Genome-wide association mapping of milk production traits in Braunvieh cattle

J. Maxa,*1 M. Neuditschko,† I. Russ,† M. Förster,* and I. Medugorac*
*Chair of Animal Genetics and Husbandry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinaerstr. 13, 80539 Munich, Germany
†Tierzuchtforschung e.V. München, Senator-Gerauer-Straße 23, 85586 Grub, Germany

ABSTRACT

A whole-genome association study of milk production traits: milk yield, protein yield, fat yield, protein percentage, and fat percentage, was performed on the population of Braunvieh cattle. Five hundred and fifty-four progeny-tested bulls and 36,219 autosomal single nucleotide polymorphism (SNP) markers on 29 Bos taurus autosomes (BTA) were included in the analysis. A principal component analysis was conducted to adjust for the effect of population stratification in the analyzed data set. For the principal component analysis, genome-wide relationships between individuals were calculated. Three different criteria (Horn’s test, Kaiser’s criterion, and Jolliffe’s criterion) were tested to determine the number of significant principal components. Estimation of putative associations between SNP and milk production traits was carried out using a linear regression model in R software (R Foundation for Statistical Computing, Vienna, Austria). Significant principal components, adjusting for population stratification separately for each criterion and family relationships and genotypes at individual SNP were included as fixed effects in the model. The inflation factor λ and quantile-quantile plots were calculated to compare how the different criteria deal with stratification in our mapping population. Based on the analyses on all of the aforementioned criteria, we can conclude that Jolliffe’s criterion deals the best with population stratification. Furthermore, significance thresholds for a given genome-wide false discovery rate of 5% were estimated and used for specific traits. Three of the analyzed traits showed genome-wide significant association with SNP. Two SNP had an effect on milk yield on BTA4, 2 SNP affected fat yield on BTA14 and BTA23, and 1 SNP was associated with fat percent on BTA1. Single nucleotide polymorphisms identified in this study as associated with milk production traits will further contribute to the mapping of corresponding quantitative trait loci and investigation of the genes responsible for polymorphisms in milk production traits in dairy cattle. Described comparison of different criteria for determination of significant principal components can provide important information for similar studies in stratified populations.

Key words: Braunvieh cattle, genome-wide association, milk, population stratification

INTRODUCTION

The population of Braunvieh cattle in Europe (BBV) is a mixture of original Braunvieh (OBV) and Brown Swiss (BSW), which in the United States is the adapted OBV type. Braunvieh is nowadays mainly spread over the Alpine regions of Austria, Germany, Italy, and Switzerland. After the first export of the OBV breed from Europe to the United States in 1869 (Yoder and Lush, 1937), a severe population bottleneck, and following stronger selection for dairy characteristics, the American Brown Swiss cattle breed was established in the United States. Since 1906, the American Brown Swiss, previously classified as a dual-purpose breed, has been judged in the United States as a dairy type. From the 1960s, the American Brown Swiss has been used for crossing with the OBV population in Europe.

This historical development of the BBV breed and significant differences in milk yield levels between the 2 initial populations of OBV and BSW should be taken into account in genome-wide association studies. Because of the stratification within the current BBV population, spurious associations between a candidate marker and a phenotype can occur, as mentioned in several studies (Lander and Schork, 1994; Pritchard et al., 2000; Zhang et al., 2009). For association mapping, samples with minimal population structure or family relatedness are ideal and result in the greatest statistical power (Yu et al., 2006). Studies using different tools to correct for population stratification or family relatedness were mainly published by human geneti-
cists (Pritchard et al., 2000; Freedman et al., 2004; Zhu et al., 2008).

In recent years, many QTL affecting production traits in dairy cattle were localized (for the review, see Khatkar et al., 2004) but most of the association studies focused on Holstein breed and only few have been implemented in the minor, but still global, breeds such as BBV or BSW (Bagnato et al., 2008).

Therefore, the main objective of this study was to set up a whole genome association scan in BBV cattle in Europe, considering family relationships within a structured population. Associations were investigated for the main production traits milk yield (MY), milk protein yield (PY), milk fat yield (FY), milk protein percent (PP), and milk fat percent (FP).

**MATERIALS AND METHODS**

**Animals and Phenotypic Information**

Five hundred and ninety-one BBV progeny-tested sires were genotyped. Sires born before the year 1985 as well as sires with paternity problems were excluded from analyses, resulting in 554 bulls involved in the association study. Furthermore, 41 BBV, 40 OBV, and 30 BSW unrelated bulls were used for calculation of genetic differentiation (FST) among these populations. From these, 29 BBV and 19 BSW sires were also included in the association study. Extraction of DNA of BBV was performed mainly from semen samples for all available progeny-tested bulls in Germany until the year 2008 and included some of the sires originating from Austria (5), France (1), Italy (9), Switzerland (5), and the United States (52). Estimated breeding values for the 5 milk production traits MY, PY, FY, PP, and FP, as well as reliability values, were obtained in the year 2011 from the joint Austria-Germany genetic evaluation of the BBV cattle (Emmerling et al., 2002). Estimated breeding values were then subsequently deregressed and corrected for parent average effect as described by Garrick et al. (2009) and deregressed EBV were used as phenotypes in the association study.

**Genotypic Information**

Single nucleotide polymorphism genotyping was carried out by Tierzuchtorschung e.V. München (Munich, Germany) using the commercial BovineSNP50 v1 BeadChip providing 54,001 SNP (Illumina Inc., San Diego, CA). The average minor allele frequency (MAF) was 0.21 across all loci. For this study, genomic DNA was extracted from blood, semen, or hair root samples. Only animals with a genotype call rate above 0.97 were considered in our study. Furthermore, markers with genotyping errors (genotypes causing sporadic paternity conflicts at pairs of individuals with confirmed paternity), unknown chromosomal position, and minimum call rates for loci lower than 90% as well as marker loci with MAF lower than 5% were omitted. A sample set of 41 unrelated BBV animals (no grandsires in common) was used for Hardy-Weinberg equilibrium test (Wigginton et al., 2005) and those SNP that departed from it at the threshold of P < 0.01 were excluded. This resulted in a total of 36,219 autosomal SNP that passed quality control and were used for further analyses. The Bos taurus genome assembly UMD 3.1 (http://www.ncbi.nlm.nih.gov/assembly/Bos_Taurus/) was used to determine the SNP position within each autosome.

**Statistical Analyses**

At first, genome-wide relationships between individuals were estimated using the method of Powell et al. (2010). Furthermore, a principal component analysis (PCA) was considered to adjust for the effect of population stratification and relatedness in the analyzed data set using a genome-wide relationship matrix based on marker information among the individuals. We applied 3 different criteria to determine the number of significant principal components (PC). For the first criterion, we used the paran package (Dinno, 2010) in R (version 2.13.2; http://www.R-project.org), which implements Horn’s technique (Horn, 1965). In this package, Horn’s method contrasts eigenvalues from PCA on several random data sets of uncorrelated variables to produce eigenvalues for components that are adjusted for the sample error-induced inflation. Finally, only PC with adjusted eigenvalues greater than 1 are retained (Dinno, 2010). The second criterion was based on the simple method of Kaiser (1956) where PC with the mean eigenvalues greater than 1 were included. The third, Jolliffe’s criterion, retained PC with an eigenvalue accounting for at least 70% of the mean eigenvalue (Jolliffe, 2002). Only these significant components (N_{SPC}) were used separately for each criterion (33, 153, and 254 PC based on Horn’s, Kaiser’s, and Jolliffe’s criterion, respectively) to correct for family relationship and stratification within the BBV population in our study.

The test of putative associations among SNP and milk production traits in BBV cattle was carried out using linear regression model in R (version 2.13.2; http://www.R-project.org). Significant PC adjusting for population stratification and family relationship were included as covariates in the model. The genotypes of sires were evaluated as fixed effect covariates where values of each covariate were continuous between
1 and 3 (Gengler et al., 2007) with elements set to 1, 2, or 3 for the homozygote, heterozygote, and other homozygote, respectively. The effect of sire genotype on each of the traits was estimated from single-trait analysis, repeated for each SNP, using the model

\[ y = Xb + e, \]

where \( y = n \times 1 \) vector of deregressed EBV for MY, PY, FY, PP, and FP for \( n = 554 \) sires; \( X = n \times (N_{\text{SPC}} + 1) \) design matrix of \( N_{\text{SPC}} \) principal components, with \( N_{\text{SPC}} = 33, 153, \) and 254 based on Horn’s, Kaiser’s, and Jolliffe’s criterion, respectively, and 1 column for genotypes of sires as fixed effect covariates (corresponding to 1, 2, or 3); \( b = \) fixed and unknown \((N_{\text{SPC}} + 1) \times 1\) vector of regression coefficients for \( N_{\text{SPC}} \) and 1 regression coefficient for genotypes of sires; and \( e = n \times 1 \) random vector of residual effects.

The \( P \)-value thresholds for a given genome-wide false discovery rate (FDR) were estimated using the method of Storey and Tibshirani (2003) as implemented in the qvalue package in R (version 2.13.2; http://www.R-project.org). In this method, \( q \)-values are calculated for the specific set of \( P \)-values. In comparison to \( P \)-values, which give a measure of significance in terms of the false positive rate, \( q \)-values are a measure in terms of FDR as explained by Storey and Tibshirani (2003). False discovery rate was shown as a sensible measure and the 5% FDR level was defined for 5 traits at first. Taking into account results based on the implementation of Jolliffe’s criterion, the stringent level of 5% FDR was applied for MY, FY, and FP, whereas 8 and 12% FDR resulted in significant SNP for PY and PP, respectively.

To evaluate genetic differentiation of the 2 initial populations BSW and OBV and the admixed BBV population, we calculated pairwise \( F_{\text{ST}} \) from population allele frequencies (Weir and Cockerham, 1984) across the set of all autosomal SNP that were tested for Hardy-Weinberg equilibrium and were informative across the set of all autosomal SNP that were tested. Pairwise \( F_{\text{ST}} \) estimation takes into account the sample of unrelated animals (41 BBV, 40 OBV, and 30 BSW), assigning these to groups according to their pedigree analysis. The animals with deep solid pedigrees going back to the beginning of the upgrading of European OBV by BSW but without any BSW ancestors were designated as OBV. The animals with both BSW parents were assigned to the BSW group. We traced the pedigree of the remaining BBV animals at least 5 generations back and estimated the proportion of genes originating from BSW within the BBV population (45–98%). To illustrate admixture gradient within BBV sample used in the association study and to put it into context of the differentiation between 2 source populations, we estimated a genome-wide relationships matrix (Powell et al. 2010) between all individuals (41 OBV, 513 BBV, and 63 BSW), transformed it into a distance matrix, and presented it as a 2-dimensional plot. Applied heuristic algorithm (Veit-Kensch et al. 2007) used the first 2 principal components as a starting configuration and further improved it by maximization of the correlation between the multidimensional distance matrix and 2-dimensional plot.

**RESULTS AND DISCUSSION**

**Characterization of the Braunvieh Population and Stratification**

A known difference exists between the production levels of the analyzed traits between the BSW population in the United States and the BBV population in Europe. Because of different yield levels (e.g., MY, PY, and FY) and significant genetic differentiation (\( F_{\text{ST}} = 0.089 \)) between the 2 initial populations OBV and BSW, and due to the admixed status of the BBV population with overemphasis of BSW genes through upgrading, we can expect spurious associations of some markers with alleles overrepresented in one of the initial populations. The genetic differentiation between BSW and BBV was relatively small (\( F_{\text{ST}} = 0.009 \)), whereas higher \( F_{\text{ST}} \) of 0.068 was estimated between OBV and BBV. For the sake of comparability, the \( F_{\text{ST}} \) between OBV and BSW (0.089) is higher than between the dairy breed Holstein and the beef breed Belgian Blue (0.071; Neuditschko 2011). The genetic differentiation between 2 source populations and admixture gradient within BBV mapping population is clearly illustrated by the 2-dimensional plot of the genomic distances matrix (Figure 1). The geometric projection of the pairwise individual distances resulted in higher cophenetic correlation (\( r = 0.818 \)) than in the plot of the first 2 PC (\( r = 0.704 \)). A severe population bottleneck and stronger artificial selection resulted in closer genetic relationship between BSW animals (i.e., little diversity space occupied by BSW in Figure 1). Braunvieh bulls with increasing proportion of OBV genes segregate more and more from this BSW core. The bulls with less than 75% BSW genes predominantly take a marginal position in the BSW-BBV cluster and most of these present an intermediary between 2 source populations.

Considering the historical development and family relationships within the BBV breed, we applied correction for population stratification based on PCA analysis in the current study. As input for the PCA analysis, the genome-wide relationships among the animals were calculated using marker genotypic information. Ances-
try information relies on DNA marker genotypes (Price et al., 2006), especially in human association studies. This information may be more accurate compared with pedigree information, especially when applied to admixed cattle (Gibbs et al., 2009). A comparison of PCA with other methods of detecting population structure and its theoretical basis was presented by Patterson et al. (2006). A PCA analysis was shown to be practical, especially when large data sets were tested (Zhang et al., 2009). When we compared the 3 different criteria to determine the number of significant PC, we concluded that applying Horn’s test resulted in the highest number of significant associations over the whole genome. Except for FY, for which there were no associations, a total of 175 significantly associated SNP were found at 5% FDR: 6, 9, 15, and 145 for PP, PY, MY, and FP, respectively. Kaiser’s criterion involving 153 PC included in the model resulted in no significant association at 5% FDR for all traits except for PP and FP. Kaiser’s criterion involving 153 PC included in the model resulted in no significant association at 5% FDR for all traits except for PP and FP, with 4 associated SNP in total. The third criterion of Jolliffe revealed 2 significantly associated SNP with MY at 5% FDR, 2 with FY, and 1 significantly associated SNP with FP. The inflation factor $\lambda (\lambda_{GC})$ was calculated for each trait and criterion to evaluate how the selected criterion dealt with stratification in the analyzed sample. High values of $\lambda_{GC}$ were obtained for associations based on Horn’s test (1.39 for MY, 1.36 for PY, 1.37 for FY, 1.33 for PP, and 1.52 for FP). The inflated test statistics indicated that the method performed unacceptably in correcting for the stratification in our mapping population. Compared with Horn’s test, Kaiser’s criterion revealed lower values of $\lambda_{GC}$ (1.18 for MY, 1.16 for PY, 1.21 for FY, 1.20 for PP, and 1.25 for FP). The most acceptable values of $\lambda_{GC}$ (1.14 for MY, 1.13 for PY, 1.16 for FY, 1.10 for PP, and 1.16 for FP) were obtained using Jolliffe’s criterion. Nevertheless, not even fitting all significant PC identified by Jolliffe’s criterion could account for all stratification within the breed. Therefore, only SNP displaying a genome-wide threshold of FDR $\leq$5% were accounted for as true positive. Despite $\lambda_{GC}$ also quantile-quantile (Q-Q) plots comparing observed distributions of $-\log(P$-value) to the expected distribution under the no-association model for each trait and criterion are shown in Figures 2, 3, 4, 5, and 6. The distribution of observed $-\log(P$-value) presented in Figures 2 to 6 is more shifted toward the null expectation in the case of Jolliffe’s criterion, especially compared with Horn’s test. Taking into account both $\lambda_{GC}$ values as well as resulting Q-Q plots, we can conclude that Horn’s and Kaiser’s criterions were incapable of dealing with observed stratification in our data set. Out of the 3 methods used, Jolliffe’s criterion dealt the best with observed stratification in the data and was, therefore, the preferred method for identification of significant PC.

Single SNP Associations

A total of 36,219 SNP on 29 Bos taurus autosomes (BTA) passed the quality control criteria and were included in the association analyses. The number of SNP per autosome varied from 655 on BTA27 to 2,397 on BTA1. Details of significant SNP associations with one of the analyzed milk production traits are presented in Table 1. Very few significantly associated SNP over the whole genome were found at 5% FDR: 2 SNP for MY, 2 SNP for FY, and 1 SNP for FP. This can be mainly due to the low number of genotyped individuals used in the whole genome association analysis. A recent study by Schwarzenbacher et al. (2012) on the population of almost 300 Brown Swiss bulls selectively genotyped for the trait protein yield (bulls with the highest and lowest protein yield EBV) supports this conclusion. Based on simulation of 300 dairy bulls and BovineSNP50 Illumina BeadChip genotypes, the above-mentioned study suggests that only effects explaining more than 10% of the phenotypic variation could be detected in the population of that size. Such effects are expected to be rare. Nevertheless, SNP detected as significant in our study explained, on average, 1.7% of the total phenotypic variation. For the other 2 traits (PY and

![Figure 1. Two-dimensional presentation of individual distances based on the proportion of shared alleles of original Braunvieh (OBV), Brown Swiss (BSW), and Braunvieh cattle (BBV) with various estimated proportion (45–98%) of genes originating from the BSW population. Each dot resembles an individual according to the respective legend. The cophenetic correlation coefficient is 0.82. Color version available in the online PDF.](image)
Figure 2. Quantile-quantile plot for milk yield where Horn’s criterion (A), Kaiser’s criterion (B), and Jolliffe’s criterion (C) was used to determine the number of significant principal components included in the whole-genome association study. Color version available in the online PDF.

Figure 3. Quantile-quantile plot for protein yield where Horn’s criterion (A), Kaiser’s criterion (B), and Jolliffe’s criterion (C) was used to determine the number of significant principal components included in the whole-genome association study. Color version available in the online PDF.

Figure 4. Quantile-quantile plot for fat yield where Horn’s criterion (A), Kaiser’s criterion (B), and Jolliffe’s criterion (C) was used to determine the number of significant principal components included in the whole-genome association study. Color version available in the online PDF.
PP), the first significant associations were revealed at FDR thresholds of 8 and 12%, respectively. The total number of significant SNP were 2 for PY and 1 for PP. However, these SNP are not further discussed in detail because of the risk that they represent false-positive signals.

On BTA1, marker Hapmap57808-rs29009859 showed significant association with FP at 88.2 Mb and, to our knowledge, no QTL affecting FP was previously mapped close to the mentioned region. Overall, BTA1 currently contains only few detected QTL for FP, as presented by Nadesalingam et al. (2001) and Liu et al. (2004).

Bos taurus autosome 4 revealed 2 significant associations for MY: Hapmap32136-BTA-160383 and BTB-01295275 at 92.7 and 92.8 Mb, respectively. Several QTL for MY on BTA4 have been reported in the literature and some of them are located very near to the region detected by this study. Bagnato et al. (2008), who focused on the Brown Swiss population in 3 European countries using selective DNA pooling and a panel of 187 microsatellite markers covering the whole genome, reported QTL for MY on BTA4 at 87.3 cM, corresponding to approximately between 93.1 and 93.3 Mb. Two other studies by Lindersson et al. (1998) and Daetwyler et al. (2008) identified QTL for MY near this region as well as at 86 cM (corresponding to approximately between 91.8 and 92.0 Mb) in Swedish Red and White and Holstein cattle.

One significant association was observed on BTA14 at 10.2 Mb for FY. This association is positioned in the region of QTL span for FY reported by Winter et
al. (2002) and Boichard et al. (2003) in different German and French cattle breeds. Nevertheless, we could not detect any significant association between marker and milk production traits in the very proximal region of BTA14 overlapping the DGAT1 gene that causes a major effect on milk fat content and other milk characteristics by changing lysine to alanine in the enzyme diacylglycerol O-acyltransferase (Grisart et al., 2002; Winter et al., 2002). It is worth mentioning that the lysine-encoding DGAT1 allele is also present in BBV and OBV animals, although to a lesser extent (even at a frequency lower than the herein-applied MAF criterion) than in the other breeds (Winter et al., 2002). Until recently, several studies, as for example, Weller et al. (2003) and Sun et al. (2009), published QTL for FY at 0.5 cM, overlapping the DGAT1 gene. The second SNP significantly associated with FY was located on BTA23 at 18.0 Mb. Compared with those reported in the literature, BTA23 contained very few detected QTL for FY. Two studies of Plante et al. (2001) and Bennewitz et al. (2004) reported QTL for FY in the region of 43.6, 63.4, and 49.2 cM. The large QTL span between approximately 5.4 and 36.2 Mb published by Bennewitz et al. (2004) overlapped the associated region mentioned in our study.

### CONCLUSIONS

In the current genome-wide association study focusing on the Braunvieh cattle population, evidence for the presence of QTL affecting milk-production traits, namely MY, FY, and FP was observed on BTA1, BTA4, BTA14, and BTA23. Based on the history, development, and current family relationships of the Braunvieh breed, correction for population stratification using significant PC was applied. Choosing the criterion for selection of significant PC was shown to be crucial. From 3 different criteria applied in this study, we can conclude that Horn’s test and Kaiser’s criterion did not deal properly with observed stratification in our data set. On the contrary, applying Jolliffe’s criterion to identify significant PC was shown to be the best in dealing with stratification in the data. The results presented in this study revealed a potential for QTL fine-mapping in the associated autosomal regions. Further investigation on the genes responsible for polymorphisms in milk production traits particular to the Braunvieh breed should follow.

### ACKNOWLEDGMENTS

The authors are grateful for funding from Tierzuchtforschung e.V. Munich (Munich, Germany) and Deutsche Forschungsgemeinschaft (DFG, Bonn, Germany), project ME3404/2-1. The Landeskuratorium der Erzeugerringe für tierische Veredelung in Bayern e.V. (LKV, Munich, Germany) kindly provided information on EBV and reliability values of sires used in this study. Furthermore, we thank the breeders and breeding associations who sent us samples free of charge to support this study as well as to Arbeitsgemeinschaft Süddeutscher Rinderzucht- und Besamungsorganisationen e. V. (Munich, Germany) and Swissgenetics (Zollikofen, Switzerland).

### REFERENCES


