

RESEARCH ARTICLE

Open Access



# Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins

Xiaoping Wu<sup>1,2</sup>, Mogens S Lund<sup>1</sup>, Goutam Sahana<sup>1</sup>, Bernt Gulbrandsen<sup>1</sup>, Dongxiao Sun<sup>2</sup>, Qin Zhang<sup>2</sup> and Guosheng Su<sup>1\*</sup>

## Abstract

**Background:** The sensitivity of genome-wide association studies for the detection of quantitative trait loci (QTL) depends on the density of markers examined and the statistical models used. This study compares the performance of three marker densities to refine six previously detected QTL regions for mastitis traits: 54 k markers of a medium-density SNP (single nucleotide polymorphism) chip (MD), imputed 777 k markers of a high-density SNP chip (HD), and imputed whole-genome sequencing data (SEQ). Each dataset contained data for 4496 Danish Holstein cattle. Comparisons were performed using a linear mixed model (LM) and a Bayesian variable selection model (BVS).

**Results:** After quality control, 587, 7825, and 78 856 SNPs in the six targeted regions remained for MD, HD, and SEQ data, respectively. In general, the association patterns between SNPs and traits were similar for the three marker densities when tested using the same statistical model. With the LM model, 120 (MD), 967 (HD), and 7209 (SEQ) SNPs were significantly associated with mastitis, whereas with the BVS model, 43 (MD), 131 (HD), and 1052 (SEQ) significant SNPs (Bayes factor > 3.2) were observed. A total of 26 (MD), 75 (HD), and 465 (SEQ) significant SNPs were identified by both models. In addition, one, 16, and 33 QTL peaks for MD, HD, and SEQ data were detected according to the QTL intensity profile of SNP bins by post-analysis of the BVS model.

**Conclusions:** The power to detect significant associations increased with increasing marker density. The BVS model resulted in clearer boundaries between linked QTL than the LM model. Using SEQ data, the six targeted regions were refined to 33 candidate QTL regions for udder health. The comparison between these candidate QTL regions and known genes suggested that *NPFFR2*, *SLC4A4*, *DCK*, *LIFR*, and *EDN3* may be considered as candidate genes for mastitis susceptibility.

## Background

Mastitis, i.e. inflammation of the mammary gland, is a common and costly disease [1, 2] that is particularly problematic in the dairy industry. It adversely affects both animal and human health, since milk from affected cattle can enter the food supply and pose a health risk [3]. Mastitis also leads to economic losses owing to reduction in milk yield and quality, increased costs associated with treatment of affected cattle, discarded milk, and culling of severely affected cattle [4]. Therefore, prevention of mastitis is an extremely important breeding goal for dairy cattle. However, the heritability of clinical mastitis is low [5–9], and its genetic correlation with

production traits is unfavorable [6, 10, 11]. Identification of specific genes associated with clinical mastitis would be extremely useful in breeding programs to reduce mastitis incidence. To that end, quantitative trait loci (QTL) mapping is a useful tool to identify genomic regions that are associated with clinical mastitis. Genome-wide association studies (GWAS) have been successfully used to identify QTL regions for a variety of traits in livestock [12]. Several studies have also attempted to detect QTL that affect clinical mastitis, with varying success [13–16].

Many factors influence the efficiency of GWAS to detect QTL. One of the most important factors is marker density. Currently, a medium-density (MD) SNP chip with ~54 000 markers is widely used for GWAS in dairy cattle [17–19]. In recent years, two high-density (HD) SNP chips with 777 962 SNPs from Illumina Inc. [20]

\* Correspondence: guosheng.su@mbg.au.dk

<sup>1</sup>Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark  
Full list of author information is available at the end of the article

and 648 874 SNPs from Affymetrix Inc. [21], and whole-genome sequencing (SEQ) data [22] have become available. Higher marker densities mean that markers are in stronger linkage disequilibrium (LD) with QTL that affect the trait of interest. In addition, genome sequencing data includes causative genetic variants. Thus, it is expected that using HD or SEQ markers will lead to the detection of more and more accurately localized QTL.

In addition to marker density, the statistical model selected for analysis has an effect on detection sensitivity. Several models have been proposed for GWAS, such as single-marker tests (SMT) [23], mixed model analysis (MMA) and linear models (LM) [24, 23], haplotype models (HM) [25], genealogy-based mixed-model (GENMIX) [25], and Bayesian variable selection models (BVS) [26]. Some studies have carried out model comparisons using simulated data. For example, Sahana et al. [23] compared the SMT, HM, LM, and BVS models, and concluded that the BVS model performed best. Dashab et al. [25] compared LM, random HM, GENMIX, and BVS models, and showed that LM and BVS were better than the other methods in terms of power and type-I error rate. However, there are very few reports on model comparisons based on real data from livestock [27, 28].

In general, LM models are performed in single-marker test analyses, for which each SNP is fitted separately in the model. In contrast, BVS models estimate the effects of all SNPs simultaneously. Usually, BVS models are implemented via Markov chain Monte Carlo (MCMC) algorithms [29–31]. For both models, false positive associations due to population structure, such as family relationships can be controlled [23, 24] by including systematic factors and taking polygenic effects into account. For LM models, significant associations are commonly established using a t-test with Bonferroni correction, assessment of false discovery rate, or permutation testing [32]. For BVS models, the number of QTL can be determined by post-MCMC analysis using Bayes factors [33] or QTL intensity profiles [29]. Previous studies have claimed that BVS models yield higher power than LM and are a solution to the problem of establishing significance of multiple-testing in simulation studies [25, 23]. Sahana et al. [23] compared the marginal posterior probability for single markers (BAYSM) and the joint posterior probabilities for intervals of 11 markers (BAYINT) to infer the presence of QTL using a BVS model based on simulated data. They reported that BAYINT resulted in higher power to detect QTL, while BAYSM was more precise in estimating QTL position.

Complex traits are likely influenced by multiple QTL that have small individual effects but a large collective effect on the phenotype. Conversely, the effects of a given QTL can be distributed over several markers that

are in LD with the QTL by using a Bayesian model which estimates effects of all SNPs simultaneously. The QTL intensity profile, which summarizes the marginal posterior probabilities of markers in a small region, was proposed to detect QTL regions [34]. In this situation, a small QTL can be detected via a marked peak, and multiple QTL can be separated [34]. QTL intensity profiles based on post-MCMC analysis provide a suitable approach to identify QTL regions using real data.

The first objective of this study was to refine six previously detected QTL regions for udder health in Danish Holstein cattle using three marker densities, i.e. MD, HD and SEQ datasets. The second objective was to compare the performance of the LM and BVS models to detect and separate QTL that are closely located.

## Methods

### Phenotype and genotype data

A total of 4496 Danish Holstein bulls with de-regressed estimated breeding values (DRP) for the udder health index were used. The index for udder health is a measure of the genetic value of the cow's resistance to mastitis. The index was calculated based on estimated breeding values (EBV) of clinical mastitis (CM) from the 1<sup>st</sup> to 3<sup>rd</sup> lactations, i.e.:

$$\text{Index of udder health} = 0.25 * CM_{11} + 0.25 * CM_{12} \\ + 0.30 * CM_2 + 0.20 * CM_3,$$

where  $CM_{11}$ ,  $CM_{12}$ ,  $CM_2$ ,  $CM_3$  are EBV for clinical mastitis in lactation 1 from days in milk (DIM) -15 (15 days before calving) to 50, lactation 1 from 51 to 305 DIM, lactation 2 from -15 to 150 DIM, and lactation 3 from -15 to 150 DIM, respectively. Cows with and without clinical mastitis were recorded as 1 and 0, respectively. Breeding values of clinical mastitis for each lactation were estimated using a multi-trait random regression test-day animal model. Somatic cell counts for the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> lactation, fore udder attachment, and udder depth from the 1<sup>st</sup> lactation were included as correlated traits to improve the accuracy of EBV for clinical mastitis [35], but EBV of these correlated traits were not included in the index of udder health. Udder health DRP [36, 37] were derived from the index of udder health evaluated in November 2010.

An association study for udder health was carried out using three marker datasets: (1) Illumina BovineSNP50 BeadChip (MD), (2) Illumina BovineHD BeadChip (HD), and (3) genome sequence data (SEQ). Six chromosome regions with significant effects on clinical mastitis that were reported in our previous study [38], were used for further association analysis. These six regions are located between 84 and 95 Mb on BTA5 (BTA for *Bos taurus* chromosome), 88 and 96 Mb on BTA6, 57 and 63 Mb on BTA13, 48 and 55 Mb on BTA16, 55 and 58 Mb on BTA19, and 32 and

40 Mb on BTA20. SNP array genotyping, sequence analysis, and imputation were performed as described by Sahana et al. [38]. Briefly, 5214 bulls were genotyped with the 50 k SNP chip. SNP chip data were edited using a number of criteria. First, individuals with a call rate higher than 85 % and SNPs with a call rate higher than 95 % were kept. SNPs with a minor allele frequency below 5 % and those that deviated from the Hardy-Weinberg equilibrium ( $P < 0.00001$ ), or with average GenCall scores (Illumina Inc.) below 0.65 were excluded. After this quality control filtering, 43 415 SNPs remained. The MD data of 5214 Holstein bulls were imputed to HD data with BEAGLE [39], using 557 HD genotyped bulls (including 138 Nordic Holstein bulls, which represent a substantial contribution of the population) in the EuroGenomics project [40] as reference data. After quality control (using criteria similar to those described above for the 50 k chip), 648 219 HD SNPs remained. Finally, six regions with 790 SNPs from the MD chip and 8260 SNPs from the HD chip were targeted for analysis. Linkage disequilibrium (average  $r^2$  between adjacent marker pairs) estimates were equal to 0.21 for the MD and 0.56 for the HD data in the six targeted regions.

Sequence data analysis was performed at the Center for Quantitative Genetics and Genomics at Aarhus University. Details on sequencing, sequence analysis, and quality control are in [41, 42, 38]. Using the 89 sequenced bulls (including 32 Nordic Holstein bulls, which represent a substantial contribution of the present population) as reference data, 5214 Holstein bulls with imputed HD data for six targeted regions were imputed to SEQ data by BEAGLE [39]. A total of 196 882 SNPs were located within the six regions for the SEQ data. In the end, using all available phenotypic and genotypic data, 4496 Holstein bulls with MD, imputed HD, and imputed SEQ data were used in the analyses. These animals were the progeny of 373 bulls with 3169 dams.

The SNPs in each of the six regions were checked for quality as follows. First, SNPs that exclusively appeared in lower density data but disappeared in higher density data were discarded, removing 201 SNPs that were in MD but not in the HD and SEQ data as well as 308 SNPs that were in the HD but not in the SEQ data. Second, SNPs with a minor allele frequency less than 0.01 were discarded, removing one SNP from the HD dataset and 41 304 SNPs from the SEQ dataset. Finally, if two markers were in complete LD, then either the SNP that was absent from the lower marker density dataset was deleted or the more downstream SNP was deleted if both markers were absent from the lower density data. This removed two SNPs in the MD, 126 SNPs in the HD, and 76 722 SNPs in the SEQ data. After these edits, 587, 7825, and 78 856 SNPs in the six regions remained

in the MD, HD, and SEQ data, respectively. The average distances between adjacent SNPs were approximately 70.0, 5.5, and 0.5 kb for the MD, HD, and SEQ data, respectively.

### Statistical analyses

A linear mixed model and a Bayesian variable selection model were used to detect associations of SNPs with phenotypic data, as described below.

#### Linear mixed model (LM)

The LM model [24] used in this study is a single-SNP regression model. The model includes a fixed regression of phenotypes using SNP genotypes as a measure of the SNP effects. In addition, a random polygenic effect that accounts for shared genetic effects of related individuals was included. The statistical model is described by the formula:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{x}\mathbf{g} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad (1)$$

where  $\mathbf{y}$  is the vector of phenotypes (DRP),  $\mathbf{1}$  is a vector of ones,  $\mu$  is the overall mean,  $\mathbf{g}$  is the additive genetic effect of the analyzed SNP,  $\mathbf{x}$  is a vector of SNP genotypes (coded as 0, 1 or 2), and  $\mathbf{u}$  is a vector of random polygenic effects, which are normally distributed  $\mathbf{u} \sim N(0, \mathbf{A}\sigma_u^2)$ , where  $\mathbf{A}$  is the pedigree-based additive relationship matrix,  $\sigma_u^2$  is the polygenic variance,  $\mathbf{Z}$  is an incidence matrix relating phenotypes to the corresponding random polygenic effects, and  $\mathbf{e}$  is a vector of residual effects, which are normally distributed  $\mathbf{e} \sim N(0, \mathbf{D}\sigma_e^2)$ , where  $\mathbf{D}$  is a diagonal matrix with elements  $d_{ii} = (1 - r_{DRP}^2)/r_{DRP}^2$  to account for heterogeneous residual variances due to different reliabilities of DRP ( $r_{DRP}^2$ ), and  $\sigma_e^2$  is the residual variance.

Significance testing of SNP effects was performed using a two-sided t-test. Our aim was to refine QTL regions from the selected large regions using the same population that was used to identify these regions. Although the segregation of QTL in the targeted regions was known, regions were large (up to 11 Mb). Therefore, we assumed that, except for a few, most of the SNPs were not associated with the trait. Accordingly, our null hypothesis was  $\mathbf{g} = 0$ . A Bonferroni correction was applied to control for false positive associations. We declared a SNP significant if the  $P$  value was less than  $0.05/N$ , where  $N$  is the number of SNPs. Therefore, the significant threshold value for  $-\log_{10}(P)$  was 4.07, 5.19, and 6.20 for the MD, HD, and SEQ data, respectively. Analyses were performed using the DMU package [43].

#### Bayesian variable selection model (BVS)

The BVS model [26, 44] used in this study describes SNP effects as a mixture distribution that estimates the

effects of all SNPs simultaneously. The model is given by the following formula:

$$\mathbf{y} = \mathbf{1}\mu + \sum_{k=1}^m \mathbf{x}_k g_k + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (2)$$

where  $\mathbf{y}$ ,  $\mathbf{1}$ ,  $\mu$ ,  $\mathbf{u}$ ,  $\mathbf{Z}$ , and  $\mathbf{e}$  are defined as in the LM model. The term  $\sum_{k=1}^m \mathbf{x}_k g_k$  fits additive SNP association effects,  $\mathbf{x}_k$  is a vector of SNP genotypes (0, 1 or 2), and  $g_k$  is the effect of SNP  $k$ , which was assumed to follow a mixture distribution that assumes that most SNPs have a small phenotypic effect and some SNPs have a large effect:

$$g_k \sim \begin{cases} N(0, \sigma_{g_0}^2) & \text{with probability } \pi_0 \\ N(0, \sigma_{g_1}^2) & \text{with probability } \pi_1 = 1 - \pi_0 \end{cases}, \quad (3)$$

where  $\pi_0$  is the probability of a SNP having a small effect (i.e., drawn from a distribution with a small variance) and  $\pi_1$  is the probability of a SNP having a large effect (i.e., drawn from a distribution with a large variance). We assumed that the proportions in the two mixture distributions had a Bernoulli distribution and the prior distribution of  $\pi_0$  and  $\pi_1$  is a Beta (100, 1) distribution. We set  $\pi_0 = 0.98, 0.998, \text{ and } 0.9998$  as priors for the MD, HD, and SEQ data, respectively. We also assumed that the priors  $\mu, \sigma_{g_0}^2, \text{ and } \sigma_{g_1}^2$  followed a uniform distribution. By assuming a small variance instead of zero for the first distribution ( $N(0, \sigma_{g_0}^2)$ ), the MCMC implementation is straightforward with recognizable conditional distributions for all model parameters, as described elsewhere [26, 44]. Analysis of the BVS model was performed using the Bayz software [45]. The Gibbs sampler was run as a single chain with 52 000 samples. The first 20 000 iterations were discarded as burn-in. After this, every 20<sup>th</sup> sample of the remaining 32 000 was saved for posterior analysis.

Within the Bayesian framework, the association of each SNP with the trait can be evaluated by a Bayes factor (BF). In this study, the BF for SNP  $i$  was calculated as the ratio between the estimated posterior probability and the average of posterior probabilities of SNPs with large effects (large variance,  $g_k \sim N(0, \sigma_{g_1}^2)$ ) [46]:

$$BF_i = \frac{\hat{p}_i / (1 - \hat{p}_i)}{\hat{\pi}_1 / (1 - \hat{\pi}_1)}, \quad (4)$$

where  $\hat{p}_i$  is the posterior probability of the effect for SNP  $i$  being drawn from the distribution with large effects, and  $\hat{\pi}_1$  is the average of posteriors probabilities of SNPs with large effects. According to commonly used

criteria [33], a BF between 3.2 and 10 was considered moderate and a BF greater than 10 as strong evidence for the segregation of a QTL affecting a trait.

### QTL regions

A QTL region was detected using the posterior QTL intensity function [29]. In a Bayesian analysis, each chromosome is divided into many small intervals of equal length (bins), i.e. for instance, 1 or 2 cM in simulation studies [29, 47]. In our study, we defined the QTL intensity for each SNP on interval  $\Delta_i$  as follows [29]:

$$I(i) = \frac{1}{N} \left[ \sum_{k=1}^N \left( \sum_{q=1}^{N_{sup}^{(k)}} \mathbf{1}_{\{snp_q^{(k)} \in \text{TSD}\}} / \|\Delta_i\| \right) \right], \quad (5)$$

where  $i$  is the  $i^{\text{th}}$  bin,  $N$  is the number of MCMC samples kept for analysis, and  $\mathbf{1}_{\{snp_q^{(k)} \in \text{TSD}\}}$  is an indicator function with value 1 if the SNP in the  $k^{\text{th}}$  sample falls in the distribution with large variance (TSD), and  $\Delta_i$  is the size of bin (constant number of markers in the current study) beginning from the  $i^{\text{th}}$  position. To avoid having too many SNPs diluting the QTL effect and to balance different marker densities, the bin size was set to 14, 18 and 18 SNPs, which correspond to 1, 0.1, and 0.01 Mb for the MD, HD, and SEQ data, respectively. The QTL intensity for each bin was calculated as the average posterior probability of SNPs having a large effect within the bin.

The QTL intensity profile within a region is expected to show a peak if the region contains a QTL. Xu et al. [34] declared that the intensity profile is not able to distinguish two linked QTL regions and that the QTL effect profile is not able to accurately locate the causative mutation within an QTL region, while an intensity profile [48] that is weighted by estimated SNP effects can distinguish intervals with QTL and also show sharp peaks within intervals. Therefore, we also used the weighted QTL intensity profile that has been used in several previous studies [49, 48, 34]. The weighted QTL intensity is defined as:

$$I_w(i) = I(i) * w(i) \quad (6)$$

where  $I(i)$  is the QTL intensity of  $i^{\text{th}}$  bin, as given in Equation (5), and  $w(i)$  is the average SNP effect for the  $i^{\text{th}}$  bin.

To detect QTL regions based on the weighted QTL intensity profile, bins that were three standard deviations above or below the mean of total bins were deleted as outliers (assuming that the bins are in a QTL region) and the mean and standard deviation of the remaining bins were re-calculated for the purpose of the t test. Then, a multiple t-test ( $t = (I(i) - \bar{I})/s$ ), where  $s$  is the standard deviation of the weighted QTL intensity of



the bins) with a Bonferroni correction, was used to identify significant peaks among all the bins, including those outlier bins. Each significant peak represented a QTL region. A QTL region was specified as a region where the sum of the QTL intensities for the bins around the significant peak exceeded a predefined threshold (0.95). Finally, the average  $r^2$  (LD) between significant SNPs ( $BF \geq 3.2$ ) within a QTL region was calculated for each significant peak and between adjacent significant peaks. Adjacent peaks with an average  $r^2$  larger than 0.5 were merged into one region.

The position of each QTL region was defined according to the *Bos taurus* genome assembly UMD3.1 [50]. Genes that were located within or overlapped with the QTL regions were determined using information from the National Center for Biotechnology Information [51].

In the following, LM models with MD, HD, and SEQ data were designated  $LM_{MD}$ ,  $LM_{HD}$ , and  $LM_{SEQ}$ ; BVS models with MD, HD, and SEQ data were designated  $BVS_{MD}$ ,  $BVS_{HD}$ , and  $BVS_{SEQ}$ ; and BVS models using the

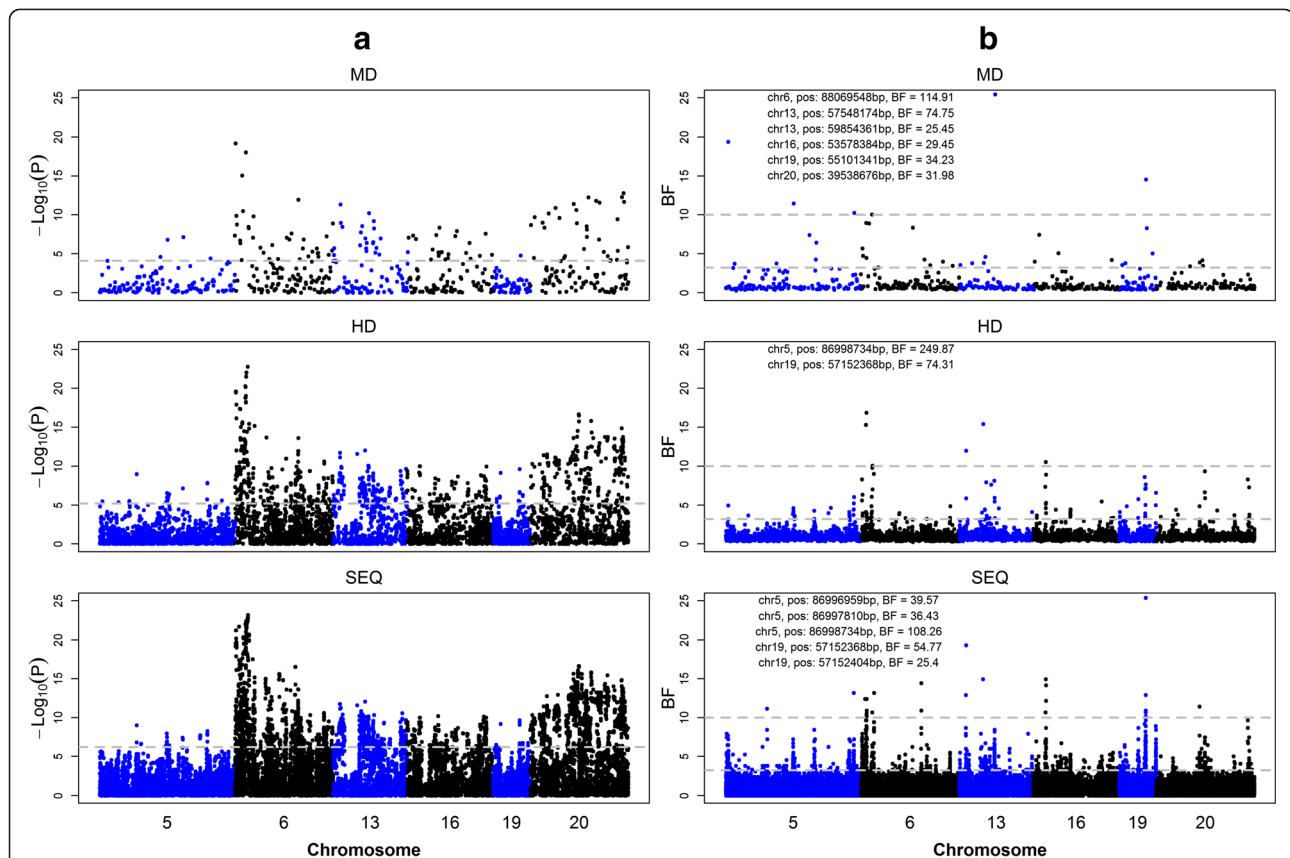
QTL intensity profile with MD, HD, and SEQ data were designated  $BVSINT_{MD}$ ,  $BVSINT_{HD}$ , and  $BVSINT_{SEQ}$ , respectively.

## Results

### Analysis using two models based on various marker densities

Figure 1 shows the associations of SNPs with mastitis using the LM and BVS models. For each model, the association patterns were similar for the three marker densities, although the number of significant SNPs decreased with decreasing marker densities, and both models showed peaks at similar locations. However, the BVS model presented clearer signals for QTL regions. In addition, with increasing marker densities, the peaks of putative QTL became sharper for the BVS model and the boundaries of adjacent QTL regions became more obvious for the LM model.

Details on the SNPs detected by  $LM_{MD}$ ,  $LM_{HD}$ ,  $LM_{SEQ}$ ,  $BVS_{MD}$ ,  $BVS_{HD}$ , and  $BVS_{SEQ}$  are in Table S1



**Fig. 1** Genome-wide plot for association of SNPs with udder health. **(a)** Plots generated by the LM model based on MD, HD, and SEQ data. **(b)** Plots generated by BVS model based on MD, HD, and SEQ data. Six chromosome regions between 84 and 95 Mb on BTA5, 88 and 96 Mb on BTA6, 57 and 63 Mb on BTA13, 48 and 55 Mb on BTA16, 55 and 58 Mb on BTA19, and 32 and 40 Mb on BTA20 are marked in alternating colors for clarity. The vertical axis is  $-\log_{10}(P)$  for the LM and BF for the BVS model, respectively; the horizontal dotted line indicates the genome-wide significance levels (for BVS, 3.2 is considered as putative and 10 is considered as strong evidence); the text on the subfigures presents the significant SNPs that were detected by the BVS model with a BF greater than 25

[See Additional file 1: Table S1]. Table 1 presents the numbers of significant SNPs that were detected with the LM and BVS models. A total of 26 (MD), 75 (HD), and 465 (SEQ) SNPs were significant for both models. The total genetic variance estimated from the BVS model was equal to 110.20 for the MD data, and the variance explained by the SNPs of the six regions was equal to 13.10 (0.32 from SNPs with small effects and 12.77 from SNPs with large effects). Based on the HD data, the total genetic variance was equal to 113.6, and the variance explained by the SNPs of the six regions was equal to 20.54 (0.35 from SNPs with small effects and 20.20 from SNPs with large effects). Based on the SEQ data, the total genetic variance was equal to 123.70, and the variance explained by the SNPs of the six regions was equal to 30.50 (0.32 from SNPs with small effects and 30.18 from SNP with large effects). Using the LM<sub>HD</sub>, there were 5, 8, 3, 19, 17, and 27 significant SNPs in the six targeted regions with a higher test statistic ( $-\log_{10}(P)$ ) than the highest test statistic obtained with the LM<sub>MD</sub>. Similarly, using LM<sub>SEQ</sub>, there were 0, 6, 0, 7, 2, and 0 SNPs with a  $-\log_{10}(P)$  greater than the highest  $-\log_{10}(P)$  achieved with the LM<sub>HD</sub> analysis for the six targeted regions.

The BVS model detected fewer significant SNPs in the six targeted regions than the LM model. Figure 2 graphically represents the position of significant SNPs on BTA6 using both models and the three marker densities. The LM model included one SNP for each run. Therefore, the  $-\log_{10}(P)$  values were consistent for the three marker densities with those of the LM model. However, when using the BVS model, the results among the three SNP datasets were not consistent. We observed three cases: (1) seven SNPs were detected by both BVS<sub>MD</sub> and BVS<sub>HD</sub>, while 70 SNPs were detected by both BVS<sub>HD</sub> and BVS<sub>SEQ</sub>; (2) some SNPs were only detected by BVS<sub>SEQ</sub>, e.g. the SNP at position 88 326 909 bp on BTA6, which is due to the fact that these SNPs were not present in MD and HD data; and (3) some SNPs were only detected by BVS<sub>MD</sub>, e.g. the SNP at position 88 656 290 bp on BTA6, which may be explained by the QTL effect being distributed over several nearby SNPs, probably because the QTL was in strong LD with many markers in the HD and SEQ data.

### Analysis of QTL intensity

QTL intensity profiles from the analysis based on the three marker datasets are in Fig. 3. The corresponding

**Table 1** Number and percentage of significant SNPs

Marker	Model	
	LM	BVS
MD	120 (20.4 %)	43 (7.3 %)
HD	967 (12.4 %)	131 (1.7 %)
SEQ	7209 (9.1 %)	1052 (1.3 %)

The analysis was implemented by the linear mixed model (LM) and the Bayesian variable selection model (BVS); figures in brackets are percentages of significant SNPs out of the total number of SNPs used in the analysis

weighted QTL intensities are in Fig. 4. The association patterns of the weighted QTL intensities were consistent across the three marker densities. BVSINT<sub>MD</sub>, BVSINT<sub>HD</sub>, and BVSINT<sub>SEQ</sub> detected one, 16, and 33 QTL intensity peaks, respectively (Fig. 5). Among the SNPs that were significant with the BVS models ( $BF > 10$ ), 36.4, 87.5, and 86.7 % were within the QTL regions identified by BVSINT<sub>MD</sub>, BVSINT<sub>HD</sub>, and BVSINT<sub>SEQ</sub>, respectively. The positions and intervals of the detected QTL intensity peaks by BVSINT<sub>MD</sub> and BVSINT<sub>HD</sub> are in Table 2. BVSINT<sub>MD</sub> detected only one QTL intensity peak on BTA6. Table 3 shows the 51 genes that were located within or adjacent to the QTL intensity peaks detected by BVSINT<sub>SEQ</sub>. Among these QTL intensity peaks, 27 were located within or overlapped with known genes, while the others were 5 to 165 kb away from the nearest known gene. The average LD ( $r^2$ ) of the 33 QTL intensity regions was equal to 0.68. The average interval length of the QTL intensity peaks became smaller as marker densities increased and was approximately 1.39, 0.22, and 0.10 Mb for BVSINT<sub>MD</sub>, BVSINT<sub>HD</sub>, and BVSINT<sub>SEQ</sub>, respectively.

### Discussion

To our knowledge, only two studies have performed sequence-based association studies for clinical mastitis [52, 38]. In this study, based on SEQ data and a BVS model for QTL intensity analysis, 33 QTL regions with an average length of 0.069 Mb were detected in the six pre-selected chromosome regions.

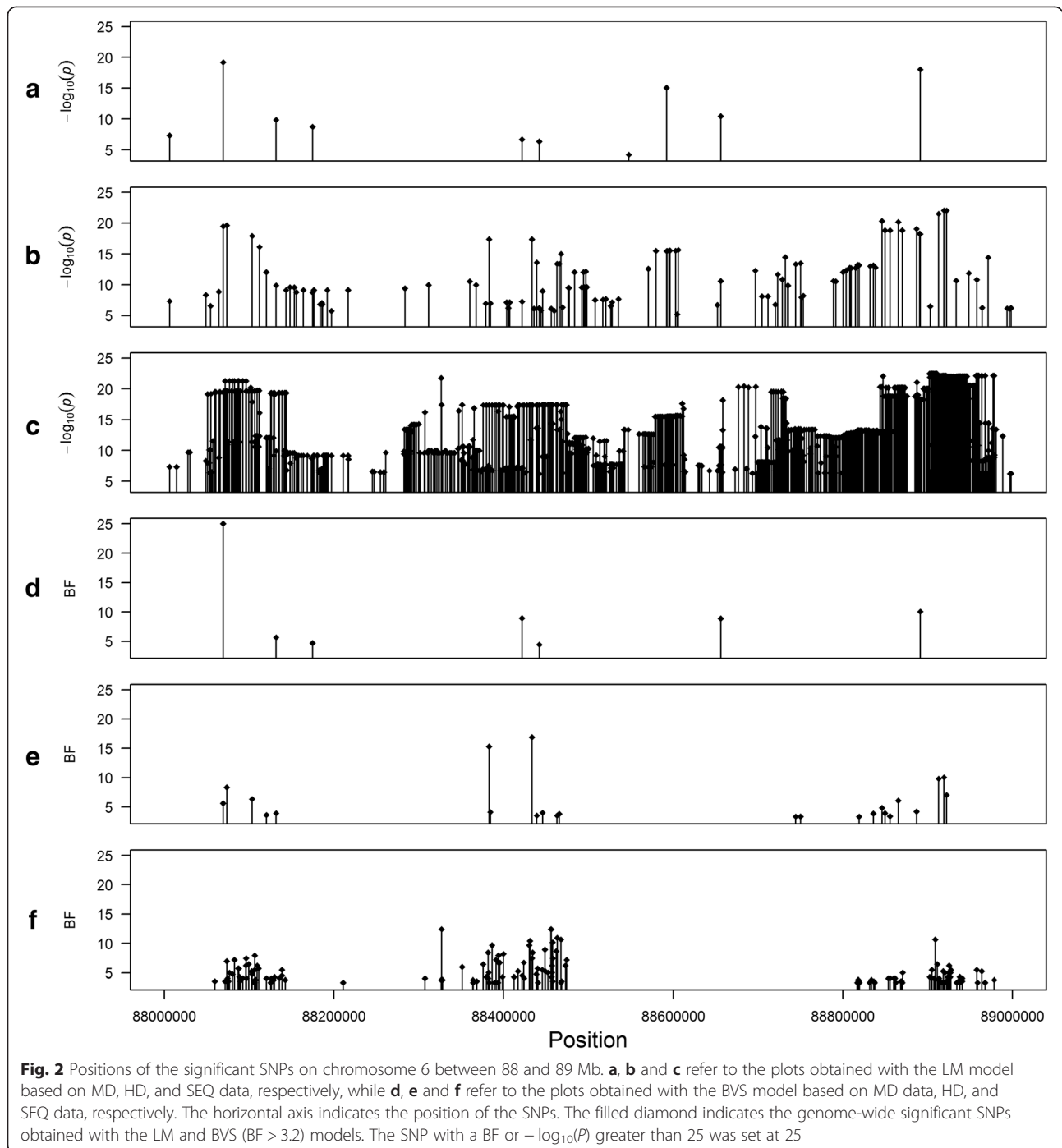
### Comparison of the two models

The association patterns between SNPs and mastitis were similar for the LM and BVS models. However, signals were much more distinct when using the BVS model (Fig. 1). This indicates that the BVS model is able to identify and narrow down regions of putative QTL and to distinguish linked QTL.

According to the significance level used to detect SNPs in this study, the number of significant SNPs identified with the BVS model was smaller than with the LM model (see Table 1). The difference in the percentage of significant SNPs between the two models increased with increasing marker densities. This is because the LM model is a single-SNP model and, thus, a QTL effect is determined by a single SNP in each run. Consequently, many SNPs in LD with the QTL can have a significant effect on the model's output. In contrast, the BVS model estimates the effects of all SNPs simultaneously, and the effect of a QTL may be represented by a single SNP or distributed over several SNPs that are in strong LD with the QTL [53] (i.e., several SNPs could together represent the effect of a single QTL).

### Comparison of the three marker densities

In a previous study, using a simulated dataset, Cleveland et al. [53] reported that Bayesian analysis showed

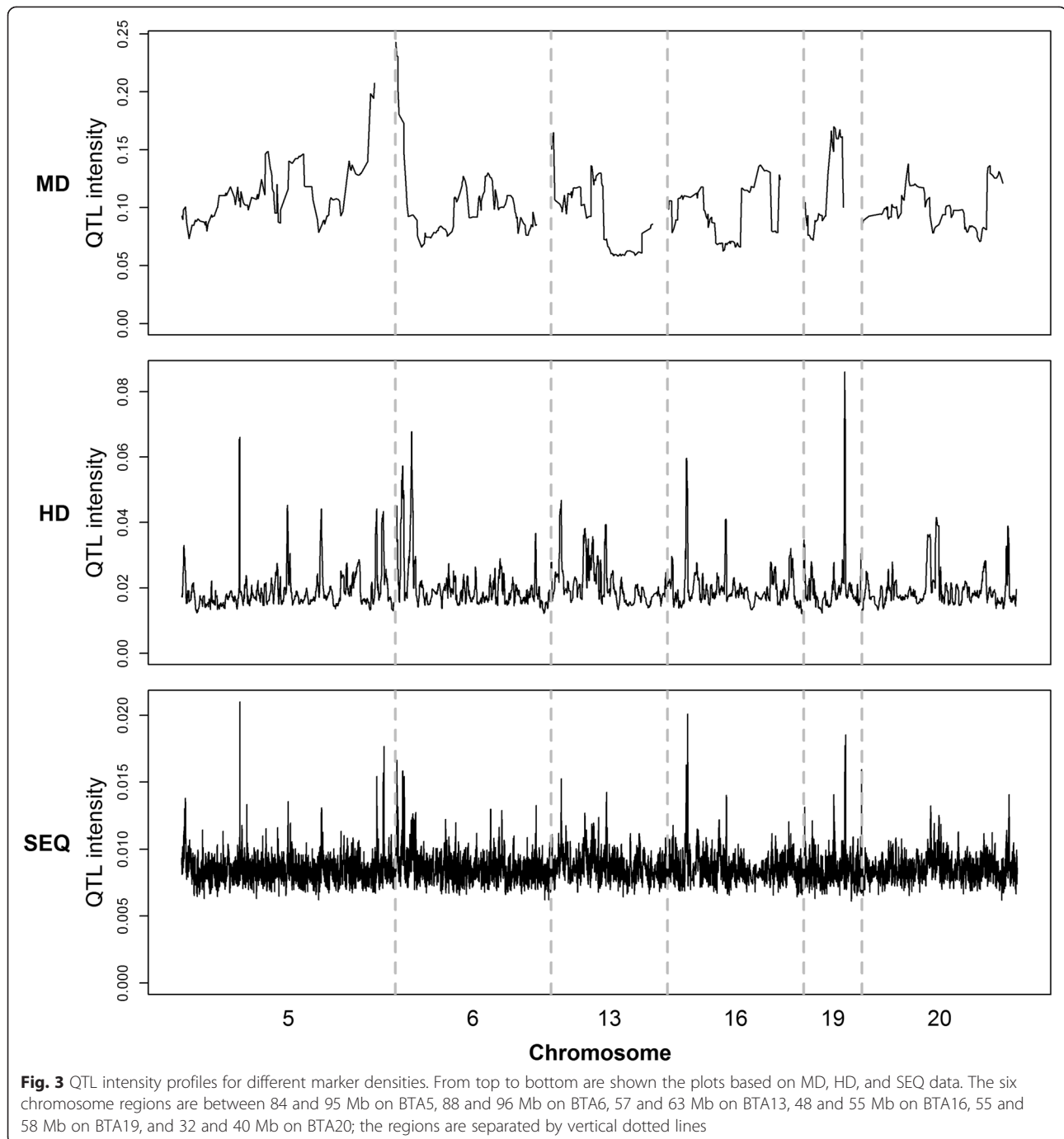


clear QTL signals when SNPs were clustered in a 0.5 to 5.0 cM region. In this study, three marker densities were analyzed, with average distances between markers of 70 kb (for MD), 5.5 kb (for HD), and 0.5 kb (for SEQ).

Our results (Fig. 1) showed that peak locations were largely consistent for the three marker densities. Spencer et al. [54] reported that marked differences in genome coverage may not translate into appreciable differences in

power to detect causative variants when using an additive model. However, it is obvious that there were fewer markers near a QTL in the MD data compared to the HD data, and therefore, there were few markers in LD with a QTL [53].

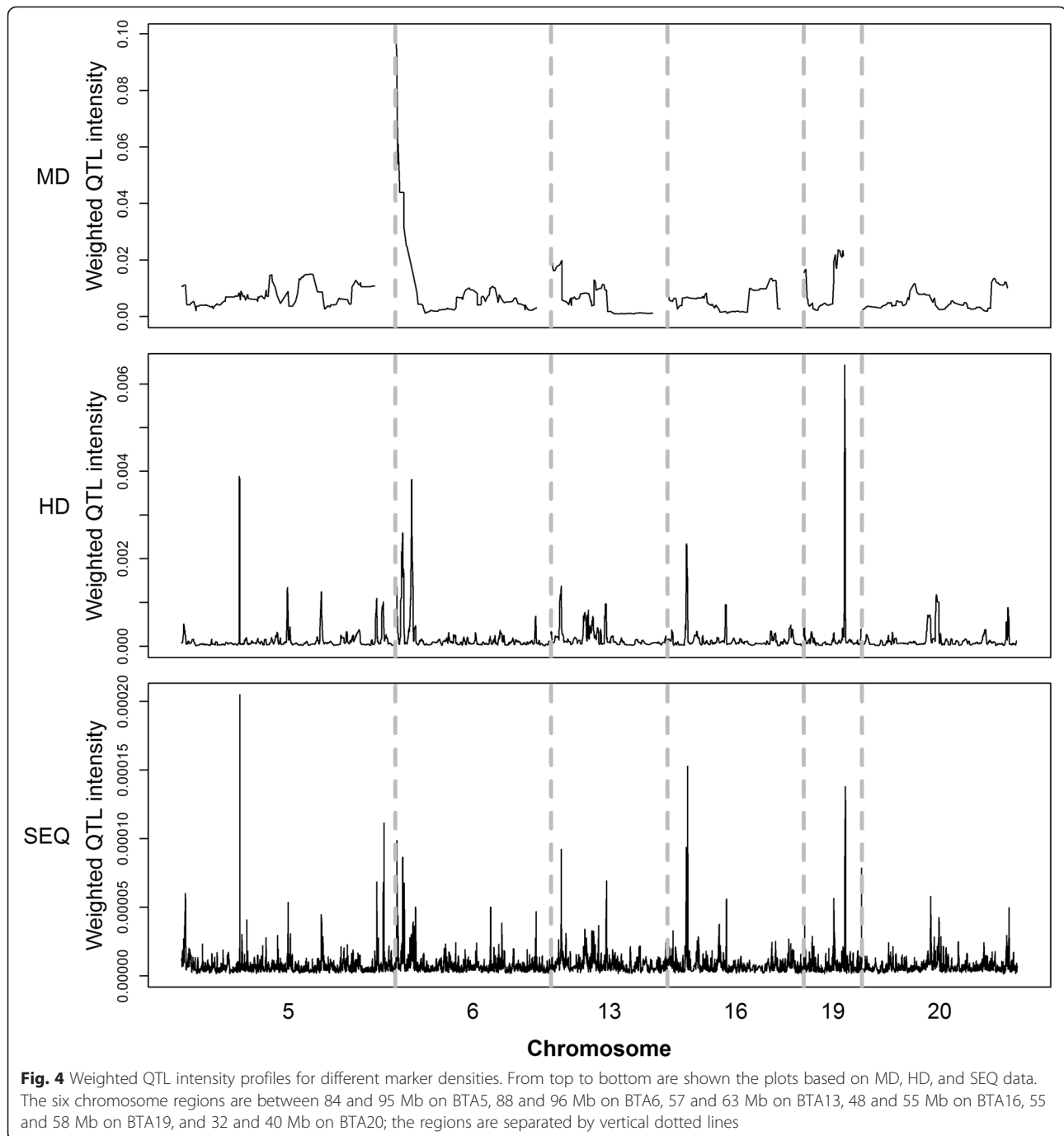
Since, in this study, a single-SNP regression was performed using the LM model and all lower-density SNPs were included in the higher-density dataset, all significant SNPs that were identified with  $LM_{MD}$  were also detected with  $LM_{HD}$ , and all significant SNPs that were identified



with  $LM_{HD}$  were also detected with  $LM_{SEQ}$  if the same significance threshold was used. However, some SNPs that were detected with  $LM_{MD}$  did not reach significance with  $LM_{HD}$ , and some SNPs detected with  $LM_{HD}$  did not reach significance with  $LM_{SEQ}$ , because of the different significance thresholds for the three marker densities as a result of Bonferroni multiple-testing correction. In addition, as density increased, the LD within a given region became stronger and a larger number of markers

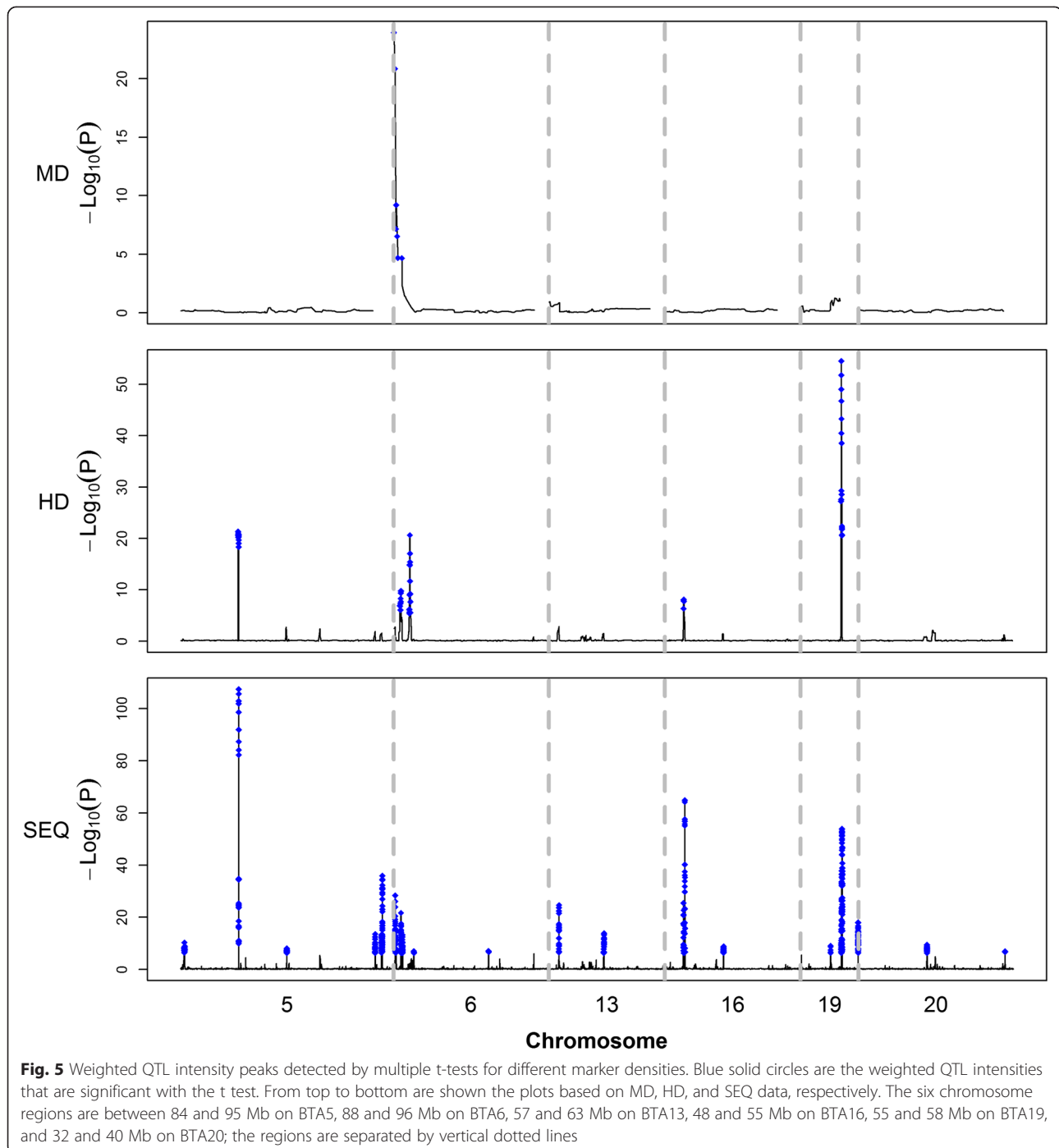
around a QTL could show an effect on the trait. Therefore, increasing marker density does not necessarily clarify the QTL region boundaries when using LM analysis. In addition, there were 79 significant SNPs for  $LM_{HD}$  with a  $-\log_{10}(P)$  that was larger than the highest  $-\log_{10}(P)$  obtained with  $LM_{MD}$  for the six regions, but only 15 significant SNPs for  $LM_{SEQ}$  with a  $-\log_{10}(P)$  that was larger than the highest  $-\log_{10}(P)$  obtained with  $LM_{HD}$  for the six regions. Thus, the relative increase





in associated SNPs from the HD to the SEQ data was less than the increase from the MD to the HD data. This may be due to imperfect imputation of sequence data due to the small reference size or may reflect that the density of the HD data was sufficient for QTL detection and further density increases only give small improvements. However, it should also be noted that both the MD and HD data have ascertainment biases towards common variants, which are more suitable for GWAS.

In contrast, some significant SNPs in the  $BVS_{MD}$  set were not detected when using  $BVS_{HD}$  and  $BVS_{SEQ}$  data, and some significant SNPs in the  $BVS_{HD}$  set were not detected in the  $BVS_{SEQ}$  set, although the same BF criterion was used for the three marker densities. Moreover, increasing marker density did not result in a higher BF. Since the BVS model fits all markers simultaneously, it is possible that when markers are very close to a QTL but are not part of low-density data, the QTL effect may be shifted to nearby markers [53], or a QTL effect may



be represented by a single SNP when using low-density data but shared by many SNPs when using high-density data. This suggests that instead of a single SNP, a summary statistic across a small region is necessary to detect a QTL when using the BVS model.

In spite of the interesting properties of high-throughput sequencing, it is necessary to take some of the limitations of imputed sequence data into consideration. Only 89 sequenced animals from three breeds were used as

imputation reference population and, therefore, imputed sequences were expected to have relatively low imputation accuracy, especially for SNPs with low minor allelic frequencies. Brøndum et al. [42] reported an imputation accuracy of around 0.90 when the reference population included 242 individuals, which is much lower than the accuracy obtained (0.97) in imputation of HD markers from 50 k data for Nordic Holstein cattle [55]. In addition, the errors in calling

**Table 2** QTL regions detected by QTL intensity analysis using the MD and HD data

Marker	Chr	Start_pos (bp)	End_pos (bp)	Reg_size (bp)
MD	6	88006286	89399736	1,393,450
HD	5	86941977	87064339	122,362
	5	89391490	89553633	162,143
	5	91099849	91359899	260,050
	5	93966483	94118706	152,223
	5	94289241	94526253	237,012
	6	88006286	88275373	269,087
	6	88283864	88476481	192,617
	6	88808116	88964217	156,101
	13	57446360	57841028	394,668
	13	59741546	60044106	302,560
	16	48939356	49090916	151,560
	16	50915500	51130913	215,413
	19	57101615	57198154	96,539
	20	35703892	36030444	326,552
	20	39473971	39663597	189,626

Chr = chromosome; Start\_pos = start position of the QTL region; End\_pos = end position of the QTL region; Reg\_size = size of the QTL region

variants may be higher with high-throughput sequencing than with SNP-array genotyping data [42]. Daetwyler et al. [22] also pointed out that results of association studies that are based on imputed sequences should be interpreted with caution, since SNPs with slightly higher P values than the most significant SNPs from multiple-testing can also be considered as potential causative mutations, particularly if there is strong supporting functional evidence. In addition, with the very high density of SNPs in sequence data, it is expected that a large number of SNPs will be in high or near perfect LD with the causal variants. This makes it difficult to distinguish causal variants from SNPs in high or near perfect LD with the causal variants using a Bayesian approach since the QTL effect may be diluted across many neighboring SNPs. Therefore, in this study, we removed one of the SNPs from a pair in which two SNPs were in complete LD although this may cause the loss of causative variants. However, it is most likely that adjacent SNPs that are in complete LD are close to each other. In other words, if a causative variant is removed by LD-based pruning, it will be replaced by another SNP that is closely associated with the causal variant. Thus in these conditions, the putative loss of causative variants should have little influence on the identification of QTL peaks. In fact, if a neutral SNP is in complete LD with a causal SNP, the latter cannot be differentiated using either the single-SNP linear mixed model or the

Bayesian model. However, the Bayesian approach is more appropriate for narrowing QTL regions and distinguishing multiple QTL regions that segregate closely. Post-analysis based on bins using the Bayesian approach without pruning the SNPs may avoid the possible loss of causal variants and dilution of QTL effects.

### QTL regions

In the BVS model, the QTL effect may be distributed over several markers that are in LD with the QTL. Therefore, combining the posterior probabilities of closely located markers can result in higher power when inferring the presence of a QTL [23]. Detection of QTL regions based on QTL density profiles was initially proposed by Sillanpaa and Arjas [29]. Using simulated data, QTL regions are easy to determine because of the simplicity of the simulation [56]. However, when using real data, the significance threshold for QTL intensity peaks is not as well defined because of the low signal to noise ratio (Fig. 4). Previous studies have detected QTL peaks by using QTL intensity, but only a few defined the interval and specific position of QTL regions [34, 56, 48]. In this study, we used multiple t-tests based on weighted QTL intensities to determine QTL regions.

Based on the present data, the association patterns of QTL intensities (Fig. 4) and BF (Fig. 1) were similar. However, the QTL intensities presented clear peaks. Analysis of the QTL intensity profiles detected only one QTL peak on BTA6. This indicates that studies with higher marker densities have greater power to detect QTL regions by QTL intensity profile analysis. In addition, the average length of the detected QTL regions became shorter as marker densities increased (Tables 2 and 3), which suggests that QTL regions can be refined by using high-density markers and applying the BVS model together with a QTL intensity profile.

Based on SEQ data, the average  $r^2$  (LD) of significant SNPs between a pair of QTL regions was lower than 0.55, while the average  $r^2$  of significant SNPs within each QTL region (average 0.68) was higher than the  $r^2$  among different QTL regions. In addition, comparisons between QTL intensity profiles and BF (>10) showed that 87 % of the significant SNPs detected by BF were located within the QTL regions determined by QTL intensity profiles based on the HD and SEQ data. Among the 33 QTL regions detected by QTL intensity profiles using the SEQ data, 27 regions contained a known gene. For the remaining six regions, the closest gene was at most 165 kb away.

Five of the detected regions have been reported by previous studies or contain genes that are known to be functionally associated with mastitis traits. One of these QTL regions (between 88 056 115 and 88 164 130 bp on BTA6) overlaps with the *deoxycytidine kinase* (*DCK*)

**Table 3** Genes located within or closest to the QTL regions

Chr	Start_pos(bp)	End_pos(bp)	Reg_size	Gene symbol	Distance(bp)	r <sup>2</sup>
5	84128853	84232545	103,692	<i>SSPN</i>	overlap	0.51
5	86979804	87132636	152,832	<i>SOX5,LOC101906017</i>	overlap	0.36
5	87331630	87387548	55,918	<i>TRNAG-CCC</i>	121,327	0.65
5	88917412	88967709	50,297	<i>LDHB</i>	overlap	0.83
5	89457697	89510420	52,723	<i>SLCO1C1</i>	Overlap	1.00
5	89581859	89643424	61,565	<i>PDE3A</i>	Overlap	0.88
5	91175414	91271746	96,332	<i>PLEKHA5</i>	Overlap	0.56
5	94025000	94448464	423,464	<i>DERA</i>	Overlap	0.31
6	88056115	88164130	108,015	<i>DCK</i>	Overlap	0.58
6	88355595	88476214	120,619	<i>SLC4A4</i>	Overlap	0.70
6	88813863	88828238	14,375	<i>GC</i>	74,566	1.00
6	88841591	89072556	230,965	<i>NPFRR2</i>	102,294	0.79
6	92877489	92929724	52,235	<i>SCARB2, FAM47E</i>	Overlap	0.62
6	93440325	93546958	106,633	<i>SOWAHB,SEPT11</i>	Overlap	0.63
6	95216638	95255224	38,586	<i>LOC101904933</i>	Overlap	0.63
13	57509023	57581780	72,757	<i>EDN3</i>	Overlap	0.75
13	57733575	57868987	135,412	<i>SLMO2, LOC100140850</i>	Overlap	0.99
13	58701953	58817166	115,213	<i>C13H20orf85</i>	within	0.78
13	59072545	59132622	60,077	<i>ZBP1</i>	within	0.25
13	59150552	59236725	86,173	<i>CTCFL,LOC101903272</i>	within	0.41
13	59421380	59480987	59,607	<i>BMP7</i>	overlap	0.53
13	59847351	59865177	17,826	<i>LOC101903611</i>	18,568	0.97
16	48274699	48326469	51,770	<i>NPHP4</i>	overlap	0.64
16	48926813	49072532	145,719	<i>LOC100297820</i>	138,439	0.61
16	50624458	50723980	99,522	<i>LOC515105,WRAP73,TPRG1L,MEGF6</i>	overlap	0.52
16	51020468	51057724	37,256	<i>PRDM16</i>	overlap	0.95
19	55035976	55078482	42,506	<i>SEPT9</i>	overlap	0.99
19	56545652	56643735	98,083	<i>SAP30BP, RECQL5, SMIM5, LOC100302389, MYO15B, LOC101907888</i>	overlap	0.97
19	57102415	57184032	81,617	<i>FDXR, USH1G, OTOP2, FADS6</i>	overlap	0.36
19	57979089	57999724	20,635	<i>RPL38</i>	164,726	0.58
20	35525215	35566004	40,789	<i>OSMR,RICTOR</i>	overlap	0.66
20	35900432	36010180	109,748	<i>LIFR,EGFLAM</i>	overlap	0.67
20	39432294	39613378	181,084	<i>RAI14</i>	overlap	0.77

QTL regions were detected by analysis of QTL intensity using SEQ data; Chr = chromosome; Start\_pos = start position of the QTL region; End\_pos = end position of the QTL region; Reg\_size = size of the QTL region; Distance = distance between gene and QTL region; r<sup>2</sup> = linkage disequilibrium between significant SNPs (BF ≥ 3.2) within each region

gene. Abdel-Shafy et al. [57] reported that this gene is associated with somatic cell score, which is a mastitis indicator [58]. Another QTL region (between 88 355 595 and 88 476 214 bp on BTA6) overlapped with the *solute carrier family 4, sodium bicarbonate co-transporter, and member 4 (SLC4A4)* gene belonging to the *SLC4* family. Sodeland et al. [59] reported a high correlation of SNPs surrounding the *SLC4A4*

gene with clinical mastitis. A third QTL region between 88 984 167 and 89 072 556 bp on BTA6, overlapped with a previously reported QTL region between 89 and 91 Mb for mastitis [60]. This region contains a part of the *neuropeptide FF receptor 2 (NPFRR2)* gene that encodes a member of a G-protein-coupled neuropeptide receptors subfamily that is activated by the neuropeptides A-18-amide

(NPAF) and F-8-amide (NPFF) [61]. Sun et al. [61] showed that NPFF is involved in anti-inflammatory effects, both *in vitro* and *in vivo*, and Minault et al. [62] showed that it modulates the proliferation of human T lymphocytes. In addition, Sodeland et al. [59] detected a QTL region around 35.5 Mb on BTA20 which is associated with clinical mastitis. A fourth QTL region, between 35 900 432 and 36 010 180 bp on BTA20, was detected and overlapped with the *leukemia inhibitory factor receptor alpha* (*LIFR*) gene. *LIFR* is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker [63]. Finally, a QTL region between 57 509 023 and 57 581 780 bp on BTA13 overlapped with the QTL at 57.54 Mb reported by Sahana et al. [64]. This region overlaps with the *endothelin 3* (*EDN3*) gene, which influences neutrophil activation [65]. In blood, neutrophils are the major leukocytes that respond to inflammatory stimuli. Therefore, *NPFFR2*, *SLC44A4*, *DCK*, *LIFR*, and *EDN3* are candidate genes for susceptibility to mastitis.

## Conclusions

The power of QTL detection can be increased by increasing marker densities and the BVS model outperforms the LM model in refining QTL locations with clear boundaries between linked QTL. Based on the results obtained with the SEQ data, six preselected regions were refined into 33 candidate QTL regions for udder health. Furthermore, the comparison between these candidate QTL regions and known genes suggests that *NPFFR2*, *SLC44A4*, *DCK*, *LIFR*, and *EDN3* may be considered as candidate genes for mastitis susceptibility. Further studies are required to validate the causative loci that underlie these QTL and to investigate the function of the candidate genes that affect udder health.

## Additional file

**Additional file 1: Table S1.** Position of detected SNPs by LM<sub>MD</sub>, LM<sub>HD</sub>, LM<sub>SEQ</sub>, BVS<sub>MD</sub>, BVS<sub>HD</sub>, and BVS<sub>SEQ</sub>. **Table S1.** contains six sheets that provide the position of significant SNPs detected by LM<sub>MD</sub>, LM<sub>HD</sub>, LM<sub>SEQ</sub>, BVS<sub>MD</sub>, BVS<sub>HD</sub>, and BVS<sub>SEQ</sub>.

## Competing interests

The results presented here are part of a patent application number PCT/EP2013/067838 entitled 'Genetic markers for mastitis resistance' (AU-2012-634-489).

## Authors' contributions

XW, GS and MSL conceived and designed the experiments. XW performed the experiments and analyzed the data. MSL, GSA, BG, DS, QZ and GSU contributed reagents, materials and analysis tools. XW wrote the paper. All authors read, revised and approved the final manuscript.

## Acknowledgments

This work was supported by the research project "Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253,

Denmark)" and the Chinese Scholarship Council (CSC) that provided a scholarship for the first author.

## Author details

<sup>1</sup>Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark. <sup>2</sup>Key Laboratory of Animal Genetics and Breeding of Ministry of Agriculture, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China.

Received: 11 September 2014 Accepted: 21 May 2015

Published online: 19 June 2015

## References

- Hogeveen H, Pyorala S, Waller KP, Hogan JS, Lam TJGM, Oliver SP, et al. Current status and future challenges in mastitis research. In Proceedings of the National Mastitis Council 50<sup>th</sup> Annual Meeting: 23–26 January 2011; Arlington: the National Mastitis Council 50<sup>th</sup> Annual Meeting; 2011:36–48.
- Awale M, Dudhata G, Avinash K, Chauhan B, Kamani D, Modi C, et al. Bovine mastitis: a threat to economy. *Sci Rep*. 2012;1:1–10.
- Hameed KGA, Sender G, Korwin-Kossakowska A. Public health hazard due to mastitis in dairy cows. *Anim Sci Pap Rep*. 2007;25:73–85.
- Bennedsgaard TW, Enevoldsen C, Thamsborg SM, Vaarst M. Effect of mastitis treatment and somatic cell counts on milk yield in Danish organic dairy cows. *J Dairy Sci*. 2003;86:3174–83.
- Nash DL, Rogers GW, Cooper JB, Hargrove GL, Keown JF, Hansen LB. Heritability of clinical mastitis incidence and relationships with sire transmitting abilities for somatic cell score, udder type traits, productive life, and protein yield. *J Dairy Sci*. 2000;83:2350–60.
- Lund T, Miglier F, Dekkers JCM, Burnside EB. Genetic-relationships between clinical mastitis, somatic-cell count, and udder conformation in Danish Holsteins. *Livest Prod Sci*. 1994;39:243–51.
- Hansen M, Lund MS, Sorensen MK, Christensen LG. Genetic parameters of dairy character, protein yield, clinical mastitis, and other diseases in the Danish Holstein cattle. *J Dairy Sci*. 2002;85:445–52.
- Heringstad B, Klemetsdal G, Ruane J. Selection for mastitis resistance in dairy cattle: a review with focus on the situation in the Nordic countries. *Livest Prod Sci*. 2000;64:95–106.
- Rupp R, Boichard D. Genetic parameters for clinical mastitis, somatic cell score, production, udder type traits, and milking ease in first lactation Holsteins. *J Dairy Sci*. 1999;82:2198–204.
- Lund MS, Jensen J, Petersen PH. Estimation of genetic and phenotypic parameters for clinical mastitis somatic cell production deviance, and protein yield in dairy cattle using Gibbs sampling. *J Dairy Sci*. 1999;82:1045–51.
- Lescouret F, Coulon JB. Modeling the impact of mastitis on milk-production by dairy cows. *J Dairy Sci*. 1994;77:2289–301.
- Georges M. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annu Rev Genomics Hum Genet*. 2007;8:131–62.
- Sahana G, Lund MS, Andersson-Eklund L, Hastings N, Fernandez A, Iso-Touru T, et al. Fine-mapping QTL for mastitis resistance on BTA9 in three Nordic red cattle breeds. *Anim Genet*. 2008;39:354–62.
- Holmberg M, Andersson-Eklund L. Quantitative trait loci affecting health traits in Swedish dairy cattle. *J Dairy Sci*. 2004;87:2653–9.
- Klungland H, Sabry A, Heringstad B, Olsen HG, Gomez-Raya L, Vage DI, et al. Quantitative trait loci affecting clinical mastitis and somatic cell count in dairy cattle. *Mamm Genome*. 2001;12:837–42.
- Kühn C, Bennewitz J, Reinsch N, Xu N, Thomsen H, Looft C, et al. Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *J Dairy Sci*. 2003;86:360–8.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*. 2009;4, e3550.
- Cole JB, Wiggins GR, Ma L, Sonstegard TS, Lawlor Jr TJ, Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics*. 2011;12:408.
- Mai MD, Sahana G, Christiansen FB, Guldbandsen B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50 K single nucleotide polymorphism chip. *J Anim Sci*. 2010;88:3522–8.



20. Matukumalli LK, Schroeder S, DeNise SK, Sonstegard T, Lawley CT, Georges M, et al. Analyzing LD blocks and CNV segments in cattle: novel genomic features identified using the BovineHD BeadChip. San Diego: Illumina Inc; 2011. Pub. No. 370-2011-002.
21. Rincon G, Weber KL, Eenennaam AL, Golden BL, Medrano JF. Hot topic: performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *J Dairy Sci.* 2011;94:6116–21.
22. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet.* 2014;46:858–65.
23. Sahana G, Gulbrandsen B, Janss L, Lund MS. Comparison of association mapping methods in a complex pedigreed population. *Genet Epidemiol.* 2010;34:455–62.
24. Yu JM, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38:203–8.
25. Dashhab GR, Kadri NK, Shariati MM, Sahana G. Comparison of linear mixed model analysis and genealogy-based haplotype clustering with a Bayesian approach for association mapping in a pedigreed population. *BMC Proc.* 2012;6:54.
26. George EI, McCulloch RE. Variable selection via Gibbs sampling. *J Am Stat Assoc.* 1993;88:881–9.
27. Sahana G, Kadlecova V, Hornshoj H, Nielsen B, Christensen OF. A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait. *J Anim Sci.* 2013;91:1041–50.
28. Pausch H, Jung S, Edel C, Emmerling R, Krogmeier D, Gotz KU, et al. Genome-wide association study uncovers four QTL predisposing to supernumerary teats in cattle. *Anim Genet.* 2012;43:689–95.
29. Sillanpää MJ, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics.* 1998;148:1373–88.
30. Satagopan JM, Yandell YS, Newton MA, Osborn TC. A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics.* 1996;144:805–16.
31. Heath SC. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet.* 1997;61:748–60.
32. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. *PLoS Comput Biol.* 2012;8, e1002822.
33. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc.* 1995;90:773–95.
34. Xu CW, Wang XF, Li ZK, Xu SZ. Mapping QTL for multiple traits using Bayesian statistics. *Genet Res (Camb).* 2009;91:23–37.
35. Nordic cattle genetics evaluation. <http://www.nordicebv.info/Forside.htm>. Accessed 20 March 2013.
36. Goddard M. A method of comparing sires evaluated in different countries. *Livest Prod Sci.* 1985;13:321–31.
37. Schaeffer LR. Model for international evaluation of dairy sires. *Livest Prod Sci.* 1985;12:105–15.
38. Sahana G, Gulbrandsen B, Thomsen B, Holm LE, Panitz F, Brondum RF, et al. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *J Dairy Sci.* 2014;97:7258–75.
39. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 2009;84:210–23.
40. Lund MS, de Roos APW, de Vries AG, Druet T, Ducrocq V, Fritz S, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol.* 2011;43:43.
41. Hoglund J, Sahana G, Brondum R, Gulbrandsen B, Buitenhuis B, Lund M. Fine mapping QTL for female fertility on BTA04 and BTA13 in dairy cattle using HD SNP and sequence data. *BMC Genomics.* 2014;15:790.
42. Brondum RF, Gulbrandsen B, Sahana G, Lund MS, Su G. Strategies for imputation to whole genome sequence using a single or multi-breed reference population in cattle. *BMC Genomics.* 2014;15:728.
43. Madsen P., Jensen J. A user's guide to DMU. University of Aarhus, Faculty Agricultural Sciences. 2010. <http://dmu.agrsci.dk/DMU/Linux/Previous/dmuv6-R5.1-EM64T-build-2012-09-21.tar.gz>. Accessed 20 June 2013.
44. Meuwissen THE. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet Sel Evol.* 2009;41:35.
45. Heuven HC, Janss LL. Bayesian multi-qtI mapping for growth curve parameters. *BMC Proc.* 2010;4 Suppl 1:S12.
46. Janss L. Genomic models in bayz. <http://www.bayz.biz/Bayz.GenomicModels.dec2010.pdf>. Accessed 20 June 2013.
47. Wang H, Zhang YM, Li XM, Masinde GL, Mohan S, Baylink DJ, et al. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics.* 2005;170:465–80.
48. Xu S, Yi N. Mixed model analysis of quantitative trait loci. *Proc Natl Acad Sci U S A.* 2000;97:14542–7.
49. Fang M. Bayesian shrinkage mapping of quantitative trait loci in variance component models. *BMC Genet.* 2010;11:30.
50. Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009;10:R42.
51. National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov/>. Accessed 20 July 2013.
52. Qanbari S, Pausch H, Jansen S, Somel M, Strom TM, Fries R, et al. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 2014;10, e1004148.
53. Cleveland MA, Deeb N. Evaluation of a genome-wide approach to multiple marker association considering different marker densities. *BMC Proc.* 2009;3:55.
54. Spencer CCA, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 2009;5, e1000477.
55. Brondum RF, Ma P, Lund MS, Su G. Short communication: genotype imputation within and across Nordic cattle breeds. *J Dairy Sci.* 2012;95:6795–800.
56. Liu JF, Liu YJ, Liu XG, Deng HW. Bayesian mapping of quantitative trait loci for multiple complex traits with the use of variance components. *Am J Hum Genet.* 2007;81:304–20.
57. Abdel-Shafy H, Bortfeldt RH, Tetens J, Brockmann GA. Single nucleotide polymorphism and haplotype effects associated with somatic cell score in German Holstein cattle. *Genet Sel Evol.* 2014;46:35.
58. Heringstad B, Gianola D, Chang YM, Odegard J, Klemetsdal G. Genetic associations between clinical mastitis and somatic cell score in early first-lactation cows. *J Dairy Sci.* 2006;89:2236–44.
59. Sodeland M, Kent MP, Olsen HG, Opsal MA, Svendsen M, Sehested E, et al. Quantitative trait loci for clinical mastitis on chromosomes 2, 6, 14 and 20 in Norwegian Red cattle. *Anim Genet.* 2011;42:457–65.
60. Sodeland M, Grove H, Kent M, Taylor S, Svendsen M, Hayes BJ, et al. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. *BMC Genet.* 2011;12:70.
61. Sun YL, Zhang XY, Sun T, He N, Li JY, Zhuang Y, et al. The anti-inflammatory potential of neuropeptide FF in vitro and in vivo. *Peptides.* 2013;47:124–32.
62. Minault M, Lecron JC, Labrousse S, Simonnet G, Gombert J. Characterization of binding-sites for neuropeptide FF on T-lymphocytes of the Jurkat cell-line. *Peptides.* 1995;16:105–11.
63. Chen DH, Sun YT, Wei YK, Zhang PJ, Rezaeian AH, Teruya-Feldstein J, et al. LIFR is a breast cancer metastasis suppressor upstream of the Hippo-YAP pathway and a prognostic marker. *Nat Med.* 2012;18:1511–7.
64. Sahana G, Gulbrandsen B, Thomsen B, Lund MS. Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle. *Anim Genet.* 2013;44:620–6.
65. Elferink JGR, De Koster BM. The involvement of protein kinase G in stimulation of neutrophil migration by endothelins. *Eur J Pharmacol.* 1998;350:285–91.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

