Short communication

Comparison of two QTL mapping approaches based on Bayesian inference using high-dense SNPs markers

G. R. Dashab*

Department of Animal Science, Faculty of Agriculture, University of Zabol.

* Corresponding author, E-mail address: dashab@uoz.ac.ir

Abstract To compare different QTL mapping methods, a population with genotypic and phenotypic data was simulated. In Bayesian approach, all information of markers can be used along with combination of distributions of SNP markers. It is assumed that most of the markers (95%) have minor effects and a few numbers of markers (5%) exert major effects. The simulated population included a basic population of 1020 non-relative cattle that was continuously crossed for 4 generations to make disequilibrium linkage among QTL position and markers. In all generations, 20 bulls were mated with 1,000 cows and each cow produced only one offspring. Whole tree family included 4100 head of livestock. Genotype of 10000 SNPs on 5 chromosomes at equal distance (0.05 cM) in the total population was simulated. The length of each chromosome was 100 cM. Simulated trait was milk production. Progeny of the first to third generation had record but the basic population and fourth generation of offspring did not have any record. Therefore, from the total population of 4100 heads, 3000 cattle had record. Two different models, Bayz A and Bayz B, were used to analyze QTL linked to the SNP markers. Analysis was conducted by BAYZ software. SNPs with more than 0.6 effect or Bayes factor (BF) greater than 5.5 were considered as QTL. The resultant analysis of two models of BAYZ A and BAYZ B were 7 and 9 QTL locations on 5 chromosomes, respectively. QTL position identified by BAYZ B method was matched on simulated location, but showed a false positive on chromosome 4. QTL positions identified by BAYZ A method were located near by the simulated positions, but with many false positive points.

Keywords: bayesian analysis, QTL mapping, bayes factor, genomics

Received: 22 Oct. 2014, *accepted:* 25 May 2015, *published online:* 16 Jun. 2015

Introduction

The purpose of genomic selection is to capture all controller stations of a quantitative trait that is conducted by tracking the adjacent markers on chromosomal fragments (Meuwissen et al., 2001). The traditional methods problems of selection based on marker data can be overcome by using the high-dense markers (Meuwissen et al., 2001). To estimate breeding values based on genomic selection, different methods have been proposed including GBLUP, BAYZ A, BAYZ B and BAYZ C or LASSO (Yi and Xu, 2008; Zhang et al., 2010; Habier et al., 2011).

Unlike complex models based on the single marker data or random haplotype, one or a limited number of markers are entered into the model at each stage while in BAYZ models, all of marker data are used simultaneously in the model with different distribution. Also, in the BAYZ B method, the combination of different distributions is used for the effects of SNP markers in the model and it is not required to conduct multiple tests. It is assumed that the most markers have minor effects in the model (about 95%) and only limited a few number of markers (approximately 5%) are with major effects (George and McCulloch, 1993; Hill et al., 2008). Therefore, the aim of the present study was to investigate the accuracy of two methods, BAYZ A and BAYZ B, in detection of QTL.

Materials and Methods

Simulated population

Basic population includes 1020 non-relative animals. This population with a combination of 1000 female animals and 20 male parents over four generations (G1-G4) (Fig. 1) were mated randomly to provide the final population for conducting genomic studies. Trait studied was production of milk. 3000 progenies resulting from the intercourses of generations 1 to 3 had milk pro-

duction record. Genotypes of 4100 animals of five generations were simulated for 10,000 SNP markers on 5 chromosomes (2,000 markers on each chromosome). Length of each chromosome was 100 cM and markers were located at equal distance (0.05 cM). Markers with frequency less than 0.05 were removed from the statistical model. Phenotypic data and marker data were simulated by the QMSim software (Sargolzaei and Schenkel, 2009).

BAYS A model

Effect of SNP marker in the final population was estimated based on the following statistical model.

$$y=1\mu+Xg+e$$

Where, g and X are vector of random effect of marker SNP and genotypic coefficients matrix, respectively (0, 1 and 2 for 11, 12 and 22 genotypes, respectively). It is assumed that all SNP data have a similar effect in the BAYZ A model but their variance is different. The parameters of the model were estimated by the BAYZ software (Janss, 2011).

BAYZ B model (Bayesian Variable Selection)

By definition, this model has different distributions of combination of SNP markers effects when is used in the analysis of all markers simultaneously. In this model, it is assumed that the majority of markers have very small effects on trait (98%) and only a minor fraction of markers (2%) are with major effects.

Analysis of the above models was conducted by the BAYZ software (Janss, 2011) and the variance components of two above combinations were estimated. SNP markers that had a Bayes factor greater than 5.5 were introduced as QTL location.

Results and Discussion

Results of analyzing BAYZ approach of different models on chromosomes 1 to 5 have been shown in Fig. 2. BAYZ method in QTL detection is better than other models because it has the minimum number of locations of false positive (FP). From 9 QTL simulated on 5 chromosomes, BAYZ A model recognized only 7 locations and the simulated QTL on chromosomes 2 and 3 have been detected as false negative points by this model (Fig. 2). However, BAYZ B method could detect all of the simulated QTL and was associated with minimum false positive points or locations that erroneously had been detected as QTL. In BAYZ models, the effects of all markers were estimated simultaneously. Therefore, because of correlation between two QTL with epistasis effect, the more contribution of the first QTL in trait variation may lead to decrease of chance detection of the second QTL (Hill et al., 2008). Wang et al. (2012) compared five different methods for estimating breeding values and QTL mapping and reported that the performance of BAYZ A, BAYZ B and BAYZ C methods were identical but BAYZ B model can be considered the most accurate model in QTL mapping studies.

Conclusion

Although two BAYZ methods used in the present study have had similar performance in detection of simulated QTL, BAYZ B model, which includes two different distributions for markers effects, identified more accurate location of QTL but also resulted in fewer errors such as false negative and false positive providing more accu-

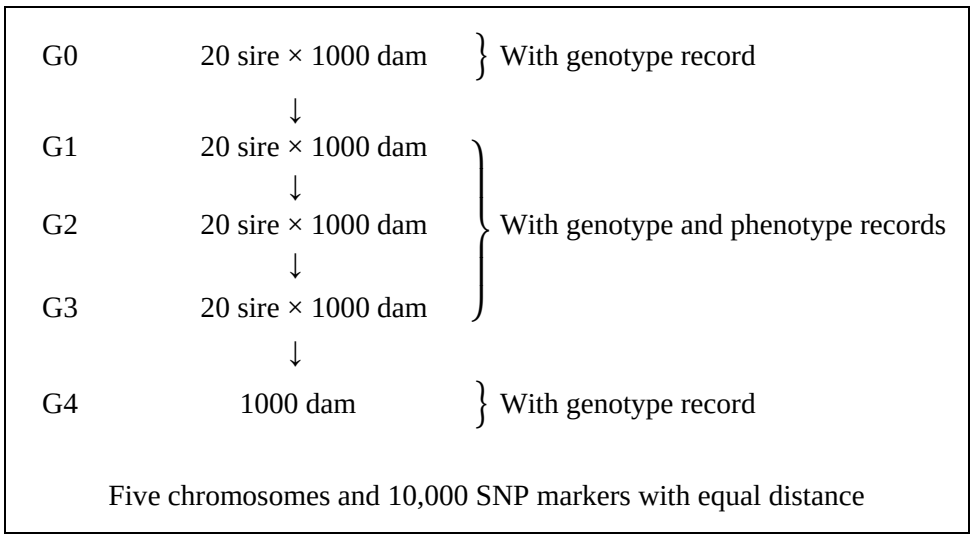

Fig. 1. Characteristic of population simulated for QTL mapping study

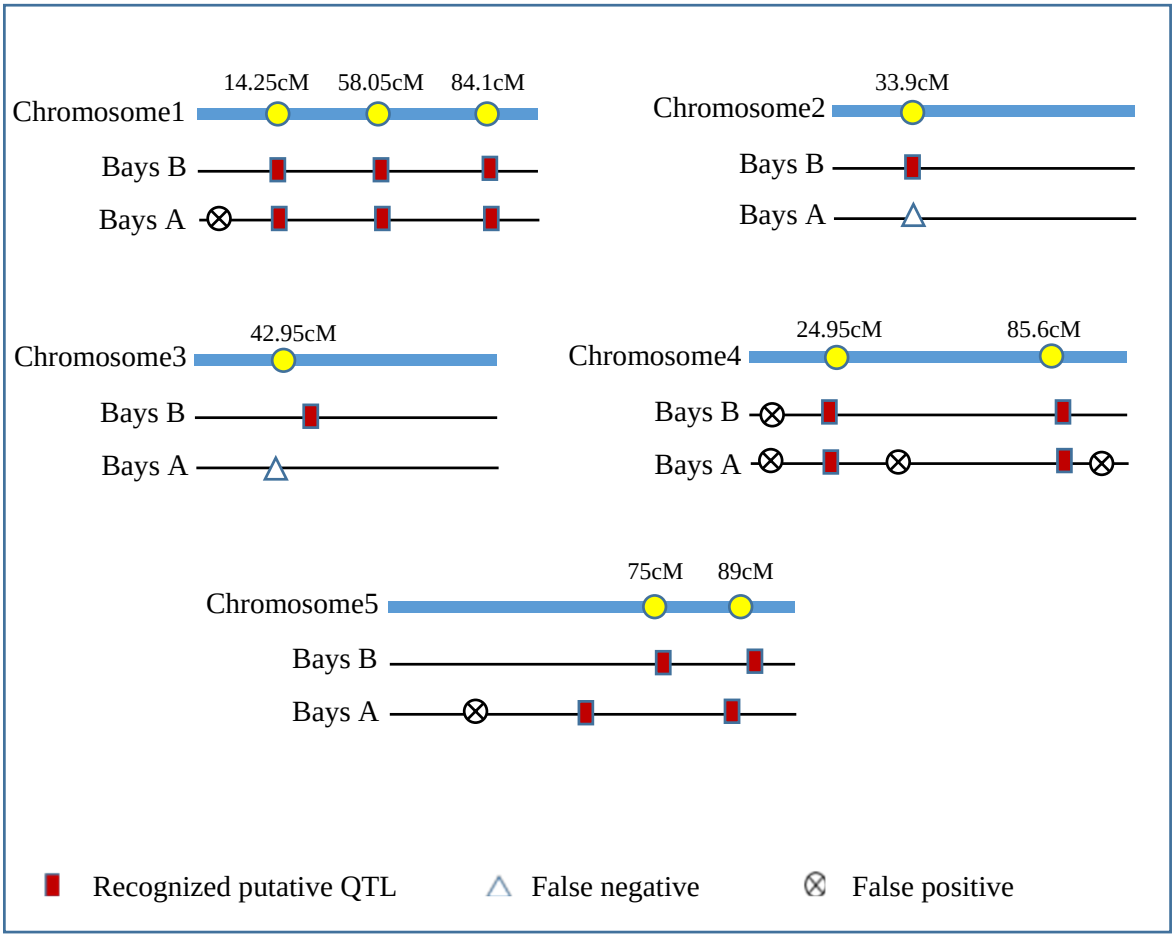

Fig. 2. Quantitative trait locus (QTL) identified by different models in comparison with the original location simulated on chromosomes 1 to 5.

rate breeding values than other methods.

Reference

George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881-889.

Habier, D., Fernando, R.L., Kizilkaya, K. and Garrick, D.J., 2011. Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186.

Hill, W.G., Goddard, M.E., Visscher, P.M., 2008. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4, e1000008.

Janss, L., 2011. BAYZ Manual version 2.02. Janss Biostatistics, Leiden, Netherlands.

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

Sargolzaei, M. and Schenkel, F.S., 2009. QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25: 680-681. First published January 28, doi:10.1093 /bioinformatics/btp045.

Wang, C.L., Ma, P.P., Zhang, Z., Ding, X.D., Liu, J.F., Fu, W.X., Weng, Z.Q. and Zhang, Q., 2012. Comparison of five methods for genomic breeding value estimation for the common dataset of the 15th QTL-MAS Workshop, *BMC Proceedings* 6(Suppl 2), S13.

Yi, N. and Xu, S., 2008. Bayesian LASSO for quantitative trait loci mapping. *Genetics* 179, 1045-1055.

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.J. and Qin, Z., 2010. Best linear unbiased prediction of genomic breeding values using a trait-specific marker derived relationship matrix. *PLoS ONE* 5(9), e12648.

Communicating editor: *Ali K. Esmailizadeh*

مقایسه دو روش مبتنی بر استنباط بیزی برای مکان یابی QTL با استفاده از نشانگرهای تک نوکلئوتیدی متراکم

غ. داشاب

نویسنده مسئول، پست الکترونیک: dashab@uoz.ac.ir

چکیده جهت مقایسه روش‌های مختلف مکان‌یابی QTL، یک مجموعه داده‌های ژنوتیپی و فنوتیپی شبیه سازی گردید. در مدل‌های بیز (BAYZ)، می‌توان علاوه بر این که از تمام اطلاعات نشانگرها استفاده نمود، ترکیبی از توزیعات مختلف برای اثرات نشانگرهای SNP در نظر گرفت. در مدل‌های فوق فرض می‌گردد که اکثر نشانگرها دارای اثرات کوچک (در مطالعه حاضر ۹۸درصد) و تنها تعداد محدودی از نشانگرها (۲درصد) دارای اثرات بزرگ هستند. جمعیت شبیه سازی شده شامل یک جمعیت پایه ۱۰۲۰ راس دام غیرخویشاوند بود که برای ایجاد عدم تعادل پیوستگی در بین جایگاه QTL با نشانگرها ۴ نسل تلاقی تصادفی ادامه پیدا کرد. در تمام نسل‌ها ۲۰ راس دام نر با ۱۰۰۰ راس گاو ماده آمیزش داده شد و هر فرد فقط یک نتاج تولید نمود. کل شجره شامل ۴۱۰۰ راس دام بود. تعداد ۱۰۰۰۰ SNP بر روی ۵ کروموزوم در فواصل مساوی(۰/۰۵cM) در کل جمعیت تعیین ژنوتیپ شدند. طول هر کروموزوم ۱۰۰ سانتی مورگان در نظر گرفته شد. صفت شبیه سازی شده مربوط به عملکرد تولید شیر بود. نتاج حاصل از نسل اول تا سوم دارای رکورد و جمعیت پایه و افراد نسل چهارم بدون رکورد بودند، لذا از کل جمعیت ۴۱۰۰ راسی، ۳۰۰۰ تا دارای رکورد بودند. از دو مدل مختلف Bayz A و Bayz B برای آنالیز پیوستگی QTL با نشانگرهای SNP استفاده گردید. آنالیز با نرم افزار BAYZ انجام گرفت. SNP های با اثرات بالاتر از ۰/۶ یا فاکتور بیز (BF) بالاتر از ۵/۵ به عنوان QTL گزارش گردید. نتایج آنالیز دو مدل بیز A و بیز B به ترتیب ۷ و ۹ مکان QTL بر روی ۵ کروموزوم بود. موقعیت QTLهای شناسایی شده با روش بیز B منطبق بر مکان شبیه سازی شده قرار داشتند، هر چند یک False positive بر روی کروموزوم ۴ نشان داد. مکان‌های QTL شناسایی شده با روش بیز A هر چند نزدیک مکان های شبیه سازی شده قرار داشتند، اما با تعداد زیادی نقاط False positive همراه بودند.