

## A meta-analysis for bovine tuberculosis resistance in dairy cattle

S. Tsairidou<sup>1</sup>, J.A. Woolliams<sup>1</sup>, A.R. Allen<sup>2</sup>, R.A. Skuce<sup>2,3</sup>, S.H. McBride<sup>2</sup>, R. Pong-Wong<sup>1</sup>, O. Matika<sup>1</sup>, E.K. Finlay<sup>4</sup>, D.P. Berry<sup>5</sup>, D.G. Bradley<sup>4</sup>, S.W.J. McDowell<sup>2</sup>, E.J. Glass<sup>1</sup>, S.C. Bishop<sup>1</sup>

<sup>1</sup>The Roslin Institute & RDVS, University of Edinburgh, Midlothian, UK, <sup>2</sup>Agri-Food and Biosciences Institute, Belfast, UK, <sup>3</sup>School of Biological Sciences, Queen's University of Belfast, Belfast, UK, <sup>4</sup>Department of Genetics, Trinity College Dublin, Dublin, Ireland, <sup>5</sup>Teagasc, Moorepark, Fermoy, County Cork, Ireland.

**ABSTRACT:** Bovine tuberculosis (bTB) has been among the most persistent diseases in cattle. Genetic selection of disease resistant individuals may be a complementary approach to assist with disease control. The aim of this study was to conduct a meta-analysis on two dairy cattle populations with bTB phenotypes and SNP chip genotypes, identifying genomic regions underlying bTB resistance and testing genomic predictions by means of cross-validation. We identified a region on chromosome 6 likely to be associated with bTB resistance and confirmed that this chromosome as a whole contributes a major proportion of the observed variation in dataset. Genomic prediction for bTB was shown to be feasible even when different populations are combined, with the chromosome heritability results suggesting that the accuracy arises from the SNPs capturing linkage disequilibrium between markers and QTL as well as additive relationships between animals. Further studies on larger populations will be needed to confirm findings.

**Keywords:** Regional Heritability; Cross Validation; Genomic Selection; Disease Resistance; Bovine Tuberculosis

### Introduction

In many countries bovine tuberculosis (bTB) is still persistent in livestock despite on-going eradication programmes. bTB control has relied on diagnosis through the tuberculin skin test and abattoir carcass inspection (de la Rua-Domenech et al. (2006)). The limitations, however, in sensitivity of these methods, along with our incomplete understanding of bTB transmission, impede successful eradication. Genetic selection of individuals resistant to bTB may offer a complementary strategy for the control of tuberculosis in cattle.

Previous studies have shown genetic variation underlying resistance to bTB (Brotherstone et al. (2010); Bermingham et al. (2012, 2014); Finlay et al. (2012)). Our hypothesis is that a meta-analysis combining populations may reveal new information concerning the genetic architecture of bTB resistance, by means of simultaneous analysis of individuals distantly related, assisted by a larger sample size. Specifically, we anticipate that meta-analyses will provide additional information on specific loci affecting resistance and it will enable enhanced genomic predictions of resistance.

Regional heritability (RH) mapping (Nagamine et al. (2012)) is a flexible means of identifying genomic regions affecting complex traits, particularly when individual SNPs contribute only a small proportion of genetic variation, but groups of SNPs may collectively be

significantly associated with the trait. RH mapping can also be an effective means of combining disparate datasets (Riggio et al. (2014a)), avoiding the need to assume the same linkage phase between markers and mutations across populations. Combining datasets in this way also allows testing of genomic predictions within and across populations, using cross-validation (CV) methods (Luan et al. 2009). Further, chromosome-level heritabilities also give insight into properties of the genomic heritability (Riggio et al. (2014b)). The aim of this study was to investigate the genomic control of bTB resistance and to explore the feasibility of genomic selection for bTB resistance. This was done by combining two independent bTB resistance datasets and analysing this joint dataset.

### Materials and Methods

**Animals and phenotypes.** Two datasets were used: dataset 1 comprised 1151 female Holstein-Friesians originating from commercial herds in Northern Ireland, constitutes of confirmed cases of bTB or controls (Bermingham et al. (2014)); dataset 2 comprised 287 Holstein bulls from the Republic of Ireland with estimated breeding values (EBVs) calculated from their daughter phenotypes (Finlay et al. (2012)). To analyse the two datasets together, an initial fixed effects model was used for the first dataset, pre-correcting for all non-genetic fixed effects and the residuals of this model were used as phenotypes in the subsequent analysis. For dataset 2, the de-regressed EBVs (i.e. EBVs divided by their reliability) were used as phenotypes, either with pedigree information included in their calculation or without. All phenotypes were standardised by their standard deviation.

**Genotypes.** The first dataset was genotyped using the Illumina high-density Bead Chip, while the second was genotyped with the Illumina Bovine50 SNP chip. A pooled dataset was constructed with the SNPs present in both datasets (i.e. 36690 SNPs) after applying quality control (MAF>0.05, call rate>95%, HWE  $p < 0.000001$ , all homozygotes, all heterozygotes, or all missing were removed).

**Regional heritability mapping.** RH mapping methodology was applied as described by Nagamine et al. (2012). The population of origin was fitted as a fixed effect whereas genomic and local IBS kinship ( $\mathbf{G}$ ,  $\mathbf{G}_L$ ) matrices (Leutenegger et al (2003)) calculated for each window were fitted as random effects i.e.  $y_i = m + \beta_i + u_i + r_i + e_i$ , where  $y_i$  is the adjusted phenotype of individual  $i$ ,  $\beta_i$  indicates the dataset from which  $i$  belong to,  $u_i$  its additive genomic effects ( $\mathbf{u} \sim \text{MVN}(0, \sigma_u^2 \mathbf{G})$ ), and  $r_i$  its additive regional effects ( $\mathbf{r} \sim \text{MVN}(0, \sigma_r^2 \mathbf{G}_L)$ ). The combined dataset

was analysed with three window sizes: a 50-SNP window size (25-SNP step), 30-SNP (15-SNP step), and 20-SNP (10-SNP step). Overall heritability was obtained in a separate ASReml analysis assuming no QTL affecting the trait (null hypothesis). The RH was calculated for every window as  $h^2_r = \sigma^2_r / \sigma^2_p$ , where  $\sigma^2_r$  was the variance explained by the window and  $\sigma^2_p$  was the phenotypic variance. A log-likelihood ratio test (LRT) was calculated for every window. The region with the maximum RH estimate and LRT ( $RH_{\max}$  and  $LRT_{\max}$ ) was identified for each chromosome across all chromosomes, and suggestive and genome-wide significance thresholds were obtained after Bonferroni correction. Analyses were repeated for EBVs calculated with and without pedigree information for dataset 2, and results compared.

**Chromosomal heritability estimation.** The heritability for each chromosome was calculated using four approaches: a) using  $\mathbf{G}$  matrices calculated separately for each chromosome ( $h_{c(sep)}^2$ ), i.e.  $y_i = \beta_i + g_c + e$ , where  $g_c$  is the vector of genetic effects of chromosome  $i$ , b) fitting separately each chromosome plus the whole  $\mathbf{G}$  matrix, c) fitting all variance components (29 chromosomes) as random effects simultaneously ( $h_c^2$ ), i.e.  $y = \beta + \sum g_c + e$ , and d) fitting simultaneously the chromosomes with non-zero variance in a), b) and c). Comparing results from a), b) and c) gives insight into how the  $\mathbf{G}$  matrix contributes to the heritability, i.e. through genome-wide additive genetic relationships or specific markers tagging mutations with large effects on the trait. If (i)  $h_{c(sep)}^2$  (a) and (ii)  $h_{c(sep)}^2 - h_c^2$  (c) are regressed on chromosome length ( $L_c$ ), then the proportion of genetic variance due to population structure can be calculated as  $b_0/b_{0(sep)}$  where  $b_0$  and  $b_{0(sep)}$  are the intercepts of the two regressions (Yang et al. (2011)).

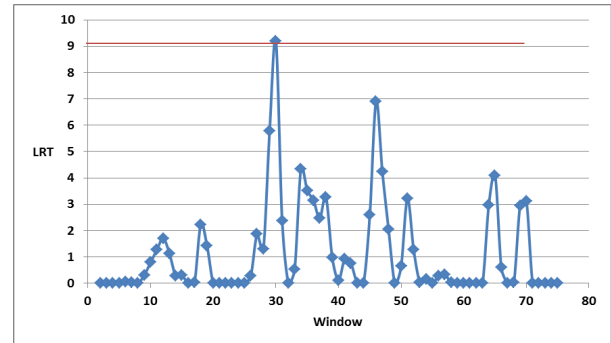
**Genomic prediction cross-validation.** The predictive ability of the genomic EBVs (DGVs) was tested through CV methodology. Two approaches were followed: (a) individuals from both populations were combined in one dataset and then randomly assigned to five groups (5-fold CV), each time using one group as the test set and the remaining four groups as training sets using the model  $y_i = m + u_i + e_i$ , where  $u_i$  is the genomic estimated breeding value with  $\mathbf{u} \sim MVN(0, \sigma_u^2 \mathbf{G})$ . This procedure was repeated 50 times, each time with a different randomisation of the individuals in the groups, and DGVs both with and without familial information (dataset 2) were tested; (b) CV was conducted across populations using the smaller population (dataset 2) as the test set with pedigree derived EBVs. The average accuracy across 50 randomisations was calculated in (a) and the expected accuracy in (b), as  $E[r(\hat{g}, \hat{g})] \approx r(\hat{y}, \hat{y})/h$ , where  $r(\hat{y}, \hat{y})$  is the correlation between the cross-validated predicted DGVs ( $\hat{y}$ ) and the adjusted phenotypes and  $h$  is the square root of the heritability. The standard error of the accuracies in (a) was calculated as the empirical standard deviation of the 50 accuracy estimates.

## Results and Discussion

**Regional heritability estimates.** The genomic heritability when the EBVs without pedigree information (for dataset 2) were used was 0.14 (0.05), and for the

EBVs estimated using pedigree was 0.11 (0.04). These values are indicative of genetic variation but care should be taken in their interpretation due to the different trait definitions. Windows contributing the maximum heritability and with the maximum LRT test value were identified for each chromosome, for the three window sizes. For the EBVs without pedigree information (dataset 2),  $RH_{\max}$  ranged from 0.005 to 0.032 for the 20-SNP window, from 0.004 to 0.029 for the 30-SNP window, and from 0.006 to 0.028 for the 50-SNP window. For the dataset 2 using pedigree-based EBVs  $RH_{\max}$  ranged from 0.172 to 0.409, from 0.009 to 0.359, and from 0.007 to 0.353 for the 20, 30 and 50-SNP windows. However, as demonstrated by Ekine et al. (2013), inclusion of pedigree information is likely to inflate the estimated effects.

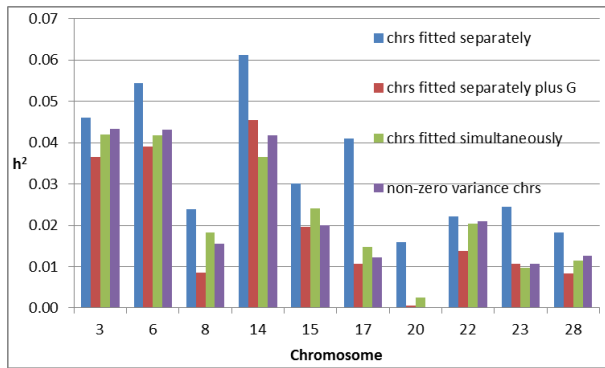
The strongest evidence for association was on chromosome 6 (BTA6) for the 50-SNP window, significant at the suggestive significance threshold for the dataset comprising EBVs without pedigree information ( $LRT_{\max} = 9.19$ , suggestive threshold=9.02), with the window explaining 2.7% ( $h^2_r = 0.027$ ) of the phenotypic variance (Fig. 1 and 2). This result was not observed by either Bermingham et al. (2014) or Finlay et al. (2012). With EBVs calculated using pedigree, BTA6 gave an  $LRT_{\max} = 7.16$  for the 50-SNP window and  $h^2_r = 0.238$ , with this value again likely to be an overestimate.



\*The red line represents the suggestive significance threshold.

**Figure 1. Log-likelihood ratio test for chromosome 6 from the Regional heritability analysis ( 50-SNP window).**

Comparing RH to a single-SNP approach, a GWAS was conducted fitting two principal components as fixed effects along with dataset of origin to address possible stratification (1438 individuals and 36461 markers passed quality control, MAF>0.05, call rate>0.95). A pattern on BTA6 was still visible, but no locus reached significance. This may indicate the increased power of RH mapping to detect genomic regions associated with the trait, compared to single-SNP GWAS, however SNP-mutation linkage phases may also differ between populations.



**Figure 2. Chromosomal heritability estimates for non-zero variance chromosomes.**

### Chromosomal heritability estimates.

Chromosomal heritability estimates confirmed findings from RH analyses for BTA6 with all four methods in good agreement (Fig. 2). Heritability estimates for BTA6 were 0.054, 0.039, 0.042 and 0.043 when every chromosome was fitted one by one, when the genomic matrix was also included in the model, when all 29 chromosomes were fitted simultaneously, and when the chromosomes with non-zero variance were fitted, respectively. From approaches (a) and (b) it is possible to identify the chromosomes contributing most of the observed variation, although heritabilities when fitting only one chromosome (a) were always overestimated. For highly polygenic traits the proportion of variance explained by each chromosome is expected to be proportional to its length (Yang et al. (2011)); here the captured genetic variation was weakly related to chromosome length ( $b_{0(sep)}=0.0149$ ,  $b_{1(sep)}=0.00004$ ,  $R^2 = 0.0053$ ,  $p$ -value: 0.7), and 9 chromosomes explained 22% of the phenotypic variance. An explanation would be that bTB resistance is a complex trait with clusters of causal variants spread over several chromosomal regions collectively controlling the trait. Regressing the difference between the heritability when the chromosomes were fitted individually (a) and when were fitted simultaneously (c) on the chromosomal length was not significantly different than zero ( $b_0 = 0.0126$ ,  $b_1 = -0.00003$ ,  $R^2 = 0.0166$ ,  $p$ -value: 0.5). From the intercepts  $b_0$  and  $b_{0(sep)}$  the proportion of genetic variance due to relatedness was estimated to be 0.85. These results suggest that although the markers capture loci effects through linkage disequilibrium (LD), additive pedigree-correlated relationships are likely to play an important role.

**Genomic prediction cross-validation.** Increasing the size of the training set is expected to be beneficial in a cross validation. Therefore, we combined the two populations in a 5-fold cross validation, and compared predictions within and across populations. The average prediction accuracies obtained on the combined populations across 50 iterations were  $r(g,\hat{g})=0.33$  (s.e. 0.007) when the EBVs without pedigrees were used, and  $r(g,\hat{g})=0.38$  (s.e. 0.007) for the pedigree based EBVs. Consistent with results mentioned above, including family information in the EBVs resulted in inflated estimates. Cross validation across populations, and despite using

EBVs with pedigree information, resulted in reduced accuracy ( $r(g,\hat{g})=0.1$ ). These results suggest that genomic prediction is feasible but less accurate when applied across disparate populations since the accuracy depends on the genetic relationship of the test set to the training set. This is due to systematic differences in allele frequencies and linkage phases across the populations. Prediction accuracy was improved when individuals from the population to be predicted were included in the training set (a).

### Conclusion

Following the regional heritability and chromosomal heritability approaches we have identified a region on BTA6, suggesting a putative association with the trait of bTB resistance. Many genes reside in the genomic region identified that could be suggested as possible candidates for further studies. Combining different populations is a challenging procedure, with the joint populations expected to behave differently than when analysed separately. The cross validation conducted in this meta-analysis has confirmed the potential feasibility of genomic selection for bTB resistance in cattle, although drawing robust inferences on across-population predictions would require more phenotyped and genotyped animals.

### Acknowledgements

We acknowledge Science Foundation Ireland 09/IN.1/B2642, access to the DAFF dataset, and financial support from the Roslin Institute, the Greek State Scholarship Foundation, and the BBSRC.

### Literature Cited

- Bermingham, M. L., Bishop, S. C., Woolliams, J. A. et al. (2014). *Heredity*.
- Bermingham, M. L., More, S. J., Good, M. et al. (2009). *J. Dairy Sci.* 92:3447-3456.
- Brotherstone, S., White, I. M., Coffey, M. et al. (2010) *J. Dairy Sci.* 93:1234-1242.
- De la Rua-Domenech, R., Goodchild, A. T., Vordermeier, H. M. et al. (2006). *Res. Vet. Sci.* 81:190-210.
- Ekine, C. C., Rowe, S. J., Bishop, S. C. et al (201). *G3: Genes|Genomes|Genetics* in Press.
- Finlay, E. K., Berry, D. P., Wickham, B. et al. (2012). *PLoS ONE* 7:e30545.
- Leutenegger A. L., Prum, B., Genin, E., et al. (2003). *Am. J. Hum. Genet.* 73:516-523.
- Luan, T., Woolliams, J. A., Lien, S. et al. (2009). *Genetics* 183:1119-1126.
- Nagamine, Y., Pong-Wong, R., Navarro, P., et al. (2012). *PLoS ONE* 7: e46501.
- Riggio, V. Pong-Wong, R., Sallé, G. et al. (2014a). *J. Anim. Breed. Genet.* doi: 10.1111/jbg.12071.
- Riggio, V., Abdel-Aziz, M., Matika, O. et al. (2014b). *Animal* 8:520-528.
- Yang, J., Manolio, T. A., Pasquale, L. R., et al. (2011). *Nat. Genet.* 43:519-525.