Application of Various Models for the Genomic Evaluation of Bovine Tuberculosis in Dairy Cattle

R. Mrode[1,4], G. Banos[1,2] , M. Winters[3] and M. Coffey[1]

[1]Scotlands Rural College, Easter Bush, Midlothian, EH25 9RG UK; [2]Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG UK ; [3]AHDB Dairy Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, UK [4]International Livestock Research Institute, Nairobi, 00100 Kenya

Abstract

Using de-regressed breeding values (DRPs) from the routine genetic evaluations for resistance to bovine tuberculosis (bTB), genomic evaluation was undertaken model with a reference and validation populations of 1,700 and 537 bulls, respectively, genotyped with the Illumina 50Kchip. The validation set of 537 bulls (VAL1) resulted from using a cut off birth year of 2007. In an attempt to equate infection rate in the validation data set to that in the reference, two additional validation data sets were created based on sample of the first 30 bulls with reliability >=89 in the reference set plus all bulls in VAL1 with the same level of reliability (VAL2) and a third validation set (VAL3) made up of a random sample of the first 30 bulls with reliability >=93 in the reference set plus all bulls in VAL1 with the same level of reliability. The models used for the analyses included SNP-BLUP and BayesCpi and a single-step (ssGBLUP) which was based on phenotypic observations. Different levels of polygenic effects were investigated and their impact on SNP effects for SNPs with different allele frequency. The accuracy of evaluations from the SNP-BLUP based on the correlation between genomic breeding values in the validation set and individual daughter deviations (IDD) for bulls was 0.20 with no polygenic effects in the model. A similar estimate was from BayesCpi. However, the estimates of accuracies increased with increasing levels of polygenic effects with values of 0.24 at 30% polygenic effects for SNP-BLUP. However, the estimates of accuracy from ssGBLUP were much higher at 0.48 or 0.54 at 0% or 30% polygenic effect. The use of VAL2 and Val3 generally increased the accuracy of genomic prediction for SNP-BLUP and BayesCpi but had very little impact in ssGBLUP. Fitting a polygenic effect in the model does not have a uniform impact on the estimates of SNP effects but its influence is dependent on the allele frequency of the SNP.

Key Words: bovine tuberculosis, SNP-BLUP, BayesCpi, ssGBLUP, polygenic effect

Introduction

Bovine tuberculosis (bTB) is a chronic bacterial disease of cattle caused *by Mycobacterium bovis* (*M. bovis*) infection primarily involving the respiratory tract. It is endemic in the UK and other countries, and presents a significant challenge to the UK cattle sector. The Department for Environment, Food and Rural Affairs (DEFRA) lists bTB as one of the four most important livestock diseases globally, incurring annual costs of about £175 million in the UK (Abernethy *et al.,* 2013). Routine genetic evaluation for resistance to bovine tuberculosis (bTB) has been implemented in the UK since January 2016. Infected animals were classified as those with a positive skin test or negative skin test but with a positive post–mortem examination result. It is trait with a low heritability of about 0.09 and from the breakdown model (Banos *et al.,* 2016) officially implemented, only about 22% of the 19315 sires represented in the data had a reliability of at least 50%. However, genotypic data is available of the some of the sires in the bTB conventional evaluations and incorporation of this information might result in an increase in the reliability of both proven and young bulls. This study therefore examines the application of genomic models for the incorporation of genotypic data to assess their impact on the accuracy of evaluations. In addition the impact of different levels of polygenic effects in the model on SNPs of different frequencies is also examined.

Materials and Method

De-regressed sire proofs were computed from evaluations of the UK official July 2016 run for bTB for 2232 bulls with at least 10 daughters, a reliability of 40% and have genotypes. The SNPs equivalent to the 50K Illumina chip were extracted for bulls which were genotyped with High density chip while those genotyped with low density chips were imputed to the 50K chip. Genomic evaluations were undertaken using these bulls and a SNP-BLUP model and BayesCpi. The bulls genotyped were born from 1990 to 2011. 1695 bulls born before 2007 were used as the reference population to estimate the SNP effects while 537 bulls born on 2007 and afterwards were included in the validation set (VAL1). However, most of the bulls in the validation set where with reliabilities of 65% or less compared with most reference bulls with reliabilities of 80% or more. This reflects the fact that the exposure time of the progeny of the younger bulls in the validation is much lower compared to those in the reference set; therefore the rate of infection between both sets is different. This makes validation on younger bulls with limited reliability and limited exposure to the disease more difficult. This implies that the validation data set based on cut off year may not be optimum in computing the accuracy of prediction in this situation. Therefore the use two additional validation sets were also investigated. The second validation set (VAL2) was created by a random sample of the first 30 bulls with reliability >=89 in the reference set plus all validation bulls with the same level of reliability; this resulted in 1888 and 344 bulls in reference and validation respectively. Finally a third validation set (VAL3) was created by a random sample of the first 30 bulls with reliability >=93 in the reference set plus all validation bulls with the same level of reliability; this resulted in 2014 and 214 bulls in reference and validation respectively.

A total of 43143 SNPs were selected for genomic evaluations after the usual SNP quality edits. The statistical model used for the estimation of SNP effects is:

$$y_i = \mu + v_i + \sum_{j=1}^{m} z_{ij} u_j + e_i$$

where y_i is the de-regressed proof of a bull, μ is the overall mean, v_i is the residual polygenic effect (10% of additive genetic variance) of i^{th} bull, z is the genotype value coded as 0 and 2 for the homozygotes and 1 for the heterozygote, u_i is the random regression coefficient for j^{th} SNP_j and e_i is the residual effect. Analyses were also carried out assuming four levels of polygenic effects (0%, 10%, 20%, and 30%) and results were compared. The same model was fitted BayesCpi with but with no polygenic effect included.

In addition, ssGBLUP analysis was also undertaken with 934987 cows with phenotypic records for bTB fitting the model of Banos *et al.* 2016. Briefly, the model fitted was:

$$Y_{ijkmn} = \mu + B_i + R_j \cdot M_k + L_m + b_1 dur + b_2 age + b_3 phol + A_n + e_{ijkmn} \quad (1)$$

where Y= bTB infection status record of animal n in breakdown i (0/1); μ= population mean, B = fixed effect of the breakdown I; R·M = fixed effect of the interaction between calendar year j and month k of breakdown onset; L = fixed effect of lactation number m (m=1 for primiparous cows, 2 for multiparous cows); dur = linear regression on duration of the breakdown (b_1=regression coefficient); age = linear regression on age of animal at breakdown onset (b_2=regression coefficient); phol = linear regression on percentage of Holstein genes of the animal (b_3=regression coefficient); A = random additive genetic effect of animal n including pedigree ; e = random residual.

A total of 5435 sires of these cows had genotypes, therefore **a G** matrix was computed for these sires from the genotypes using VanRaden (2008) method one. The $\mathbf{G}_{22}$ matrix was then computed as $\mathbf{G}_{22} = (1-w)\,\mathbf{G} + w\mathbf{A}_{22}$, with w set at 4 levels of 0, 10, 20 and 30%. The $\mathbf{H}^{-1}$ was then computed for cows and bulls incorporating the $\mathbf{G}_{22}$ for the genotype animals. The same set of 537 validation bulls were used and the 47616 observations for their daughters

were set missing. The accuracy of evaluations were computed as the correlation between the GEBVs of the validation bulls and the mean of the bull individual daughter deviations (IDD-; yield of daughters corrected for all effects include half of dam breeding value) or de-regressed proofs.

Usually, the inclusion of polygenic effects is mainly to account for the fact that SNPs may not account for all genetic variance of the traits. However, it is not clear whether varying levels of polygenic effect will have proportionate effects on all SNPs independent of their allele frequencies. In order to investigate the effect of differing levels of polygenic effect on SNP solutions of varying allele frequencies, SNPs were classified into 5 levels based on the allele frequencies. These were 0.05-0.10, 0.11-0.40, 0.41-0.70, 0.71-0.90 and > 0.90. The mean SNP solutions from models with different levels of polygenic effects were computed and compared.

The software MiX99 (Lidauer *et al.,* 2014) was used for the de-regression, SNP-BLUP and the ssGBLUP00 analyses. The BayesCpi software used was developed in-house as described in Mrode (2014).

Results and Discussions

The accuracies of genomic evaluation estimated from the different models are presented in Figure 1. The accuracies from the SNP-BLUP model when no polygenic effects was fitted was about 0.20 and was very similar to the value of 0.22 obtained from BayesCpi. The estimate of π obtained from the analysis was 60%, indicating a high proportion of SNPs seems to have little or no effect on bTB. A subsequent analysis using BayesCpi with the value of π fixed at 30% resulted in a similar estimate of accuracy and a correlation of about unity between SNP solutions from both Bayesian models. However, as the level of polygenic effects increases the accuracy of SNP-BLUP increased from 0.20 to 0.24 at 30% polygenic effect. Compared with results from the ssGBLUP, estimates from the latter were about twice from those from the SNP-BLUP ranging from 0.48 to 0.54 as the proportion of $\mathbf{A}_{22}$ incorporated in $\mathbf{G}_{22}$ increases. These accuracies were obtained using the mean of IDDs of bulls as the validation variable for the validation bulls. However, these estimates of accuracies varied from 0.56 to 0.62 if DRPs were used for the validation. The increase in accuracy with the ssGBLUP could be attributed to the fact that the number bulls with genotypes was about 3 times that included in the SNP-BLUP model or Bayesian methods. The requirement for bulls with at least 10 daughters and a reliability of 40% in the SNP-BLUP meant only 1695 could be utilized for the analysis given the low reliability of the trait. In addition, the SS method incorporated all phenotype and pedigree information.

The predictive ability of the models based on the regression estimates (Figure 2) indicates that with no polygenic effect, BayesCpi resulted in a similar estimate as ssGBLUP but higher than the estimate from SNP-BLUP. Similar to the estimates of accuracies, the regression estimates increased with increase in the level of polygenic effects varying from 0.46 to 0.70 at 30% polygenic effect for the SNP-BLUP models. While the estimates of regressions from the ssGBLUP were generally higher than those from SNP-BLUP, however this difference decreased as the level of polygenic effects increases with estimates of 0.70 (SNP-BLUP) and 0.73 (ssGBLUP) at 30% polygenic level.

The impact of using validation sets based on random selecting of bulls (VAL2 and VAL3) rather than cut off based on birthdate is shown in Table 1. For SNP-BLUP, the accuracy of genomic prediction increased from 0.24 (VAL1) to 0.34 (VAL2) and 0.42) (VAL3) with the random inclusion of bulls with high reliability in the validation sets. This was also accompanied with higher predictive ability of the model for VAL3 but a slight decrease in VAL2. Slightly higher increases were observed for BayesCpi in VAL2 and VAL3 both in terms of accuracy and predictive ability. However, in the case of ssGBLUP, this resulted in lower predictive ability of the model in both VAL2 & VAL3 compared with VAL1 and accuracy of prediction only increased in VAL3 relative to VAL1.

The above results indicate that as the percentage of polygenic effects increases, the accuracy increased correspondingly for all methods used in the analysis. The impact of the increasing levels of polygenic effects on SNPs of different allele frequencies is shown in Figure 3. For SNPs with allele frequencies of 0.05 to 0.10, the mean SNP effects increased with increasing level of polygenic effect until 10%; thereafter it decreased. However for SNPs with frequencies of 0.11-0.40 or > 0.90; the mean SNP effects increased with increasing levels of polygenic effects. However, the opposite was true for SNPs with frequencies between 0.41-0.70 or 0.71-0.80. This implies that fitting a polygenic effect does not have a uniform impact on the estimates of SNP effects and its influence is dependent on the allele frequency of the SNP.

Conclusion

Given the data structure and size, ssGBLUP seems the most appropriate model to apply for the genomic prediction of bTB in this study as it uses all available information. However attempts to define validation data sets that capture similar rate of infection as in the reference sets resulted in more accurate genomic predictions for SNP-BLUP and BayesCpi. Fitting a polygenic effect in the model does not have a uniform impact on the estimates of SNP effects but its influence is dependent on the allele frequency of the SNP.

References

Abernethy, D.A., Upton, P., Higgins, I.M., McGrath, G., Goodchild, A.V., Rolfe, S.J., Broughan, J.M., Downs, S.H., Clifton-Hadley, R., Menzies, F.D., de la Rua-Domenech, R., Blissitt, M.J., Duignan, A. & More, S.J. 2013. Bovine tuberculosis trends in the UK and the Republic of Ireland, 1995–2010. *Veterinary Record 172:312.* doi:10.1136/vr.100969.

Banos, G., Winters, M., Mrode, R., Michell, A.P., Bishop, S.C., Woolliams, J.A. & Coffey, M.P. 2016. Genetic evaluation for bovine tuberculosis resitance in dairy cattle. *Journal daoty science,* (accepted)

Lidauer, M., Matilainen, K., Mäntysaari, E.A., Pitkänen, T., Taskinen, M. & Strandén, I. 2014. *Technical Reference Guide for MiX99 Solver.* Release VII/2014. MTT Agrifood Research Finland.

Mrode, R.A. 2014. *Linear models for the prediction of animal breeding values.* Third edition, CAB International, Wallingford, Oxon, UK. 360 pp.

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

Table 1. The accuracies of genetic prediction for bTb using validation data set 2 (Val2) and set 3 (VAL3).

Validation Set	SNP-BLUP (30% polygenic)		BayesCpi		ssGBLUP	
	r	b	R	b	r	B
VAL2	0.32	0.65	0.34	0.70	0.51	0.53
VAL3	0.41	0.77	0.42	0.80	0.56	0.54

r = correlations and b = regressions

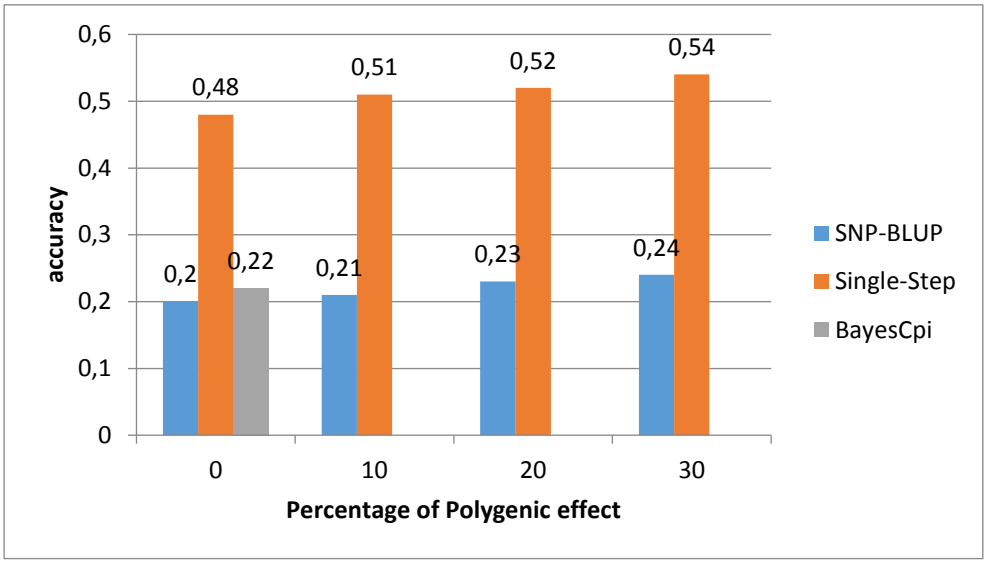

Figure 1. The accuracy of genomic evaluations as correlations between DGVs and mean of IDD of validation bulls.

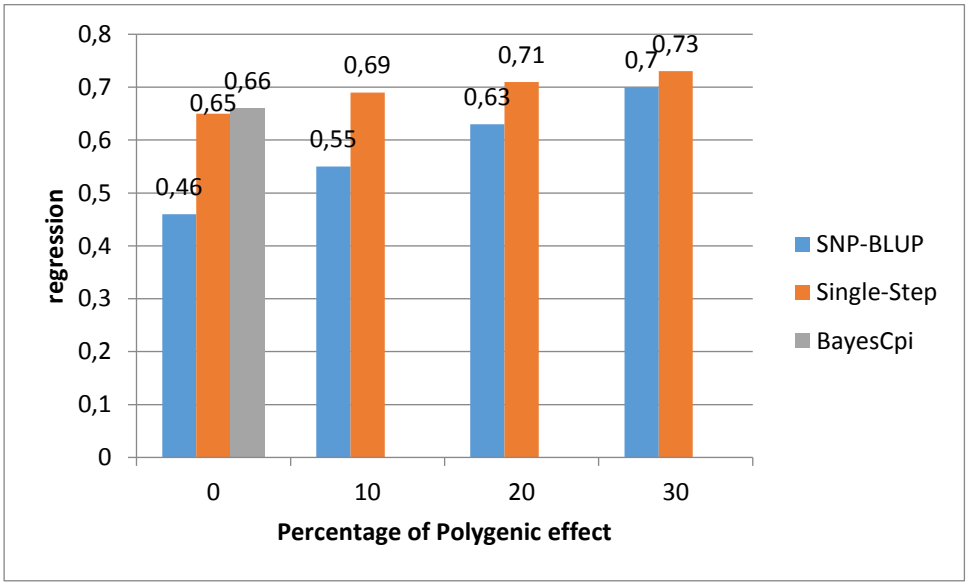

Figure 2. Regression coefficients of DGVs on mean of IDD of validation bulls.

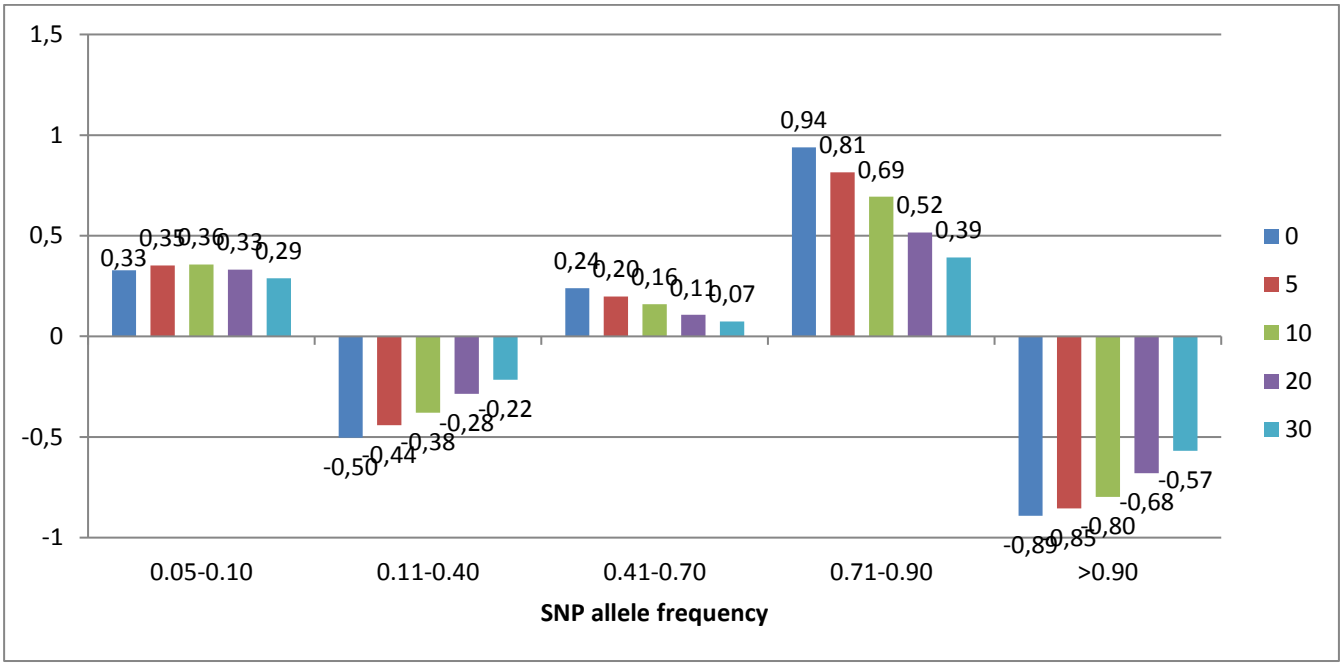

Figure 3. Mean SNP effects at various levels of polygenic effects ion SNPs of different allele frequency.