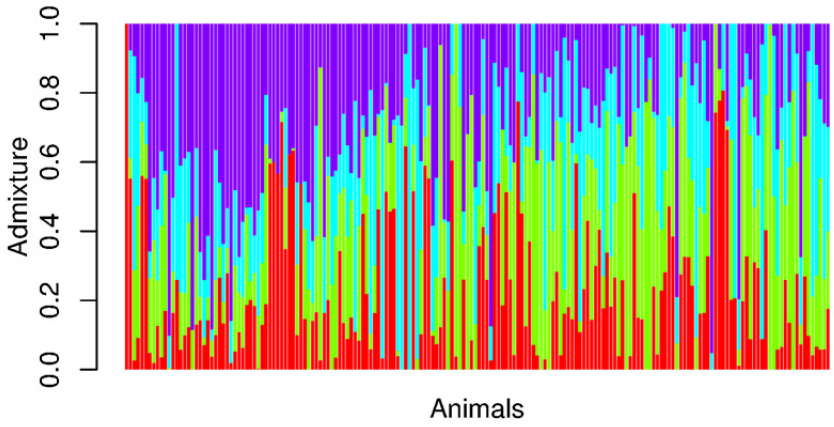


Figure 12: The proportion of subpopulations in each BSW bull estimated by ADMIXTURE analysis
Slika 12: Delež podpopulacij v vsakem biku BSW pasme ocenjeno z ADMIXTURE analizo

The distribution of observed distributions of -log(*p-value*) for each SNP was compared to the expected distribution under no-association (red line) in a quantile-quantile (Q-Q) plot for each dairy trait (Figures 13, 14, and 15). Q-Q plot of data without accounting population stratification (black line) for all analysed traits showed an early separation of the observed from the expected distribution (red line) indicating that the model which neglected family relationship was problematic. The fact that the black dots are above the diagonal axis shows an elevated amount of statistical significance for a huge number of variants, which is probably the result of population stratification. After correction, the data were well-behaved (green line) however for the majority of SNPs there was no more statistical significance.

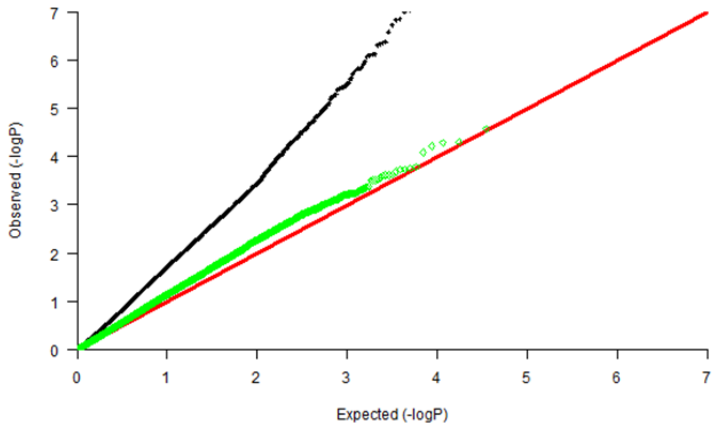


Figure 13: Q-Q plot for milk yield. Black dots are GWAS without stratification, green dots are GWAS with stratification added to analysis
Slika 13: Q-Q grafični prikaz za količino mleka. Črne pike predstavljajo GWAS brez stratifikacije, zelene pike predstavljajo GWAS z upoštevano stratifikacijo v analizi

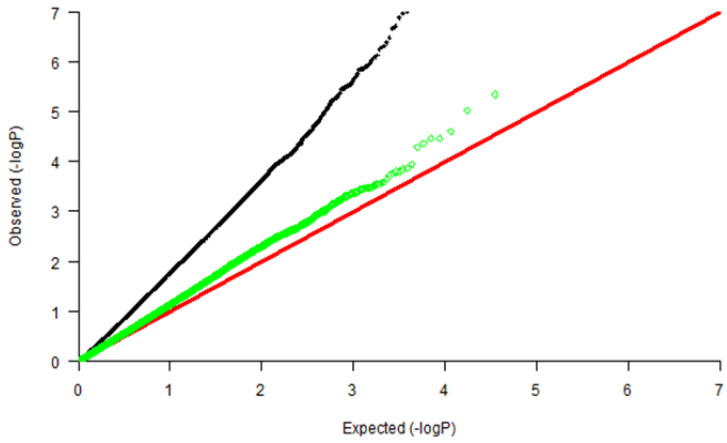


Figure 14: Q-Q plot for fat yield (PY). Black dots are GWAS without stratification, green dots are GWAS with stratification added to analysis
Slika 14: Q-Q grafični prikaz za količino maščobe. Črne pike predstavljajo GWAS brez stratifikacije, zelene pike predstavljajo GWAS z upoštevano stratifikacijo v analizi

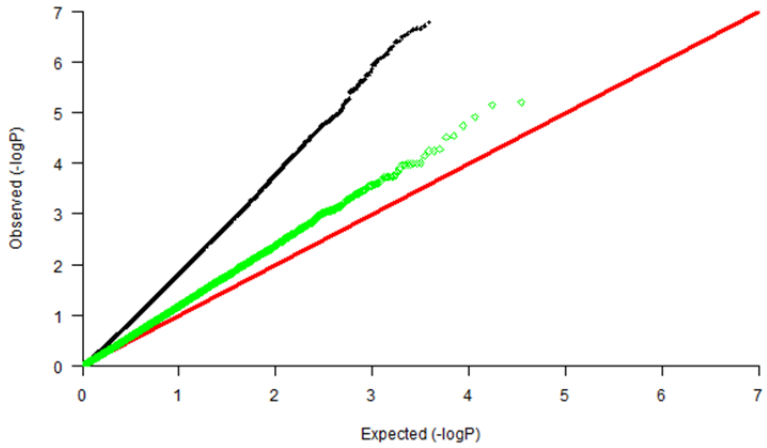


Figure 15: Q-Q plot for protein yield. Black dots are GWAS without stratification, green dots are GWAS with stratification added to analysis
Slika 15: Q-Q grafični prikaz za količino beljakovin. Črne pike predstavljajo GWAS brez stratifikacije, zelene pike predstavljajo GWAS z upoštevano stratifikacijo v analizi

2.3.3.3 Single SNP analysis without accounting for population stratification

In total, 52 SNPs with a significant effect on MY, FY, and PY were identified by single trait analysis. Among these, 11 were significant for all milk traits. The association was spread over 14 BTA. The estimated effects transformed to negative logarithm of *p*-values of all significant effects are shown in Table 6. The majority of the SNPs associated with

MY were on BTA21 followed by BTA9. The majority of genome-wide significant SNPs for FY and PY were also located on these BTA.

Table 11: Genome-wide significant SNPs for milk (MY), fat (FY), and protein yield (PY)
Tabela 11: Na (celem) genomu značilni označevalci SNP za količino mleka, maščobe in beljakovin

SNP Name	BTA	Position (bp)	-log(P-value)		
			MY	FY	PY
ARS-BFGL-NGS-34153	5	29,998,702	6.12		
ARS-BFGL-NGS-26592	5	10,803,982		6.39	
ARS-USMARC-201	5	10,846,184		6.64	
BTB-02050304	5	111,394,023			5.85
ARS-BFGL-NGS-1801	6	83,815,542	8.47	6.66	8.57
BTA-68275-no-rs	6	90,184,758	6.99		6.70
ARS-BFGL-NGS-28041	6	90,415,521		6.28	
Hapmap44452-BTA-22099	6	90,564,545		6.41	
ARS-BFGL-NGS-59916	7	22,514,886		6.07	
Hapmap38264-BTA-96587	7	101,536,840		6.15	
ARS-BFGL-NGS-115573	9	408,418	6.35	7.06	7.00
ARS-BFGL-NGS-79944	9	25,010	6.33		
BTA-103211-no-rs	9	608,084	7.26	7.38	7.65
BTB-00379127	9	4,238,169	6.72		6.87
BTB-01696182	9	4,183,032	5.91	6.78	6.06
BTB-01991280	9	4,091,763		6.65	5.83
BTB-01736853	9	33,283,957	6.10		
ARS-BFGL-NGS-111248	11	28,357,295	6.09	6.04	5.96
ARS-BFGL-NGS-27959	11	23,656,239		6.52	
Hapmap44191-BTA-109695	11	28,192,599		6.10	
ARS-BFGL-NGS-117303	13	57,425,302	6.37	7.74	6.30
BTA-32343-no-rs	13	36,510,636	6.10		6.12
ARS-BFGL-NGS-40188	13	41,245,438		6.23	
ARS-BFGL-NGS-101990	16	60,703,872			6.09
Hapmap23443-BTA-161049	17	7,913	6.58		
ARS-BFGL-NGS-114223	17	18,463,501		6.59	6.48
BTB-01315978	19	30,381,362	6.85	6.71	
Hapmap40740-BTA-45574	19	7,770,388		6.07	6.97
ARS-BFGL-NGS-101953	19	35,627,984			5.93
ARS-BFGL-BAC-33968	21	26,006,913	6.06		
ARS-BFGL-NGS-105488	21	41,098,410	6.83	7.70	7.51
ARS-BFGL-NGS-106159	21	56,399,834	8.36	8.34	8.19
BTA-01762-rs29012716	21	41,873,571	7.05	6.30	6.65
BTA-19227-no-rs	21	39,823,726	7.02	6.48	5.85
BTB-00818234	21	39,025,869	9.13	10.35	9.35
BTB-01718760	21	9,907,818	5.96	6.55	
Hapmap49357-BTA-87614	21	43,174,463	7.47	8.61	7.52
BTB-00818514	21	39,382,860		5.96	5.91
BTB-01303847	21	36,815,473		7.55	7.17
BTB-01383642	21	41,209,679		6.40	
BTB-01781411	21	39,254,989		6.16	

SNP Name	BTA	Position (bp)	-log(P-value)		
			MY	FY	PY
Hapmap43190-BTA-16332	21	44,201,493		6.02	
BTB-00818375	21	39,081,011			6.03
ARS-BFGL-NGS-56282	22	61,579,817		5.96	6.25
ARS-BFGL-NGS-1679	22	46,477,281			6.38
Hapmap47322-BTA-54813	22	51,965,282			5.85
ARS-BFGL-NGS-95117	23	43,028,783	5.85		7.13
ARS-BFGL-NGS-17146	23	52,583,723			5.87
Hapmap54312-rs29012588	23	44,574,227			6.28
ARS-BFGL-NGS-79347	27	43,658,924		5.86	
BTB-00969310	27	46,944,279		5.95	
BTB-01713276	28	37,494,090	6.30	6.11	

Manhattan plots for the GWAS of dairy traits were ploted based on chromosomes (x-axis) and the negative logarithm of the association *p-value* for every SNP (y-axis) (Figures 16, 17, and 18). SNP with the highest association with the trait has the smallest *p-value.* Therefore, the negative logarithm value will be the greatest and the scatter plot will be the highest peak in the Manhattan plot. The horizontal dashed line indicates the Bonferroni-corrected significance level.

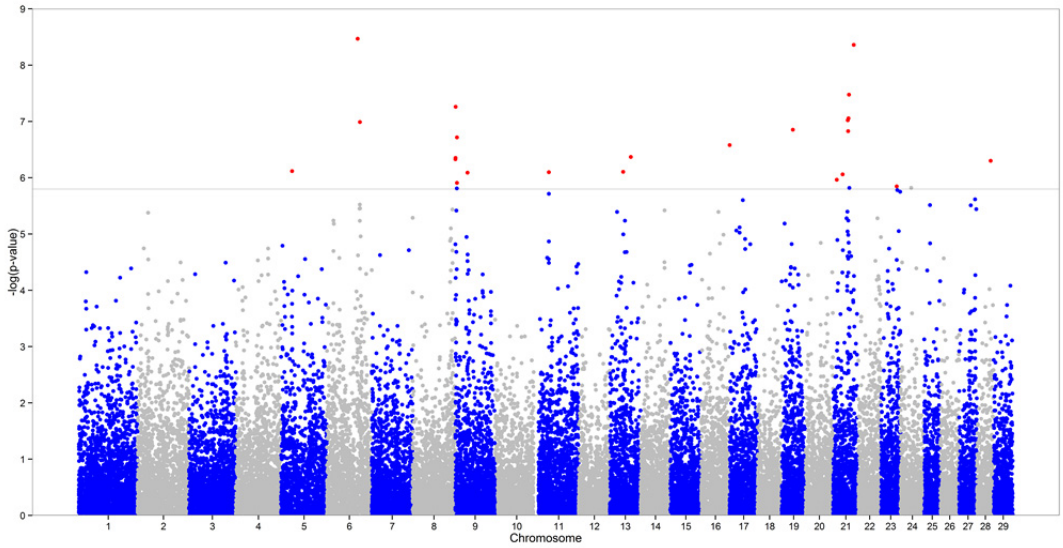


Figure 16: Manhattan plot for milk yield using single SNP analysis
Slika 16: Grafični prikaz Manhattan za količino mleka z uporabo analize na posameznem SNP označevalcu

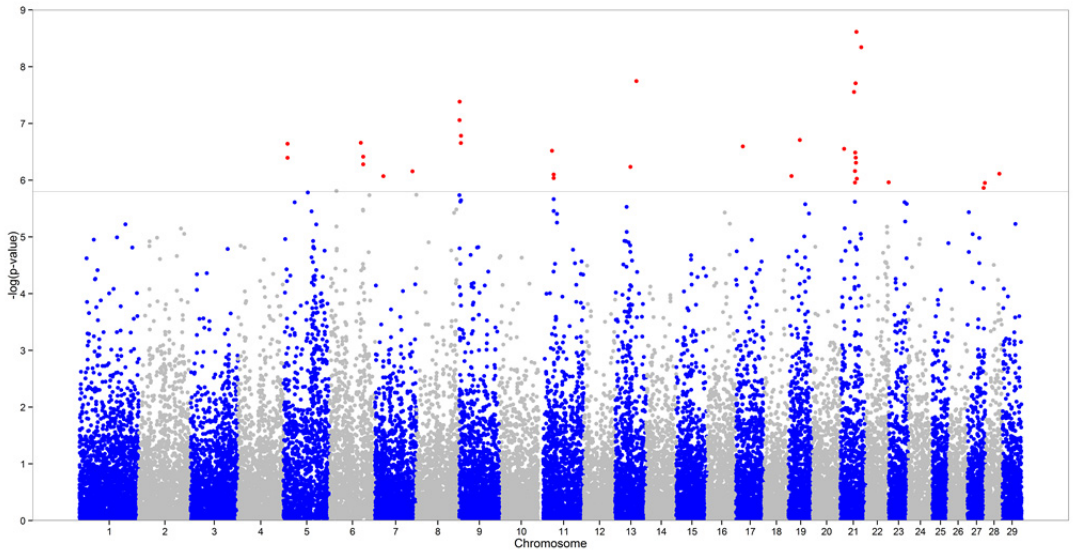


Figure 17: Manhattan plot for fat yield using single SNP analysis
Slika 17: Grafični prikaz Manhattan za količino maščobe z uporabo analize na posameznem SNP označevalcu

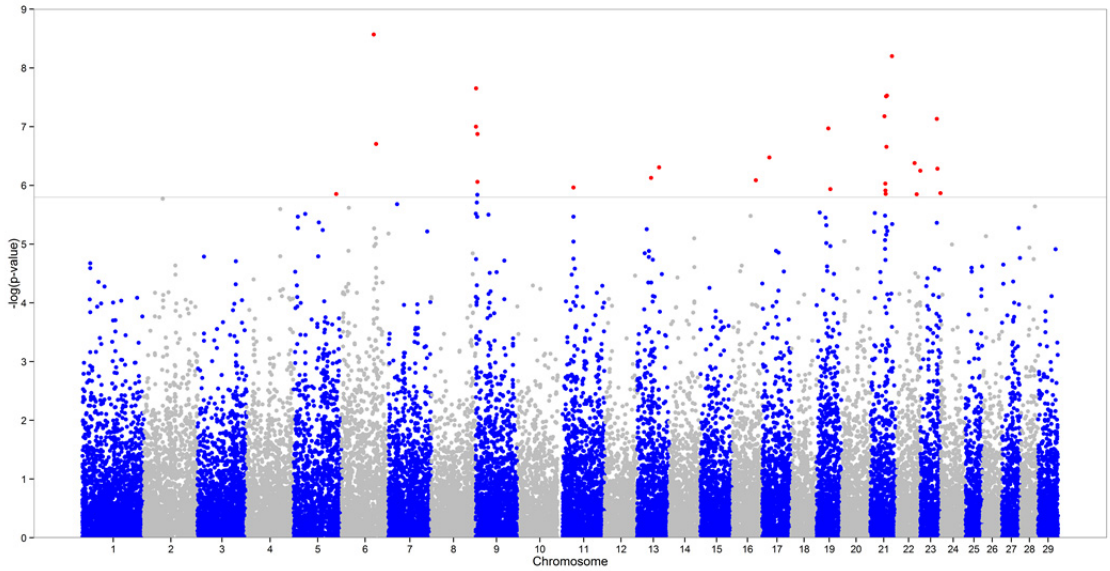


Figure 18: Manhattan plot for protein yield using single SNP analysis
Slika 18: Grafični prikaz Manhattan za količino beljakovin z uporabo analize na posameznem SNP označevalcu

2.3.3.4 Single SNP analysis with accounting for population stratification

Correction for population stratification based on the selected criteria was shown to be crucial. No significant associations between SNPs and dairy traits have been detected

(Figure 19) showing that the model without correction on the family relationship was not dealing properly with the analysed data set.

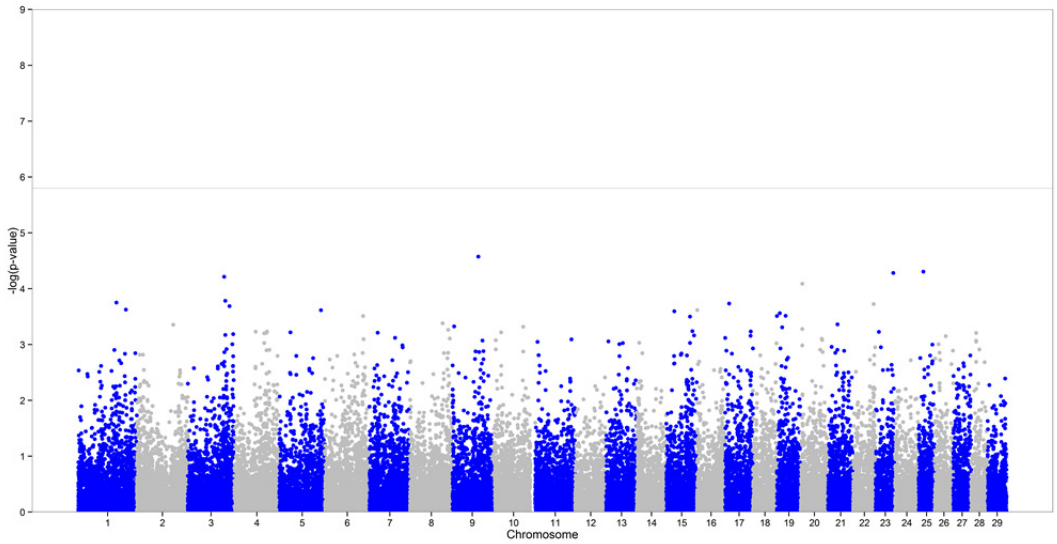


Figure 19: Manhattan plot for milk yield using single SNP analysis and accounting for population stratification
Slika 19: Grafični prikaz Manhattan za količino mleka z uporabo analize na posameznem SNP označevalcu upoštevajoč stratifikacijo populacije

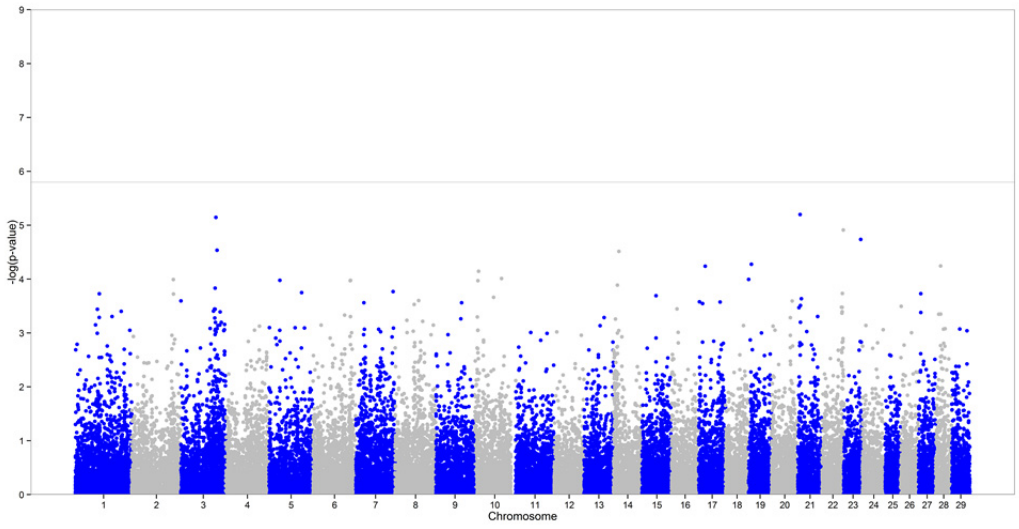


Figure 20: Manhattan plot for fat yield using single SNP analysis and accounting for population stratification
Slika 20: Grafični prikaz Manhattan za količino maščobe z uporabo analize na posameznem SNP označevalcu upoštevajoč stratifikacijo populacije

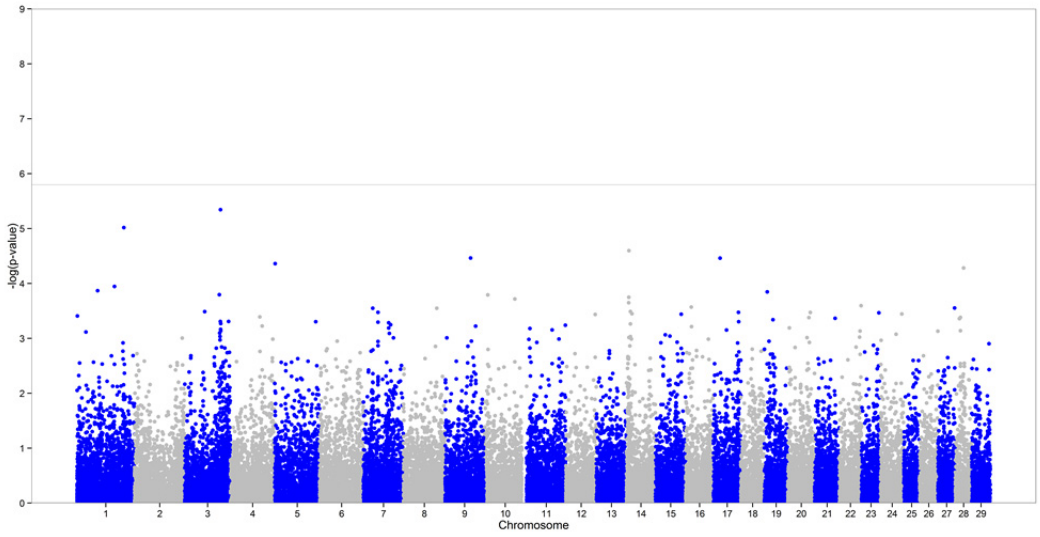


Figure 21: Manhattan plot for protein yield using single SNP analysis and accounting for population stratification
Slika 21: Grafični prikaz Manhattan za količino beljakovin z uporabo analize na posameznem SNP označevalcu upoštevajoč stratifikacijo populacije

Single trait genome-wide approach for dairy traits was presented to address the aspects of the GWAS for implementation of genomic selection in Slovenian BSW breed. The approach was used to estimate genome-wide associations between phenotypes and SNP markers based on data from national genetic evaluation for dairy traits. The effect of population stratification was taken into account since it was known from earlier publications (Guo et al., 2012; Maxa et al., 2012) that BSW breed is dealing with population stratification which has an impact on the results of GWAS. In the mentioned studies, principal component analysis was used to adjust the effect of population stratification. In this study, after correction for population stratification, the data were well-behaved (see Figure 13, green line) meaning that SNPs were no more statistical significant than would be randomly expected. That was observed for all dairy traits. One of possible reason of no significant associations was the small number of bulls included in the analysis. However, higher number of bulls (554) was used in the analysis of German Braunvieh (Maxa et al., 2012). Altogether, five SNPs with the significant effect on dairy traits were detected. Two SNPs which affected MY were placed on the BTA4, two SNPs associated FY were on the BTA14 and BTA23, while significant SNP with an effect on the fat content was on the BTA1. The largest data set (7,038 bulls) considering BSW breed (Guo et al., 2012) was based on large scale data from the international genetic evaluation. The total number of significant SNPs was 16 that affected dairy traits, with the strongest signal on the BTA25 (Guo et al., 2012). SNP associations for MY and PY were located on the same chromosome likewise in Finnish Ayrshire (Viitala et al., 2003) and Holstein (Harder et al., 2006) breed. GWAS in the dairy cattle mainly focused on the Holstein breed

due to worldwide spreading (Daetwyler et al., 2008; Pryce et al., 2010). For MY, significant SNPs were located on BTA8, BTA9, BTA10, BTA11, BTA13, BTA25, and BTA29 (Hayes et al., 2009b; Bolormaa et al., 2010; Jiang et al., 2010; Mai et al., 2010). BTA14 has also been in focus since many studies reported DGAT1 as a major gene affecting milk traits (Grisart et al., 2002; Ashwell et al., 2004; Jiang et al., 2010; Mai et al., 2010). In this study, the association between SNPs and dairy traits were not detect in the DGAT1 region.

2.3.4 Conclusion

In the present study, the preliminary results from GWAS analysis were reported for dairy traits based on the small population of genotyped bulls of Slovenian BSW breed. Altogether, 52 SNPs with a significant effect on MY, FY, and PY were identified using single SNP analysis. After adjusting for population stratification, no significant associations between SNPs and dairy traits have been detected. Based on the analysis that considered aforementioned criteria we could conclude that population stratification exists in the analysed data set and the model without correction on data were not dealing properly with the analysed data set. Further improvements should be made on the enlargement of the number of genotyped bulls in order to detect association signals and the identification of genes associated with dairy traits. Results from GWAS will facilitate the understanding of the genetic architecture of complex traits.

2.4 GENOME PARTITIONING OF GENETIC COVARIANCE BETWEEN DAIRY AND BEEF PRODUCTION TRAITS IN DUAL PURPOSE POPULATIONS

2.4.1 Introduction

Some cattle breeds are specialized either for dairy (e.g., Holstein Friesian, Ayrshire, and Jersey) or beef production (e.g., Angus, Charolais, and Limousin). In contrast, dual-purpose breeds such as Fleckvieh have breeding emphasis on both dairy and beef performances as well as on other traits important also in specialized breeds, e.g., fertility, longevity, etc. Efficiency of such multi-trait selection depends on the sign, size, and architecture of genetic covariation between the traits of interest. In particular the major limitation for the improvement of a multi-trait objective is posed by large genetic covariances/correlations with the unfavourable sign. An example of such a relationship studied widely in cattle is between milk yield and fertility that has low to medium negative genetic correlation, -0.12 to -0.41 (e.g., Berger et al., 1981; Oltenascu et al., 1991; Pryce et al., König et al., 2008). Most estimates of genetic correlations between milk yield and growth or carcass conformation are close to zero or slightly positive, ranging from 0.02 to 0.29 (Mason, 1964; Langlet, 1965; Bar-Anan, 1965; Calo et al., 1973; Pirchner, 1984; Liinamo et al., 2001). These low correlations suggest that dual-purpose selection for both dairy and beef performance is possible, though some efficiency is lost due to a broader selection objective in comparison to the specialized dairy or beef breeds. Most of the estimates of genetic covariance between dairy and beef performance are based on out-dated and/or small-scale experiments and should be re-estimated with more recent data collected in larger numbers.

However, breeders are not interested only in the total genetic covariance but also in the architecture of covariance along the genome. Such knowledge would empower breeders with the ability to construct optimized selection objectives that would take into account favourable, neutral, and unfavourable genome-wide covariances, to search for recombined haplotypes in regions with unfavourably linked alleles, and to optimize matings (Snelling et al., 2012). Developments in molecular genetics have provided a rich set of genome-wide markers and with that an opportunity to dissect total genetic parameters such as additive genetic variance into its genome-wide components (Goddard and Hayes, 2009; Bush and Moore, 2012; Eggen, 2012; Montaldo et al., 2012). While there are numerous single-trait genome-wide association (GWA) studies reporting significant associations (e.g., Pryce et al., 2010; Mai et al., 2010; Bolormaa et al., 2011; Murdoch et al., 2011; Becker et al., 2013; Wang et al., 2014), analysis of covariation between two or more traits due to either linkage or pleiotropy is scarce. Some single-trait GWA studies have found antagonistic associations by selecting associations for one trait and looking at associations with other traits. For example, in Norwegian Red cattle a significant association on the BTA12 was fine mapped to a locus having an antagonistic effect on non-return rate and dairy traits (Olsen et al., 2011). Antagonistic associations have also been identified for calving ease

and postnatal growth traits on the BTA14 and BTA21 in the German Fleckvieh breed (Pausch et al., 2011). An example of a multi-trait approach looking into the covariation at the genome, chromosome, and chromosome region level has been reported among the mastitis resistance related traits in the Danish Holstein breed (Sørensen et al., 2012); where the amount of covariance due to markers on a chromosome was primarily a function of the chromosome size though some chromosomes explained more covariance indicating harboring of closely linked or pleiotropic genes. Similarly to the GWA studies, genomic selection is mainly based on single trait models though several studies showed advantage of the multi-trait approach (Calus and Veerkamp, 2011; Jia and Jannink, 2012; Guo et al., 2014; Tsuruta et al., 2011). Further understanding of genetic covariation at the genome, chromosome, and chromosome region level would provide new insight into the biological processes between dairy and beef traits, which could in turn be utilized through genomic selection in applied breeding.

There were three objectives of this study. First, to estimate parameters describing total genetic covariation between the dairy and beef production traits in two well-known dual-purpose populations using comprehensive large scale routinely collected data from progeny testing. Second, to analyse genome-wide covariation between the pairs of dairy and beef traits in order to describe its architecture in the terms of sign, size, and distribution along the genome. Third, to utilize bioinformatics resources to enrich the obtained results with the existing knowledge and to propose potential pleiotropic loci (pleiotropici) that underlie a part of genetic covariation of analysed traits.

2.4.2 Methods

The study involved two populations of dual-purpose cattle with available phenotype data for dairy and beef traits and genome-wide marker genotype data. These data were quality controlled and analyzed using different models to estimate total genetic covariation and genome-wide covariation between traits, which was further subject to bioinformatic analyses. Basic data handling and analyses were done in SAS (SAS Inst., Inc., Cary, NC, 2011) and R (R Development Core Team, 2013), while other programs were used for specific tasks as described below.

2.4.2.1 Animals and phenotypic data

Phenotypic, pedigree, and genotypic data were obtained for 4,131 German Fleckvieh (GFV) bulls born between the years 1975 and 2004 provided by the Bavarian State Research Centre for Agriculture (http://www.lfl.bayern.de) and 800 Italian Pezzata Rossa (IPR) bulls born between the years 1982 and 2007 provided by the Italian Simmental Breeders Association (ANAPRI; http://www.anapri.eu). Phenotypes considered were milk

yield (MY) and fat yield (FY) as dairy traits and net daily gain (NG) and carcass quality (CQ) as beef traits. In the GFV data the CQ trait was represented with the EUROP carcass conformation score, while in the IPR data the CQ trait was represented with muscling score that is very highly correlated with the EUROP carcass conformation score. Phenotypes values for these traits were available as summarized phenotypes with high heritability (reliability) for each bull; obtained from the routine national genetic evaluations (Interbull, 2014) in the form of progeny phenotype deviations for the GFV population (Emmerling et al., 2002; Schild et al., 2003) and deregressed estimates of breeding values (Garrick et al. 2009) for the IPR population. In both populations the procedures of preparing summarized phenotypes involved only univariate models to avoid the influence of prior genetic covariances. Descriptive statistics for the used phenotypes and the corresponding reliabilities are shown in Table 14, which shows differences in means due to different genetic bases set in the genetic evaluation of each population. In addition, difference in the definition of CQ trait between the populations is manifested by substantially different standard deviations. Only bulls with the phenotypic data for all traits were used in further analysis, which is the 4,122 bulls of the GFV population and the 592 bulls of the IPR population.

Table 12: Descriptive statistics of phenotypic values (PV) for dairy and beef traits with the corresponding reliabilities (R^2) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) bulls
Tabela 12: Opisna statistika za fenotipske vrednosti (PV) za lastnosti mlečnosti in mesnatosti z ustreznimi zanesljivostmi (R^2) pri bikih nemške lisaste (GFV) in italijanske lisaste (IPR) pasme

Population	Trait	Bulls	Mean		Std.		Minimum		Maximum	
			PV	R^2 (%)	PV	R^2 (%)	PV	R^2 (%)	PV	R^2 (%)
GFV[1]	MY	4,131	1601.9	94	549.2	4	-690.5	56	3390.6	99
	FY	4,131	57.5	93	20.0	4	-39.8	52	117.1	99
	NG	4,122	6.9	88	14.3	13	-71.0	8	82.1	99
	CQ	4,122	1.9	85	3.0	14	-22.1	5	17.5	99
IPR[2]	MY	592	-204.3	90	564.4	5	-1925.1	75	1581.3	99
	FY	592	-7.6	89	23.7	5	-86.6	75	60.0	99
	NG	681	109.7	55	11.0	7	76.7	34	144.5	84
	CQ	681	106.6	68	11.5	6	63.5	39	136.7	90

[1]Phenotypic values are progeny yield deviations; [2] Phenotypic values are deregressed estimates of breeding values; MY = milk yield; FY = fat yield; NG = net daily gain; CQ = carcass quality

2.4.2.2 Marker genotypes

The GFV bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1) comprising of 54,001 SNP markers (3,734 bulls) and Illumina BovineSNP50K BeadChip® (version 2, 54Kv2) comprising of 54,609 SNP markers (379 bulls), where 52,340 SNP markers were shared between the two versions. The IPR bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1). The initial number of SNP markers for the IPR bulls was 41,809 because the obtained data already passed some

of the quality controls. The number of SNP markers or bulls excluded due to quality control is shown in Table 13. In addition to the standard quality control, SNP markers that were in high linkage disequilibrium (LD) were pruned using the variance inflation factor (VIF) analysis option in PLINK (Purcell et al, 2007); which iteratively excludes the individual markers that have a VIF>2 with other markers. VIF is defined as $1/(1-\mathrm{R}^2)$, where R^2 is the multiple correlation coefficient between a SNP marker and other SNP markers in a window (Purcell et al, 2007). After the quality control and pruning any missing genotypes were imputed with the gpig program (Strandén, 2010) using the linear approximation method (Gengler et al., 2007).

Table 13: The number of SNP markers or bulls excluded due to quality control criteria and pruning in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) bulls
Tabela 13: Število SNP označevalcev ali bikov izključenih na podlagi meril za kontrolo kakovosti in izluščenja pri bikih nemške lisaste (GFV) in italijanske lisaste (IPR) pasme

Criteria/Breed	GFV	IPR
Call rate ≥ 90% by SNP marker	707	0
Call rate ≥ 90% by bull	17	0
Minor allele frequency < 0.01	7,880	1,943
Departure from the HWE ($P < 0.001$)	1,517	612
Unmapped and Chr X SNP	1,579	1,575
In the analysis before pruning	41,774	37,675
In the analysis after pruning	37,116	33,425

2.4.2.3 Statistical analyses

Two types of analyses were conducted to quantify genetic covariation between traits. Initially, a bivariate mixed linear model was used to estimate the total genetic covariances and correlations between the pairs of dairy and beef traits. Then, a single trait genome-wide association study was performed to estimate genome-wide associations between phenotypes and SNP markers for each of the dairy and beef traits and use these results to estimate genome-wide covariances and correlations between the pairs of dairy and beef traits as described in the following.

2.4.2.3.1 Total genetic covariation

The bivariate mixed linear model for estimating total genetic covariation for all the pairs of dairy and beef traits was defined as:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Z}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \end{bmatrix} \quad \ldots (10)$$

$$\begin{bmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \boldsymbol{G}_0 \otimes \boldsymbol{H}\right) \boldsymbol{G}_0 = \begin{bmatrix}\sigma^2_{a_1} & \sigma_{a_1,a_2}\\ \sigma_{a_1,a_2} & \sigma^2_{a_2}\end{bmatrix} \quad \ldots (11)$$

$$\begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \sim \boldsymbol{N}\left(\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}, \boldsymbol{R}_0 \otimes \boldsymbol{I}\right) \boldsymbol{R}_0 = \begin{bmatrix}\sigma^2_{e_1} & \sigma_{e_1,e_2}\\ \sigma_{e_1,e_2} & \sigma^2_{e_2}\end{bmatrix} \quad \ldots (12)$$

where $\boldsymbol{y_i}$ is a vector of phenotype values for the i-th trait, $\boldsymbol{b}_i$ is a vector of unknown parameters for fixed effects (only intercept), $\boldsymbol{a}_i$ is a vector of unknown additive genetic (polygenic) effects, $\boldsymbol{e}_i$ is a vector of environmental residuals, while $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices linking $\boldsymbol{y}$ respectively with $\boldsymbol{b}$ and $\boldsymbol{a}$, $\boldsymbol{H}$ is an relationship matrix between the phenotyped individuals based either on the pedigree (Colleau, 2002) or SNP markers (VanRaden, 2008), $\otimes$ is the direct Kronecker product, $\boldsymbol{G_0}$ is an additive genetic covariance matrix representing the total genetic covariation between the analyzed traits due to the whole genome, and $\boldsymbol{R_0}$ is an environmental residual covariance matrix between traits. Total genetic correlations were estimated from the estimated elements of $\boldsymbol{G_0}$. Heterogeneous variances of the summarized phenotype values were accommodated in the model with individual and trait specific weights in the residual covariance matrix. This model was fitted using the Residual Maximum Likelihood (REML) method as implemented in the WOMBAT software (Meyer, 2007).

2.4.2.3.2 Genome-wide associations

Genome-wide associations between phenotypes and SNP markers for each dairy or beef trait were estimated using the following univariate mixed linear model to estimate association for one marker at a time taking into account the additive genetic (polygenic) effects to account for background genetic variation and population structure:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \quad \ldots (13)$$

where $\boldsymbol{y}$ is a vector of phenotype values, $\boldsymbol{b}$ holds the intercept and allele substitution effect for the SNP marker of interest, $\boldsymbol{a} \sim N(\boldsymbol{0}, \boldsymbol{H}\sigma^2_a)$ is a vector of additive genetic (polygenic) effects, while $\boldsymbol{X}$ and $\boldsymbol{Z}$ are incidence matrices linking $\boldsymbol{y}$ respectively with $\boldsymbol{b}$ and $\boldsymbol{a}$, and $\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}\sigma^2_e)$ is a vector of residuals. As with the bivariate model (10-12) heterogeneous variances of the summarized phenotypes were accommodated in the model with individual and trait specific weights in the residual variance matrix. The matrix $\boldsymbol{H}$ is a relationship matrix between the phenotyped individuals based either on pedigree (Colleau, 2002) or SNP markers (VanRaden, 2008). This model was fitted using the Residual Maximum Likelihood (REML) method as implemented in the WOMBAT software (Meyer, 2007) using the SNP Snappy option to estimate allele substitution effects (one at a time) quickly (Meyer and Tier, 2012).

2.4.2.3.3 Genome-wide covariation

Estimated allele substitution effects from (13) were used as an input for the estimation of genome-wide covariances and correlations for the pairs of dairy and beef traits. Variance and covariance of additive genetic effects for the trait x and y at the i-th SNP marker are:

$$Var(a_{x,i}) = V_{a_{x,i}} = 2p_i q_i b_{x,i}^2 \quad \ldots (14)$$

$$Var(a_{y,i}) = V_{a_{y,i}} = 2p_i q_i b_{y,i}^2, \quad \ldots (15)$$

$$Cov(a_{x,i}, a_{y,i}) = C_{a_{x,y,i}} = 2p_i q_i b_{x,i} b_{y,i}, \quad \ldots (16)$$

where p_i and q_i are the allele frequencies and $b_{x,i}$ and $b_{y,i}$ are the allele substitution effects (Falconer and Mackay, 1996). The correlation at the i-th SNP marker is then by definition equal to:

$$Cor(a_{x,i}, a_{y,i}) = \frac{Cov(a_{x,i}, a_{y,i})}{\sqrt{Var(a_{x,i}) Var(a_{y,i})}} = \frac{b_{x,i} b_{y,i}}{|b_{x,i} b_{y,i}|}, \quad \ldots (17)$$

which is equal to 1 if covariance is positive, to 0 if covariance is zero, or to -1 if covariance is negative (Daniel Gianola, personal communication). Due to this phenomenon, further analyses of genome-wide covariation were performed only with covariances.

Assessing the significance of estimated genome-wide covariances (16) in the absence of a formal test for a genome-wide application that would take into account the population structure (13) is possible with the use of permutation testing (Churchill and Doerge, 1994). However, such an approach would be computationally very demanding as it would require running the model (13) with permuted data $n_p \times n_s$ times, where n_p is the number of permutations, say $n_p = 10{,}000$, and n_s is the number of SNP markers (Table 12). Instead of hunting for the significant hits, the focus was put on describing the architecture of genome-wide covariances by assessing the contribution of covariance at the i-th SNP marker in the distribution of all genome-wide covariances; which was measured with the ratio of squared covariance $C_{a_{x,y,i}}^2$ divided by the sum of all squared covariances $\sum_i C_{a_{x,y,i}}^2$. Distributions of these contributions were summarized according to the sign of covariances and size (described bellow) for all markers as well as for each chromosome by dividing the sum of squared covariances for all markers in a group (sign, size, or chromosome) by the sum of squared covariances for all markers. Under the assumption of a very large number of loci with small effects and some loci with moderate to large effects, an arbitrary threshold of 0.1% was chosen to group SNP markers by size in two groups with small or large contributions. Markers with large contribution represent pleiotropic loci (regions) that manifest pleiotropy either due to harboring pleiotropic genes or closely linked genes

affecting different traits. Such a marker is hereafter referred to as "pleiotropicus" after pleiotropic locus (plural pleiotropici). The other markers are hereafter referred to as polygenes. In the absence of formal significance testing, the results were "validated" by comparing the results from the analysis of complete data of the GFV population with the analysis of a random subset of 1,000 bulls from the GFV population and all the bulls from the IPR population.

2.4.2.3.4 Characterization of the most pronounced pleiotropici

To augment the results of this study with the already existing knowledge the estimated genome-wide covariances were first mapped to their genomic positions and plotted with the Manhattan plots and then the selected pleiotropici were overlaid with a) the reported quantitative trait loci (QTL) intervals for traits related to dairy and beef production and b) known genes. Based on the overlap of these three sources of information for both breeds together a set of candidate genes with pleiotropic effects for the analysed traits was proposed. The QTL intervals were downloaded from the Animal QTL Database, release 23 (http://www.animalgenome.org/cgi-bin/QTLdb/index/) (Hu et al., 2013) for the following twelve dairy and beef traits: milk and fat yield, body weight at birth, weaning, yearling, mature stage, and slaughter, average daily gain, pre-weaning average daily gain, carcass weight, fat thickness at the 12th rib, and longissimus muscle area. The gene information was downloaded from the Ensembl, release 75 (http://www.ensembl.org/index.html). In total, 19.994 protein-coding genes, and 3.825 non-coding genes were used. All three sources of information (pleiotropici, QTL intervals, and known genes) were graphically represented with the web-based interactive visualization tool Flash GViewer (http://gmod.org/wiki/Flashgviewer/) developed by the GMOD project.

2.4.3 Results

2.4.3.1 Total genetic covariation

Estimates of covariance components for MY and FY were in general smaller in the analysis using the genomic relationship matrix than with the pedigree relationship matrix, while they were comparable for NG and CQ (Annexes A1 and A2); suggesting that in the GFV and IPR populations the used SNP marker array allows capturing less genetic variance for dairy traits than for beef traits. Correspondingly, the residual variances and covariances differed between the pedigree and genomic models. However, the genetic correlations were very similar between the two models (Annex A2). With the genomic model, high positive total genetic correlations were estimated between MY and FY (0.72±0.01 in GFV and 0.81±0.04 in IPR) and between NG and CQ (0.53±0.03 in GFV and 0.51±0.09 in IPR); which is in agreement with the estimates obtained using the traditional pedigree model (e.g., Fürst et al., 2013; Roman and Wilcox, 2000; Welper and

Freeman, 1992). Total genetic correlations between MY and NG were slightly positive or close to zero (0.09±0.03 in GFV and 0.10±0.09 in IPR) and similarly between FY and NG (0.04±0.03 in GFV and -0.08±0.11 in IPR). Total genetic correlations between MY and CQ were slightly negative or close to zero (-0.11±0.03 in GFV and -0.25±0.09 in IPR) as well as between FY and CQ (-0.15±0.03 in GFV and -0.17±0.11 in IPR). These estimates are concordant with estimates from pedigree models (Mason, 1964; Langlet, 1965; Pirchner, 1984).

2.4.3.2 Distribution of genome-wide covariances by sign

The proportion of positive and negative genome-wide covariances determined total genetic correlations in both populations for all the pairs of dairy and beef traits (Table 14). The high positive total genetic correlations estimated between MY and FY (0.72±0.01 in GFV and 0.81±0.04 in IPR) were determined by the dominant proportion of positive genome-wide covariances (90.2% in GFV and 99.2% in IPR). The same pattern was also observed between MY and NG, MY and CQ, FY and NG, and FY and CQ, though with different proportions (Table 14). For example, the total genetic correlation between FY and NG was not different from zero (0.04±0.03 in GFV and -0.08±0.11 in IPR) and was roughly determined by the equal proportion of positive and negative genome-wide covariances, i.e., the contribution of positive genome-wide covariances was 53.0% in GFV and 49.1% in IPR. The exception to this pattern was observed between NG and CQ, where medium positive total genetic correlation (0.53±0.03 in GFV and 0.51±0.09 in IPR) was determined by the dominant proportion of positive genome-wide covariances (94.8% in GFV and 92.1% in IPR); though this exception can be explained by the different distribution of covariances (Table 14). Concordant results were observed on a random subset taken from the complete GFV dataset (Annex A3).

Table 14: Total genetic and genome-wide covariation for the pairs of dairy and beef traits* in the German Fleckvieh, Italian Pezzata Rossa population and in the German Fleckvieh population based on a random subset

Tabela 14: Skupna genetska in genomska kovariacija za pare lastnosti mlečnosti in mesnatosti* pri nemški in italijanski lisasti populaciji in v populaciji nemške lisaste pasme na osnovi naključnega podseta podatkov

Trait combination	MY : FY	MY : NG	MY : CQ	FY : NG	FY : CQ	NG : CQ
Total genetic covariation			**German Fleckvieh (n=4,105)**			
Correlation	0.72 ± 0.01	0.09 ± 0.03	-0.11 ± 0.03	0.04 ± 0.03	-0.15 ± 0.03	0.54 ± 0.03
Genome-wide covariation						
Positive versus negative, %	90.2 : 9.8	58.4 : 41.6	43.1 : 56.9	53.0 : 47.0	39.9 : 60.1	94.8 : 5.2
Large versus small, %	17.1 : 82.9	9.8 : 90.2	5.7 : 94.3	11.6 : 88.4	6.6 : 93.4	15.9 : 84.1
Positive large, % / no. markers	8.8 / 55	3.9 / 25	2.1 / 15	3.7 / 23	1.7 / 12	15.9 / 66
Positive small, % / no. markers	81.4 / 28,319	54.5 / 19,620	41.0 / 18,181	49.3 / 19,126	38.2 / 17,790	78.9 / 24,768
Negative small, % / no. markers	1.5 / 8,726	35.7 / 17,448	53.3 / 18,885	39.1 / 17,931	55.2 / 19,265	5.2 / 12,262
Negative large, % / no. markers	8.3 / 13	5.9 / 19	3.6 / 24	7.9 / 29	4.9 / 30	0.0 / 0
Total genetic covariation			**Italian Pezzata Rossa (n=511)**			
Correlation	0.81 ± 0.04	0.10 ± 0.09	-0.25 ± 0.09	-0.08 ± 0.11	-0.17 ± 0.10	0.51 ± 0.09
Genome-wide covariation						
Positive versus negative, %	99.2 : 0.8	52.3 : 46.7	44.9 : 55.1	49.1 : 50.9	50.6 : 49.4	92.1 : 7.9
Large versus small, %	9.8 : 90.2	5.3 : 94.7	4.7 : 95.3	4.3 : 95.7	3.9 : 96.1	8.6 : 91.4
Positive large, % / no. markers	9.8 / 64	2.1 / 15	1.3 / 9	1.6 / 11	1.7 / 12	8.6 / 60
Positive small, % / no. markers	89.4 / 25,664	51.2 / 17,159	43.6 / 16,525	47.5 / 16,676	48.9 / 16,979	83.5 / 21,623
Negative small, % / no. markers	0.8 / 7,697	43.5 / 16,231	51.7 / 16,869	48.2 / 16,738	47.2 / 16,419	7.9 / 11,802
Negative large, % / no. markers	0.0 0	3.2 / 20	3.4 / 22	2.7 / 19	2.2 / 15	0.0 / 0

*MY = milk yield; FY = fat yield; NG = net daily gain; CQ = carcass quality

2.4.3.3 Distribution of genome-wide covariance contributions by size and sign

Distribution of genome-wide covariance contributions by size indicated a number of SNP markers with large contributions (pleiotropici) though the majority of SNP markers had small (polygenic) contributions (Table 14). The total contribution (positive and negative) from the pleiotropici ranged from 5.7% between MY and CQ to 17.1% between MY and FY in the GFV population (Table 14). Somewhat lower total contribution of the pleiotropici was found in the IPR population; ranging from 3.9% between FY and CQ to 9.8% between MY and FY (Table 14). Summarizing covariance contributions by size and sign showed nearly symmetric distribution between the pairs of traits MY and NG, MY and CQ, FY and NG, and FY and CQ, while distributions were not symmetric between the pairs of traits MY and FY and NG and CQ (Table 14). Between MY and FY in the GFV population the total contribution of positive pleiotropici was 8.8% and of negative pleiotropici was 8.3%, while the contribution of small positive covariances was 81.4% and of small negative covariances was 1.5%. Between NG and CQ in the GFV population the total contribution of positive pleiotropici was 15.9% and of negative pleiotropici 0.0%, while the contribution of small positive covariances was 78.9% and of small negative covariances was 5.2% This difference in the distribution of contributions by size and sign between these two pairs of traits is likely the reason for different total genetic correlation. Similar results were observed also in the IPR population, though without the large negative covariances between MY and FY.

2.4.3.4 Distribution of genome-wide covariance contributions by size, sign, and chromosome

Distribution of genome-wide covariance contributions by size, sign, and chromosome indicated gross location of pleiotropici and polygenes across chromosomes and their contribution to the total genetic covariance between the analyzed traits (Figure 22). Most chromosomes had contributions in line with the overall distribution for each pair of traits (Table 14, Figure 22). Overall the contribution of each chromosome was associated with its physical length as measured by the regression of contributions of each chromosome onto its physical length (Annex A4) - all regressions had significant slope ($p<0.05$) either prior to or after the removal of the most deviating chromosome. Such a chromosome in the GFV population was the BTA14 (Table 14) with sizeable contribution of negative covariances due to 13 or 14 pleiotropici between MY and FY (8.3%), MY and NG (4.3%), and FY and NG (5.0%). The observed trend in this analysis is in agreement with the results reported for genetic architecture of single traits by Visscher et al. (2007) for human height. Similarly, Yang et al. (2011), who analyzed data on human height, body mass index (BMI), von Willebrand factor, and QT interval and found statistically highly significant relationship between the contribution of additive genetic variance explained by a chromosome and its genetic length. Pimentel et al. (2011) came to the same conclusion by

explaining a contribution of each chromosome to genetic variation of milk yield and composition traits in Holstein.

The BTA14 had similar contributions due to negative and positive covariances between MY and CQ and FY and CQ, which were mostly due to polygenes (Figure 22). Finally, the same chromosome had predominantly positive covariances between NG and CQ with roughly equal contributions due the 16 pleiotropici (2.9%) and polygenes (3.6%). The same pair of traits had even larger contribution of positive covariances on the BTA10 due to 23 pleiotropici (9.0%) and polygenes (5.4%). In addition to the BTA14 and BTA10 the BTA6 and BTA11 also showed deviations from the overall trend indicating that they might harbor regions with pleiotropic effects. These deviations were not as strongly expressed in the smaller set of bulls from the IPR population.

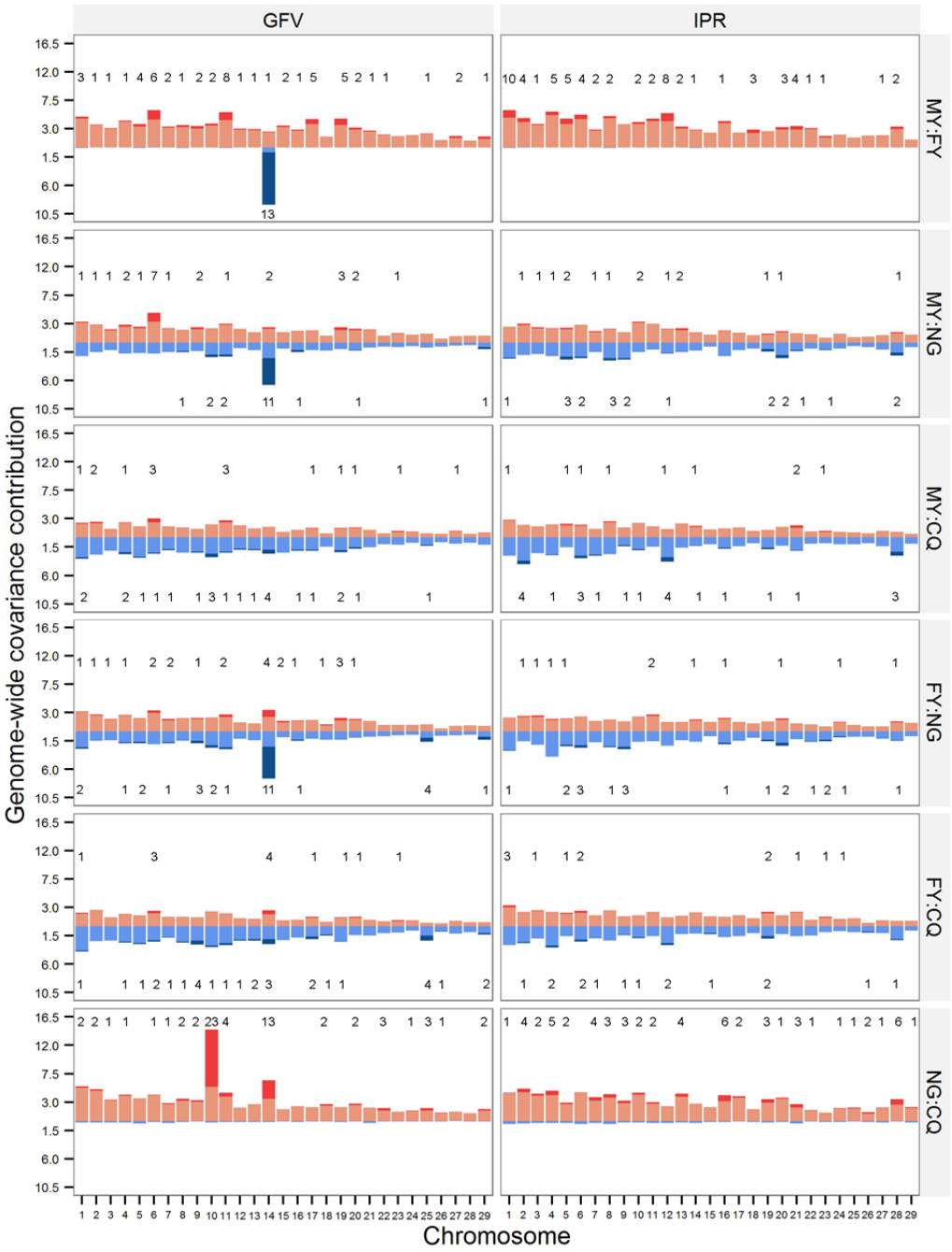


Figure 22: Distribution of genome-wide covariance contributions by size, sign, and chromosome for the pairs of dairy and beef traits in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Light bars indicate contribution due to polygenes, while dark bars stacked on light bars indicate contribution due to pleiotropici. Red color for positive and blue color for negative covariances. Numbers indicate the number of pleiotropici

Slika 22: Porazdelitev genomske kovariance po velikosti, predznaku in kromosomu za pare lastnosti mlečnosti in mesnatosti pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Svetli stolpci prikazujejo prispevek, ki ga povzroča poligenija, temni stolpci na vrhu svetlih pa prispevek pleiotropije. Rdeča barva označuje pozitivno, modra pa negativne kovariance. Številka prikazuje število pleiotropov

2.4.3.5 Characterization of the most pronounced pleiotropici

Individual estimates of all genome-wide covariances plotted with the Manhattan plots between all pairs of dairy and beef traits are shown in Annexes A5-A10. The most pronounced pleiotropic SNP markers with respect to their synergistic or antagonistic effects, selected based on the relative covariance contribution higher than 0.1%, were further analyzed for their functional and biological characteristics. We performed genomic overlap analysis comparing the genomic location of SNP markers, genes, and QTL for all dairy and beef traits. The genetic views and catalogue of genomic regions of overlapping SNP markers, genes, and QTL are available at http://www.integratomics-time.com/GSE. The SNP markers with a high number of overlaps could potentially be stronger molecular markers. The total of 164 SNP markers, 98 with positive and 66 with negative covariance, overlaps QTL for dairy and beef quality traits are shown in Annex A11. In addition, 59 of them also overlap with genes. The highest number of overlapping genomic elements was identified on BTA14 ranging from 16.51 to 20.88 Mb. This region contains five SNP markers having negative covariance (rs109146371, rs109968515, rs109421300, rs41256919, rs109350371), five protein-coding genes (*FOXH1*, *CYHR1*, *DGAT1*, *KIAA1875*, and *PLEC*) and 17 QTL for eight traits: milk and fat yield, milk and fat yield (daughter deviation), milk and fat yield (EBV), body weight at weaning, and fat thickness at the 12th rib, see Figure 23).

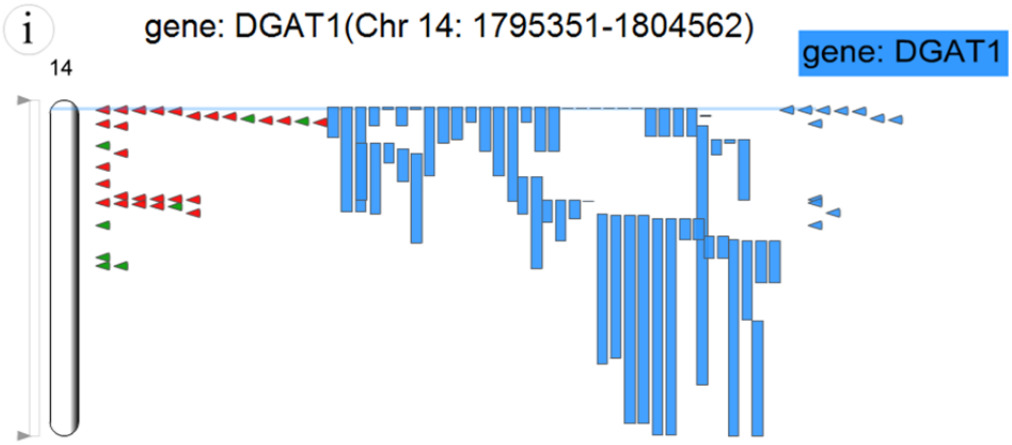


Figure 23: The close up of the BTA14 region presenting overlaps of SNPs, genes, and QTL. Blue line represent QTL, blue triangles represent genes. Red triangles represent position of SNPs having negative effect between dairy and beef traits. Green triangles represent position of SNPs having positive effect between dairy and beef traits

Slika 23: Pregled prekrivanja SNP označevalcev, genov, and QTL-ov regiji BTA14. Modra črta predstavlja QTL, modri trikotniki gene. Rdeči trikotniki predstavljajo pozicijo SNP označevalcev z negativno povezavo med lastnostmi mlečnosti in mesnatosti. Zeleni trikotniki predstavljajo pozicijo SNP označevalcev s pozitivno povezavo med lastnostmi mlečnosti in mesnatosti

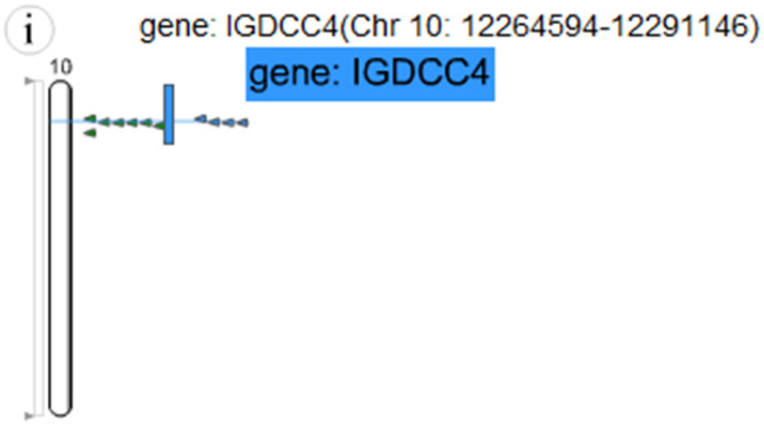


Figure 24: The close up of the BTA10 region presenting overlaps of SNPs, genes, and QTL. Blue line represent QTL, blue triangles represent genes. Red triangles represent position of SNPs having positive effect between dairy and beef traits. Green triangles represent position of SNPs having positive effect between dairy and beef traits
Slika 24: Pregled prekrivanja SNP označevalcev, genov, and QTL-ov regiji BTA10. Modra črta predstavlja QTL, modri trikotniki gene. Rdeči trikotniki predstavljajo pozicijo SNP označevalcev z negativno povezavo med lastnostmi mlečnosti in mesnatosti. Zeleni trikotniki predstavljajo pozicijo SNP označevalcev s pozitivno povezavo med med lastnostmi mlečnosti in mesnatosti

The most pronounced positive covariance between NG and CQ with the highest number of genomic overlaps of SNPs and genes was placed on the BTA10 ranging from 10.90 to 13.12 Mb. This region contains seven SNPs having positive covariance (rs43616186, rs43615966, rs110752655, rs41664401, rs110168357, rs43611427, and rs110447449), four protein-coding genes (*THBS4*, PDCD7, IGDCC3, and *IGDCC4*) and one QTL for carcass weight (EBV), see Figure 24).

2.4.4 Discussion

We presented an approach where estimated allele substitution effects from GWA were used as an input in evaluating covariance contribution of individual or group of SNP markers. This approach enabled the discovery of SNPs associated with multiple traits and genes with either synergistic or antagonistic pleiotropic effects. We evaluated the proposed approach on correlations between milk and beef traits of dual purpose breeds, here separated in three data sets. In the GFV full data set, we developed and showed its appropriateness on a large data set. Loci with the synergistic and antagonistic effect were reported. The general results, with few local differences, for example differences between

DGAT1 effects, were further confirmed in a subset of GFV as well as on another independent dataset, here on similar but different breed (IPR).

Progeny derived phenotypes in the form of yield deviations or deregressed EBVs are mainly used as phenotypes for GWA of complex traits (Thomsen et al., 2001). In the GFV data sets, yield deviations were used to account for the information of progeny. Pausch et al. (2011) reported that there was no significant difference in applying either progeny yield deviation or deregressed EBVs as phenotypes due to high reliability of traits (Table 12). Therefore, in order to assess the effect of using different progeny phenotypes, GWA with the deregressed EBVs was performed in IPR data. Furthermore, Pausch et al. (2011) pointed that progeny-derived phenotypes act purely additively and they do not allow unravelling of non-additive effects. So, all results presented are valid under assumption of pure additive models while at the same time we were aware that non-additive effects might be present. However, the simplifications were made to reduce complexity of the models while concentrating on the concept. As far as we are aware, this is the first time, where the total genetic covariance was further partitioned to synergistic and antagonistic covariation as well as to the polygenic and large (pleiotropici) marker effects. Results obtained support the evidence that genetic covariation of quantitative traits involves a large number of genes (or QTLs) with small effect, evidence of large contribution of the polygenic component, with the presence of a number of markers with large effect. This reflects the skewed distribution of effects due to limited number of QTL that actually control the trait. Thus, our results are supportive multi-trait extension of recent single trait analyses deducing genetic architecture of domestic animals as a polygenic with a presence of moderate to large gene effects.

An example of strong contribution from a single gene was explained by Grisart et al. (2002) where mutation in *DGAT1* considerably increases milk and protein yield and decreases fat content. Moderate to large effects have been also demonstrated for the myostatin mutation on the meat and carcass quality in beef cattle (Wiener et al., 2009). However, the number of detected QTL depends on the genetic architecture of the trait (Chamberlain et al., 2007; Hayes and Goddard, 2001). For example, a mutation of *STX17* gene explains large proportions of the phenotypic variance of melanoma and pigmentation of coat and skin in Lipizzan horse (Curik et al., 2013). A similar genetic architecture was presented for coat colour in Holstein (Hayes et al., 2010) in a study where *KIT*, *MITF*, and a locus on chromosome 8 together explained 24% of the variation in the proportion of black coat. A significant correlation between chromosome length and the proportion of explained variation for a particular single trait was illustrated in several studies (Visscher et al, 2007; Yang et al, 2011; Pimentel et. al, 2011; Abdollahi-Arpanahi et al, 2014) suggesting that larger chromosomes explained more genetic variance. However, results from production traits in cattle showed some exceptions compared to human data. A large proportion of variance for milk yield and fat and protein content was attributed by the

strong effect of *DGAT1* gene (Pimentel et al., 2011). Similarly, Pausch et al. (2011) reported that particular chromosomes contain loci with a large effect on paternal calving ease, udder clearness (Pausch et al., 2012a), and ambilateral circumocular pigmentation (Pausch et al., 2012b) in Fleckvieh population. Chromosomes BTA3, BTA5, BTA6, BTA14, and BTA19 explained more variance for the pathogen-specific mastitis traits in Danish cattle (Sørensen et al, 2012). Conversely, results for human height provided evidence that many loci across the whole genome had small effects without any strong effect of individual locus (Visscher et al, 2007; Yang et al, 2011). The same is true for somatic cell count in Holstein (Pimentel et al, 2011). Several studies attempted to quantify the number of genes which have influence on complex traits in livestock populations. However, the estimates showed a lot of variation among species and strongly depend on the effective population size (Hayes et al., 2010; Daetwyler et al., 2010).

The proportion of explained variance by the SNPs, extending here to explained covariance, depends on the number of SNPs close to causative genes. The results from cattle studies are different from human, where the increased number of markers explained larger amount of genetic variation (Makowsky et al., 2011). In cattle, a number of around 40,000 SNPs (McKay et al., 2007) is sufficient for GWAS analysis implying that the increase of sample size is more important for QTL detection. Furthermore, linkage disequilibrium (LD) between the SNPs and the causative mutations also affects the amount of genetic variance captured by the SNPs. In theory, LD between markers and causative genes depends on effective population size which determines the number of independently segregated chromosome segments (Goddard 2009; Daetwyler et al, 2010). Cattle populations have a lower effective population size than human populations (Kemper and Goddard, 2012) with the lower recombination rate. Artificial selection is intense in livestock compared to natural selection in human populations. Therefore haplotype blocks in the cattle may be longer than those in humans.

Although genome-wide association studies give a possibility to gain insights into the genetic architecture of complex traits, some limitations still exist. Most of researches were performed using single SNP analysis to test each SNP independently. Gianola et al. (2013) describe the lack of GWA and show how single step GWAS can be violated for multiple traits, when a trait is complex and is affected by sets of many genes in LD. Even though GWA analysis is powerful, it can be misleading when a trait is complex and affected by many genes in LD. Gianola et al. (2013) concluded that the partition of the variance into locus-specific contributions is not straightforward when LD exists. In order to avoid that, we excluded all SNPs that were in LD. Furthermore, Bayesian genomic selection approaches were marked as successful methods of in trait mapping (Sun et al., 2011; Veerkamp et al., 2010). In the analysis, we applied ridge regression method (results not

shown) to test multiple SNP associations, which gained similar results to the analysis presented in the results.

With the concept applied we were able to identify "pleiotropici" as single genes with large pleiotropic effects. The positional candidate region for milk and beef traits on BTA14 was identified, with overlapping genomics regions of the SNPs, genes, and QTL. That region contains the diacylglycerol O-acyltransferase 1 (*DGAT1*) gene, which was independently indicated as a strong candidate gene for milk production traits (Winter et al., 2002 and Grisart et al. 2002). The *DGAT1* gene encodes an enzyme catalyzing the final step of triglyceride syntesis (Winter et al., 2002; Grisart et al., 2002) and showed a major effect on milk yield and milk composition (Grisart et al. 2002). These gene was also associated with intramuscular and backfat tissue deposition in cattle (Barendse, 1997; Thaller et al., 2003; Anton et al., 2010). The DGAT enzyme might be regulated post-translationally by a tyrosine kinase (Haagsman et al., 1982, Lau et al., 1996). In addition, the neighbouring gene *PTK2* (protein tyrosine kinase 2) was associated with milk production traits in Chinese Holstein dairy cattle (Wang et al., 2013).

Activation of this gene may be important in early step of the cell growth and their function (Corsi et al., 2006). Intracellular signal transduction pathways triggered in response to certain neural peptides or to cell interactions with the extracellular matrix through their receptors. It is also described the critical role of this gene during embryogenesis through migration and survival of the cells (Corsi et al., 2006). *FOXH1* gene (Forkhead Activin Signal Transducer 1, Attisano et al., 2001) together with *SMAD2* (SMAD Family Member 2) and *SMAD3* (SMAD Family Member 3; Hoodless et al., 1999) are associated with signal transmission of cellular proteins important in maintain the pregnancy. Specific interaction were made between surface receptor gene Plectin (*PLEC*), with α-actin 2 (*ACTA2*), α-actinin 2 (*ACTN2*) that were involved in the protein synthesis in prostate epithelial cells and growth of epithelial and endothelial cell types (Cho et al., 2005; Ibaragi et al., 2009). Two genes were identified as a non-protein coding transcript: *CYHR1* (Cysteine and histidine-rich cytoplasmic protein) which is involved in enzyme catalyses, cellular trafficking of galectin 3 and protein interactions (Menon et al., 2000) and *KIAA1875*. Hu et al. (2005) described that *TRAPPC9* (Trafficking Protein Particle Complex 9) gene plays a role in signalling transcription factors that encodes a proteins and is a synonyms to *KIAA1882* gene. Possible role is related to the activity of genes in the nervous system developing cortical plate. Higher expression is indicating in postmitotic neurons than in progenitor cells, but some researches show involving of the trafficking protein particle complex in embryo development of bovine oocytes (Pfeffer et al., 2007). *COL22A1* (Collagen, type XXII, alpha 1) is a member of the *FACIT* (fibrillar-associated collagens with interrupted triple helices) subgroup of the collagen protein family, specifically localizes to tissue junctions. Is a protein coding gene that acts as a cell adhesion ligand for skin epithelial cells and fibroblasts (Koch et al., 2004).

The positional candidate region for NG and CQ was identified on BTA10 containing four protein-coding genes (*THBS4*, *PDCD7*, *IGDCC3*, and *IGDCC4*). *THBS4* gene (Thrombospondin 4) belongs to the thrombospondin protein family. The gene regulates protein production and binds numerous proteins like glycoproteins that mediate cell-to-cell, cell-to-matrix interactions as interaction in extracellular matrix. *PDCD7* (Programmed cell death 7) gene encodes a protein that is associated with apoptosis, endocytosis, adhesion and cytokinesis. This gene also influences the metabolic function that regulates calcium trafficking of proteins and regulates the cell viability (Tian et al., 2014). *IGDCC3* (Immunoglobulin Superfamily, DCC Subclass, Member 3) and *IGDCC4* (Immunoglobulin Superfamily, DCC Subclass, Member 4) belongs to immunoglobulin superfamily, DCC subclassis. Important paralog of protein-coding gene *IGDCC3* is *ROBO2* (Roundabout, axon guidance receptor, homolog 2) which is a transmembrane receptor with functions in axon guidance and cell migration. Important paralog to *IGDCC4* gene this gene is *CHL1* (Cell Adhesion Molecule L1-Like). It is a neural recognition molecule that may be involved in signal transduction pathways like neuronal survival, migrations and formation of synapses in damage tissue (Jakovcevski et al., 2007).

We believe that approach presented will provide additional knowledge to enhance understanding of genetic basis of covariation. While the current model is quite simplified, further extensions are expected. The estimation of regional and chromosomal genetic covariances among traits seems to be promising direction.

2.4.5 Conclusions

The results from this study showed that estimated allele substitution effects could be used as an input for the estimation of contribution partition of genetic covariances. To the best of our knowledge, this is the first study demonstrating the partition of genome genetic covariance. The genetic architecture of the covariation contribution is mostly polygenic with the presence of a number of genes with large effects. Genetic correlations are largely, with exception of correlations between NG and CQ, determined by the summation of positive and negative covariances. Furthermore, larger chromosomes proportionally explain more covariance compared to smaller chromosomes.

Information about genomic regions involving aforementioned effects could contribute to the identification of genes and pathways associated with dairy and beef traits in dual purpose cattle breeds and will provide additional knowledge for the Simmental breeding program with the aim to optimize matings and to enhance breeding decisions.

3 GENERAL DISCUSSION

3.1 IMPROVEMENT OF GENETIC EVALUATIONS USING ADDITIONAL INFORMATION

Selection of the 'best' animals was based on inferred (estimated) breeding values (EBV) derived from phenotypic records and pedigree information. Best linear unbiased prediction (BLUP, Henderson, 1984) is a standard procedure that has been implemented for estimation of BV. The idea of improving the accuracy of EBV using genetic markers (blood group in cattle) known to be associated with phenotype was already implementeds in the sixties (Neimann-Sorensen and Robertson, 1961). However, the systematic use of molecular information became feasible only some decades after, due to advances in molecular techniques. The availability of dense maps thousands of SNP markers and technology of microarrays made feasible genotyping individuals across the whole genome for tens or even hundreds of thousands genetic markers. The genomic selection as defined by Meuwissen et al. (2001) showed through simulations that the marker information can be used to estimate genomic BV with a considerably high accuracy. Marker based BV is often called direct genomic value (DGV) since it is estimated from the direct use of marker data.

The advantage of using genomic selection compared to conventional genetic evaluation is increased accuracy of EBV for young or non-phenotyped animals (Goddard and Hayes, 2007) and especially for low heritable traits. If the accuracy of GEBV is high enough, early use of young bulls will shorten generation interval and increase genetic merit (Schaeffer, 2006). Accuracy of genomic prediction depends greatly on the size of the reference population (Goddard, 2009; Daetwyler et al., 2008). The larger the reference population, the more accurate genomic prediction is. In Brown Swiss cattle, pooling genotypes for bulls from seven countries (Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States) via Intergenomics project operated at the Interbull formed a common reference population which helped to increase the reliability of predictions (Jorjani et al., 2011; VanRaden et al., 2012).

The integration of genomic information in the national evaluation system based on a combined reference population from several countries provided more accurate inference of breeding values, especially for young or non-phenotyped animals with genotype information in comparison to conventional evaluation. The accuracy of progeny and genomic evaluations in a small population of Slovenian Brown bulls was evaluated using univariate repeatability test-day model based on the national phenotypic and pedigree data and the bivariate model combining the univariate approach and DGV as a correlated trait into a national estimate. For that purpose, Slovenian phenotypic data on milk production,

pedigree information, and SNP genotypes (BovineSNP50 BeadChip, Illumina, San Diego, CA) of Slovenian Brown Swiss bulls were used. For the purpose of validation, the complete data was divided into training subset (701 bulls born before the year 2004) and validation subset (35 genotyped bulls born in the years between 2004 and 2007) by removing daughter phenotypes in the validation subset. Validation of predictions was assessed with the analysis of theoretical and empirical accuracies of EBV for univariate and bivariate evaluations for proven and validation bulls. Theoretical accuracies was evaluated from the the variance of prediction errors and additive genetic variance in base generation, while empirical accuracies was based on the correlation between EBV from different evaluations.

The average theoretical accuracy of proven bulls was 0.98 for all milk traits using univariate approach and the inclusion of DGV via the bivariate approach did not lead to the significant increase in the accuracy. This can be attributed to the sizeable number of daughters per bull. In proven bulls there were also a high empirical correlations between EBVs from different approaches. The correlations between the national phenotype based evaluation and genomic prediction using DGV were 0.98 for milk and protein yield and 0.99 for fat yield. With the bivariate approach correlations were similar. Since proven bulls had a sufficient number of daughters, blending of DGV with phenotype evaluation did not change their EBV significantly. The average theoretical accuracy for bulls in the validation set was 0.58 for milk and protein yield, and 0.52 for fat yield based on parent average information. The inclusion of DGV in the bivariate analysis led to an increase in the accuracy up to 0.89. Further increases in the theoretical accuracy up to 0.96 were achieved using both, univariate and bivariate analyses when these validation bulls were progeny tested. Empirical correlations show the benefit of combining all the available information via the bivarite approach in the analysed population characterized by a small population size and a small number of daughters per tested bull. In the set of validation bulls, empirical correlation between parent average and progeny evaluation was 0.49 for milk yield and increased to 0.92 when combining both progeny and DGV data. Similar trends in increased empirical correlations were observed in fat yield (from 0.50 to 0.95) and protein yield (from 0.56 to 0.91).

Very succesfull was bivarite approach which is not present in the literature for dairy traits. On contrary, in beef cattle, mentioned approach is very common. Reference populations are usually much smaller than in dairy cattle due to more disconnected local populations of the same breed over several regions and/or countries. Kachman (2008) presented an alternative method of using genomic information in such setting. In these populations, prediction equation for DGV could be developed on some experimental population of reasonable size and later used to compute the DGV of other animals in national population that is later blended with national EBV via bivariate analysis. Similar approach was also

suggested by Mäntysaari and Strandén (2010). Examples of applying these methods for beef populations are MacNeil et al. (2010a) and Johnston et al. (2010; 2012). Benefit of the bivariate approach is the unified system integrating all the available data in a single EBV instead of reporting several EBV from different data sources for the same animal.

3.2 THE REASONS FOR BIAS IN CLASSICAL AND GENOMIC EVALUATIONS

The methods of validation for genomic evaluation became an important topic since many countries beside within-country genomic evaluations, have been included in multi-country genomic evaluations. Mäntysaari and Strandén (2010) proposed a method for the correction of double counting since data in training and validation data sets overlapped. Amer and Banos (2010) showed the consequences of overlapping the training and validation data sets using a single national genetic evaluation run pointing that is best to not overlap data in the mentioned data sets.

One of the possibilities to avoid double counting of shared data at both the national and international levels is based on Bayesian methods. Quaas and Zang (2006) and Legarra et al. (2007) proposed 2 different Bayesian procedures where external or international information (EBV and associated accuracies) were integrated into the internal or national evaluation as prior information. Improvements of these methods were given by Vandenplas and Gengler (2012) and Vandenplas et al. (2014) in order to take into account the double counting among related external animals. In the first paper was shown that the number of Slovenian BSW bulls used for the development of DGV equation was small compared to the total number of bulls in the InterGenomics consortium leading to the negligible amount of double counting. However, the integration of the international information (EBV or GEBV and associated reliabilities) into internal evaluation leads to improved ranking and more reliable EBV for animals with external evaluation (Vandenplas and Gengler, 2012). Bayesian approach (Vandenplas et al., 2014) was performed in order to integrate and blend several external sources of information (MACE EBV and GEBV) into an internal genetic evaluation for BSW breed in Slovenia.

3.3 GENOME-WIDE ASSOCIATION ANALYSES AND DECOMPOSITION OF GENETIC CORRELATIONS

3.3.1 Using progeny-derived phenotypes for genome-wide association analyses

Progeny derived phenotypes in the form of yield deviations or deregressed EBVs are mainly used as phenotypes for GWA of complex traits (Thomsen et al., 2001). In the BSW and GFV data sets, yield deviations were used to account for the information of progeny. Pausch et al. (2011) reported that there was no significant difference in applying either

progeny yield deviation or deregressed EBVs as phenotypes due to high reliability of traits. Therefore, in order to assess the effect of using different progeny phenotypes, GWA with the deregressed EBVs was performed in IPR population. Furthermore, Pausch et al. (2011) pointed that progeny-derived phenotypes act purely additively and they do not allow unraveling of non-additive effects. So, all results presented are valid under assumption of pure additive models while at the same time we were aware that non-additive effects might be present. However, the simplifications were made to reduce complexity of the models while concentrating on the concept.

In the third paper of the present thesis, GWAS was performed in BSW population in order to detect SNP association signals across the whole genome for dairy traits. In the forth paper, the mapping of QTL affecting dairy and beef traits was facilitated by using EBVs or DYDs based on data set of 4,105 artificial insemination bulls of the German FV population and 511 bulls of the Italian PR population. However, there was no significant difference in applying either DYDs or EBVs as phenotypes, most likely because of the high reliability of the EBVs for traits used in the analysis. EBVs with high accuracy are appropriate phenotypes for GWAS as the major part of the EBV information results from progeny information and pedigree information contributes less.

3.3.2 Detected associations for dairy traits in BSW population

Single SNP analysis was performed to detect significant associations among SNPs and dairy traits across the genome. Two models were considered. The first was linear regression model considering one marker at a time. The second model differed from first only by admixture components included for accounting the population stratification. The Bonferroni correction resulted in 52 significant SNPs for dairy traits using the first model. Among these, 11 were significant for all milk traits. The association was spread over 14 BTA. The majority of the genome-wide significant SNPs associated with dairy traits were located on BTA21 followed by BTA9. Correction for population stratification based on selected criteria was shown to be crucial. No significant associations among SNPs and dairy traits have been detected using the second model. This analysis indicate that the first model was not dealing properly with the analysed data set. Further improvements should be made on the enlargement of number of genotyped bulls in order to detect association signals and the identification of genes associated with dairy traits in Slovenian BSW breed.

To date, the number of GWAS related to dairy traits in BSW population is limited. Guo et al. (2012) reported 16 significant SNPs associated with dairy traits, with the strongest signal on BTA25. SNP associations for milk and protein yield were also found at the same BTA in Finnish Ayrshire (Viitala et al., 2003) and Holstein (Harder et al., 2006). In German Braunvieh Maxa et al. (2012) reported five SNPs with the significant effect on

dairy traits. Two SNPs which affected MY were placed on the BTA4, two SNPs associated FY were on the BTA14 and BTA23, while significant SNP with an effect on the fat content was on the BTA1. However, GWAS in dairy cattle have mainly focused on the Holstein breed due to worldwide use of this breed (Daetwyler et al., 2008; Pryce et al., 2010). These significant SNPs were located on BTA8, 9, 10, 11, 13, 25 and 29 (Hayes et al., 2009b; Bolormaa et al., 2010; Jiang et al., 2010; Mai et al., 2010). BTA 14 has also been in focus since many studies reported DGAT1 as a major gene affecting milk production traits (Grisart et al., 2002; Ashwell et al., 2004; Jiang et al., 2010; Mai et al., 2010). In Spanish Churra sheep, Garcia-Gomez et al. (2012) reported significant association for fat and protein content located on chromosome 3.

3.3.3 Genetic architecture of the analysed traits

Results of the GWAS support the evidence that total genetic covariation of quantitative traits involves a large number of genes (or QTLs) with small effect, evidence of large contribution of the polygenic component, with the presence of a number of markers with large effect. Several SNP markers with a large positive covariance were placed on the BTA6, BTA17, and BTA19 chromosome. SNP markers with a negative covariance were located only on BTA14 chromosome. Among these, 14 SNP markers with a large effect dominated and were placed in the region of the *DGAT1* gene. Thus, our results are supportive multi-trait extension of recent single trait analyses deducing genetic architecture of domestic animals as a polygenic with a presence of moderate to large gene effects.

The contribution of total genetic covariance attributed by each chromosome was found to be associated with its physical length. This was shown by linear relationship observed between the contribution of covariance explained by each chromosome and its chromosome length as larger chromosomes accounted for more genetic covariance. Partially, this could be explained by the number of markers since larger chromosomes harbour more SNP markers. These results correspond to the hypotheses that the traits are influenced by a large number of genes distributed throughout the whole genome. The exceptions are genes with the large effect (pleiotropici). For example, the determination coefficient (R^2) for MY and FY pair-wise was 3% compared to R^2 of 7.5% when BTA14 was excluded from the data set. Results for pair-wise NG and CQ were similar to the previously mentioned pair-wise and the value of R^2 was increased from 8% to 13% when BTA10 chromosome was removed from the data set. The observed trend in this analysis is in agreement with the results reported for genetic architecture of single traits by Visscher et al. (2007) for human height. Similarly, Yang et al. (2011), who analyzed data on human height, body mass index (BMI), von Willebrand factor, and QT interval and found statistically highly significant relationship between the contribution of additive genetic variance explained by a chromosome and its genetic length. Pimentel et al. (2011) came to

the same conclusion by explaining a contribution of each chromosome to genetic variation of milk yield and composition traits in Holstein.

3.3.4 Biological relevance of the identified QTL based on covariance

Information about genomic regions involving aforementioned effects could contribute to the identification of genes and pathways associated with dairy and beef traits in dual purpose cattle populations and will provide additional knowledge for the Simmental breeding program with the aim to optimize matings and to enhance breeding decisions.

The positional candidate region for milk and beef traits on BTA14 was identified, with overlapping genomics regions of the SNPs, genes, and QTL. That region contains the diacylglycerol O-acyltransferase 1 (*DGAT1*) gene, which was independently indicated as a strong candidate gene for milk production traits (Winter et al., 2002 and Grisart et al. 2002). The *DGAT1* gene encodes an enzyme catalyzing the final step of triglyceride syntesis (Winter et al., 2002; Grisart et al., 2002) and showed a major effect on milk yield and milk composition (Grisart et al. 2002). These gene was also associated with intramuscular and backfat tissue deposition in cattle (Barendse, 1997; Thaller et al., 2003; Anton et al., 2010). The DGAT enzyme might be regulated post-translationally by a tyrosine kinase (Haagsman et al., 1982, Lau et al., 1996). In addition, the neighbouring gene *PTK2* (protein tyrosine kinase 2) was associated with milk production traits in Chinese Holstein dairy cattle (Wang et al., 2013). Activation of this gene may be important in early step of the cell growth and their function (Corsi et al., 2006). Intracellular signal transduction pathways triggered in response to certain neural peptides or to cell interactions with the extracellular matrix through their receptors.

It is also described the critical role of this gene during embryogenesis through migration and survival of the cells (Corsi et al., 2006). *FOXH1* gene (Forkhead Activin Signal Transducer 1, Attisano et al., 2001) together with *SMAD2* (SMAD Family Member 2) and *SMAD3* (SMAD Family Member 3; Hoodless et al., 1999) are associated with signal transmission of cellular proteins important in maintain the pregnancy. Specific interaction were made between surface receptor gene Plectin (*PLEC*), with α-actin 2 (*ACTA2*), α-actinin 2 (*ACTN2*) that were involved in the protein synthesis in prostate epithelial cells and growth of epithelial and endothelial cell types (Cho et al., 2005; Ibaragi et al., 2009). Two genes were identified as a non-protein coding transcript: *CYHR1* (Cysteine and histidine-rich cytoplasmic protein) which is involved in enzyme catalyses, cellular trafficking of galectin 3 and protein interactions (Menon et al., 2000) and *KIAA1875*. Hu et al. (2005) described that *TRAPPC9* (Trafficking Protein Particle Complex 9) gene plays a role in signalling transcription factors that encodes a proteins and is a synonyms to *KIAA1882* gene. Possible role is related to the activity of genes in the nervous system

developing cortical plate. Higher expression is indicating in postmitotic neurons than in progenitor cells, but some researches show involving of the trafficking protein particle complex in embryo development of bovine oocytes (Pfeffer et al., 2007). *COL22A1* (Collagen, type XXII, alpha 1) is a member of the *FACIT* (fibrillar-associated collagens with interrupted triple helices) subgroup of the collagen protein family, specifically localizes to tissue junctions. This is a protein coding gene that acts as a cell adhesion ligand for skin epithelial cells and fibroblasts (Koch et al., 2004).

The positional candidate region for NG and CQ was identified on BTA10 containing four protein-coding genes (*THBS4*, *PDCD7*, *IGDCC3*, and *IGDCC4*). *THBS4* gene (Thrombospondin 4) belongs to the thrombospondin protein family. The gene regulates protein production and binds numerous proteins like glycoproteins that mediate cell-to-cell, cell-to-matrix interactions as interaction in extracellular matrix. *PDCD7* (Programmed cell death 7) gene encodes a protein that is associated with apoptosis, endocytosis, adhesion and cytokinesis. This gene also influences the metabolic function that regulates calcium trafficking of proteins and regulates the cell viability (Tian et al., 2014). *IGDCC3* (Immunoglobulin Superfamily, DCC Subclass, Member 3) and *IGDCC4* (Immunoglobulin Superfamily, DCC Subclass, Member 4) belongs to immunoglobulin superfamily, DCC subclassis. Important paralog of protein-coding gene *IGDCC3* is *ROBO2* (Roundabout, axon guidance receptor, homolog 2) which is a transmembrane receptor with functions in axon guidance and cell migration. Important paralog to *IGDCC4* gene this gene is *CHL1* (Cell Adhesion Molecule L1-Like). It is a neural recognition molecule that may be involved in signal transduction pathways like neuronal survival, migrations and formation of synapses in damage tissue (Jakovcevski et al., 2007).

3.3.5 Impacts for practical animal breeding

Brown Swiss bulls have several evaluations at the national and at the international level. All these evaluations complicate the publication of results and inhibit the use of all the available data. Integration of DGV from a large Intergenomics consortium into the national evaluation as a correlated trait enabled combination of all the available data. In addition, this approach provides a way to automatically blend all the results in a single value avoiding the need to publish several estimates of breeding values per animal.

Evidence for a genetic covariation between milk and beef traits estimated directly by interrogation of the genome could have an important impact on the design of future genetic and functional studies and may provide new insights for novel traits. Understanding genetic covariation at the genome, chromosome, and chromosome region level would provide new insight into the biological processes between dairy and beef traits, which could in turn be utilized through genomic selection in applied breeding. Information about

genomic regions harbouring pleiotropic effects will provide additional knowledge for the Simmental breeding program with the aim to optimize mattings to enhance breeding decisions of selecting a dual purpose breed with good dairy and beef performance.

4 CONCLUSIONS

The following general conclusions can be drawn from this thesis.

- Different approaches of integrating genomic information into a national evaluation system for small population of Brown breed were evaluated in comparison to the conventional evaluation based on phenotype and pedigree data.

- Results indicate that integration of DGV from a large consortium into the national evaluation as a correlated trait enabled combination of all the available data.

- Bivariate approach provides a way to automatically blend all the results in a single value avoiding the need to publish several estimates of breeding values per animal.

- The proposed Bayesian method blended and integrated several sources of external information (MACE EBV and GEBV) into an internal genetic evaluation for BSW bulls in Slovenia.

- The results showed that the method was able to avoid double counting of contributions due to records and due to relationships.

- Results from GWAS for dairy traits in BSW bulls without correction for population stratification showed 52 SNPs with a significant effect on MY, FY, and PY using single SNP analysis. The majority of the SNPs associated with milk traits were located on BTA21 and BTA9.

- After adjusting for population stratification, no significant associations between SNPs and dairy traits have been detected. Population stratification exists in the analysed data set and the model without correction on data were not dealing properly with the analysed data set.

- Further improvements should be made on the enlargement of the number of genotyped bulls in order to detect association signals and the identification of genes associated with dairy traits. Results from GWAS will facilitate the understanding of the genetic architecture of complex traits and information from these studies may be used to improve breeding programmes.

- The results from the study of genome-wide covariation between the pairs of dairy and beef traits showed that estimated allele substitution effects could be used as an input for the estimation of contribution partition of genetic covariances.

- To the best of our knowledge, this is the first study demonstrating the partition of genome genetic covariance. The genetic architecture of the covariation contribution is mostly polygenic with the presence of a number of genes with large effects.

- Genetic correlations are largely, with exception of correlations between NG and CQ, determined by the summation of positive and negative covariances. Furthermore, larger chromosomes proportionally explain more covariance compared to smaller chromosomes.

- Information about genomic regions involving aforementioned effects could contribute to the identification of genes and pathways associated with dairy and beef traits in dual purpose cattle breeds and will provide additional knowledge for the Simmental breeding program with the aim to optimize matings and to enhance breeding decisions.

5 SUMMARY (POVZETEK)

5.1 SUMMARY

Most of economically important traits in livestock have a quantitative expression. These traits are influenced by many genes of predominantly small effect and environment which results in continuous variation of phenotypes (Falconer and Mackay, 1996). Genetic improvement of quantitative traits is commonly based on phenotypic and pedigree information that are used to estimate additive genetic effect of all the genes of an individual or the so called breeding value. Statistically the estimation is carried out using the mixed model methodology (Henderson, 1984). In dairy cattle, phenotypic data are collected through various recording schemes (milk and fertility recording, type classification, etc.) on daughters of progeny tested bulls. The latter are four to six years old when the first crop of data on their daughters is collected. At that time the accuracy of bull estimated breeding value (conventional breeding value - CBV) based on progeny data (estimated breeding value - EBV) is 0.90 or more. Without progeny testing which is a source of phenotypic information that pertains to bulls, the accuracy of EBV using only pedigree information (i.e., average parental breeding values - PA) is around 0.60 (e.g., Schefers and Weigel, 2012).

In recent years, the availability of affordable high-density panels of thousands of single-nucleotide polymorphisms (SNP) has led to the abundant use of this information in livestock breeding programs. This is commonly referred to as genome-wide or genomic selection (Meuwissen et al., 2001). Genomic selection is based on the inference of breeding values based on the sum of SNP association effects with phenotype across the whole genome (Meuwissen et al., 2001; Solberg et al., 2008). Statistically the SNP association effect is a regression coefficient of phenotype on SNP genotype, while genetically the SNP association effect is an average allele substitution effect of a particular SNP. Both representations describe the additive genetic effect of a SNP association between phenotypic values and genomic region of a SNP. The sum of all SNP associations represents an estimate of total additive genetic (breeding) value of an individual. In dairy cattle this estimate is often called direct genomic value (DGV) and is based solely on genotypic information of an individual animal and population based SNP associations. For the implementation of genomic selection, phenotyped and genotyped reference population is needed to estimate the SNP associations and assemble them in the prediction equation of a DGV which is in turn used to estimate DGV for non-phenotyped individuals (Meuwissen et al., 2001; Goddard and Hayes, 2007). Different types of breeding value estimates (PA, EBV, and DGV) can be blended in one value often called genomically enhanced breeding value (GEBV) using various approaches (VanRaden 2008; Kachman, 2008; Aguilar et al., 2010). The accuracy of genomic evaluation mainly depends on the number of genotyped and phenotyped individuals in the reference population (Goddard, 2009; Schefers and Weigel, 2012). Use of genomic data generally provides less information than phenotypic

data from a progeny test and consequently lower accuracy of EBV. However, the advantage of using genomic data is to increase the accuracy of EBV of young or non-phenotyped animals from about 0.60 (PA) to 0.80 (Hayes et al., 2009a). Even though these accuracies are lower than with progeny test the early use of young bulls in artificial insemination shortens generation interval considerably and therefore increases genetic gain per unit of time (Schaeffer, 2006).

The highest gains of using genomic selection could be expected in populations with large a number of genotyped and phenotyped animals, especially progeny tested bulls. An example of such large international reference population is the InterGenomics project for Brown Swiss breed (BSW) operated by the Interbull (Zumbach et al., 2010; Jorjani et al., 2012). The reference population is consisted of almost 8,000 bulls from seven countries (Austria, France, Germany, Italy, Slovenia, Switzerland, and the United States of America) (Zumbach et al., 2010; Jorjani et al., 2012). The need for large reference populations has therefore steered a lot of international cooperation, which is essential for introduction of genomic selection in small populations. Within the scope of the InterGenomics consortium specific prediction equations for each country are developed based on all available SNP data and multiple across country evaluation (MACE) EBV on a country's specific scale (Jorjani et al., 2012). Each bull has therefore several evaluations at the national level (PA and EBV) and at the international level (DGV and a blended GEBV of DGV and PA for young bulls and DGV and MACE EBV for proven bulls). Reference populations in beef cattle are usually much smaller than in dairy cattle due to more disconnected (sub)populations in one or several countries. In these populations SNP prediction equations could be developed on experimental populations of reasonable size and later used to compute the DGV of other animals in whole national population. In such cases DGVs can be blended with the national EBV from pedigree and/or phenotype inference via the use of bivariate analysis (MacNeil et al., 2010; Johnston et al., 2010; Johnston et al., 2012).

Slovenian BSW bulls have results of genetic (genomic) evaluation from several sources of information at the national level (PA for young bulls and EBV for proven bulls) and at the international level (MACE EBV, DGV and GBV). Different approaches of combining all available data information have been shown with the emphisis to evaluate the accuracy of EBV for young animals. Integration of DGV based on a combined reference population provides significantly more accurate estimates of breeding values for routine national genomic evaluation than PA. The accuracy of evaluations was assessed using: univariate national evaluation (conventional EBV), international DGV, and bivariate national evaluation incorporating DGV as a correlated trait into conventional EBV.

For the purpose of validation, the complete data was divided into training subset (701 bulls born before the year 2004) and validation subset (35 genotyped bulls born in the years

between 2004 and 2007) by removing daughter phenotypes in the validation subset. Validation of predictions was assessed with the analysis of theoretical and empirical accuracies of EBV for univariate and bivariate evaluations for proven and validation bulls. Theoretical accuracies was evaluated from the the variance of prediction errors and additive genetic variance in base generation, while empirical accuracies was based on the correlation between EBV from different evaluations.

The average theoretical accuracy of proven bulls was 0.98 for all milk traits using univariate approach and the inclusion of DGV via the bivariate approach did not lead to the significant increase in the accuracy. In proven bulls there were also high empirical correlations between EBVs between the national phenotype based evaluation and genomic prediction using DGV (0.98 for milk and protein yield and 0.99 for fat yield). The average theoretical accuracy for bulls in the validation set was 0.58 for milk and protein yield, and 0.52 for fat yield based on PA information. The inclusion of DGV in the bivariate analysis led to an increase in the accuracy to 0.85 (protein yield), 0.86 (fat yield), and 0.89 (milk yield). Further increases in the theoretical accuracy up to 0.96 were achieved using bivariate analysis. Empirical correlations also showed the benefit of combining all the available information via the bivariate approach in the analyzed population characterized by a small population size and a small number of daughters per tested bull.

As already mentioned the bivariate approach was very successful in combining all the available information in a small population of BSW breed. Slovenian BSW bulls were also used for the development of DGV equation at the InterGenomics consortium, which potentially leads to double counting of data in the presented analysis. However, in the current study, the number of such bulls used for the development of DGV equation was small compared to the total number of bulls in the InterGenomics consortium leading to the negligible amount of double counting. Therefore, the issue of double counting was neglected in the analysis of Slovenian BSW bulls.

Appropriate statistical methods can correct for the double counting of national data in the process of integration of the external information (EBV or GEBV and associated reliabilities) into national genomic evaluation. One of the possibilities to avoid double counting of shared data at both the national and international levels is based on Bayesian methods. Vandeplas et al. (2014) proposed the Bayesian method in order to take into account the double counting among related external animals. Therefore, several external sources of information (MACE EBV and GEBV) were blended and integrated into an internal genetic evaluation for BSW breed in Slovenia to avoid double counting of contributions due to records and due to relationships.

Comparisons were performed for two groups of bulls: The first group was related to internally used bulls that were associated with either Slovenian and MACE information (400 bulls) or Slovenian and GEBV information (320 bulls) and had daughters with records in the Slovenian internal dataset. The second group included internally unused bulls that had either Slovenian and MACE information (5,360 bulls) or Slovenian and GEBV information (5,344 bulls) but had no daughters with records in the Slovenian internal dataset. All comparisons were performed based on the following parameters: Spearman's rank correlation coefficients (r) of EBV_{MACE} or EBV_{GEBV} with corresponding evaluations, MSE of evaluations (i.e. mean squared errors expressed as a percentage of average MSE of EBV_S or EBV_{NGNDC}), regression coefficients (a), R^2 of the regressions of $EVAL_{MACE}$ or $EVAL_{GEBV}$ on the four other evaluations, RE_{tot}, and the average REL for the corresponding evaluations.

These evaluations showed that double counting of contributions due to records ($EVAL_{DCR}$) affected the prediction of EBV_{MACE} for internally used bulls slightly according to all parameters. $EVAL_{DCR}$ had small advantage compared to $EVAL_{DCP}$ based on r, MSE, and R^2 and was more reliable based on MSE than $EVAL_{DCRP}$ for analysed trait. Based on these results, it can be stated that double counting of contributions due to relationships and due to records had little effect on EBV for internally used bulls. The results indicate that Bayesian evaluation that blended internal and external information and avoided most double counting of contributions due to records and due to relationships was successful in integrating MACE information for 400 internally used bulls. The same is true for evaluations based on EBV_{GEBV} for the 320 internally used bulls. Slightly better prediction was found when contributions due to records (EBV_{NGDCR}) were considered. However, double counting of contributions due to relationships and due to records had little effect on EBV_{GEBV} for internally used bulls. For group of internally unused bulls double counting is only based on contributions due to relationships (EBV_{DCP} or EBV_{NGDCP}) and no contributions due to records. These evaluations showed that double counting of contributions due to relationships and due to records had little effect on either EBV_{MACE} or EBV_{GEBV} for internally unused bulls.

Development of SNP marker panels also enabled genome wide association studies (GWAS) with a purpose to determine associations between DNA polymorphisms and phenotypes (Hirschhorn and Daly, 2005). In the GWAS studies, association analysis was used to identify chromosome regions that harbor the genes or regulatory elements related to the traits of interest (e.g., Goddard and Hayes, 2009; Bush and Moore, 2012; Eggen, 2012; Montaldo et al., 2012). The first use of GWAS was in the analysis of human diseases in order to reveal the genetic basis of disease (Massey et al., 2007). Association studies in the field of livestock species are used to understand the genetic control of economically important traits such as milk (e.g., Grisart et al., 2002; Pryce et al., 2010; Mai et al., 2010),

meat (e.g., Bolormaa et al., 2011), and health (e.g., Murdoch et al., 2011). Such knowledge can be used to enhance biological understanding as well as to enhance the methods of genomic selection due to greater resolution of genome. For example in dairy cattle, significant associations using genome-wide data were detected for milk yield on BTA 8, 9, 10, 11, 13, 25, and 29 (Hayes et al., 2009b; Bolormaa et al., 2010). Milk fat and protein content was associated with SNPs on BTA 5, 6, 11, 14, 19, and 26 (Pryce et al., 2010; Schopen et al., 2011). In beef cattle, Du et al. (2013) reported SNPs with a significant effect placed on BTA6 for birth weight and weaning weight. Furthermore, associations between SNP and residual feed intake, birth weight and hip height were also reported on BTA 3, 5, 7, and 8 (Bolormaa et al., 2011).

GWAS for dairy traits (milk yield - MY, fat yield -FY, and protein yield –PY) were introduced for Slovenian BSW bulls with the aim to consider they possibility of implementation the GWAS aspects into genomic selection for dairy traits. Steps in preparation of genomic data were shown together with SNPs associated with dairy traits which could lead to identify candidate genes and genetic variants underlying these studied traits. In the analysis, phenotypic, pedigree, and genotypic data were obtained. Phenotypes considered were estimated breeding values (EBVs) for MY, FY, and PY obtained from the routine national genetic evaluations with the corresponding reliabilities. Altogether, EBVs of 182 progeny tested and genotyped BSW bulls born between 1990 and 2007 were included in the analysis. The DNA was extracted from the blood samples and BSW bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1) comprising of 54,001 SNP markers. Genotypes of the bulls were scored as 0 (BB), 1 (AB) and 2 (AA). Several quality controls of genotypic data were applied to investigate the SNPs integrity and usefulness. These controls included genotyping call rate (CR) by SNP and animal, minor allele frequency (MAF), departure from Hardy–Weinberg equilibrium (HWE), and parent–progeny conflicts. The number of SNPs per chromosome was also checked. The SAS statistical package (SAS Institute, 2011) was used to make edits leaving a final set of 34,450 SNPs placed on 29 *Bos taurus* autosomes (BTA). After the quality control any missing genotypes were imputed using the gpig program (Strandén, 2010) using the linear approximation method (Gengler et al., 2007). Single SNP analysis that tests one SNP at a time was used to identify SNP effects on dairy traits based on ordinary linear model one time for each SNP using R software package (R Development Core Team, 2011). In the GWAS, significance threshold was calculated using Bonferroni correction for multiple testing. Correction based on *p-value* of 0.05 resulted in Bonferroni threshold of 1.45×10^{-6}. SNPs with the smaller *p-values* than the Bonferroni threshold are considered as significant. The results of GWAS were plotted using Manhattan plot based on the obtained *p*-value of each SNP. The *p-values* of the association test were transformed to –log10 (*p*-values) for each SNP versus its chromosomal locations (Ding et al., 2009).

In total, 52 SNPs with a significant effect on MY, FY, and PY were identified based on linear regression model without considering population stratification. Among these, 11 were significant for all dairy traits. The association was spread over 14 BTA. The majority of the SNPs associated with MY, FY and PY were on BTA21 followed by BTA9. Correction for population stratification using four ancestral populations was shown to have an impact on the results of analysis. No significant associations among SNPs and dairy traits have been detected using mentioned model. This analysis indicate that the model without considering population stratification was not dealing properly with the analysed data set. QTLs affecting milk traits in Brown Swiss breed were reported in study of Guo et al. (2012) and 16 significant SNPs associated with milk traits had the strongest signal on BTA 25. Viitala et al. (2003) and Harder et al. (2006) also found the SNP association for milk and protein yield located on the same chromosome in Finnish Ayrshire and Holstein. Five SNPs with the significant effect on dairy traits were reported in German Braunvieh (Maxa et al., 2012). Two SNPs affected MY were placed (BTA4), two SNPs were associated FY (BTA14 and BTA23), and one SNP with significant effect on fat content (BTA1). However, GWAS in dairy cattle have been mainly focused on the Holstein breed due to worldwide use of this breed (Daetwyler et al., 2008; Pryce et al., 2010). For milk yield, Zang et al. (2012) reported a total of 734 SNPs with significant effects. These SNPs were located on BTA 8, 9, 10, 11, 13, 25 and 29 (Hayes et al., 2009b; Bolormaa et al., 2010; Jiang et al., 2010; Mai et al., 2010). Chromosome 14 has been in focus since many studies reported DGAT1 as a major gene affecting milk production traits (Grisart et al., 2002; Ashwell et al., 2004; Jiang et al., 2010; Mai et al., 2010).

One of the practical problems in breeding programs is selection on multiple traits with variable genetic correlations between traits. The limitation is posed by (strongly) negative genetic correlations limiting the space for improvement due to closely linked genes or pleiotropy. Classical example of antagonistic traits in dairy cattle is milk yield and fertility. For example in Norwegian Red cattle Olsen et al. (2011) reported QTL on BTA 12 having opposite effect on non-return rate and milk traits. Another example with antagonistic effect has been reported for calving ease and postnatal growth traits where two QTLs were identified on BTA 14 and 21 in the German Simmental breed (Pausch et al., 2011). Due to polygenic architecture for economically important traits there can be genomic regions with favourable, unfavourable, and no genetic correlations that all jointly result in a global genetic correlation between traits. Genome-wide data provides an opportunity to dissect the genome in such regions to facilitate further gene discovery as well as to enhance breeding. Most studies were focused on milk production and fertility traits in dairy cattle. However, these aspects have not yet been studied in dual purpose breeds, where the antagonism is not only present between milk production and fertility, but also between milk and beef production.

GWAS was performed in order to characterize genomic regions with associations having different magnitude and direction of effect on dairy and beef traits in dual purpose population providing the possibility to dissect genetic correlations between traits. There were three objectives regarding to the decomposition of genetic correlations using genomic data. First, to estimate total genetic covariation between the dairy and beef traits in two well-known dual-purpose populations using comprehensive large scale routinely collected data from progeny testing. Second, to analyze genome-wide covariation between the pairs of dairy and beef traits in order to describe its architecture along the genome due to close linkage and/or pleiotropy in the terms of sign, size, and distribution along the genome. Third, to utilize bioinformatics resources to enrich the obtained results with the existing knowledge and to propose potential candidate genes (‘pleiotropic loci‘) that underlie a part of genetic covariation of analyzed traits.

The study was done on German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) bulls, with phenotypic records for dairy (milk yield - MY and fat yield - FY) and beef traits (net daily gain - NG and carcass quality - CQ) and with available Illumina BovineSNP50K genotypes. Phenotypes values for these traits were available as summarized phenotypes with high heritability (reliability) for each bull; obtained from the routine national genetic evaluations (Interbull, 2014) in the form of progeny phenotype deviations for the GFV population (Emmerling et al., 2002; Schild et al., 2003) and deregressed estimates of breeding values (Garrick et al. 2009) for the IPR population. The GFV bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1) comprising of 54,001 SNP markers (3,734 bulls) and Illumina BovineSNP50K BeadChip® (version 2, 54Kv2) comprising of 54,609 SNP markers (379 bulls), where 52,340 SNP markers were shared between the two versions. The IPR bulls were genotyped with the Illumina BovineSNP50K BeadChip® (version 1, 54Kv1). Quality control criteria that includes call rate ($\geq$ 90%) by SNP marker and by bull, minor allele frequency ($<$ 0.01), departure from the HWE ($P < 0.001$) and exclusion of unmapped SNPs and those placed on Chr X were applied to genotypes using the SAS software (SAS Inst., Inc., Cary, NC, 2011). To remove any multi-collinearity between SNP markers, we performed linkage disequilibrium (LD) based pruning of genotypes in PLINK (Purcell et al, 2007). The bivariate mixed linear model was fitted to estimate genetic correlations between the pairs of dairy and beef traits. Single trait genome-wide association study was performed to estimate genome-wide associations between phenotypes and SNP markers for each of the dairy and beef traits. All computations were performed in WOMBAT software (Meyer, 2007). The R package (R Development Core Team, 2011) was used to compute the inverse of $\boldsymbol{A}$ or $\boldsymbol{G}$ and to test a distribution of covariance on uniformity. Bioinformatic analyses were based on the Animal QTL Database, release 23.

Genetic correlations estimated between MY and FY were and high positive in GFV (0.72±0.01) and in IPR (0.81±0.04) and between NG and CQ (0.54±0.03 in GFV and 0.51±0.09 in IPR) which is in agreement with the estimates obtained using the traditional bivariate model with the pedigree relationship matrix in other studies (e.g., Fürst et al., 2013; Roman and Wilcox, 2000; Welper and Freeman, 1992). Total genetic correlations between MY and NG were slightly positive or close to zero (0.09±0.03 in GFV and 0.10±0.10 in IPR) and similarly between FY and NG (0.04±0.03 in GFV and -0.08±0.11 in IPR). Total genetic correlations between MY and CQ were slightly negative or close to zero (-0.11±0.03 in GFV and -0.17±0.09 in IPR) as well as between FY and CQ (-0.15±0.03 in GFV and -0.17±0.10 in IPR). These estimates are concordant with estimates in previous studies (Mason, 1964; Langlet, 1965; Pirchner, 1984).

The proportion of positive and negative genome-wide covariance contributions determined total genetic correlations in both populations for all the pairs of dairy and beef traits. The high positive total genetic correlations estimated between MY and FY were determined by the dominant proportion of positive genome-wide covariances (90.2% in GFV and 99.2% in IPR). The same pattern was also observed between MY and NG, MY and CQ, FY and NG, and FY and CQ, though with different proportions. The exception to this pattern was observed between NG and CQ, where medium positive total genetic correlation (0.54±0.03 in GFV and 0.51±0.09 in IPR) was determined by the dominant proportion of positive genome-wide covariances (94.8% in GFV and 92.1% in IPR), though this exception can be explained by the different distribution of contributions.

Distribution of covariance indicated that there were number of SNP markers with large but prevalent number of SNP markers with low contribution. While the genetic component of covariation between observed traits in two populations was mainly polygenic there were a number of SNP markers with sizeable contribution. The markers with large contribution represent pleiotropic loci either due to pleiotropic or closely linked genes and could be called as “pleiotropicus” after pleiotropic locus (plural pleiotropici). The total contribution (positive and negative) from these pleiotropici ranged from 5.7% for MY and CQ to 17.1% for MY and FY in the GFV population. Somewhat lower total contribution of pleiotropici was estimated in the IPR population; ranging from 3.9% for FY and CQ to 9.8% for MY and FY.

The contribution, magnitude and chromosomal location of genes with large contribution to the covariation (pleiotropici) differed between pairs of traits. For MY and FY pair-wise, several SNP markers with a large positive covariance were placed on the BTA6, BTA17, and BTA19 chromosome. SNP markers with a negative covariance were located only on BTA14 chromosome. Among these, 14 SNP markers with a large effect dominated and were placed in the region of the *DGAT1* gene. Grisart et al. (2002) reported that a polymorphism in *DGAT1* explained 18% of the variance in milk yield. A positive

covariance dominated the NG and CQ, with a total of 62 SNP markers that had a large covariance effect placed on the BTA10 and BTA14. The remaining four pair-wise combinations of traits represented approximately equal contribution of a positive and negative covariance. Combinations of MY and NG pair-wise and FY and NG contained a prevalent contribution of covariance with a small effect. The number of SNP markers which have a negative sign is slightly larger than the number of those with a positive large effect. Most of them were located on BTA10 and BTA14. For covariation between MY and NG the largest number of positive pleiotropici was determined on BTA6 and BTA19, while for the combination FY and NG BTA14 had the largest number of positive pleiotropici followed by BTA11, BTA18, and BTA19. A similar distribution of small and large covariance was determined for the other two covariation, between MY and CQ and between FY and CQ. While large negative pleiotropici were mainly concentrated on a few chromosomes (BTA1, BTA6, and BTA14), large positive pleiotropici were more often distributed all over the genome.

The contribution of total genetic covariance attributed by each chromosome was found to be associated with its physical length. This was shown by linear relationship observed between the contribution of covariance explained by each chromosome and its chromosome length as larger chromosomes accounted for more genetic covariance. Partially, this could be explained by the number of markers since larger chromosomes harbor more SNP markers. These results correspond to the hypotheses that the traits are influenced by a large number of genes distributed throughout the whole genome. The exceptions are genes with the large effect (pleiotropici). The observed trend in this analysis is in agreement with the results reported for genetic architecture of single traits by Visscher et al. (2007) for human height. Similarly, Yang et al. (2011), who analyzed data on human height, body mass index (BMI), von Willebrand factor, and QT interval and found statistically highly significant relationship between the contribution of additive genetic variance explained by a chromosome and its genetic length. Pimentel et al. (2011) came to the same conclusion by explaining a contribution of each chromosome to genetic variation of milk yield and composition traits in Holstein.

Genomic overlap analysis was performed by comparing the genomic location of SNP markers, genes, and QTL for all dairy and beef traits. The SNP markers with a high number of overlaps could potentially be stronger molecular markers. The total of 164 SNP markers, 98 with positive and 66 with negative covariance, overlaps QTL for dairy and beef quality. In addition, 59 of them also overlap with genes. The highest number of overlapping genomic elements was identified on BTA14 ranging from 16.51 to 20.88 Mb. This region contains five SNP markers having negative covariance (rs109146371, rs109968515, rs109421300, rs41256919, rs109350371), five protein-coding genes (*FOXH1*, *CYHR1*, *DGAT1*, *KIAA1875*, and *PLEC*) and 17 QTL for eight traits: milk and

fat yield, milk and fat yield (daughter deviation), milk and fat yield (EBV), body weight at weaning, and fat thickness at the 12th rib. The most pronounced positive covariance between NG and CQ with the highest number of genomic overlaps of SNPs and genes was placed on the BTA10 ranging from 10.90 to 13.12 Mb. This region contains seven SNPs having positive covariance (rs43616186, rs43615966, rs110752655, rs41664401, rs110168357, rs43611427, and rs110447449), four protein-coding genes (*THBS4*, *PDCD7*, *IGDCC3*, and *IGDCC4*) and one QTL for carcass weight (EBV).

In this thesis, the integration of DGV in the national evaluation system based on a combined reference population from several countries will provide more accurate inference of breeding values, especially for young or non-phenotyped animals with genotype information in comparison to conventional evaluation. Methodological developments have been provided new knowledge useful for implementation of genomic selection in small populations and improvement of breeding programs. Integration of genomic information based on a combined reference population from several countries leads to double counting, which can be corrected with the appropriate statistical methods leading to appropriate integration procedure. New methods of correcting for the double counting were developed and applied to real data. Double counting of contributions due to relationships and due to records had little effect on MACE EBV and GEBV for internally used bulls. The results indicate that Bayesian evaluation that blended internal and external information and avoided most double counting of contributions due to records and due to relationships was successful in integrating inernational information for internally used bulls. These evaluations also showed that double counting of contributions due to relationships and due to records had little effect on either EBV_{MACE} or EBV_{GEBV} for internally unused bulls.

Genome-wide association studies for dairy traits performed in Slovenian BSW bulls enabled important contribution especially to people working with Brown Swiss breed. Furthermore, the issue is dealing with actual research topic and contains information that is important for the scientific community. Genome partitioning of genetic covariance between dairy and beef traits in dual purpose populations showed that estimated allele substitution effects could be used as an input for the estimation of contribution partition of genetic covariances. To the best of our knowledge, this is the first study demonstrating the partition of genome genetic covariance. The genetic architecture of the covariation contribution is mostly polygenic with the presence of a number of genes with large effects. Genetic correlations are largely, with exception of correlations between NG and CQ, determined by the summation of positive and negative covariances. The approach will provide additional knowledge to enhance understanding of genetic basis of covariation and will contribute to the identification of genes and pathways associated with dairy and beef traits in dual purpose cattle populations. The approach will provide additional knowledge

for the Simmental breeding program with the aim to optimize matings and to enhance breeding decisions.

5.2 POVZETEK

Večina gospodarsko pomembnih lastnosti v živinoreji je kvantitativnih. Na te lastnosti vpliva več genov s pretežno majhnim vplivom in okolje, ki se kaže v stalnem spreminjanju fenotipa (Falconer in Mackay, 1996). Genetsko izboljševnje kvantitativnih lastnosti običajno temelji na fenotipskih informacijah in rodovniku, ki služijo za oceno aditivnega genetskega vpliva vseh genov osebka oziroma za oceno plemenske vrednosti (PV). Statistično se za oceno PV uporablja metodologija mešanega modela (Henderson, 1984). Pri mlečnih pasmah goveda se fenotipski podatki zbirajo iz različnih kontrol o prireji (kontrola mlečnosti, spremljanje plodnosti, ocenjevanje zunanjosti, itd) hčera bikov v progenem testu. Biki so štiri do šest let stari, ko se zberejo prvi podatki o hčerah. Takrat je točnost ocenjene PV (konvencionalna PV ali EBV) bika na osnovi progenega testa hčera 0,90 ali več. Brez progenega testa je točnost PV, ki je ocenjena samo na podlagi podatkov iz rodovnika (povprečna starševska PV ali PA), okoli 0,60 (npr. Schefers in Weigel, 2012).

V zadnjih letih je cenovna dostopnost čipov z veliko gostoto SNP označevalcev (Single Nucleotide Polymorphisms = polimorfizem na posameznem označevalcu) pripeljala do masovne uporabe teh informacij v rejskih programih pod skupnim izrazom genomska selekcija (Meuwissen in sod., 2001). Genomska selekcija temelji na oceni PV na osnovi vplivov povezav SNP označevalcev s fenotipom na celotnem genomu (Meuwissen in sod., 2001; Solberg in sod., 2008). Statistično je vpliv povezav SNP označevalcev enak regresijskemu koeficientu med fenotipom in SNP genotipom, medtem ko je genetsko vpliv povezav SNP označevalcev enak povprečnemu vplivu zamenjave alel določenega SNP označevalca. Oba opisa predstavljata aditivni genetski vpliv povezav SNP označevalcev med fenotipskimi vrednostmi in genomsko regijo SNP označevalca. Vsota vseh povezav SNP označevalcev predstavlja oceno skupne aditivne (plemenske) vrednosti osebka. Pri mlečnih pasmah goveda se ta ocena pogosto imenuje direktna genomska vrednost (DGV) in temelji izključno na genotipskih informacijah posamezne živali ter na povezavah SNP označevalcev izračunanih v populaciji. Za uvedbo genomske selekcije je potrebna referenčna populacija živali z znanim fenotipom in genotipom za oceno povezav SNP označevalcev izraženih v enačbi za napoved DGV, ki se uporablja tudi za oceno DGV pri živalih brez fenotipske vrednosti (Meuwissen in sod., 2001; Goddard in Hayes, 2007). Različne vrste ocen plemenskih vrednosti (PA, EBV in DGV) se lahko združujejo v eno vrednost, ki je običajno imenovana razširjena plemenska vrednost (GEBV) z uporabo različnih pristopov (VanRaden 2008; Kachman, 2008; Aguilar in sod., 2010). Točnost genomske napovedi je odvisna predvsem od števila živali z znanim genotipom in fenotipom v referenčni populaciji (Goddard, 2009; Schefers in Weigel, 2012). Uporaba genomskih podatkov na splošno daje manj informacij kot fenotipski podatki iz progenega testa in posledično manjšo točnost napovedane PV. Vendar pa je prednost uporabe genomskih podatkov v povečanju točnosti napovedi PV za mlade živali ali za živali brez znanega fenotipa, iz okoli 0,60 (PA) na 0,80 (Hayes in sod., 2009a). Čeprav so te točnosti

manjše kot pri progenem testu, zgodnja uporaba mladih bikov v osemenjevanju bistveno skrajša generacijski interval in posledično povečuje genetski napredek na enoto časa (Schaeffer, 2006).

Največji napredek uporabe genomske selekcije je mogoče pričakovati pri populacijah z velikim številom genotipiziranih živali, še posebno bikov s progenim testom. Primer takšne velike mednarodne referenčne populacije je projekt InterGenomics za rjavo pasmo, ki ga izvaja Interbull (Zumbach in sod., 2010; Jorjani in sod., 2012). Referenčna populacija je sestavljena iz skoraj 8.000 bikov iz sedmih držav (Avstrija, Francija, Nemčija, Italija, Slovenija, Švica in Združene države Amerike) (Zumbach in sod., Jorjani in sod., 2012). Potreba po velikih referenčnih populacijah je vodila v večje mednarodno sodelovanje, ki je bistvenega pomena za uvedbo genomske selekcije predvsem v majhnih populacijah. V okviru InterGenomics konzorcija so bile razvite posebne enačbe za napoved PV za vsako državo posebej na podlagi vseh razpoložljivih SNP informacij in na podlagi napovedi PV iz združenega konvencionalnega mednarodnega obračuna t.i. Multiple Accross Country Evaluation ali MACE (Jorjani in sod., 2012). Vsak bik ima tako več napovedi na nacionalnem nivoju (PA in EBV) in na mednarodnem nivoju (DGV, združena GEBV sestavljena iz DGV in PA za mlade bike, DGV in MACE EBV za bike s progenim testom). Podobno združujejo informacije pri mesnih pasmah goveda, kjer so referenčne populacije navadno precej manjše kot pri mlečnih pasmah, zaradi bolj nepovezanih (pod)populacij v eni ali več državah. V teh populacijah bi lahko enačbe za napoved SNP razvili na poskusnih populacijah primerne velikosti in jih kasneje uporabili za izračun DGV drugih živali v celotni nacionalni populaciji. V takšnih primerih se poleg EBV iz rodovnika in/ali fenotipa (PA/EBV) vključijo tudi DGV-ji z dvolastnostno analizo (MacNeil in sod., 2010; Johnston in sod., 2010; Johnston in sod., 2012).

Za slovensko populacijo bikov rjave pasme obstajajo rezultati genetskega vrednotenja iz večih virov na nacionalnem nivoju (PA za mlade bike in EBV za bike s progenim testom) in na mednarodnem nivoju (MACE EBV, DGV in GBV). V nadaljevanju smo prikazali različne možnosti kombiniranja vseh razpoložljivih informacij predvsem z vidika napovedovanja točnosti ocen PV pri mladih bikih. Povezovanje DGV na podlagi združenih referenčnih populacij omogoča značilno bolj točne napovedi plemenskih vrednosti za rutinsko nacionalno genomsko napoved v primerjavi s PA. Primerjali smo točnosti napovedi z uporabo: enolastnostne nacionalne napovedi (konvencionalna EBV), mednarodna DGV in dvolastnostna nacionalna napoved, ki vključuje DGV kot korelirano lastnost v konvencionalni EBV.

Za potrebe validacije je bil celoten set podatkov razdeljen na podset podatkov za ocenjevanje (701 bik rojen pred letom 2004) in na podset za validacijo (35 genotipiziranih bikov rojenih med letoma 2004 in 2007) pri katerem so bili odstranjeni fenotipski podatki

hčera. Validacijo napovedi smo naredili z analizo teoretične in empirične točnosti za napoved PV z enolastnostnim in dvolastnostnim modelom za bike s progenim testom in za bike v podsetu za validacijo. Teoretična točnost je bila ocenjena na podlagi variance napake napovedi in aditivne genetske variance v izhodiščni generaciji, medtem ko so empirične točnosti temeljile na korelaciji med napovedmi PV iz različnih vrednotenj.

Povprečna teoretična točnost za bike s progenim testom je bila 0,98 za vse lastnosti mlečnosti z uporabo enolastnostnega modela. Vključitev DGV preko dvolastnostnega modela ni pripeljala do značilnega povečanja točnosti. Pri bikih s progenim testom so bile tudi velike empirične korelacije med napovedmi PV iz nacionalnega vrednotenja na podlagi fenotipa in genomskimi napovedmi z uporabo DGV (0,98 za količino mleka in količino beljakovin ter 0,99 za količino maščobe). Povprečna teoretična točnost za bike v validacijskem setu je bila 0,58 za količino mleka in količino beljakovin ter 0,52 za količino maščobe na osnovi PA informacij. Vključitev DGV v dvolastnostno analizo je povečala točnost napovedi za te bike na 0,85 (količina beljakovin), 0,86 (količina maščob) in 0,89 (količina mleka). Nadaljnja povečanja teoretične točnosti do 0,96 so bila dosežena s pomočjo dvolastnostne analize. Tudi empirične korelacije so pokazale koristi združevanja vseh informacij, ki jih pridobimo z dvolastnostnim modelom v analizirani majhni populaciji in majhnim številom hčera na testiranega bika.

Kot smo že omenili je bil dvolastnostni pristop zelo uspešen pri združevanju vseh razpoložljivih informacij v majhni populaciji rjave pasme. Slovenski biki rjave pasme so bili vključeni tudi v razvoj DGV enačbe pri InterGenomics konzorciju, kar bi lahko privedlo do dvojnega štetja podatkov v prikazani analizi. Vendar je bilo v našem primeru število takih bikov, ki so bili vključeni v razvoj DGV enačbe, majhno v primerjavi s celotnim številom bikov v InterGenomics konzorciju. Posledično je bil prispevek dvojnega štetja zanemarljiv pri analizi slovenskih bikov rjave pasme.

Ustrezne statistične metode lahko popravijo dvojno štetje nacionalnih podatkov v postopku integracije zunanjih informacij (EBV ali GEBV in ustrezne zanesljivosti) v nacionalno genomsko napoved. Ena od možnosti, da bi se izognili dvojnemu štetju deljenih podatkov, tako na nacionalni kot tudi na mednarodni ravni, temelji na Bayesovih metodah. Vandeplas in sod. (2014) so predlagali Bayesovo metodo, da bi upoštevali dvojno štetje podatkov med sorodnimi živali iz zunanjih setov. Zaradi tega smo združili več zunanjih virov informacij (MACE EBV in GEBV) in jih vključili v sistem nacionalnega genetskega vrednotenja za rjavo pasmo v Sloveniji, da bi se izognili dvojnemu štetju informacij iz fenotipskih zapisov in iz podatkov o sorodstvu.

Primerjave smo opravili na dveh skupinah bikov. Prva skupina so bili biki ki so imeli slovensko in MACE informacijo (400 bikov) ali slovensko in GEBV informacijo (320 bikov) ter so imeli hčere z zapisi (fenotipom) v slovenskem internem setu podatkov. Drugo skupino sestavljajo biki, ki niso bili v interni uporabi. Ti biki imajo bodisi slovensko in MACE informacijo (5360 bikov) ali slovensko in GEBV informacijo (5344 bikov), vendar nimajo hčera z zapisi v slovenskem internem setu podatkov. Vse primerjave so bile narejene na osnovi naslednjih parametrov: Spearmanova rank korelacija (r) med rezultati referenčne metode vrednotenja (EBV_{MACE} ali EBV_{GEBV}) in vsako od testnih vrednotenj, povprečna kvadratna napaka ocen (MSE) prikazana kot odstotek MSE za EBV_S ali EBV_{NGNDC}, regresijski koeficient (a), koeficient determinacije (R^2) $EVAL_{MACE}$ ali $EVAL_{GEBV}$ na vsako od testnih vrednotenj, relativni efektivni faktor (RE_{tot}) in povprečne točnosti ocen (REL) za posamezna vrednotenja.

Ta vrednotenja so pokazala, da je dvojno štetje informacij zaradi zapisov ($EVAL_{DCR}$) po vseh parametrih nekoliko vplivalo na napoved EBV_{MACE} za bike v interni uporabi. $EVAL_{DCR}$ je bil malo boljši v primerjavi z $EVAL_{DCP}$ na osnovi r, MSE in R^2 in je bil bolj zanesljiv na osnovi MSE kot z $EVAL_{DCRP}$ za analizirano lastnost. Na podlagi teh rezultatov lahko potrdimo, da je imelo dvojno štetje informacij iz sorodstva in iz zapisov majhen vpliv na EBV pri bikih za interno uporabo. Rezultati so pokazali, da je bilo Bayesovsko vrednotenje, ki združuje notranje in zunanje informacije ter preprečuje največ dvojnega štetja informacij iz zapisov in iz sorodstva, uspešno pri vključevanju MACE informacij za 400 bikov v interni uporabi. Enako velja tudi za vrednotenja na osnovi EBV_{GEBV} za 320 bikov v interni uporabi. Nekoliko boljša napoved je bila ugotovljena, ko so bile upoštevane informacije iz zapisov (EBV_{NGDCR}). Vendar pa ima dvojno štetje informacij iz sorodstva in iz zapisov majhen vpliv na EBV_{GEBV} pri bikih v interni uporabi. V skupini bikov, ki niso bili v interni uporabi, je bilo dvojno štetje le na osnovi informacij iz sorodstva (EBV_{DCP} ali EBV_{NGDCP}) in brez informacij iz zapisov. Ta ovrednotenja so pokazala, da je imelo dvojno štetje informacij iz sorodstva in iz zapisov majhen vpliv tako na EBV_{MACE} kot tudi na EBV_{GEBV} za bike, ki niso bili v interni uporabi.

Razvoj čipov s SNP označevalci omogoča genomske asociacijske študije (GWAS) z namenom, da bi ugotovili povezave med polimorfizmi DNK in fenotipi (Hirschhorn in Daly, 2005). V GWAS študijah se asociacijska analiza uporablja za identifikacijo kromosomskih regij, ki vsebujejo gene ali regulatorne elemente, ki so povezani s proučevanimi lastnostmi (npr. Goddard in Hayes, 2009; Bush in Moore, 2012; Eggen, 2012; Montaldo in sod., 2012). GWAS so prvič naredili pri analizi človeških bolezni z namenom, da bi odkrili genetsko osnovo bolezni (Massey in sod., 2007). Asociacijske študije na področju živalskih vrst se uporabljajo za razumevanje genetskega vpliva na gospodarsko pomembne lastnosti kot so mlečnost (npr Grisart in sod., 2002;. Pryce in sod., 2010; Mai in sod., 2010), mesnatost (npr., Bolormaa in sod., 2011) in zdravje (npr

Murdoch in sod., 2011). Takšno znanje lahko uporabimo za povečanje tako biološkega razumevanja, kakor tudi za izboljšanje metod genomske selekcije zaradi večje resolucije genoma. Na primer pri mlečnih pasmah govedi so bile odkrite značilne povezave z uporabo genomskih podatkov za lastnosti mlečnosti na avtosomalnih *Bos taurus* kromosomih (BTA) 8, 9, 10, 11, 13, 25 in 29 (Hayes in sod., 2009b; Bolormaa in sod., 2010). Vsebnost maščobe in beljakovin v mleku je bila povezana s SNP označevalci na BTA 5, 6, 11, 14, 19, in 26 (Pryce in sod., 2010; Schopen in sod., 2011). Pri mesnih pasmah govedi so Du in sod. (2013) odkrili SNP označevalce z značilnimi vplivi na BTA 6 za rojstno maso in maso ob odstavitvi. Poleg tega so poročali tudi o povezavah med SNP označevalci in konzumacijo krme, rojstno maso in višino križa na BTA 3, 5, 7 in 8 (Bolormaa in sod., 2011).

GWAS za lastnosti mlečnosti (količina mleka - MY, količina maščobe - FY in količina beljakovin - PY) so bile narejene za bike rjave pasme v Sloveniji z namenom proučiti možnost implementacije GWAS pristopa v genomsko selekcijo za lastnosti mlečnosti. Prikazali smo korake pri pripravi genomskih podatkov skupaj s SNP-i povezanimi z lastnostmi mlečnosti, ki lahko vodijo do identifikacije kandidatnih genov in genetskih variant. V analizi smo uporabili podatke iz fenotipa, rodovnika in genotipa. Vključeni fenotipi so bile napovedi plemenskih vrednosti (EBV) za MY, FY, in PY iz rutinskega nacionalnega genetskega obračuna z odgovarjajočimi točnostmi. V analizo smo vključili skupno 182 genotipiziranih bikov s progenim testom, ki so bili rojeni med 1990 in 2007. Iz krvi bikov rjave pasme je bila izolirana DNK in vzorci so bili genotipizirani z Illumina BovineSNP50K čipom (verzija 1, 54K v1), ki vključuje 54.001 SNP označevalec. Genotipi bikov so bili zabeleženi kot 0 (BB), 1 (AB) in 2 (AA). Naredili smo več kontrol genotipskih podatkov, da bi proučili integriteto in uporabnost SNP označevalcev. Kontrole so vključevale uspešnost genotipizacije po SNP-u in živali (call rate ali CR), najmanjšo frekvenco alel (MAF), odstopanje od Hardy-Weinbergovega ravnotežja (HWE) in preverjanje neskladij med starši in potomci. Preverili smo tudi število označevalcev po kromosomu. Za urejanje genomskih podatkov smo uporabili statistični paket SAS (SAS Institute, 2011). Končni set podatkov je vključeval 34.450 označevalcev na 29 BTA. Po kontroli kakovosti smo vstavili (imputirali) morebitne manjkajoče genotipe s programom gpig (Strandén, 2010), ki uporablja linearno aproksimacijsko metodo (Gengler in sod., 2007). Za identifikacijo vplivov SNP označevalcev na lastnosti mlečnosti smo uporabili analize po posameznem SNP označevalcu, ki testirajo po en SNP na enkrat in temeljijo na navadnem linearnem modelu. Uporabili smo programski paket R (R Development Core Team, 2011). Prag statistične značilnosti smo za GWAS izračunali z Bonferronijevo korekcijo za multiplo testiranje. Korekcija, ki je temeljila na p-vrednosti 0,05, je imela Bonferonijev prag $1{,}45 \times 10^{-6}$. SNP označevalci z manjšimi p-vrednostmi od Bonferronijevega praga so bili značilni. Rezultate GWAS smo prikazali z Manhattan grafikonom na osnovi pridobljenih p-vrednostih vsakega SNP označevalca. p-vrednosti iz

asociacijskega testa so bile transformirane v $-\log_{10}$ (p-vrednosti) za vsak SNP označevalec glede na njihovo lokacijo na kromosomih (Ding in sod., 2009). Prag statistične značilnosti smo za GWAS izračunali z Bonferronijevo korekcijo za multiplo testiranje. Korekcija, ki je temeljila na p-vrednosti 0,05, je imela Bonferonijev prag $1{,}45 \times 10^{-6}$. SNP označevalci z manjšimi p-vrednostmi od Bonferronijevega praga so bili značilni. Rezultate GWAS smo prikazali z Manhattan grafikonom na osnovi pridobljenih p-vrednostih vsakega SNP označevalca. p-vrednosti iz asociacijskega testa so bile transformirane v $-\log_{10}$ (p-vrednosti) za vsak SNP označevalec glede na njihovo lokacijo na kromosomih (Ding in sod., 2009).

Identificirali smo 52 SNP označevalcev z značilnim vplivom na MY, FY in PY na osnovi linearnega modela z regresijo brez upoštevanja razdeljenosti (razslojenosti) populacije. Med njimi jih je bilo 11 značilnih za vse tri lastnosti mlečnosti. Povezave so bile porazdeljene na 14 kromosomih. Večina SNP-ov, ki so povezani z mlečnostjo, količino maščobe in beljakovin je bilo na kromosomih BTA21 in BTA9. Korekcija na razdeljenost populacije, ki je upoštevala štiri populacije prednikov, je pokazala, da je razdeljenost populacije vplivala na rezultate analize. Z uporabo tega modela nismo ugotovili značilnih povezav med SNP označevalci in lastnostmi mlečnosti. Zaradi tega so bili rezultati občutljivi na razdeljenost populacije na podpopulacije in skupaj z majhnim številom podatkov niso veliko pripomogli k bolj natančni oceni povezav. Guo in sod. (2012) so poročali o QTL-ih (angl. quantitative trait loci; lokusi za kvantitativne lastnosti), ki vplivajo na lastnosti mlečnosti pri rjavi pasmi in navedli 16 SNP označevalcev z značilno povezavo z lastnostmi mlečnosti, ki so imeli najmočnejši signal na kromosomu BTA25. Prav tako so tudi Viitala in sod. (2003) in Harder in sod. (2006) na istem kromosomu odkrili SNP povezave za količino mleka in beljakovin pri finski ayrshire in holštajn pasmah. Pet SNP označevalcev z značilnim vplivom na lastnosti mlečnosti so pri nemški rjavi pasmi našli Maxa in sod. (2012). Dva SNP označevalca, ki sta vplivala na količino mleka sta bila na kromosomu BTA4, dva SNP označevalca sta bila povezana s količino maščobe (BTA14 in 23) in en SNP označevalec z značilnim vplivom na vsebnost maščobe je bil na kromosomu BTA1. Večino genomskih asociacijskih študij pri mlečnemu govedu so opravili pri holštajnski pasmi zaradi njene razširjenosti (Daetwyler in sod., 2008; Pryce in sod., 2010). Zhang in sod. (2012) so poročali o skupno 734 SNP označevalcih, ki so imeli značilen vpliv na količino mleka. Locirani so bili na kromosomih BTA 8, 9, 10, 11, 13, 25 in 29 (Hayes in sod., 2009b; Bolormaa in sod., 2010; Jiang in sod., 2010; Mai in sod., 2010). Kromosom BTA14 je bil v središču številnih študij kot nosilec pomembnega *DGAT1* gena, ki velja za glavni gen z vplivom na lastnosti mlečnosti (Grisart in sod., 2002; Ashwell in sod., 2004; Jiang in sod., 2010; Mai in sod., 2010).

V rejskih programih je eden od praktičnih problemov selekcija na več lastnosti z različno genetsko korelacijo med njimi. Omejitev predstavlja (močna) negativna genetska

korelacija, ki zmanjšuje prostor za napredek, zaradi tesno povezanih genov ali pleiotropije. Klasičen primer antagonističnih lastnosti pri mlečnih pasmah goveda sta količina mleka in plodnost. Na primer, pri norveškem rdečem govedu so Olsen in sod. (2011) opisali QTL-e na BTA12, ki so imeli nasproten vpliv na uspešnost osemenitve in lastnosti mlečnosti. Drug primer z antagonističnim vplivom je bil odkrit med potekom telitve in lastnostmi postnatalne rasti, kjer so identificirali dva QTL-a na kromosomih BTA14 in 21 pri nemški lisasti pasmi (Pausch in sod., 2011). Zaradi poligene arhitekture gospodarsko pomembnih lastnosti obstajajo genomske regije z zaželenimi in neželenimi genetskimi korelacijami ter brez njih, ki skupaj tvorijo globalno genetsko korelacijo med lastnostmi. Genomski podatki dajejo možnost, da se podrobno prouči genom v teh regijah in s tem omogočijo nadaljnje odkrivanje genov kakor tudi izboljšanje selekcije. Pri mlečnih pasmah goveda je bilo največ raziskav opravljenih na lastnostih mlečnosti in plodnosti. Vendar ta vidik še ni bil raziskan pri kombiniranih pasmah, kjer je prisoten antagonizem, ne samo med mlečnostjo in plodnostjo, ampak tudi med pomembnimi lastnostmi za prirejo mleka in mesa.

Naredili smo GWAS z namenom, da bi okarakterizirali genomske regije s povezavami, ki vplivajo z različno jakostjo in v različni smeri na lastnosti mlečnosti in mesnatosti pri kombiniranih pasmah, kar nam omogoča natančno določitev genetskih korelacij med lastnostmi. Z uporabo genomskih podatkov pri dekompoziciji genetskih korelacij smo imeli tri cilje. Prvi cilj je bil oceniti celotno genetsko kovariacijo med lastnostmi mlečnosti in pomembnimi lastnostmi za prirejo mesa pri dveh dobro poznanih kombiniranih populacijah s pomočjo velike količine podatkov iz progenega testa. Drugi cilj je bil analizirati kovariacije v celotnem genomu med pari lastnosti mlečnosti in mesnatosti, da bi opisali arhitekturo vzdolž genoma zaradi tesne vezave in/ali pleiotropije kot smer, jakost in porazdelitev na genomu. Tretji cilj je bil uporabiti bioinformacijske podatke, da bi obogatili dobljene rezultate z obstoječim znanjem in predlagali potencialne kandidatne gene ("pleiotropske lokuse"), ki določajo del genetskih kovariacij analiziranih lastnosti.

Študija je bila narejena na nemški lisasti (German Fleckvieh; GFV) in italijanski lisasti (Italian Pezzata Rossa; IPR) populaciji bikov, ki so imeli fenotipske zapise za lastnosti mlečnosti (količina mleka - MY in količina maščobe - FY) in mesnatosti (dnevni neto prirast - NG in klavna kakovost - CQ) ter znane Illumina BovineSNP50K genotipe. Fenotipske vrednosti za te lastnosti so bile izražene kot ocene PV (seštete fenotipske vrednosti) z visokim dednostnim deležem in z zanesljivostjo za vsakega bika. Iz rutinskega nacionalnega genskega vrednotenja (Interbull, 2014) smo jih dobili v obliki fenotipskega odstopanja hčera (DYD - daugther yield deviation) pri GFV populaciji (Emmerling in sod., 2002; Schild in sod., 2003) in kot deregresirane ocene PV (Garrick in sod., 2009) pri IPR populaciji. GFV biki so bili genotipizani z Illumina BovineSNP50K BeadChip® (verzija 1, 54Kv1), ki vsebuje 54.001 SNP označevalec (3.734 bikov) in z Illumina BovineSNP50K BeadChip® (verzija 2, 54Kv2), ki vsebuje 54.609 SNP označevalcev (379 bikov). V obeh

verzijah čipov je bilo skupnih 52.340 SNP označevalcev. Biki IPR pasme so bili genotipizirani z Illumina BovineSNP50K BeadChip® (verzija 1, 54Kv1). S pomočjo programskega paketa SAS (SAS Institute, 2011) smo upoštevali naslednje kriterije za kontrolo kakovosti genotipov: uspešnost genotipizacije po SNP označevalcu in po živali (≥ 90%), frekvenca redkih alelov (< 0.01), odstopanje od Hardy-Weinbergovega ravnotežja ($p < 0.001$) in izločitev SNP označevalcev, ki niso mapirani na kromosomih ali pa so na spolnemu kromosomu X. Za odstranitev kakršnekoli multi-kolinearnosti med SNP označevalci smo naredili čiščenje na podlagi neravnotežja zaradi vezave (LD - linkage disequilibrium) s programu Plink (Purcell in sod., 2007). Za oceno genetskih korelacij med pari lastnosti mlečnosti in mesnatosti smo uporabili dvolastnostni mešani linearni model. Za oceno povezav med fenotipi in SNP označevalci smo naredili genomske asociacijske analize na osnovi enolastnostnega modela za vsako lastnost posebej. Vsi izračuni so bili narejeni v programu WOMBAT (Meyer, 2007). Za izračun inverze matrike (A ali G) in za testiranje normalnosti porazdelitve kovariance smo uporabili programski paket R (R Development Core Team, 2011). Bioinformacijske analize smo naredili z orodjem Animal QTL Database 23.

Ocenjene genetske korelacije med MY in FY so bile visoke in pozitivne pri GFV (0,72 ± 0,01) in pri IPR (0,81 ± 0,04) populaciji, ocenjene genetske korelacije med NG in CQ pa so bile 0,54 ± 0,03 pri GFV in 0,51 ± 0,09 pri IPR populaciji, kar je v skladu z ocenami izračunanimi z uporabo tradicionalnih dvolastnostnih modelov z matriko sorodstva v drugih raziskavah (npr. Fürst in sod., 2013; Roman in Wilcox, 2000; Welper in Freeman, 1992). Skupne ocenjene genetske korelacije med MY in NG so bile zelo majhne in pozitivne (0,09 ± 0,03 pri GFV in 0,10 ± 0,10 pri IPR). Podobno je bilo med FY in NG, 0,04 ± 0,03 pri GFV in -0,08 ± 0,11 pri IPR populaciji. Skupne ocenjene genetske korelacije so bile zelo majhne in negativne med MY in CQ (-0,11 ± 0,03 pri GFV in -0,17 ± 0,09 pri IPR populaciji) in prav tako med FY in CQ (-0,15 ± 0,03 pri GFV in -0,17 ± 0,10 pri IPR populaciji). Te ocene so bile primerljive z ocenami v drugih raziskavah (Mason, 1964; Langlet, 1965; Pirchner, 1984).

Deleži pozitivnih in negativnih genomskih kovarianc so določili skupne genetske korelacije pri obeh populacijah za vse pare lastnosti mlečnosti in mesnatosti. Velike pozitivne skupne genetske korelacije ocenjene med MY in FY so bile določene s prevladujočimi deleži pozitivnih genomskih kovarianc (90,2 % pri GFV in 99,2 % pri IPR populaciji). Enak vzorec je bil ugotovljen tudi med MY in NG, MY in CQ, FY in NG ter med FY in CQ, vendar z različnimi deleži. Izjema je bila med NG in CQ, kjer je bila skupna genetska korelacija zmerno pozitivna (0,54 ± 0,03 pri GFV in 0,51 ± 0,09 pri IPR) določena s prevladujočim deležem pozitivnih genomskih kovarianc (94,8 % pri GFV in 92,1 % pri IPR). To izjemo lahko pojasnimo z različno porazdelitvijo deležev.

Porazdelitev kovarianc je pokazala, da je bilo majhno število SNP označevalcev z velikim prispevkom in prevladujoče veliko število SNP označevalcev z majhnim prispevkom. Medtem, ko je bila genetska komponenta kovariacije med vključenimi lastnostmi pri dveh populacijah predvsem poligena, so imeli številni SNP označevalci pomemben prispevek. Označevalci z velikim prispevkom predstavljajo pleiotropske lokuse bodisi zaradi pleitropičnih ali zaradi tesno povezanih genov. Po pleiotropskem lokusu bi jih lahko poimenovali "pleiotropikus" (v množini "pleiotropici"). Njihov skupni prispevek (pozitivni in negativni) iz "pleiotropici" je bil od 5,7 % za MY in CQ do 17,1 % za MY in FY pri GFV populaciji. Nekoliko manjši skupni prispevek "pleiotropici" je bil ocenjen v razponu od 3,9 % za FY in CQ do 9,8 % za MY in FY pri IPR populaciji.

Prispevek, velikost in kromosomske lokacije genov z velikim prispevkom h kovariaciji (pleiotropici) so se razlikovale med pari lastnosti. Za pare MY in FY je bilo večje število SNP označevalcev z veliko pozitivno kovarianco lociranih na kromosomih BTA6, 17 in 19. SNP označevalci z negativno kovarianco so bili le na kromosomu BTA14. Med njimi je prevladovalo 14 SNP označevalcev z velikim vplivom v regiji gena *DGAT1*. Grisart in sod. (2002) so ugotovili, da je polimorfizem na *DGAT1* pojasnil 18 % variance pri količini mleka. Pozitivna kovarianca je prevladovala med parom lastnosti NG in CQ s skupno 62 SNP označevalci, ki so imeli velik vpliv kovariance na kromosomih BTA10 in 14. Preostale štirje pari kombinacij med lastnostmi so predstavljali približno enak prispevek pozitivne in negativne kovariance. Kombinaciji parov MY in NG ter FY in NG sta vsebovali prevladujoč prispevek kovariance z majhnim vplivom. Število SNP označevalcev z negativnim vplivom je bilo nekoliko večje kot število tistih z velikim pozitivnim vplivom. Večina jih je bila locirana na kromosomih BTA10 in 14. Največje število pozitivnih pleiotropikusov za kovariacijo med MY in NG je bilo določenih na kromosomih BTA6 in 19 medtem, ko je bilo za kombinacijo FY in NG največje število pozitivnih pleiotropikusov na kromosomu BTA14, ki so mu sledili kromosomi 11, 18 in 19. Podobna porazdelitev majhnih in velikih kovarianc, je bila določena za druga dva para kovariacije, torej med MY in CQ ter med FY in CQ. Medtem, ko so se veliki negativni pleiotropikusi nahajali v glavnem samo na nekaterih kromosomih (BTA1, 6 in 14), so bili veliki pozitivni pleiotropikusi pogosteje razporejeni po celotnem genomu.

Ugotovljeno je bilo, da je prispevek skupne genetske kovariance po vsakem kromosomu povezan z njegovo fizično dolžino. To je pokazala linearna povezava med prispevkom kovariance pojasnjene za vsak kromosom in dolžino kromosoma, tako da so daljši kromosomi pojasnili več genetske kovariance. Delno lahko to pojasnimo s številom SNP označevalcev, saj daljši kromosomi vsebujejo več SNP označevalcev. Ti rezultati potrjujejo hipotezo, da so lastnosti pod vplivom številnih genov, ki so razporejeni po celotnem genomu. Izjeme so geni z velikim vplivom (pleiotropici). Opažen trend v tej analizi je bil v skladu z že objavljenimi rezultati za genetsko arhitekturo posameznih

lastnosti npr. višina pri ljudeh (Visscher in sod., 2007). Do podobnih ugotovitev so prišli Yang in sod. (2011), ki so analizirali višino pri ljudeh in indeks telesne mase po von Willebrandovem faktorju in QT intervalu ter ugotovili statistično značilno povezavo med prispevkom aditivne genetske variance pojasnjene s kromosomom in njegovo genetsko dolžino. Pimentel in sod. (2011) so prišli do podobnega zaključka, ko so pojasnili prispevek vsakega kromosoma na genetsko variabilnost količine in sestave mleka pri holštajnski pasmi.

Genomska analiza prekrivanja je bila narejena s primerjanjem genomskih lokacij SNP označevalcev, genov in QTL-ov za vse lastnosti mlečnosti in mesnatosti. SNP označevalci z velikim številom prekrivanj bi lahko bili potencialno močnejši molekularni označevalci. Skupno 164 SNP označevalcev, 98 s pozitivno in 66 z negativno kovarianco, je prekrivalo QTL-e za lastnosti mlečnosti in mesnatosti. Poleg tega se jih je 59 izmed njih prekrivalo tudi z geni. Največje število prekrivajočih genomskih elementov je bilo ugotovljeno na kromosomu BTA14 na območju od 16,51 do 20,88 Mb. Ta regija je imela pet SNP označevalcev, ki so imeli negativno kovarianco (rs109146371, rs109968515, rs109421300, rs41256919, rs109350371), pet genov, ki kodirajo beljakovine (*FOXH1, CYHR1, DGAT1, KIAA1875*, *PLEC*) in 17 QTL-ov za osem lastnosti: MY in FY, MY in FY (DYD), MY in FY (EBV), telesna masa ob odstavitvi in debelina loja v predelu 12 rebra. Najbolj izražena pozitivna kovarianca med NG in CQ, z največjim številom genomskih prekrivanj med SNP označevalci in geni, je bila na kromosomu BTA10 na območju od 10,90 do 13,12 Mb. Ta regija je vsebovala sedem SNP označevalcev s pozitivno kovarianco (rs43616186, rs43615966, rs110752655, rs41664401, rs110168357, rs43611427, rs110447449), štiri gene, ki kodirajo beljakovine (*THBS4, PDCD7, IGDCC3*, *IGDCC4*) in en QTL za maso klavnih polovic (EBV).

Integracija DGV v sistem nacionalnega obračuna na osnovi kombinirane referenčne populacije iz večjega števila držav, bo omogočala bolj točno oceno PV v primerjavi s konvencionalnim vrednotenjem, še posebej za mlade živali ali za živali brez fenotipskih zapisov, ki imajo znane informacije genotipa. Metodološki razvoj je omogočil nova znanja koristna za vpeljavo genomske selekcije v majhnih populacijah in izboljšanje rejskih programov. Vključevanje genomskih informacij na osnovi kombiniranih referenčnih populacij iz večjega števila držav vodi do dvojnega štetja podatkov, ki je mogoče odpraviti z ustreznimi statističnimi metodami. Nove metode za korekcijo dvojnega štetja so bile razvite in preverjene na podatkih bikov rjave pasme. Dvojno štetje iz sorodstva in iz zapisov je imelo majhen vpliv na MACE EBV in GEBV za bike v interni uporabi. Rezultati kažejo, da je bilo Bayesovsko vrednotenje, ki je združilo notranje in zunanje informacije, in se je izognilo večini dvojnega štetja, uspešno pri vključevanju mednarodnih informacij za bike v interni uporabi. Poleg tega so ta vrednotenja pokazala, da je imelo

dvojno štetje iz sorodstva in iz zapisov majhen vpliv na EBV_{MACE} ali na EBV_{GEBV} za skupino bikov, ki niso bili v interni uporabi.

Asociacijske analize na celotnem genomu za lastnosti mlečnosti, ki so bile narejene pri bikih slovenske rjave pasme, predstavljajo pomemben prispevek še posebej za delo strokovnih služb na področju selekcije rjave pasme. Obravnavana raziskovalna tematika je aktualna in vsebuje informacije, ki so pomembne za znanstveno raziskovalno skupnost. Porazdelitev genetskih kovarianc na ravni genoma, med lastnostmi mlečnosti in mesnatosti pri kombiniranih pasmah goveda, je pokazala, da lahko z uporabo ocene učinka zamenjave alelov uspešno ocenimo prispevke genetskih kovarianc. Po našem poznavanju, je pričujoče delo prvo, ki obravnava porazdelitev genetskih kovarianc na ravni genoma. Genetska arhitektura prispevkov kovariacij je poligena, sestavljana iz več genov, izmed katerih imajo nekateri tudi večji vpliv. Genetske korelacije so v veliki večini, z izjemo NG in CQ, določene z vsoto pozitivnih in negativnih kovarianc. Prepričani smo, da bo opisani pristop zagotovil nova znanja za boljše razumevanje genetske osnove kovarianc, kar bo pripomoglo k identifikaciji genov in njihovih povezav z lastnostmi mlečnosti in mesnatosti pri kombiniranih pasmah govedi. Uporabljeni pristop bo zagotovil tudi dodatna znanja za načrtovanje parjenja in za izboljšavo selekcijskih odločitev pri selekciji lisaste pasme govedi.

6 REFERENCES

Abdollahi-Arpanahi R., Pakdel A., Nejati-Javaremi A., Moradi-Shahrbabak M., Morota G., Valente B.D., Kranis A., Rosa G.J.M., Gianola D. 2014. Dissection of additive genetic variability for quantitative traits in chickens using SNP markers. Journal of Animal Breeding and Genetics, 131: 183–193

Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., Lawlor T.J. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science, 93: 743–752

Alexander D.H., Novembre J., Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19: 1655–1664

Andersson L. 2001. Genetic dissection of phenotypic diversity in farm animals. Nature Reviews Genetics, 2: 130–138

Amer P.R., Banos G. 2010. Implications of avoiding overlap between training and testing data sets when evaluating genomic predictions of genetic merit. Journal of Dairy Science, 93: 3320–3330

Anton I., Kovács K., Holló G., Farkas V., Lehel L., Hajda Z., Zsolnai A. 2011. Effect of leptin, DGAT1 and TG gene polymorphisms on the intramuscular fat of Angus cattle in Hungary. Livestock Science, 135: 300–303

Ashwell M.S., Heyen D.W., Sonstegard T.S., Van Tassell C.P., Da Y. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. Journal of Dairy Science, 87: 468–475

Bar-Anan R., Levi U., Shilo A., Soller M. 1965. Progeny testing Israeli-Friesian AI sires for rate of gain. World Review of Animal Production, 1: 53

Barendse W., Vaiman D., Kemp S.J., Sugimoto Y., Armitage S.M., Williams J.L., Sun H.S., Eggen A., Agaba M., Aleyasin S.A., Band M., Bishop M.D., Buitkamp J., Byrne K., Collins F., Cooper L., Coppettiers W., Denys B., Drinkwater R.D., Easterday K., Elduque C., Ennis S., Erhardt G., Ferretti L., Flavin N., Gao Q., Georges M., Gurung R., Harlizius B., Hawkins G., Hetzel J., Hirano T., Hulme D., Jorgensen C., Kessler M., Kirkpatrick B.W., Konfortov B., Kostia S., Kuhn C., Lenstra J.A., Leveziel H., Lewin H.A., Leyhe B., Lil L., Burriel M.I. McGraw R.A., Miller J.R., Moody D.E., Moore S.S., Nakane S., Nijman I.J., Olsaker I., Pomp D., Rando A., Ron M., Shalom A., Teale A.J., Thieven U., Urquhart B.G.D., Vage D.I., Van de Weghe A., Varvio S., Velmala R., Vilkki J., Weikard R., Woodside C., Womack J.E., Zanotti M., Zaragoza P. 1997. A medium-density genetic linkage map of the bovine genome. Mammalian Genome, 8: 21–28

Barrett J.C., Hansoul S., Nicolae D.L., Cho J.H., Duerr R.H., Rioux J.D., Brant S.R., Silverberg M.S., Taylor K.D., Barmada M. M. 2008. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nature Genetics, 40: 955–962

Becker D., Wimmers K., Luther H., Hofer A., Leeb T. 2013. A Genome-Wide Association Study to Detect QTL for Commercially Important Traits in Swiss Large White Boars. PLoS One, 8: e55951, doi: 10.1371/journal.pone.0055951: 6 p.

Benjamini Y., Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B, 57, 1: 289–300

Berger P.J., Shanks R.D., Freeman A.E., Laben R.C. 1981. Genetic aspects of milk yield and reproductive performance. Journal of Dairy Science, 64: 114–122

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y., Leveziel H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. Genetics Selection Evolution, 35: 77–101

Boichard D., Guillaume F., Baur A., Croiseau P., Rossignol M.N., Boscher M.Y., Druet T., Genestout L., Eggen A., Journaux L., Ducrocq V., Fritz S. 2010. Genomic selection in French dairy cattle. In: 9th World Congres on Genetics Applied to Livestock Production, Leipzig, 1-6 Aug. 2010. Leipzig, German Society for Animal Science, Manuscript 0716: 8 p.

Bolormaa S., Neto L.R., Zhang Y.D., Bunch R.J., Harrison B.E., Goddard M.E., Barendse W. 2011. A genome-wide association study of meat and carcass traits in Australian cattle. Journal of Animal Science, 89: 2297–2309

Bolormaa S., Pryce J.E., Hayes B.J., Goddard M.E. 2010. Multivariate analysis of a genome-wide association study in dairy cattle. Journal of Dairy Science, 93: 3818–3833

Bonaiti B., Boichard D. 1995. Accounting for foreign information in genetic evaluation. Interbull Bulletin, 11: 4 p.

Bush W.S., Moore J.H. 2012. Chapter 11: Genome-Wide Association Studies. PLoS Computational Biology, 8: e1002822, doi: 10.1371/journal.pcbi.1002822: 11 p.

Calo L.L., McDowell R.E., Van Vleck L.D., Miller P.D. 1973. Genetic aspects of beef production among Holstein–Friesians pedigree selected for milk production. Journal of Animal Science, 37: 676– 682

Calus M.P.L., Veerkamp R.F. 2011. Accuracy of multi-trait genomic selection using different methods. Genetics Selection Evolution, 43: 26, doi: 10.1186/1297-9686-43-26: 14 p.

Cardon L., Bell J. 2001. Association study designs for complex diseases. Nature Reviews Genetics, 2: 91–99

Chamberlain A.J., McPartlan H.C., Goddard M.E. 2007. The number of loci that affect milk production traits in dairy cattle. Genetics, 177: 1117–1123

Charlier C., Coppieters W., Farnir F., Grobet L., Leroy P.L., Michaux C., Mni M., Schwers A., Vanmanshoven P., Hanset R., 1995. The mh gene causing double-muscling in cattle maps to bovine Chromosome 2. Mammalian Genome, 6: 788–792

Churchill G.A., Doerge R.W. 1994. Empirical threshold values for quantitative trait mapping. Genetics, 138: 963–71

Colleau J.J. 2002. An indirect approach to the extensive calculation of relationship coefficients. Genetics Selection Evolution, 34: 409–421, doi: 10.1051/gse:2002015: 13 p.

Concannon P., Rich S.S., Nepom G.T. 2009. Genetics of type 1A diabetes. The New England Journal of Medicine, 360, 16: 1646–1654

Coppieters W., Riquet J., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Nezer C., Simon P., Vanmanshoven P., Wagenaar D., Georges M. 1998. A QTL with major effect on milk yield and composition maps to bovine Chromosome 14. Mammalian Genome, 9: 540–544

Couzin-Frankel J. 2010. Major heart disease genes prove elusive. Science, 328, 5983: 1220–1221

Curik I., Druml T., Seltenhammer M., Sundström E., Pielberg G.R., Andersson L., Sölkner J. 2013. Complex Inheritance of Melanoma and Pigmentation of Coat and Skin in Grey Horses. PLoS Genetics, 9, 2: e1003248, doi: 10.1371/journal.pgen.1003248: 9 p.

Daetwyler H.D., Pong-Wong R., Villanueva B., Woolliams J.A. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185: 1021–1031

Daetwyler H.D., Schenkel F.S., Sargolzaei M., Robinson J.A. 2008. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. Journal of Dairy Science, 91: 3225–3236

Darvasi A., Weinreb A., Minke V., Weller J.I., Soller M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. Genetics, 134: 943–951

De Roos A.P.W., Hayes B.J, Spelman R., Goddard M.E. 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey and Angus cattle. Genetics, 179: 1503–1512

Dekkers J.C., Hospital F. 2002. The use of molecular genetics in the improvement of agricultural populations. Nature Reviews Genetics, 3: 22–32

Dekkers J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. Journal of Animal Science, 82: 313–328

Devlin B., Risch N. 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics, 29: 311–322

Devlin B., Roeder K. 1999. Genomic control for association studies. Biometrics, 55: 997–1004

Distefano J.K., Taverna D.M. 2011. Technological issues and experimental design of gene association studies. In: Disease Gene Identification. Methods and Protocols. Distefano J.K. (ed.). New York, Dordrecht, Heidelberg, London, Springer: 1–13

Duijvesteijn N., Knol E.F., Merks J.W., Crooijmans R.P., Groenen M.A., Bovenhuis H., Harlizius B. 2010. A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. BMC Genetics, 11: 42

Easton D., Eeles R. 2008. Genome-wide association studies in cancer. Human Molecular Genetics, 17: 109–115

Eggen A. 2012. The development and application of genomic selection as a new breeding paradigm. Animal Frontiers, 2: 10–15

Emmerling R., Lidauer M., Mantysaari E.A. 2002. Multiple lactation random regression test-day model for Simmental and Brown Swiss in Germany and Austria. Interbull Bulletin, 29: 111–117

Falconer D.S., Mackay T.F.C. 1996. Introduction to Quantitative Genetics. 4th ed. Harlow, Essex, Longman: 480 p.

Ferčej J., Pogačar J., Lovšin N., Nendl B. 1980. Oplemenjevanje rjave pasme z ameriško rjavo. Sodobno kmetijstvo, 4: 136–139

Fernando R.L., Grossman M. 1989. Marker-assisted selection using best linear unbiased prediction. Genetics Selection Evolution, 21: 467–477

Fischer G., Ibrahim S.M., Brockmann G.A., Pahnke J., Bartocci E., Thiesen H.J., Serrano-Fernández P., Möller S. 2003. Expressionview: visualization of quantitative trait loci and gene-expression data in Ensembl. Genome Biology, 4: R77, doi: 10.1186/gb-2003-4-11-r77: 6 p.

Fisher R.A. 1935. The Design of Experiments, 3rd ed. London, Oliver and Boyd: 11–25

Frayling T. 2007. Genome-wide association studies provide new insights into type 2 diabetes aetiology. Nature Review Genetics, 8: 657–662

Fürst C., Dodenhoff J., Egger-Danner C., Emmerling R., Henning H., Krogmeier D., Schwarzenbacher H. 2013. Zuchtwertschätzung beim Rind - Grundlagen, Methoden und Interpretationen. Wien, ZuchtData EDV-Dienstleistungen GmbH, 31–73

Garcia-Gomez E, Gutierrez-Gil B, Sahana G, Sanchez J-P, Bayon Y, Arranz J.J. 2012. GWA Analysis for Milk Production Traits in Dairy Sheep and Genetic Support for a QTN Influencing Milk Protein Percentage in the LALBA Gene. PLoS One, 7, 10: e47782, doi: 10.1371/journal.pone.0047782: 9 p.

Garrick D.J., Taylor J.F., Fernando R.L. 2007. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution, 41: 55, doi: 10.1186/1297-9686-41-55: 8 p.

Gengler N., Mayeres P., Szydlowski M. 2007. A simple method to approximate gene content in large pedigree populations: application to the myostatin gene in dual-purpose Belgian Blue cattle. Animal, 1: 21–28

Gianola D., Hospital F., Verrier E. 2013. Contribution of an additive locus to genetic variance when inheritance is multi-factorial with implications on interpretation of GWAS. Theoretical and Applied Genetics, 126: 1457–1472

Goddard M.E. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica, 136: 245–257

Goddard M.E., Hayes B.J. 2007. Genomic selection. Journal of Animal Breeding and Genetics, 124: 323–330

Goddard M.E., Hayes B.J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Review Genetics, 10: 381–391

Grindflek E., Lien S., Hamland H., Hansen M.H., Kent M., van Son M., Meuwissen T.H. 2011. Large scale genome-wide association and LDLA mapping study identifies QTLs for boar taint and related sex steroids. BMC Genomics, 12: 362, doi: 10.1186/1471-2164-12-362: 16 p.

Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M., Snell R. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Research, 12: 222–231

Gu X., Feng C., Ma L., Song C., Wang Y., Da Y., Li H., Chen K., Ye S., Ge C., Hu X., Li N. 2011. Genome-wide association study of body weight in chicken F2 resource population. PLoS One, 6: e21872, doi: 10.1371/journal.pone.0021872: 5 p.

Guo G., Zhao F., Wang Y., Zhang Y., Du L., Su G. 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genetics, 15: 30, doi: 10.1186/1471-2156-15-30: 7 p.

Guo J., Jorjani H., Carlborg O. 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genetics, 13: 82–92, doi: 10.1186/1471-2156-13-82: 10 p.

Haagsman H.P., De Haas C.G., Geelen M.J., Van Golde L.M. 1982. Regulation of triacylglycerol synthesis in the liver. Modulation of diacylglycerol acyltransferase activity in vitro. The Journal of Biological Chemistry, 257: 10593–10598

Harder B., Bennewitz J., Reinsch N., Thaller G., Thomsen H., Kuhn C., Schwerin M., Erhardt G., Förster M., Reinhardt F., Kalm E. 2006. Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. Journal of Animal Breeding and Genetics, 123: 89–96

Harlizius B., Lopes M.S., Duijvesteijn N., Van de Goor L.H.P., Van Haeringen W.A., Panneman H., Guimarães S.E.F., Merks J.W.M., Knol E.F. 2011. A SNP set for paternal identification to reduce the costs of trait recording in commercial pig breeding. Journal of Animal Science, 89: 1661–1668

Hayes B., Goddard M.E. 2001. The distribution of the effects of genes affecting quantitative traits in livestock. Genetics Selection Evolution, 33: 209–230

Hayes B.J., Bowman P.J, Chamberlain A.J., Goddard M.E. 2009a. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science, 92: 433–443

Hayes B.J., Bowman P.J., Chamberlain A.J., Savin K., van Tassell C.P., Sonstegard T.S., Goddard M.E. 2009b. A validated genome wide association study to breed cattle

adapted to an environment altered by climate change. PLoS One, 4, 8: e6676, doi:10.1371/journal.pone.0006676: 8 p.

Hayes B.J., Kinghorn B.P., Ruvinsky A. 2005. Genome scanning for quantitative trait loci. In: Mammalian genomics. Ruvinsky A., Marshall Graves J.A. (eds.). Oxon, CAB International: 507–537

Hayes B.J., Pryce J., Chamberlain A.J., Bowman P.J., Goddard M. E. 2010. Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. PLoS Genetics, 6: e1001139, doi:10.1371/journal.pgen.1001139: 11 p.

HGSC (Human Genome Sequencing Center), 2014. Bovine Genome Project. https://www.hgsc.bcm.edu/other-mammals/bovine-genome-project (15th May 2014)

Henderson C.R. 1984. Applications of linear models in animal breeding. Guelph, University of Guelph: 462 p.

Hirschhorn J.N., Daly M.J. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics, 6: 95–108

Hoerl A.E., Kennard R.W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12: 55–67

Hosking L., Lumsden S., Lewis K., Yeo A., McCarthy L., Bansal A., Riley, J., Purvis I., Xu C.F. 2004. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. European Journal of Human Genetics, 12: 395–399

Interbull 2014. Genetic Evaluations. Information of National and International Evaluations. Description of GES as applied in member countries. http://www.interbull.org/ib/geforms (15th April 2014)

Jia Y., Jannink J.L. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. Genetics, 192: 1513–1522

Jiang L., Liu J., Sun D., Ma P., Ding X., Yu Y., Zhang Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One, 5, 10: e13661, doi: 10.1371/journal.pone.0013661: 12 p.

Johnston D.J., Jeyaruban G.J., Graser H.U. 2010. Evaluation of Pfizer Animal Genetics HD 50K MVP calibration. http://agbu.une.edu.au/pdf/Pfizer_50K_September%202010.pdf (28th March 2012)

Johnston D.J., Tier B., Graser H.U. 2012. A Beef cattle breeding in Australia with genomics: opportunities and needs. Animal Production Science, 52: 100–106

Jorjani H., Jakobsen J., Hjerpe E., Palucci V., Dürr J. 2012. Status of genomic evaluation in the Brown Swiss populations. Interbull Bulletin, 46: 46–54

Jorjani H., Jakobsen J., Nilforooshan M.A., Hjerpe E., Zumbach B., Palucci V., Dürr J. 2011. Genomic evaluation of BSW populations interGenomics: results and deliverables. Interbull Bulletin, 43: 5–8

Kachman S.D. 2008. Incorporation of marker scores into national genetic evaluations. In: Proceedings of the 9th Genetic Prediction Workshop Beef Improvement Federation, Kansas City, National beef cattle evaluation consortium: 92–98

Kang H.M., Zaitlen N.A., Wade C.M., Kirby A., Heckerman D., Daly M.J., Eskin E. 2008. Efficient control for population structure in model organism association mapping. Genetics, 178: 1709–1723

Kappes S.M., Keele J.W., Stone R.T., McGraw R.A., Sonstegard T.S., Smith T.P.L., Lopez-Corrales N.L., Beattie C.W. 1997. A second generation linkage map of bovine genome. Genome Research, 7: 235–249

Kemper K.E., Goddard M.E. 2012. Understanding and predicting complex traits: knowledge from cattle. Human Molecular Genetics, 21: 45–51

König S., Chang Y.M., Borstel U.U., Gianola D., Simianer H. 2008. Genetic and phenotypic relationships among milk urea nitrogen, fertility, and milk yield in Holstein cows. Journal of Dairy Science, 91: 4372–4382

Kovač M., Groeneveld E., Garcia Cortes L. 2002. VCE-5, a package for the estimation of dispersion parameters. In: 7th World Congress on Genetics Applied to Livestock Production, Montpellier, 19-23 Aug. 2002. Castanet-Tolosan, INRA: 741–742

Langlet J. 1965. Genetic relations between beef and milk performance – a survey of knowledge and investigations in this field. World Review of Animal Production, 1: 31–34

Lau T.E., Rodriguez M.A. 1996. A protein tyrosine kinase associated with the ATP-dependent inactivation of adipose diacylglycerol acyltransferase. Lipids, 31: 277–283

Legarra A., Bertrand J.K., Stradel T., Sanchez J.P., Misztal I. 2007. Multi-breed genetic evaluation in a Gelbvieh population. Journal of Animal Breeding and Genetics, 124: 286–295

Lemmon Z.H., Doebley J.F. 2014. Genetic Dissection of a Genomic Region with Pleiotropic Effects on Domestication Traits in Maize Reveals Multiple Linked QTL. Genetics, 198: 345–353

Liinamo A.E., Ojala M., Van Arendonk J. 2001. Genetic relationship of meat and milk production in Finnish Ayrshire. Livestock Production Science, 69: 1–8

Liu W., Li D., Liu J., Chen S., Qu L., Zheng J., Xu G., Yang N. 2011. A genome-wide SNP scan reveals novel loci for egg production and quality traits in white leghorn and brown-egg dwarf layers. PLoS One, 6, 12: e28600, doi:10.1371/journal.pone.0028600: 8 p.

Liu Z., Seefried F.R., Reinhardt F., Rensing S., Thaller G., Reents R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genetics Selection Evolution, 43: 19–28

Logsdon B.A., Carty C.L., Reiner A.P., Dai J.Y., Kooperberg C. 2012. A novel variational Bayes multiple locus Z-statistic for genome-wide association studies with Bayesian model averaging. Bioinformatics, 28: 1738–1744

Looft C., Reinsch N., Karall-Albrecht C., Paul S., Brink M., Thomsen H., Brockmann G., Kuhn C., Schwerin M. Kalm E. 2001. A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14. Mammalian Genome, 12: 646–650

Lu S., Miller S., Sargolzaei M., Kelly M., Vander Voort G., Caldwell T., Wang Z., Plastow G. Moore S. 2013. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. Journal of Animal Science, 91: 3612-3633

Ma L., Wiggans G.R., Wang S, Sonstegard T.S, Yang J., Crooker B.A, Cole J.B, Van Tassell C.P., Lawlor T.J, Da Y. 2012. Effect of sample stratification on dairy GWAS results. BMC Genomics, 13: 536, doi: 10.1186/1471-2164-13-536: 17 p.

MacNeil M.D., Nkrumah J.D., Northcutt S.L. 2010. Genetic evaluation of Angus cattle for carcass marbling using ultrasound and genomic indicators. Journal of Animal Science, 88: 517–522

MacNeil M.D., Northcutt S.L., Schnabel R.D., Garrick D.J., Woodward B.W., Taylor J.F. 2010. Genetic correlations between carcass traits and molecular breeding values in Angus cattle. In: 9th World Congres on Genetics Applied to Livestock Production, Leipzig, 1-6 Aug. 2010. Leipzig, German Society for Animal Science, Manuscript 0482: 4 p.

Maher B. 2008. Personal genomes: The case of the missing heritability. Nature, 456: 18–21

Mai M.D., Sahana G., Christiansen F.B., Guldbrandtsen B. 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. Journal of Animal Science, 88: 3522–3528

Makowsky R., Pajewski N.M., Klimentidis Y.C., Vazquez A.I., Duarte C.W., Allison D.B., De los Campos G. 2011. Beyond missing heritability: prediction of complex traits. Plos Genetics, 7: e1002051, doi: 10.1371/journal.pgen.1002051: 9 p.

Manolio T.A., Collins F.S., Cox N.J., Goldstein D.B., Hindorff L.A., Hunter D.J., McCarthy M.I., Ramos E.M., Cardon L.R., Chakravarti A., Cho J.H., Guttmacher A.E., Kong A., Kruglyak L., Mardis E., Rotimi C.N., Slatkin M., Valle D., Whittemore A.S., Boehnke M., Clark A.G., Eichler E.E., Gibson G., Haines J.L., Mackay T.F., McCarroll S.A. Visscher P.M. 2009. Finding the missing heritability of complex diseases. Nature, 461, 7265: 747-753

Mäntysaari E.A., Strandén I. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding values evaluations. In: 9th World Congres on Genetics Applied to Livestock Production, Leipzig, 1-6 Aug. 2010. Leipzig, German Society for Animal Science, Manuscript 0353: 4 p.

Mason I.L. 1964. Genetic relations between milk and beef characters in dual-purpose cattle breeds. Animal Production, 6: 31–45

Massey D.C., Parkes M. 2007. Genome-wide association scanning highlights two autophagy genes, ATG16L1 and IRGM, as being significantly associated with Crohn's disease. Autophagy, 3: 649–651

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4, 4: e5350, doi: 10.1371/journal.pone.0005350: 13 p.

Maxa J., Neuditschko M., Russ I., Förster M., Medugorac I. 2012. Genome-wide association mapping of milk production traits in Braunvieh cattle. Journal of Dairy Science, 95: 5357–5364

McCarthy M.I., Abecasis G.R., Cardon L.R., Goldstein D.B., Little J., Ioannidis J.P. Hirschhorn J.N. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics, 9, 5: 356–369

McKay S.D., Schnabel R.D., Murdoch B.M., Matukumalli L.K., Aerts J., Coppieters W., Crews D., Neto E.D., Gill C.A., Gao C., Mannen H., Stothard P., Wang Z., Van Tassell C.P., Williams J.L., Taylor J.F., Moore S.S. 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genetics, 8: 74

Meng J.F., Fingerlin T.E. 2008. Linear models for analysis of multiple single nucleotide polymorphisms with quantitative traits in unrelated individuals. Annales Zoologici Fennici, 45: 429–440

Meurs K.M., Mauceli E., Lahmers S., Acland G.M., White S.N., Lindblad-Toh K. 2010. Genome-wide association identifies a deletion in the 3' untranslated region of striatin in a canine model of arrhythmogenic right ventricular cardiomyopathy. Human Genetics, 2128: 315–324

Meuwissen T.H.E., Goddard M.E. 1996. The use of marker haplotypes in animal breeding schemes. Genetics Selection Evolution, 28: 161–176, doi: 10.1186/1297-9686-28-2-161: 16 p.

Meuwissen T.H.E., Hayes B.H., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157: 1819–1829

Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2013. Accelerating improvement of livestock with genomic selection. Annual Review of Animal Biosciences, 1: 221–237

Meyer K. 2007. WOMBAT: A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). Journal of Zhejiang University Science, 8: 815–821

Meyer K., Tier B. 2012. SNP Snappy: A strategy for fast genome wide association studies fitting a full mixed model. Genetics, 190: 275–277

Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. 2002. Blupf90 and related programs. In: 7th World Congress on Genetics Applied to Livestock Production, Montpellier, 19-23 Aug. 2002. Castanet-Tolosan, INRA: 28: 1-2 http://nce.ads.uga.edu/wiki/doku.php (15th April 2013)

Misztal I., Wiggans G.R. 1988. Approximation of prediction error variance in large-scale animal models. Journal of Dairy Science, 71: 27–32

Mogensen M.S., Karlskov-Mortensen P., Proschowsky H.F., Lingaas F., Lappalainen A., Lohi H., Jensen V.F., Fredholm M. 2011. Genome-wide association study in Dachshund: identification of a major locus affecting intervertebral disc calcification. Journal of Heredity, 1: 81–86

Montaldo H.H., Casas E., Sterman Ferraz J.B., Vega-Murillo V.E., Román-Ponce S.I. 2012. Opportunities and challenges from the use of genomic selection for beef cattle breeding in Latin America. Animal Frontiers, 2: 23–29

Murdoch B.M., Murdoch G.K., Settles M., McKay S., Williams J.L., Moore S.S. 2011. Genome-wide scan identifies loci associated with classical BSE occurrence. PLoS One 6, 11: e26819, doi: 10.1371/journal.pone.0026819: 7 p.

Neimann-Sorensen A., Robertson A. 1961. The Association between Blood Groups and Several Production Characteristics in Three Danish Cattle Breeds. Acta Agriculturae Scandinavica, 11: 163–196

Olsen H.G., Hayes B.J., Kent M.P., Nome T., Svendsen M., Larsgard A.G., Lien S. 2011. Genome-wide association mapping in Norwegian Red cattle identifies quantitative trait loci for fertility and milk production on BTA12. Animal Genetics, 42: 466–474

Oltenacu P.A., Frick A., Lindhé B. 1991. Relationship of fertility to milk yield in Swedish cattle. Journal of Dairy Science, 74: 264–268

Patterson N., Price A.L., Reich D. 2006. Population structure and eigen analysis. PLoSGenetics, 2, 12: e190, doi: 10.1371/journal.pgen.0020190: 20 p.

Pausch H., Flisikowski K., Jung S., Emmerling R., Edel C., Götz K.-U., Fries R. 2011. Genome-wide association study identifies two major loci affecting calving ease and growth-related traits in cattle. Genetics, 187: 289–297

Pausch H., Jung S., Edel C., Emmerling R., Krogmeier D., Götz K.-U., Fries R. 2012. Genome-wide association study uncovers four QTL predisposing to supernumerary teats in cattle. Animal Genetics, 43: 689–695

Pausch H., Wang X., Jung S., Krogmeier D., Edel C., Emmerling R., Götz K.-U., Fries R. 2012. Identification of QTL for UV-protective eye area pigmentation in cattle by progeny phenotyping and genome-wide association analysis. PLoS One, 7: e36346, doi: 10.1371/journal.pone.0036346: 10 p.

Pausch H.H. 2013. Genome-wide analysis of complex traits in cattle. PhD thesis. München, Techische Universität, Lehrstuhl für Tierzucht: 141 p.

Pearson T.A., Manolio T.A. 2008. How to Interpret a Genome-Wide Association Study Journal of American Medical Association, 299, 11: 1335–1344

Pimentel E.C.G., Erbe M., König S., Simianer H. 2011. Genome partitioning of genetic variation for milk production and composition traits in Holstein cattle. Frontiers in Genetics, 2: 19

Pirchner F. 1986. Evaluation of industry breeding programs for dairy cattle milk and meat production. In: 3rd World Congress on Genetics Applied to Livestock Production, 16-22 July 1986, Lincoln, University of Nebraska, 9: 153–164

Potočnik K. 1999. Napovedi plemenskih vrednosti za lastnosti mlecnosti pri govedu na podlagi dnevne kontrole. (Prediction of breeding values for dairy production traits using test day model). Master of science thesis. Ljubljana, University of Ljubljana, Biotechnical Faculty, Zootehnical Department: 90 p.

Pritchard J.K., Stephens M., Rosenberg N., Donnelly P. 2000. Association mapping instructured populations. American Journal of Human Genetics, 67: 170–181

Pryce J.E., Bolormaa S., Chamberlain A.J., Bowman P.J., Savin K., Goddard M.E., Hayes B.J. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science, 93: 3331–3345

Pryce J.E., Esslemont R.J., Thompson R., Veerkamp R.F., Kossaibati M.A., Simm G. 1998. Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. Animal Science, 66: 577–584

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., De Bakker P.I.W., Daly M.J., Sham P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81: 559–575

Quaas R.L., Zhang Z. 2006. Multiple-breed genetic evaluation in the US beef cattle context: methodology. In: 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, 13-18 Aug. 2006. Belo Horizonte, Brazilian Society of Animal Breeding: 13–28

R Development Core Team 2011. R: A language and environment for statistical computing. Vienna, R Foundation for Statistical Computing. http://www.R-project.org (28th April 2012)

Ramos A.M., Duijvesteijn N., Knol E.F., Merks J.W., Bovenhuis H., Crooijmans R.P., Groenen M.A., Harlizius B. 2011. The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. BMC Genetics, 12: 35, doi: 10.1186/1471-2156-12-35: 7 p.

Roman R., Wilcox C. 2000. Bivariate animal model estimates of genetic, phenotypic, and environmental correlations for production, reproduction, and somatic cells in Jerseys. Journal of Dairy Science, 83: 829–835

Saatchi M., Schnabel R.D., Rolf M.M., Taylor J.F., Garrick D.J. 2012. Accuracy of direct genomic breeding values for nationally evaluated traits in US Limousin and Simmental beef cattle. Genetics Selection Evolution, 44: 38, doi: 10.1186/1297-9686-44-38: 9 p.

Saatchi M., McClure M.C., McKay S.D., Rolf M.M., Kim J.W., Decker J.E., Taxis T.M., Chapple R.H., Ramey H.R., Northcutt S.L., Bauck S., Woodward B., Dekkers J.C.M., Fernando R.L., Schnabel R.D., Garrick D.J., Taylor J.F. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution, 43: 40, doi: 10.1186/1297-9686-43-40: 16 p.

Sadar M., Jenko J., logar B., Opara A., Perpar T., Podgoršek P. 2013. Rezultati kontrole prireje mleka and mesa (Results of dairy and beef recording). Ljubljana, Kmetijski institut Slovenije: 89 p. http://www.govedo.si/files/cpzgss/knjiznica/porocila/kontrola_porocila/REZULTATI_KONTROLE_2013.pdf (14th April 2014)

Sahana G., Guldbrandtsen B., Lund M.S. 2011. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. Journal of Dairy Science, 94: 479–486

SAS Institute 2011. SAS/STAT Software, Version 9.3. Cary, NC

Schaeffer L.R. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics, 123: 218–223

Schaeffer L.R. 1994. Multiple-country comparison of dairy sires. Journal of Dairy Science, 77: 2671–2678

Schefers J.M., Weigel K.A. 2012. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Animal Frontiers, 1: 4–9

Schild H.J., Niebel E., Götz K.-U. 2003. Across country genetic evaluation of beef traits in middle European dual purpose breeds. Interbull Bulletin, 31: 158–162

Schopen G.C., Visker M.H., Koks P.D., Mullaart E., Van Arendonk J.A., Bovenhuis H. 2011. Whole genome association study for milk protein composition in dairy cattle. Journal Dairy Science, 94: 3148–3158

Seidel G.E. 2010. Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms. Reproduction Fertility and Development, 22: 138–144

Silverman EK, Palmer LJ. 2000. Case-control association studies for the genetics of complex respiratory diseases. American Journal of Respiratory Cell and Molecular Biology, 22: 645–648

Smith J.G., Newton-Cheh C. 2009. Genome-wide association study in humans. Methods in Molecular Biology, 573: 231–258

Snelling W., Cushman M., Fortes R.A., Reverter M.R., Bennett A., Keele G.L., Kuehn L.A., McDaneld T.G., Thallman R.M., Thomas M.G. 2012. Physiology and endocrinology symposium: how single nucleotide polymorphism chips will advance our knowledge of factors controlling puberty and aid in selecting replacement beef females. Journal of Animal Science, 90: 1152–1165

Solberg T.R., Sonesson A.K., Woolliams J.A., Meuwissen T.H.E. 2008. Genomic selection using different marker types and densities. Journal of Animal Science, 86: 2447–2454

Solovieff N., Cotsapas C., Lee P.H., Purcell S.M., Smoller J.W. 2013. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics, 14: 483–495

Sonstegard T.S., Ma L., Van Tassell C.P., Kim E.S., Cole J.B., Wiggans G.R., Crooker B.A., Mariani B.D., Matukumalli L.K., Garbe J.R., Fahrenkrug S.C., Liu G., Da Y. 2010. Forty years of artificial selection in U.S. Holstein cattle had genome-wide signatures. In: Poster presentation at 9th World Congres on Genetics Applied to Livestock Production, Leipzig, 1-6 Aug. 2010. Leipzig, German Society for Animal Science.
http://aipl.arsusda.gov/publish/presentations/WC9_10/WC9_10_yang_da.pdf (28th October 2014)

Sørensen L.P., Janss L., Madsen P., Mark T., Lund M.S. 2012. Estimation of (co)variances for genomic regions of flexible sizes: application to complex infectious udder diseases

in dairy cattle. Genetics Selection Evolution, 44: 18, doi: 10.1186/1297-9686-44-18: 14 p.

Spelman R., Garrick D. 1997. Utilisation of marker assisted selection in a commercial dairy cow population. Livestock Production Science, 47: 139–147

Strandén I. 2010. Manual for gpig program – pedigree based imputation of genotypes. Biotechnology and Food Research, Biometrical Genetics. Jokioinen, Finland, MTT Agrifood Research: 13 p.

Sun X., Habier D., Fernando R.L., Garrick D.J., Dekkers J.C.M. 2011. Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. BMC Proceedings, 5: S13, doi: 10.1186/1753-6561-5-S3-S13: 8 p.

Thaller G., Kühn C., Winter A., Ewald G., Bellmann O., Wegner J., Zühlke H., Fries R. 2003. DGAT1, a new positional and functional candidate gene for intramuscular fat deposition in cattle. Animal Genetics, 34: 354–357

The Bovine Genome Sequencing and Analysis Consortium 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science, 324: 522–528

The Bovine HapMap Consortium 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science, 324: 528–532

The Wellcome Trust Case Control Consortium 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature, 447: 661–678

Thomsen H., Reinsch N., Xu N., Bennewitz J., Looft C., Grupe S., Kühn C., Brockmann G.A., Schwerin M., Leyhe-Horn B., Hiendleder S., Erhard G., Medjugorac I., Russ I., Förster M., Brening B., Reinhard F., Reents R., Blümel J., Averdunk G., Kalm E. 2001. Comparison of estimated breeding value, daughter yield deviations and de-regressed proofs within a whole genome scan for QTL. Journal of Animal Breeding and Genetics, 118: 357–370

Thomsen H., Reinsch N., Xu N., Looft C., Gruppe S., Kuhn C., Brockmann G.A., Schwerin M., Leyhe-Horn B., Hiendleder S., Erhardt G., Medjugorac I., Russ I., Forster M., Brenig B., Reinhardt F., Reents R., Blumel J., Averdunk G., Kalm E. 2000. A male linkage map for the ADR granddaughter design. Journal of Animal Breeding and Genetics, 117: 289–306

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. Journal of Statistical Society. Series B, 58: 267–288

Van Tassell C.P., Smith T.P.L., Matukumalli L.K., Taylor J.F., Schnabel R.D., Lawley C.T., Haudenschild C.D., Moore S.S., Warren W.C., Sonstegard T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods, 5: 247-252

Vandenplas J., Colinet F.G., Gengler N. 2014. Unified method to integrate and blend several, potentially related, sources of information for genetic evaluation. Genetics Selection Evolution, 46: 59, doi: 10.1186/s12711-014-0059-3. 15 p.

Vandenplas J., Janssens S., Buys N., Gengler N. 2013. An integration of external information for foreign stallions into the Belgian genetic evaluation for jumping horses. Journal of Animal Breeding and Genetics, 130: 209–217

Vandenplas J., Gengler N. 2012. Comparison and improvements of different Bayesian procedures to integrate external information into genetic evaluations. Journal of Dairy Science, 95: 1513–1526

VanRaden P.M., Tooker M.E. 2012. Methods to include foreign information in national evaluations. Journal of Dairy Science, 95, Suppl. 2: 446-447

VanRaden P.M., Olson K.M., Null D.J., Sargolzaei M.M.W., van Kaam J.B.C.H.M. 2012. Reliability increases from combining 50,000- and 777,000-marker genotypes from four countries. Interbull Bulletin, 46: 75-79

VanRaden P.M., Van Tassell C.P., Wiggans G.R., Sonstegard T.S., Schnabel R.D., Taylor J.F., Schenkel F. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 92: 16–24

VanRaden P.M. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science, 91: 4414–4423

VanRaden P.M. 2001. Methods to combine estimated breeding values obtained from separate sources. Journal of Dairy Science, 84: E47–E55

Veerkamp R.F., Verbyla K.L., Mulder H.A., Calus M.P.L. 2010. Simultaneous QTL detection and genomic breeding value estimation using high density SNP chips. BMC Proceedings, 4: S9, doi: 10.1186/1753-6561-4-S1-S9: 4 p.

Viitala S.M., Schulman N.F., Koning D.J., Elo K., Kinos R., Virta A., Virta J., Tanila A.M., Vilkki J.H. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. Journal of Dairy Science, 86: 1828–1836

Visscher P.M., Macgregor S., Benyamin B., Zhu G., Gordon S., Medland S., Hill W.G., Hottenga J.J., Willemsen G., Boomsma D.I., Liu Y.Z., Deng H.W., Montgomery G.W., Martin N.G..2007. Genome partitioning of genetic variation for height from 11,214 sibling pairs. American Journal of Human Genetics, 81: 1104–1110

Visscher PM, Goddard ME 2004. Prediction of the Confidence Interval of Quantitative Trait Loci Location. Behavior Genetics, 34: 477–482

Wang H., Jiang L., Liu X., Yang J., Wei J., Xu J., Zhang Q., Liu J.F. 2013. A post-GWAS replication study confirming the PTK2 gene associated with milk production traits in Chinese Holstein. PloS one, 8: e83625, doi: 10.1371/journal.pone.0083625: 12 p.

Wang Z., Zhang H., Yang H., Wang S., Rong E., Pei W., Li H., Wang N. 2014. Genome-Wide Association Study for Wool Production Traits in a Chinese Merino Sheep Population. PLoS One, 9: e107101, doi: 10.1371/journal.pone.0107101: 8 p.

Weller J.I., Glick G., Ezra E., Zeron Y., Seroussi E., Ron M. 2011. Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. Animal Genetics, 41: 551–553

Weller J.I., Kashi Y., Soller M. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. Journal of Dairy Science, 73, 9: 2525-37

Welper R., Freeman A. 1992. Genetic parameters for yield traits of Holsteins including lactose and somatic cell score. Journal of Dairy Science, 75: 1342–1348

Wiener P., Wooliams J.A., Frank-Lawale A., Ryan M., Richardson R.I., Nute G.R., Wood J.D., Homer D., Williams J.L. 2009. The effects of a mutation in the myostatin gene on meat and carcass quality. Meat Science, 83: 127–134

Wiggans G.R., Sonstegard T.S., VanRaden P.M., Matukumalli L.K., Schnabel R.D., Taylor J.F., Schenkel F.S., Van Tassell C.P. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. Journal of Dairy Science, 92: 3431–3436

Winter A., Kramer W., Werner F.A.O., Kollers S., Kata S., Durstewitz G., Buitkamp J., Womack J.E., Thaller G., Fries R. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl- CoA: diacylglycerol acyltransferase (DGAT1) with variation at a quantitative trait locus for milk fat content. Proceedings of the National Academy of Sciences of the USA, 16: 9252–9257

Wood S.H., Ke X., Nuttall T., McEwan N., Ollier W.E., Carter S.D. 2009. Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. Immunogenetics, 61:765–772

Xie L., Luo C., Zhang C., Zhang R., Tang J., Nie Q., Ma L., Hu X., Li N., Da Y., Zhang X. 2012. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. PLoS One, 7, 2: e30910, doi: 10.1371/journal.pone.0030910: 9 p.

Yang J., Manolio T.A., Pasquale L.R., Boerwinkle E., Caporaso N., Cunningham J.M., De Andrade M., Feenstra B., Feingold E., Hayes M.G., Hill W.G., Landi M.T., Alonso A., Lettre G., Lin P., Ling H., Lowe W., Mathias R.A., Melbye M., Pugh E., Cornelis M.C., Weir B.S., Goddard M.E., Visscher P.M. 2011. Genome partitioning of genetic variation for complex traits using common SNPs. Nature Genetics, 43: 519–525

Yoder D.M., Lush J.L. 1937. A genetic history of the Brown Swiss cattle in the United States. Journal of Heredity, 28: 154–160

Zhang L, Liu J., Zhao F., Ren H., Xu L., Lu J., , Zhang S., Zhang X., Wei C., Lu G., Zheng Y. , Du L. 2013. Genome-Wide Association Studies for Growth and Meat Production Traits in Sheep. PLoS ONE 8, 6: e66569, doi: 10.1371/journal.pone.0066569: 12 p.

Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D. 2009. A whole genome assembly of the domestic cow, Bos taurus. Genome Biology, 10, 4: R42, doi: 10.1186/gb-2009-10-4-r42: 10 p.

Zou H., Hastie T. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B, 67, 2: 301–320

Zumbach B., Jorjani H., Dürr J. 2010. Brown Swiss Genomic Evaluation. Interbull Bulletin, 42: 45–51

ACKNOWLEDGEMENTS

First and foremost I want to thank my supervisor Assist. Prof. Dr. Gregor Gorjanc for his guidance and motivation during the research, inspirational discussions and comments about writing. His wide knowledge and his logical way of thinking have been of great value to me. I gratefully acknowledge my co-supervisor Prof. Dr. Ino Čurik for his advices, helpful comments, and for his outstanding support during the research. I appreciate his contribution, especially for his 'biological way' of thinking. Both of them encouraged me during the research, from which I have learned valuable things that enriched me professionally and personally. Thanks for being my mentors.

For this dissertation, I would like to thank my reading committee members: Prof. Dr. Peter Dovč, Assist. Prof. Dr. Klemen Potočnik, and Prof. Dr. Johann Söelkner for their time, interest, and helpful comments.

I would like to acknowledge my employment company, the Croatian Agricultural Agency, and the people who made this thesis possible: former director Dr. Zdravko Barać who allowed me to spend a time in Slovenia and current Dr. Marija Vukobratović. I gratefully thank Dr. Maja Dražić for her inspirational and technical support and her enthusiasm which was contagious and a precious motivational for me. Thanks to my co-workers in the Department of breeding value estimation, Mandica Lučić B.Sc. and Dr. Anamarija Smetko who were always ready to help. I am grateful to Dr. Milan Zjalić for his useful advices and comments about breeding and selection trends at the international level.

For providing data which enabled the research, I would like to acknowledge Biotechnical Faculty, Department for Animal Science for providing phenotypic, pedigree, and SNP marker data for Brown Swiss breed. I also want to express my gratitude to Dr. Kay-Uwe Götz, Dr. Reiner Emmerling, Dr. Christian Edel from the Bavarian State Research Center, Institute of Animal Breeding (LfL-ITZ) for providing phenotypic, pedigree, and SNP marker data for the German Fleckvieh breed in front of German breeding association and for fruitful discussions during our visits in Grub. Furthermore, I would also like to acknowledge Dr. Daniele Vicario from ANAPRI (Anapri-Associazione Nazionale Allevatori Pezzata Rossa Italiana) for the provision of phenotypic, pedigree, and SNP marker data for Italian Pezzata Rossa breed. Furthermore, I am thankful to Dr. Daša Jevšinek Skok and Prof. Dr. Tanja Kunej for their help with the bioinformatics analyses.

I am grateful to all co-workers in Domžale, and I wish to extend my warmest thanks to Vladimir Štepec for useful advices and SAS support. I would also like to acknowledge Mojca Simčič B.Sc. for inspirational discussions and friendly support during my research. I am very grateful to them in translating the summary into Slovenian language. I am also deeply thankful to Dr. Miran Kastelic for his advices and assistance in R programing.

The last, but not the least, I would like to thank my family for all their love and encouragement. A special thanks goes to my parents who supported me in all my pursuits.

I am grateful to my closest friends who were always prepared for long conversations.

ANNEXES

Annex A1: Estimates of total genetic and residual variance components and ratios (± standard errors)* for the pairs of dairy and beef traits** in the German Fleckvieh and Italian Pezzata Rossa population based on the analysis with the pedigree or genomic relationship matrix

Priloga A1: Ocene komponent skupne genetske variance in variance ostanka ter razmerja (± standardne napake)* za pare lastnosti mlečnosti in mesnatosti** pri nemški lisasti in italijanski lisasti populaciji na osnovi analize z rodovniki ali z genomsko matriko sorodstva

Trait	Variance				Ratio			
	Genetic, σ_a^2		Residual, σ_e^2		Genetic, h^2		Residual, e^2	
	Pedigree	Genomic	Pedigree	Genomic	Pedigree	Genomic	Pedigree	Genomic
German Fleckvieh (n=4,105)								
MY	24872.3±7415.8	207677.0±8296.0	5458.0±2850.4	25127±2627.4	0.98±0.01	0.89±0.01	0.02±0.01	0.11±0.01
FY	337.0±9.3	290.0±11.6	22.3±3.5	31.7±3.6	0.99±0.01	0.90±0.01	0.01±0.01	0.10±0.01
NG	133.4±10.1	131.6±6.9	53.8±5.7	48.8±2.8	0.71±0.04	0.73±0.02	0.29±0.04	0.27±0.02
CQ	5.0±0.4	4.9±0.3	2.7±0.2	2.5±0.1	0.65±0.04	0.66±0.02	0.35±0.03	0.34±0.02
Italian Pezzata Rossa (n=511)								
MY	224303.6.0±2567.3	199404.7±2161.8	4835.3±5818.9	24726.5±8084.6	0.98±0.07	0.89±0.03	0.02±0.02	0.11±0.03
FY	377.6±42.4	337.0±42.7	28.2±28.2	51.3±28.8	0.93±0.07	0.87±0.08	0.07±0.07	0.13±0.08
NG	112.2±19.6	87.6±17.1	33.8±27.5	73.4±22.6	0.77±0.16	0.54±0.12	0.23±0.16	0.46±0.12
CQ	95.8±14.9	89.1±13.8	27.2±17.3	30.7±15.7	0.78±0.13	0.78±0.12	0.22±0.13	0.26±0.12

*values represent averages from multiple bivariate analyses; **MY = milk yield; FY = fat yield; NG = net daily gain; CQ = carcass quality

Annex A2: Estimates of total genetic and residual covariance components and correlations (± standard errors) for the pairs of dairy and beef traits* in the German Fleckvieh and Italian Pezzata Rossa population based on the analysis with the pedigree or genomic relationship matrix

Priloga A2: Ocene komponent skupne genetske kovariance in kovariance ostanka ter korelacije (± standardne napake) za pare lastnosti mlečnosti in mesnatosti* v populacijah nemške lisaste in italijanske lisaste pasme na osnovi analize z rodovnikom ali z genomsko matriko sorodstva

Trait comb.	Covariance				Correlation			
	Genetic, σ_{a_1,a_2}		Residual, σ_{e_1,e_2}		Genetic, r_{a_1,a_2}		Residual, r_{e_1,e_2}	
	Pedigree	Genomic	Pedigree	Genomic	Pedigree	Genomic	Pedigree	Genomic
				German Fleckvieh (n=4,105)				
MY:FY	6677.4±228.4	5556.4±269.0	-64.1±79.2	624.3±83.8	0.74±0.01	0.72±0.01	-0.41±0.68	0.72±0.04
MY:NG	541.2±204.9	484.2±169.3	93.9±104.9	109.3±59.0	0.09±0.03	0.09±0.03	0.20±0.21	0.10±0.05
MY:CQ	-171.7±38.9	-108.7±34.0	59.8±20.2	23.3±12.0	0.16±0.04	-0.11±0.03	0.52±0.15	0.09±0.04
FY:NG	17.4±7.2	7.2±6.2	-3.0±3.6	3.8±2.1	0.08±0.04	0.04±0.03	-0.28±0.33	0.10±0.05
FY:CQ	-6.6±1.4	-5.5±1.3	1.4±0.7	0.9±0.4	-0.16±0.04	-0.15±0.03	0.51±0.23	0.11±0.05
NG:CQ	14.9±1.6	13.7±1.1	4.6±0.9	5.1±0.5	0.57±0.04	0.53±0.03	0.39±0.05	0.46±0.03
				Italian Pezzata Rossa (n=511)				
MY:FY	7268.6±520.8	6438.8±520.5	-170.4±530.9	512.1±460.1	0.90±0.03	0.91±0.04	-0.25±0.96	0.40±0.25
MY:NG	510.9±378.7	411.8±403.8	-290.9±228.9	-122.3±387.0	0.10±0.08	0.10±0. 09	-0.95±0.27	-0.09±0.27
MY:CQ	-1030.9±367.8	-1032.6±329.8	610.9±244.4	833.6±358.1	-0.24±0.09	-0.25±0.09	0.95±0.16	0.97±0.03
FY:NG	-4.4±20.2	-13.3±18.0	5.3±18.0	18.9±16.1	-0.02±0.10	-0.08±0.11	0.22±0.73	0.31±0.26
FY:CQ	-48.5±16.8	-28.7±16.8	39.7±17.0	29.6±13.5	-0.27±0.11	-0.17±0.11	0.97±0.08	0.70±0.30
NG:CQ	46.8±13.3	47.2±11.9	15.8±14.5	12.5±10.7	0.42±0.09	0.51±0.09	0.71±0.39	0.31±0.25

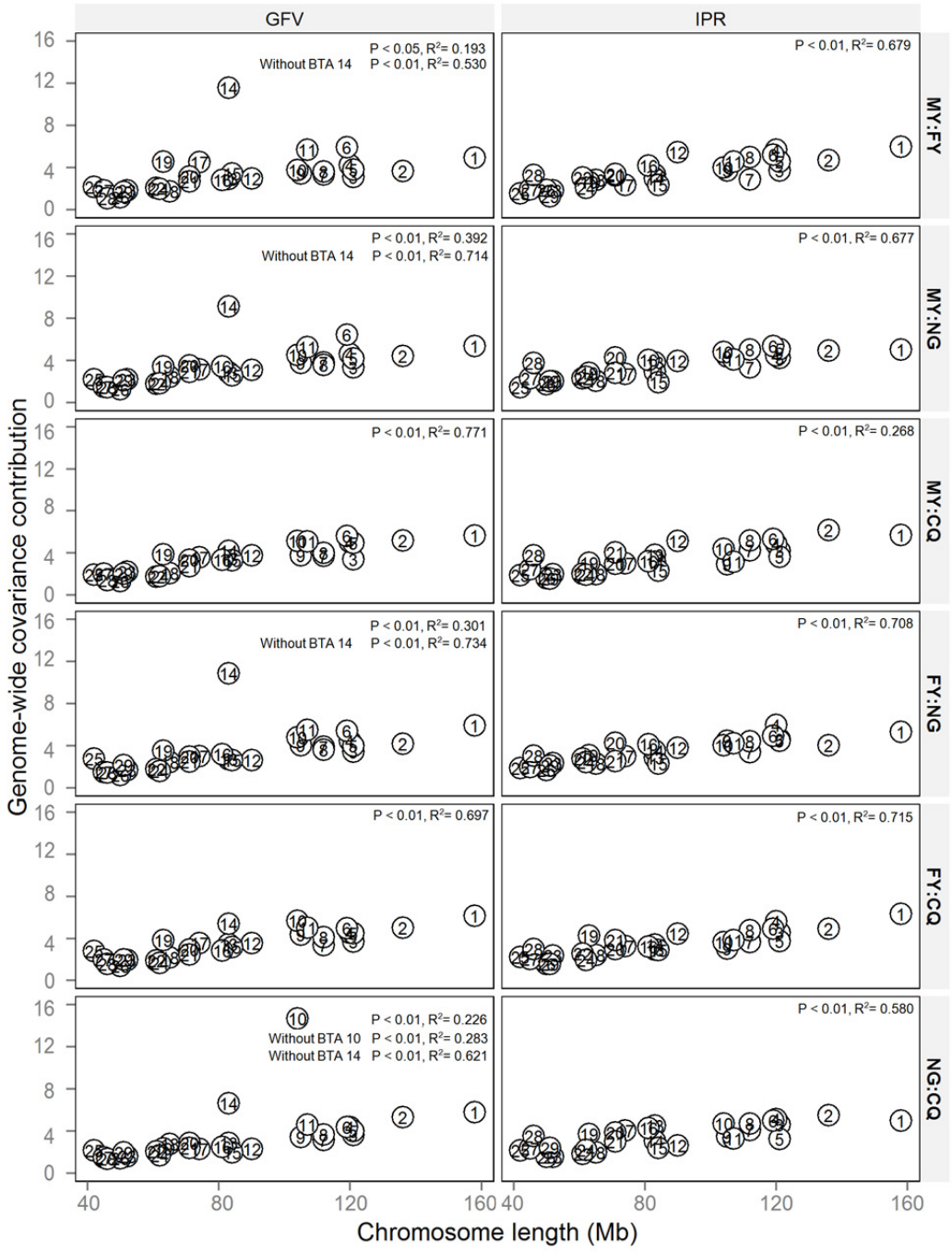
*MY = milk yield; FY = fat yield; NG = net daily gain; CQ = carcass quality

Table A3: Total genetic and genome-wide covariation for the pairs of dairy and beef traits* in the German Fleckvieh population based on a random subset

Tabela A3: Skupna genetska in genomska kovariacija za pare lastnosti mlečnosti in mesnatosti* v populaciji nemške lisaste pasme na osnovi naključnega podseta podatkov

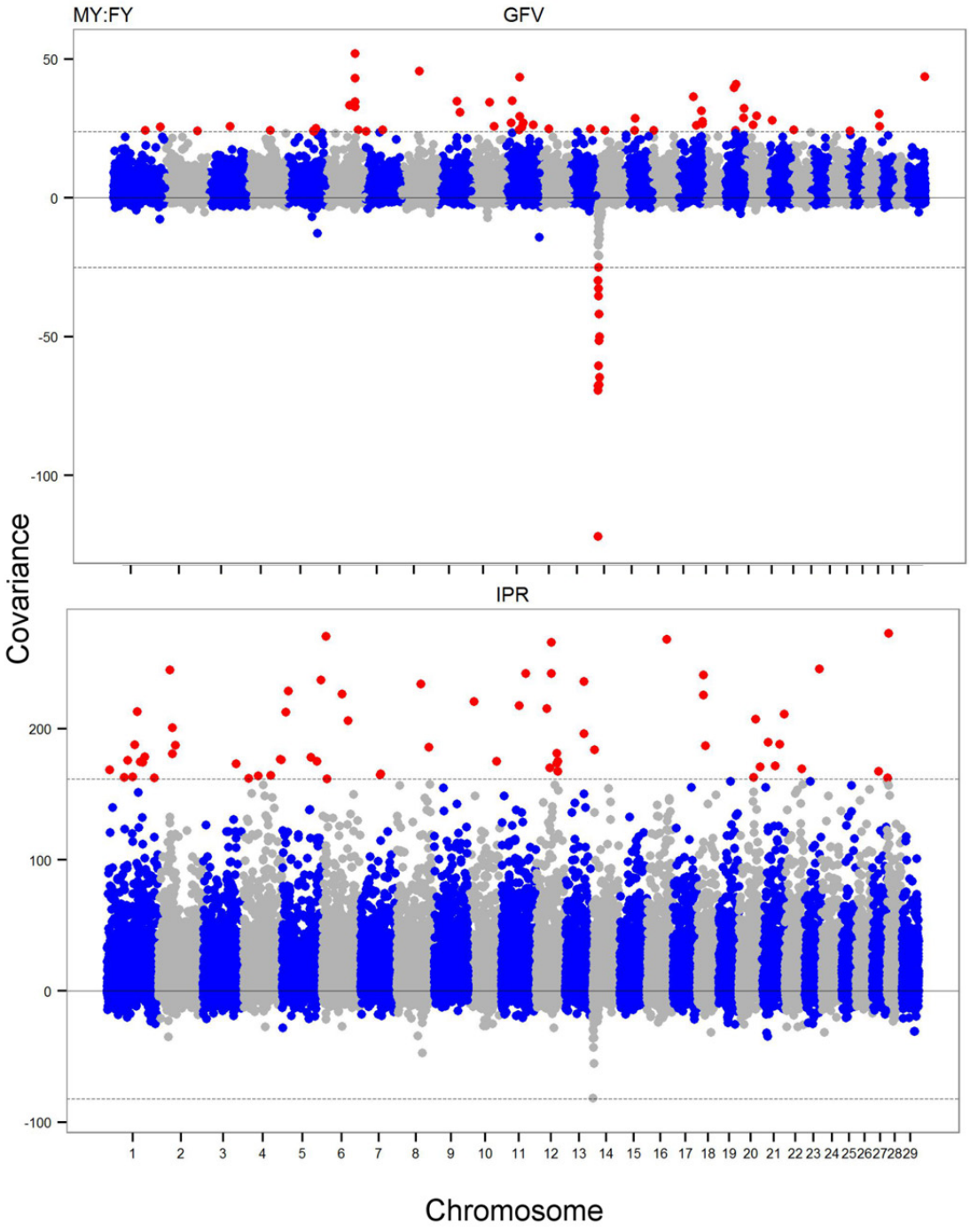
Trait combination	MY : FY	MY : NG	MY : CQ	FY : NG	FY : CQ	NG : CQ
Total genetic covariation			**German Fleckvieh (n=1,000)**			
Correlation	0.77 ± 0.03	0.18 ± 0.03	-0.06 ± 0.05	0.15 ± 0.05	-0.12 ± 0.05	0.50 ± 0.09
Genome-wide covariation						
Positive versus negative, %	97.6 : 2.4	67.9 : 32.1	49.9 : 50.1	62.6 : 37.4	45.5 : 54.5	92.0 : 8.0
Large versus small, %	8.8 : 91.2	6.6 93.4	3.6 96.4	4.3 95.7	3.9 : 96.1	9.5 : 90.5
Positive large, % / no. markers	7.7 / 44	4.8 / 31	2.2 / 14	3.3 / 20	2.1 / 15	9.5 / 6
Positive small, % / no. markers	89.9 / 28,901	63.2 / 20,377	47.7 / 18,482	59.3 / 19,696	43.4 / 17,965	82.5 / 23,914
Negative small, % / no. markers	1.3 / 8,156	30.2 / 16,694	48.7 / 18,608	36.4 / 17,391	52.7 / 19,122	8.0 / 13,146
Negative large, % / no. markers	1.1 / 6	1.8 / 14	1.4 / 12	1.0 / 9	1.8 / 14	0.0 / 0

*MY = milk yield; FY = fat yield; NG = net daily gain; CQ = carcass quality



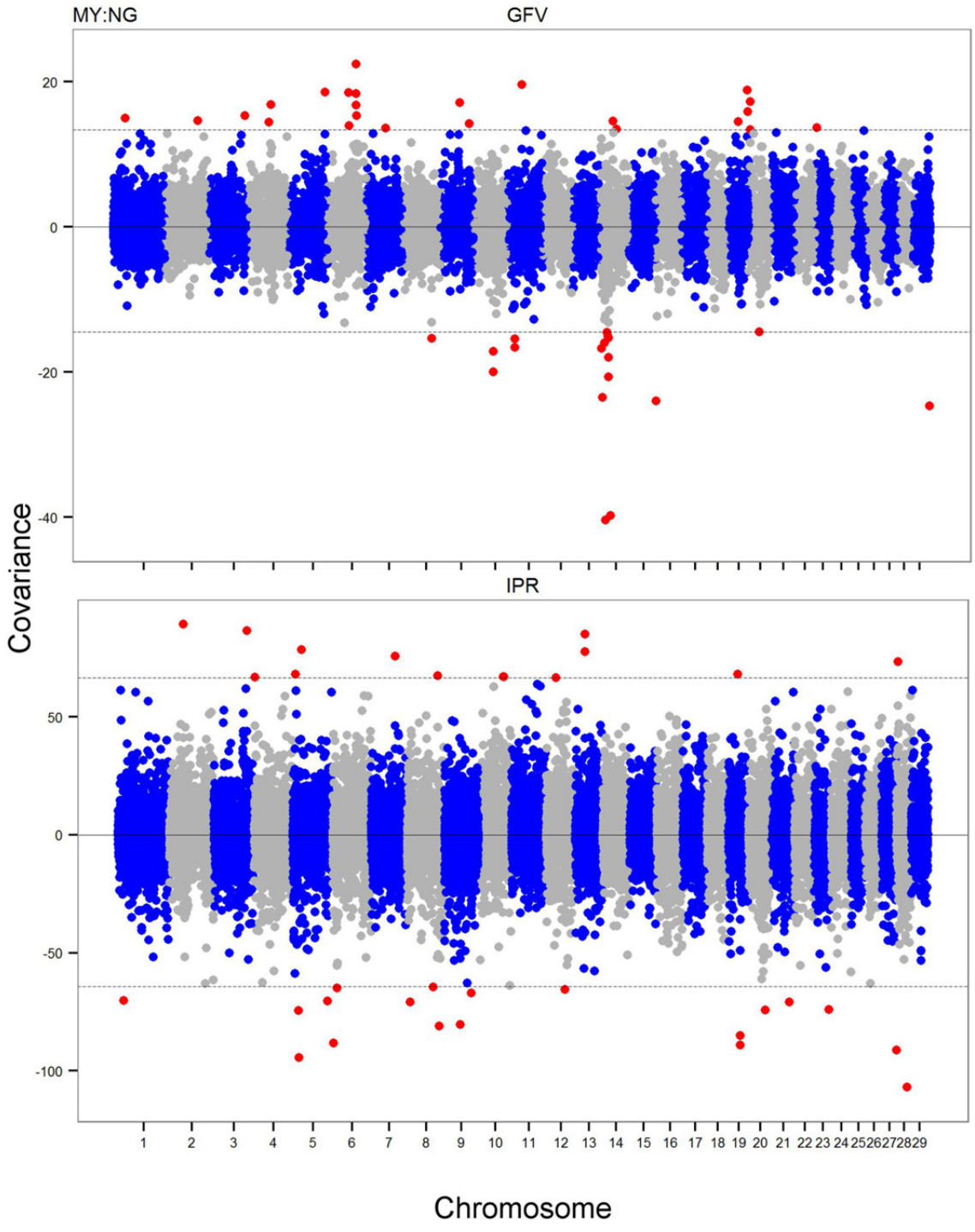
Annex A4: Genome-wide covariance contribution of each chromosome against its physical length for pairs of dairy and beef traits in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population

Priloga A4: Prispevek posameznega kromosoma k skupni genetski kovarianci v primerjavi s fizično dolžino za pare lastnosti mlečnosti in mesnatosti pri nemški (GFV) in italijanski (IPR) lisasti populaciji



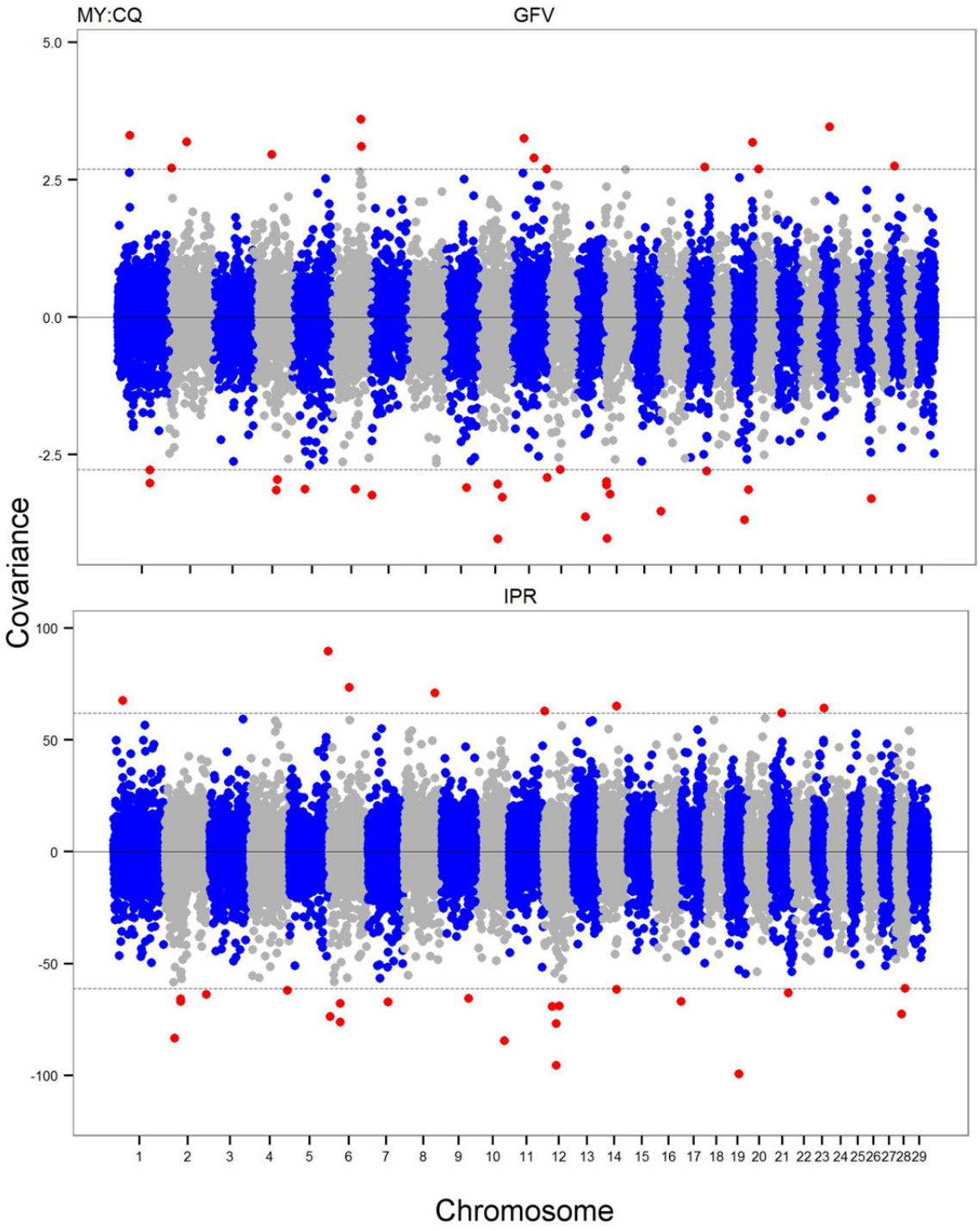
Annex A5: Genome-wide covariance plot between milk yield (MY) and fat yield (FY) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A5: Grafični prikaz genomske kovariance med količino mleka (MY) in količino maščobe (FY) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



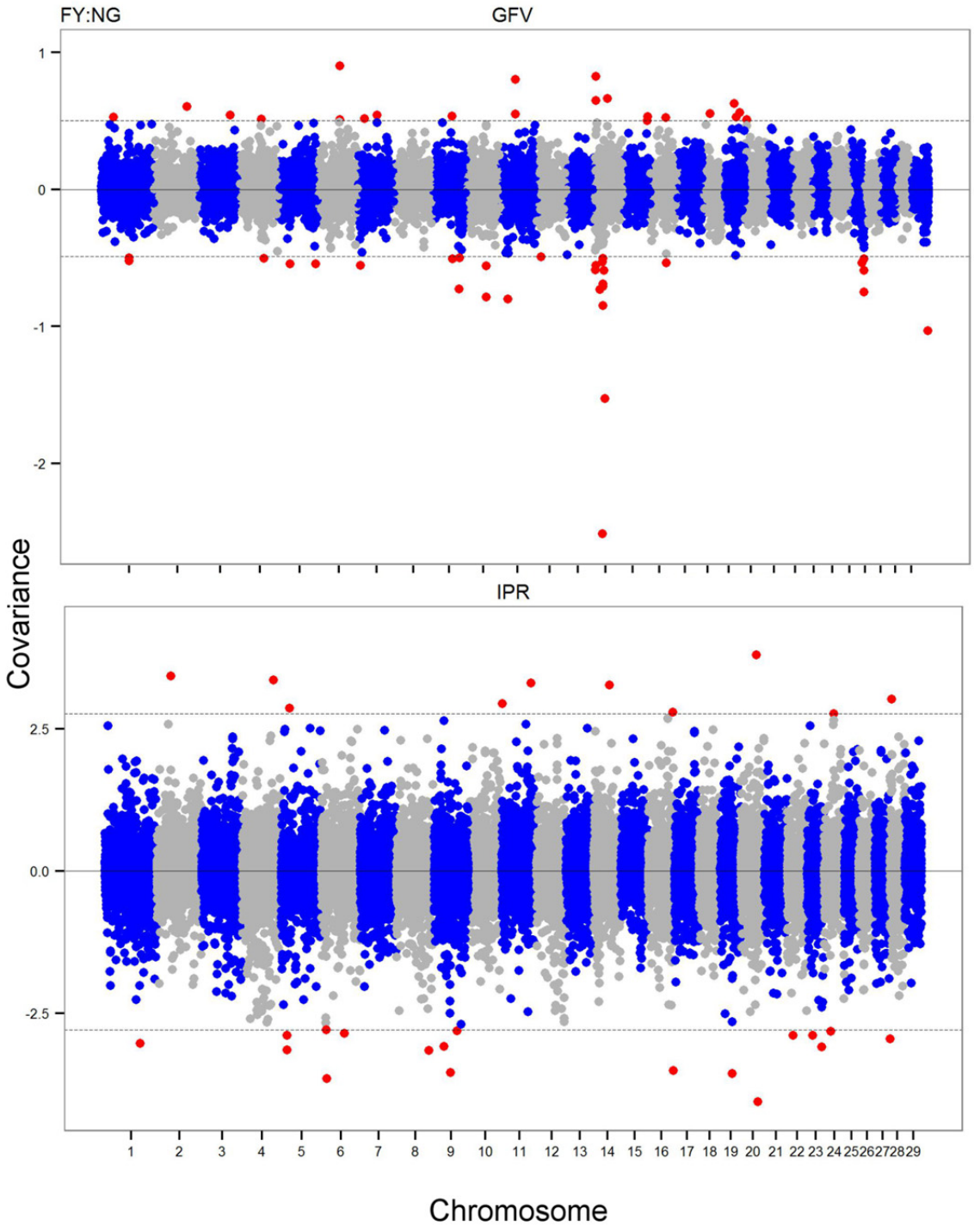
Annex A6: Genome-wide covariance plot between milk yield (MY) and net daily gain (NG) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A6: Grafični prikaz genomske kovariance med količino mleka (MY) in dnevnim neto prirastom (NG) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



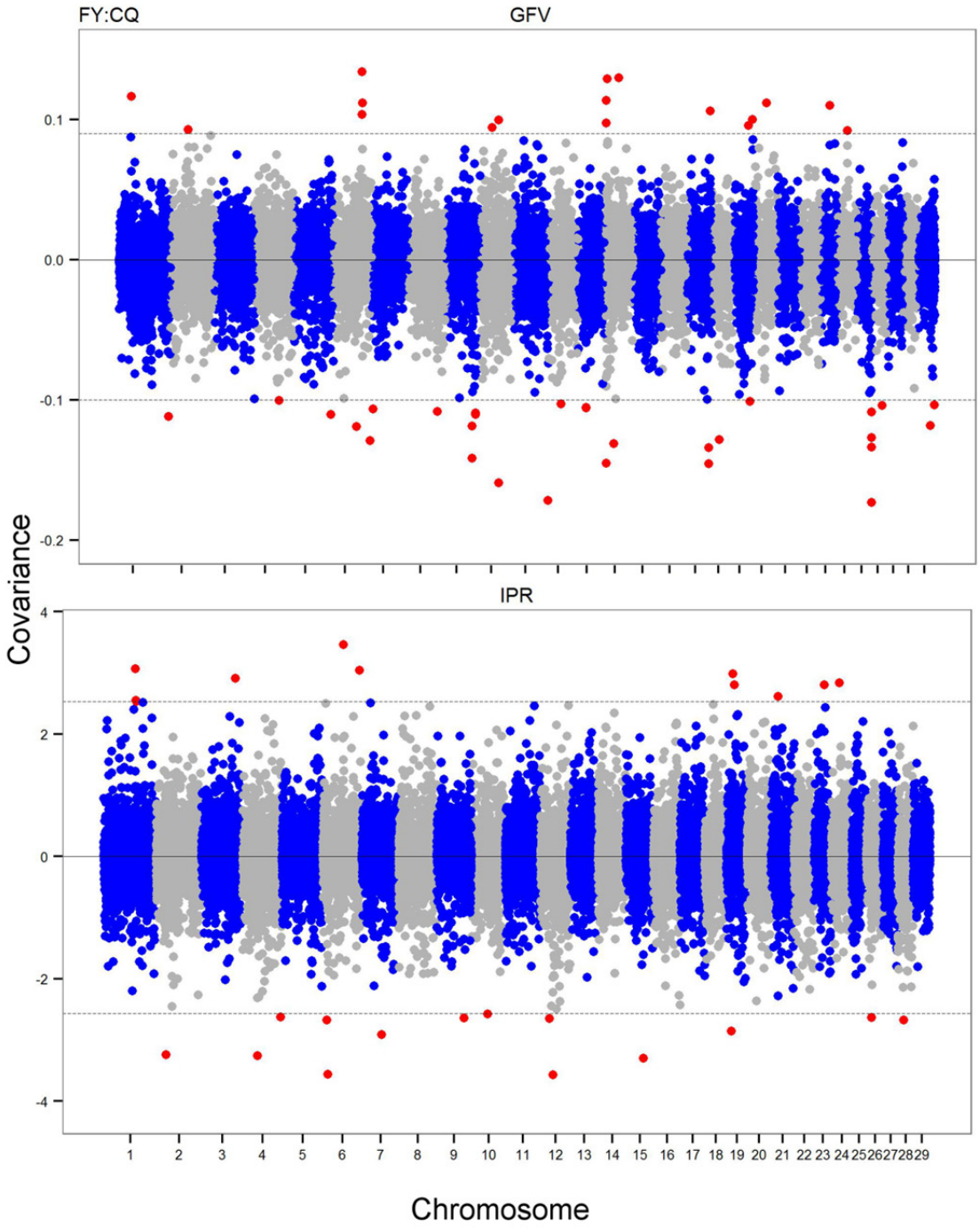
Annex A7: Genome-wide covariance plot between milk yield (MY) and carcass quality (CQ) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A7: Grafični prikaz genomske kovariance med količino mleka (MY) in klavno kakovostjo (CQ) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



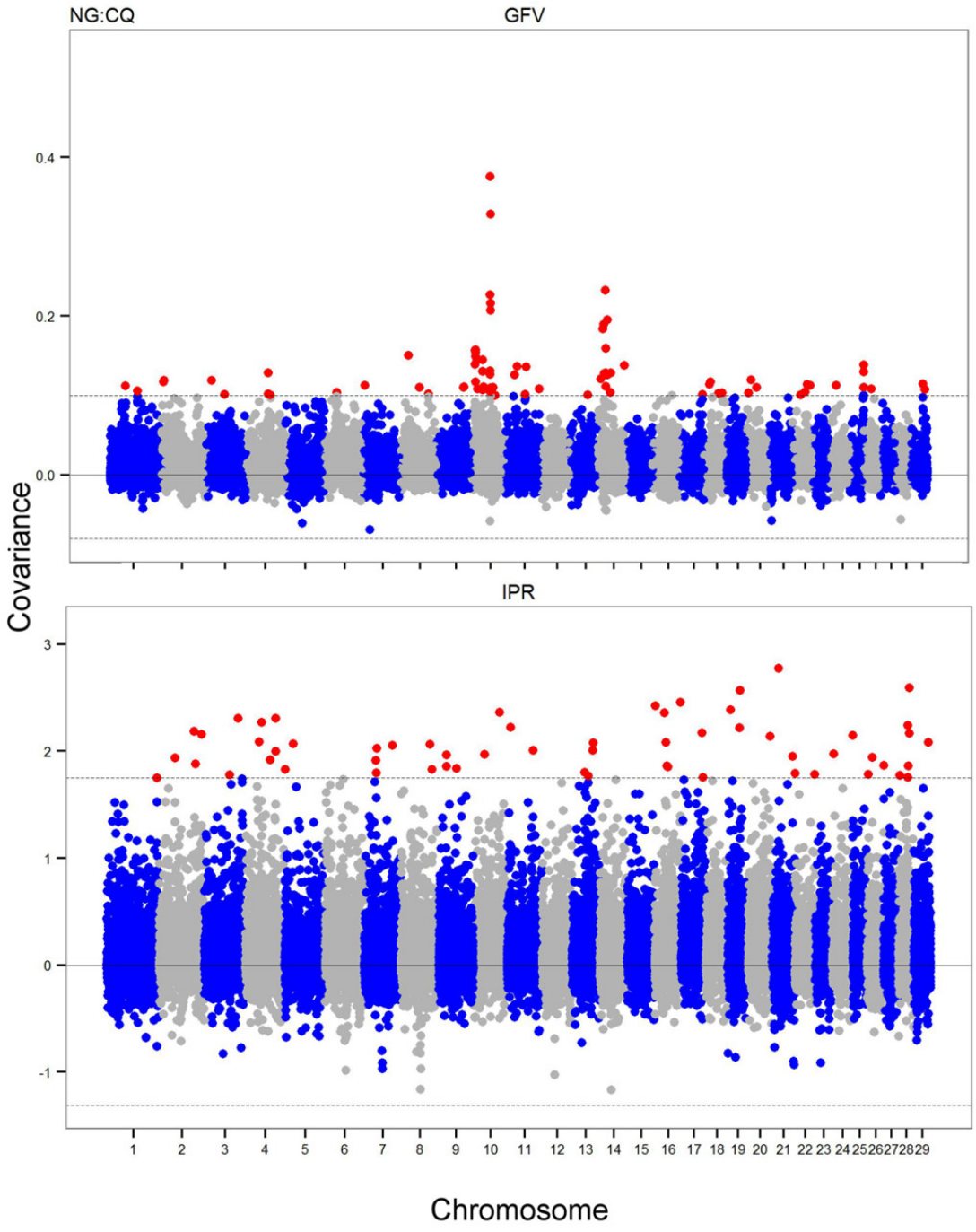
Annex A8: Genome-wide covariance plot between fat yield (FY) and net daily gain (NG) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A8: Grafični prikaz genomske kovariance med količino maščobe (FY) in dnevnim neto prirastom (NG) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



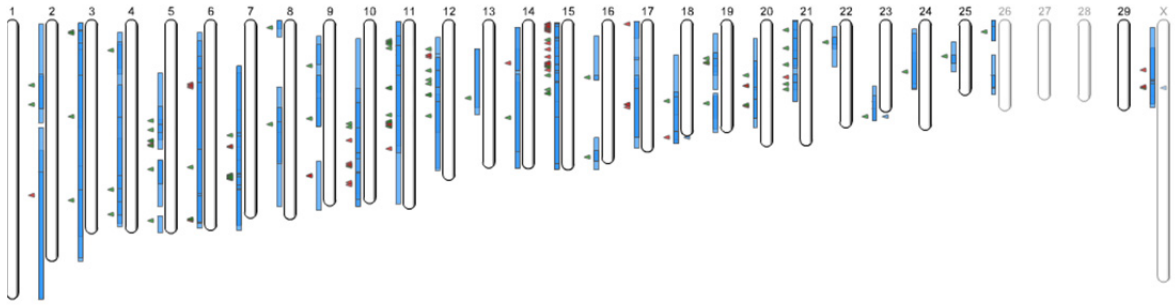
Annex A9: Genome-wide covariance plot between fat yield (FY) and carcass quality (CQ) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A9: Grafični prikaz genomske kovariance med količino maščobe (FY) in klavno kakovostjo (CQ) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



Annex A10: Genome-wide covariance plot between net daily gain (NG) and carcass quality (CQ) in the German Fleckvieh (GFV) and Italian Pezzata Rossa (IPR) population. Red dots indicate SNP markers with sizeable genetic covariance

Priloga A10: Grafični prikaz genomske kovariance med neto dnevnim prirastom (NG) in klavno kakovostjo (CQ) pri nemški lisasti (GFV) in italijanski lisasti (IPR) populaciji. Rdeče pike označujejo SNP označevalce z večjo genetsko kovarianco



Annex A11: Genomic location of SNP markers (red triangle for negative covariance and green triangle for positive covariance), QTL (blue line), and known genes (blue triangle) affecting dairy and beef traits by chromosome

Priloga A11: Lokacija SNP označevalcev na genomu po kromosomih (rdeč trikotnik za negativno kovarianco in zeleni trikotnik za pozitivno kovarianco), QTL (modra črta) in znani geni (moder trikotnik), ki vplivajo na lastnosti mlečnosti in mesnatosti