An across-breed genome wide association analysis of susceptibility to paratuberculosis in dairy cattle

Ahmed M Sallam[1,2,3], Yalda Zare[1]†, Fazli Alpay[1]‡, George E Shook[4], Michael T Collins[5], Samir Alsheikh[2], Mahmoud Sharaby[3] and Brian W Kirkpatrick[1,4]*

[1] Department of Animal Sciences, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA
[2] Animal and Poultry Production Division, Desert Research Center, Al-Mataria 11753, Cairo, Egypt
[3] Animal and Fish Production Department, Faculty of Agriculture, University of Alexandria, El-Shatby, Alexandria, Egypt
[4] Department of Dairy Science, College of Agricultural and Life Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA
[5] Department of Pathobiological Science, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI 53706, USA

Paratuberculosis is a chronic disease of ruminants caused by *Mycobacterium avium* subspecies *paratuberculosis* (*MAP*). It occurs worldwide and causes a significant loss in the animal production industry. There is no cure for *MAP* infection and vaccination is problematic. Identification of genetics of susceptibility could be a useful adjunct for programs that focus on management, testing and culling of diseased animals. A case-control, genome-wide association study (GWAS) was conducted using Holstein and Jersey cattle in a combined analysis in order to identify markers and chromosomal regions associated with susceptibility to *MAP* infection across-breed. A mixed-model method (GRAMMAR-GC) implemented in the GenABEL R package and a Bayes C analysis implemented in GenSel software were used as alternative approaches to conduct GWAS analysis focused on single SNPs and chromosomal segments, respectively. After conducting quality control, 22 406 SNPs from 2157 individuals were available for the GRAMMAR-GC (Bayes C) analysis and 45 640 SNPs from 2199 individuals were available for the Bayes C analysis. One SNP located on BTA27 (8·6 Mb) was identified as moderately associated ($P < 5 \times 10^{-5}$, FDR = 0·44) in the GRAMMAR-GC analysis of the combined breed data. Nine 1 Mb windows located on BTA 2, 3 (3 windows), 6, 8, 25, 27 and 29 each explained ≥1% of the total proportion of genetic variance in the Bayes C analysis. In an analysis ignoring differences in linkage phase, two moderately significantly associated SNPs were identified; ARS-BFGL-NGS-19381 on BTA23 (32 Mb) and Hapmap40994-BTA-46361 on BTA19 (61 Mb). New common genomic regions and candidate genes have been identified from the across-breed analysis that might be involved in the immune response and susceptibility to *MAP* infection.

Keywords: Gene, Johne's disease, cow.

Paratuberculosis, also known as Johne's disease (JD) is a chronic infection of the small intestine of ruminants with *Mycobacterium avium* subsp. *paratuberculosis* (*MAP*). According to the National Animal Health Monitoring System (NAHMS), 68·1% of dairy herds in the US are infected with JD (68% of US dairy herds examined had one or more animals infected by *MAP*). This study was based on the prevalence of viable *MAP* in the environmental faecal samples from herds in the top 17 dairy states (NAHMS, 2007). Economic costs of JD were estimated to be about US$200 to $250 million annually for the U.S. dairy industry alone (Ott et al. 1999). Paratuberculosis is responsible for an economic loss worldwide and is also a public health concern for its potential zoonotic association with Crohn's disease (CD) in humans (Bentley et al. 2008).

The clinical manifestations of JD are diarrhoea, weight loss, decreased milk production and eventually death. Cattle are thought to be typically infected as calves and clinical signs do not appear until second or third lactation (in dairy cattle), usually 4–5 years after *MAP* infection. During

*For correspondence; e-mail: bwkirkpa@wisc.edu
†Present address: ABS Global Inc., Deforest, WI 53532, USA.
‡Present address: Department of Animal Science, Faculty of Veterinary Medicine, Uludag University, Bursa, 16059, Turkey.

this period, infected animals may spread the disease in the herd. *MAP* can infect animals through contamination of colostrum, milk or faeces; the major route of *MAP* transmission is faecal-oral (Nielsen et al. 2008).

Johne's Disease as a complex disease is a multifactorial trait and is affected by both genetics and environment. Several studies have reported that the susceptibility to JD has a genetic basis with heritability estimates for susceptibility to *MAP* infection ranging from 0·03 to 0·28 (Kirkpatrick & Shook, 2011) and averaging approximately 10%. Identifying the genetic basis for variation in susceptibility to *MAP* infection and development of tools for genomic selection will be a useful adjunct to JD control programs.

In previous work, we have conducted GWAS studies independently in the Holstein and Jersey populations. A combined analysis of these two populations was conducted in an attempt to identify SNPs in linkage disequilibrium with loci and chromosomal regions contributing to susceptibility to *MAP* infection across breeds. Regarding single SNP analysis our hypothesis is that there are causative polymorphisms whose origin predates development of these two breeds and for which SNP alleles with common association could be identified in an across-breed analysis. Regarding the identification of common chromosomal regions, an association could be detected for the same reason or from independent causative polymorphisms in the two breeds, so long as differences in linkage phase do not offset the associations.

Material and methods

Phenotype definition

Holstein cattle. Animal resources and phenotypes have been described in detail in previous reports (Kirkpatrick et al. 2011; Alpay et al. 2014). Briefly, two resource populations of approximately 5000 cows each were used. For population 1, blood and faecal samples were obtained from cows in commercial dairy herds throughout the US. Cooperating herds were identified by the owners' responses to a survey that provided information indicative of the likelihood of the herd having JD. Cows were specifically chosen to be in the second or later lactation to increase the likelihood of identifying cows manifesting *MAP* infection. Samples were obtained from approximately 300 herds from across the US, with Wisconsin and California herds accounting for 48% of the total samples obtained. Our samples were collected in a cross sectional design, i.e. all samples were obtained from the herd at one time. All blood samples were tested in the Johne's Testing Center of the School of Veterinary Medicine at the University of Wisconsin Madison within 7 d by a serum ELISA (JTC-ELISA) (Shin et al. 2008) with 30% sensitivity and >99% specificity relative to faecal culture (the evidence of *MAP* infection in faeces), which is a bacterial culture of faecal samples that

detects viable *MAP* in faeces (Collins et al. 2006). For population 2, cows were sampled from six commercial Holstein herds in Wisconsin that were co-operators in a JD control project. Blood samples for disease testing and DNA extraction were obtained from all cows in these herds over a period of 15 months in 2006–2007.

Diagnosis of *MAP* infection was based on evidence of antibody titre to *MAP* using a serum ELISA test (Shin et al. 2008). Optical density values of the serum from project animals and from positive and negative controls were converted to sample to positive ratios (S/P). Based on the value of S/P ratio, results of ELISA tests were categorised as negative (0–0·09), suspect (0·10–0·24), low positive (0·25–0·39), positive (0·40–0·99), and strong positive (≥1·00) as previously suggested (Collins, 2002). Animals categorised as strong positive, positive or low-positive were all considered to be ELISA-positive. Case definition for population 1 was cows that were positive by either faecal culture (FC) or serum ELISA tests. Case definition for population 2 was based only on serum ELISA test result. Controls were cows testing negative to ELISA and FC (if available) and matched with cases on herd and proximal birth date.

Jersey cattle. Blood and faecal samples were obtained from ~5000 mature cows (minimum age of 20 months) from 30 commercial Jersey herds throughout the US. Nomination of herds for this study was based on the prevalence of JD as evidenced by the herd's owner or veterinarian. Serum ELISA and FC of *MAP* testing were performed as described above. Cases consisted of animals with positive results to either ELISA or FC, controls were cows testing negative to both ELISA and FC and matched with cases on herd and proximal birth date (Zare et al. 2014).

The cross-sectional nature of this experiment means that animals will be sampled at various stages relative to progression of Johne's disease. Some animals will have been sampled at a stage where a humoral response will have not yet been elicited despite infection. Other animals may be in advanced stages of the disease in which a humoral response or faecal shedding is evident. Our case definition is intended to reflect that animals defined as cases are liable to infection based on a positive test result, regardless of the test employed. Short of serial testing in a longitudinal study, the phenotypes are inadequate for further interpretation as evidence of tolerance or resistance.

Genotyping

Genotyping for this study has been described in detail previously (Alpay et al. 2014; Zare et al. 2014). Briefly 856 Holstein and 898 Jersey samples were genotyped with the Illumina Bovine SNP50 beadchip and an additional 263 and 182 samples, respectively were genotyped with the Illumina LD beadchip (7 K) and imputed to 50 K using

BEAGLE version 3.3 (Browning & Browning, 2009) with an accuracy of 0·96.

The data from both Holstein and Jersey populations (1045 controls and 1154 cases) were merged together for the purpose of GWAS joint analysis across the two populations. Quality control criteria were conducted using the R GenABEL package (Aulchenko et al. 2007b). A total of 7091 SNPs were excluded due to low minor allele frequency (MAF < 0·05). For the GenABEL/GRAMMAR-GC analysis a total of 15 762 SNPs were excluded because they were monomorphic in either of the two breeds. Some individuals ($n = 42$) were excluded due to high identity by state (IBS > 0·99). A total of 22 406 markers and 2157 individuals were included in the GenABEL analysis, and 45 640 markers and 2199 individuals were used in the Bayes C analysis.

Statistical analysis

Single SNP model. Single marker analysis was conducted using GRAMMAR-GC (Genome-wide association using Mixed Model and Regression – Genomic Control) within the GenABEL R package (Aulchenko et al. 2007a). The advantage of this approach is that it accounts for cryptic population substructure caused by the presence of closely related animals in the absence of pedigree information. Relationships are inferred by GRAMMAR-GC using genomic marker data (Aulchenko et al. 2007b).

Phenotypes are regressed on genotypes one SNP at a time in GRAMMAR-GC in three steps. Initially, phenotypes were corrected by accounting for familial dependence among individuals. These familial correlation-free residuals were then used as dependent quantitative traits for association analysis of each SNP using a linear regression model. Finally, genomic control (GC) was applied to correct the test statistic using the genomic inflation factor (λ), which is the regression coefficient of the observed statistic on the expected statistic.

A nominal significance threshold of $P < 5 \times 10^{-5}$ was used following the Welcome Trust Case Control Consortium suggestion (WTCCC, 2007). Additionally, false discovery rate (FDR) was used to determine probability that associations were not false positives (Bengamini & Hochberg, 1995). A quantile-quantile (Q–Q) plot of observed *P*-values *vs* expected *P*-values was created to evaluate potential inflation of the GWAS associations. A Manhattan plot of minus $\log_{10}$ of SNP *P*-values *vs* chromosomal location was drawn using an in-house script in R (R_Core_Team, 2013).

Combined within breed analysis. The analyses described above assume the same linkage phase between SNP and causative allele across breeds. To account for the possibility of a SNP being in linkage disequilibrium with the same causative SNP in both breeds, but differing in linkage phase, single SNP analyses with GRAMMAR-GC were conducted separately within Jersey and Holstein breeds with the

respective *P*-values combined subsequently. Combining *P*-values was conducted using the weighted *Z*-test, sometimes called 'Stouffer's method' which has more power and more precision than does Fisher's test (Whitlock, 2005).

Bayesian analysis. The Bayes C threshold model implemented in GenSel (Fernando & Garrick, 2009) was used in this study for mapping quantitative trait loci (QTL) of *MAP* infection as an alternative approach to single-marker GWAS. The Bayes C analysis was conducted assuming a fraction $\pi = 0.999$ of SNP markers having no effect on the phenotype. Cumulative effects of SNP across consecutive 1 Mb windows (Fernando & Garrick, 2009) was calculated to detect chromosomal regions contributing most greatly to trait variation. A high value of π was chosen to allow only regions with strongest associations to be identified. A threshold model was used in Bayes C with a probit link function for categorical binary traits (Kizilkaya et al. 2010).

Assuming an average heritability of 0·1 for susceptibility to *MAP*-infection in dairy cattle, 0·11 was used as a prior for genetic variance and was assumed to have a scaled inverse chi-squared distribution with 4 degrees of freedom and scale parameter. The genetic variance was computed from allele frequency and posterior mean of substitution effect for each SNP in 41 000 Markov Chain Monte Carlo (MCMC) iterations with 1000 burn-in cycles.

Because of LD between markers in the vicinity of a QTL, the effect of a QTL may be distributed over nearby markers. Consecutive 1 Mb non-overlapping windows (genome windows) along the bovine genome were used to calculate the cumulative effects of markers within windows (Garrick & Fernando, 2013). SNPs were assigned to consecutive genome windows using physical map order derived from bovine genome assembly UMD 3·1. The window effect in GenSel is expressed as the percentage of total genetic variance contributed by each window. Genetic variance explained by each window as a part of the total variance explained by all windows was used to identify the most informative windows. Posterior probability of inclusion (PPI), the proportion of MCMC samples in which a given SNP window was included in the model with non-zero effect, was also calculated as an indicator of contribution of the window. Percentages of explained genetic variance by windows were plotted against chromosomes using an in-house script in R (R_Core_Team, 2013).

For the most significant markers or chromosomal regions identified, potential positional candidate genes were identified considering two factors, relevance of the candidate gene to the immune system or disease resistance in mammals and proximity between the associated SNP and candidate gene. Median physical distance for pairs of markers with moderate LD ($r^2 = 0.4–0.6$) is ~1·1 Mb in the Holstein breed (Kim & Kirkpatrick, 2009), and consequently, a 2 Mb region centred on the marker or window of interest was evaluated when considering potential positional candidate genes.
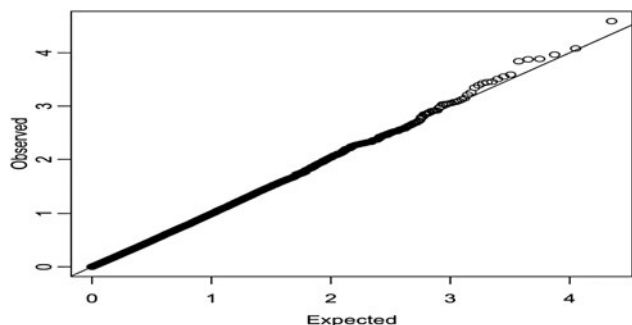
Fig. 1. Quantile-quantile plots of *P*-values for genome wide association analysis for the susceptibility to *MAP* infection. The solid black line is the slope expected under no inflation and no true associations, the y-axis represents the observed $-\log_{10}$ of *P*-values, and the x-axis represents the expected $-\log_{10}$ of *P*-values under the null hypothesis of no association.

Results

Single SNP model

The GC-corrected *P*-values for the majority of SNPs exhibit a good correspondence to the expected *P*-values under the null hypothesis of no association, with a limited number indicating association with the trait under study (Figs 1 and 2). The lambda inflation factor (λ) was estimated to be 0·95 ($\text{SE} = 1\cdot1 \times 10^{-4}$) indicating no issue of population substructure or cryptic relatedness inflating *P*-values (McCarthy et al. 2008). One SNP on chromosome 27 (Hapmap51513-BTA-88215) surpassed the significance threshold of 5×10^{-5} (Table 1).

Combined within breed analysis

Combining *P*-values from separate within-breed analyses identified two moderately ($P < 5 \times 10^{-5}$) significantly associated SNPs (Table 1, Fig. 3); ARS-BFGL-NGS-19381 on BTA23 (32 Mb) and Hapmap40994-BTA-46361 on BTA19 (61 Mb) with estimated FDR values of 0·67 and 0·44, respectively.

Bayesian analysis

A total of 2664 genome windows containing an average of 17 SNPs per window were examined. The mean posterior estimate of genetic variance was 0·082 (95% High posterior density (HPD): 0·02–0·22)). The mean posterior estimate of h^2 was 0·077 with 95% HPD interval from 0·02 to 0·18. There were nine windows that each individually explained >1% of the genetic variance, ranging from 1·13–3·26% of the total genetic variance (Table 2). Three 1 Mb windows were included in the model in more than 20% of the MCMC iterations (PPI > 20%) and these were located on chromosomes 2, 3 and 6 (Table 2). A window on BTA3 (100–101 Mb) included 27 SNPs and accounted for the highest proportion of genetic variance (3·26%). Nine

windows, each accounting for more than 1% of genetic variance were located on BTA2, 3 (3 windows), 6, 8, 25, 27 and 29; together they explained more than 16% of the total genetic variance (Table 2, Fig. 2). The window at 96 Mb on BTA2 contains 12 SNPs and had the highest posterior probability of inclusion (PPI); it was included in 27% of samples of MCMC iterations with a non-zero effect. This was the highest proportion of MCMC sampling among all windows. SNP ARS-BFGL-NGS-103501 is located within that window and it was among the most significant SNPs in both GenABEL and analysis of combined *P*-values from the within breed analysis. This SNP was also included in the model in 27% of MCMC iterations and this was the highest model frequency (proportion of MCMC iterations that included the corresponding SNP) among all SNPs in that window (Tables 1 and 2).

Correspondence between results of the two GWAS approaches and the Bayesian analysis is not apparent from the results presented, as only marker Hapmap51513-BTA-88215 on BTA27 (Table 1) and the 8 Mb window (Table 2) appear in common. However, considering individual SNP results that fell just below the 5×10^{-5} threshold, six chromosomal regions identified by the Bayesian analysis (1 Mb windows on BTA2, 3, 8, 23, 25 and 27) contained SNPs that were among the ten most significant from either of the GRAMMAR-GC analyses. Windows that explained large proportions of genetic variance in the Bayes C analysis (Table 2) but did not include the most strongly associated SNPs in the single marker analyses (Table 1) were located on BTA3 (100, 101 Mb), BTA6 (1 Mb), BTA29 (31 Mb) and BTA8 (110 Mb).

Discussion

The current study builds on previous independent analyses in Holstein and Jersey. Results from the independent analyses of these two breeds (Kirkpatrick et al. 2011; Alpay et al. 2014; Zare et al. 2014) suggested little overlap between breeds in the chromosomal regions contributing most significantly to susceptibility to *MAP* infection. A useful way of looking at the results of the current study is to identify regions not previously identified as significant which are significant in the combined analysis.

Several of the most significant associations in the current study were for SNPs or chromosomal regions not identified in previous within-breed analyses. The most significant SNP association in the current study was for Hapmap51513-BTA-88215 on BTA27 (8 Mb), a SNP, which had not been previously reported to show association with *MAP* infection. Similarly the association on BTA2 (96 Mb) identified in the current study was not previously observed in the within-breed analyses. This suggests that these associations were marginally significant in previous analyses such that combining the datasets increased statistical significance. Considering potential candidate genes, *CREB1* (Camp response element-binding protein) which is associated
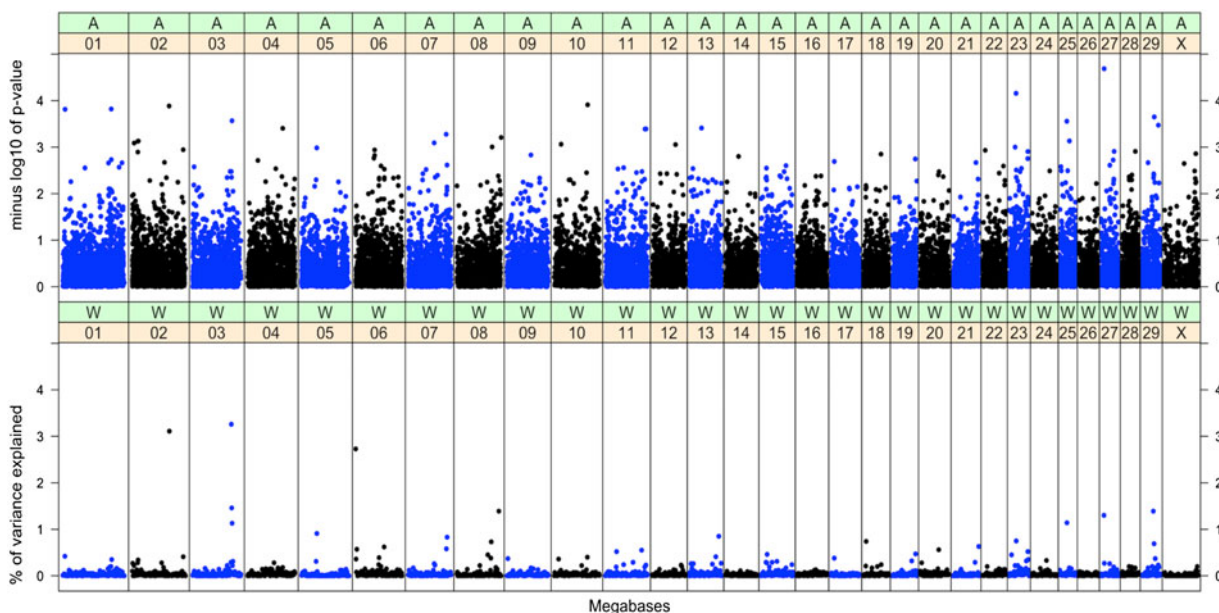
Fig. 2. Manhattan plots displaying the results of both single-marker genome-wide association analysis (GRAMMAR-GC) and Bayesian analysis for susceptibility to *MAP* infection. The upper panel displays the results of single-marker association (A) GWAS (22 406 SNPs) with the y-axis representing the $-\log_{10}$ of *P*-values corrected by genomic control (GC) and the x-axis representing megabase distance along chromosomes. The lower panel displays the results of Bayesian analysis in which each dot corresponds to the cumulative effects of SNPs within a 1 Mb window (W). The y-axis represents the percent of genetic variance, and the x-axis represents megabase distance along chromosomes.

Table 1. Results of single-marker (GenABEL) GWAS analysis for susceptibility to *MAP* infection across breed

SNP	BTA†	Position‡	*P*-value	FDR§
Analysis of combined Holstein and Jersey data				
Hapmap51513-BTA-88215	27	8 689 042	$2{\cdot}05 \times 10^{-5}$	0·44
Combined *P*-values from within-breed analysis¶				
ARS-BFGL-NGS-19381	23	32 621 505	$3{\cdot}07 \times 10^{-5}$	0·67
Hapmap40994-BTA-46361	19	61 630 610	$4{\cdot}83 \times 10^{-5}$	0·44

†*Bos taurus* chromosome.
‡Location of SNP in basepairs based on bovine genome assembly UMD 3·1.
§False discovery rate.
¶*P*-values from independent within-breed analysis of Holstein and Jersey combined, irrespective of linkage phase (Whitlock, 2005).
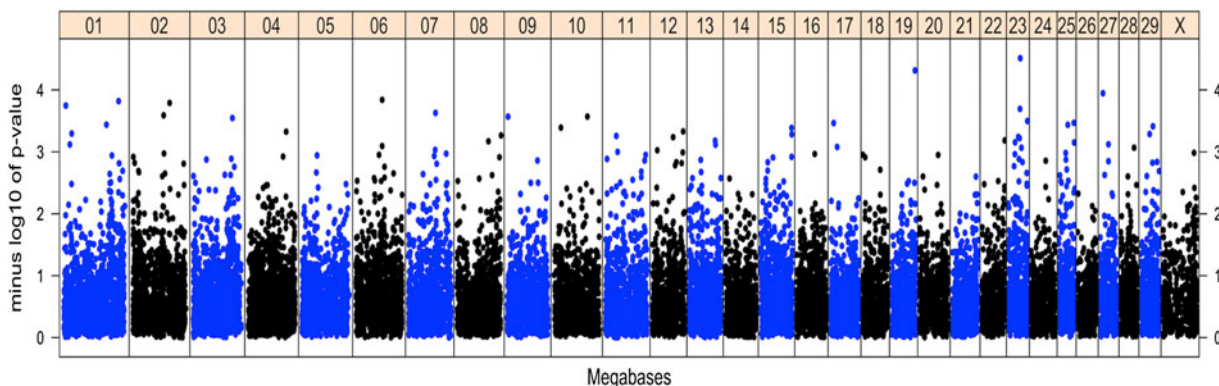

Fig. 3. Manhattan plots displaying the results of single-marker genome-wide association analysis (GRAMMAR-GC) for susceptibility to *MAP* infection GWAS (22 406 SNPs) within breed disregarding the linkage phase among the two breeds, with the y-axis representing the $-\log_{10}$ of *P*-values corrected by genomic control (GC) and the x-axis representing megabase distance along chromosomes.

Table 2. One megabase windows from Bayes C analysis of combined Holstein and Jersey data accounting for greater than one per cent of genetic variance

BTA†	Starting location (bp)‡	SNPs§	% Genetic variance¶	PPI > 0††
3	100 009 142	27	3·26	0·20
2	96 021 111	12	3·11	0·27
6	202 769	20	2·73	0·25
3	101 076 596	19	1·46	0·14
29	31 026 439	16	1·39	0·10
8	110 015 684	17	1·39	0·13
27	8 003 102	22	1·30	0·18
25	19 173 921	15	1·14	0·12
3	102 026 561	18	1·13	0·12

†*Bos taurus* chromosome.
‡Location of first SNP within the 1 Mb window.
§Number of SNPs within the 1 Mb window.
¶Percentage of total genetic variance explained by the 1 Mb window.
††Posterior probability of inclusion (PPI): The proportion of samples of the Markov chain in which a given SNP window was included in the model with a non-zero effect.

with T-cell activation and interleukin-2 production (Solomou et al. 2001) was identified for the 96 Mb region of BTA2.

Several studies have examined the genetic basis of *MAP* infection in cattle using genome-wide marker analyses (reviewed by (Kirkpatrick & Shook, 2011)). The correspondence between the results of these studies and our joint analysis is low potentially due to examining different stages of disease progression or use of different diagnostic methods and hence different disease phenotype definitions (Minozzi et al. 2012). A more fundamental concern and explanation is the modest to low power that characterises all of these studies.

Based on the results from combining *P*-values of the two independent within-breed analyses, two potential candidate genes were observed within 1 Mb of the most significant SNP. *BTN1A1* (*Butyrophilin*), is a member of the immunoglobulin superfamily and is located 1·3 Mb proximal to ARS-BFGL-NGS-19381. Human *BTN1A1* is localised in the major histocompatibility complex (MHC I) region. A second gene, tyrosyl-DNA phosphodiesterase 2 (*TDP2*) is located 0·2 Mb distal to ARS-BFGL-NGS-19381. The encoded protein associates with CD40 (a member of the tumour necrosis factor receptor family that plays a critical role in many immunological processes).

The proportion of genetic variance explained by all SNPs across the genome (0·086) was in the range of previous pedigree-based heritability estimates (0·08–0·27). This suggests that markers captured most of the additive genetic variation through LD with QTL (Zare et al. 2014). The relatively high FDR (~0·5) estimates for the most significant markers suggest a high probability of type I error in these associations, so caution in interpretation of these results is warranted. The moderate marker density employed in the current study may have limited the ability to identify markers in common linkage phase across breed, thus compromising statistical significance in the across-breed analysis. Re-analysis of the combined breed data after imputation of genotypes to high density may reveal new associations or strengthen some of the identified associations.

References

Alpay F, Zare Y, Kamalludin MH, Huang X, Shi X, Shook GE, Collins MT & Kirkpatrick BW 2014 Genome-wide association study of susceptibility to infection by Mycobacterium avium subspecies paratuberculosis in Holstein cattle. *PLoS ONE* **9** e111704

Aulchenko YS, de Koning DJ & Haley C 2007a Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics* **177** 577–585

Aulchenko YS, Ripke S, Isaacs A & van Duijn CM 2007b GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23** 1294–1296

Bengamini Y & Hochberg Y 1995 Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57** 289–300

Bentley RW, Keenan JI, Gearry RB, Kennedy MA, Barclay ML & Roberts RL 2008 Incidence of Mycobacterium avium subspecies paratuberculosis in a population-based cohort of patients with Crohn's disease and control subjects. *American Journal of Gastroenterology* **103** 1168–1172

Browning BL & Browning SR 2009 A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics* **84** 210–223

Collins MT 2002 Interpretation of a commercial bovine paratuberculosis enzyme-linked immunosorbent assay by using likelihood ratios. *Clinical and Diagnostic Laboratory Immunology* **9** 1367–1371

Collins MT, Gardner IA, Garry FB, Roussel AJ & Wells SJ 2006 Consensus recommendations on diagnostic testing for the detection of paratuberculosis in cattle in the United States. *Journal of the American Veterinary Medical Association* **229** 1912–1919

Fernando RL & Garrick D 2009 GenSel – User Manual for a Portfolio of Genomic Selection Related Analyses. Animal Breeding and Genetics, Iowa State University, Ames. http://bigs.ansci.iastate.edu/bigsgui/login.html (Accessed February 2014)

Garrick DJ & Fernando RL 2013 Implementing a QTL detection study (GWAS) using genomic prediction methodology. *Methods in Molecular Biology* **1019** 275–298

Kim ES & Kirkpatrick BW 2009 Linkage disequilibrium in the North American Holstein population. *Animal Genetics* **40** 279–288

Kirkpatrick BW & Shook GE 2011 Genetic susceptibility to paratuberculosis. *Veterinary Clinics of North America Food Animal Practice* **27** 559–571

Kirkpatrick BW, Shi X, Shook GE & Collins MT 2011 Whole-Genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Animal Genetics* **42** 149–160

Kizilkaya K, Fernando RL & Garrick DJ 2010 Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *Journal of Animal Science* **88** 544–551

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP & Hirschhorn JN 2008 Genome-wide association studies for complex

traits: consensus, uncertainty and challenges. Nature reviews. *Genetics* **9** 356–369

Minozzi G, Williams JL, Stella A, Strozzi F, Luini M, Settles ML, Taylor JF, Whitlock RH, Zanella R & Neibergs HL 2012 Meta-analysis of two genome-wide association studies of bovine paratuberculosis. *PLoS ONE* **7**, e32578

NAHMS 2007 *Part I: Reference of Dairy Health and Management in the United States*. Fort Collins, CO: USDA: APHIS: VS: CEAH

Nielsen SS, Bjerre H & Toft N 2008 Colostrum and milk as risk factors for infection with Mycobacterium avium subspecies paratuberculosis in dairy cattle. *Journal of Dairy Science* **91** 4610–4615

Ott SL, Wells SJ & Wagner BA 1999 Herd-level economic losses associated with Johne's disease on US dairy operations. *Preventive Veterinary Medicine* **40** 179–192

R_Core_Team 2013 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Shin SJ, Cho D & Collins MT 2008 Diagnosis of bovine paratuberculosis by a novel enzyme-linked immunosorbent assay based on early secreted antigens of Mycobacterium avium subsp. paratuberculosis. *Clinical and Vaccine Immunology* **15** 1277–1281

Solomou EE, Juang YT, Gourley MF, Kammer GM & Tsokos GC 2001 Molecular basis of deficient IL-2 production in T cells from patients with systemic lupus erythematosus. *Journal of Immunology* **166** 4216–4222

Whitlock MC 2005 Compining probability from indepenent tests: the weighted Z-method is a superior to Fisher's approach. *Journal of Evolutionary Biology* **18** 1368–1373

WTCCC 2007 Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447** 661–683

Zare Y, Shook GE, Collins MT & Kirkpatrick BW 2014 Genome-wide association analysis and genomic prediction of Mycobacterium avium Subspecies paratuberculosis infection in US Jersey cattle. *PLoS ONE* **9**, e88380