

A whole genome association analysis identifies loci associated with *Mycobacterium avium* subsp. *paratuberculosis* infection status in US holstein cattle

M. Settles*, R. Zanella*, S. D. McKay[†], R. D. Schnabel[†], J. F. Taylor[†], R. Whitlock[‡], Y. Schukken[§], J. S. Van Kessel[¶], J. M. Smith** and H. Neibergs*

*Department of Animal Sciences, Washington State University, Pullman, WA 99164, USA. [†]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA. [‡]School of Veterinary Medicine, University of Pennsylvania, Kennett Square, PA 19348, USA. [§]College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA. [¶]USDA-ARS, Beltsville, MD 20705, USA. **Department of Animal Science, University of Vermont, Burlington, VT 05405, USA

Summary

The purpose of this study was to identify loci associated with *Mycobacterium avium* subspecies *paratuberculosis* (*Map*) infection status in US Holsteins using the *Illumina* BovineSNP50 BeadChip whole genome single nucleotide polymorphism (SNP) assay. Two hundred forty-five cows from dairies in New York, Pennsylvania and Vermont enrolled in longitudinal herd studies between January 1999 and November 2007 were assessed for the presence of *Map* in both faecal and tissue samples. An animal was considered tissue infected if any sample contained at least one colony forming unit of *Map* per gram of tissue (CFU/g) and the same definition was employed for faecal samples. Each animal was genotyped with the *Illumina* BovineSNP50 BeadChip and after quality assurance filtering, 218 animals and 45 683 SNPs remained. We sought to identify loci associated with four different case/control classifications: presence of *Map* in the tissue, presence of *Map* in faeces, presence of *Map* in both tissue and faeces and presence of *Map* in tissue but not faeces. A case-control genome wide association study was conducted to test the four different classifications of *Map* infection status (cases) when compared with a *Map*-negative control group (control). Regions on chromosomes 1, 5, 7, 8, 16, 21 and 23 were identified with moderate significance ($P < 5 \times 10^{-5}$). Two regions, one on chromosome 3 (near *EDN2*) and another on chromosome 9 (no positional gene candidates), were identified with a high level of association to the presence of *Map* in tissue and both tissue and faeces respectively ($P < 5 \times 10^{-7}$, genome-wide Bonferonni $P < 0.05$).

Keywords *Bos taurus*, BovineSNP50, genome wide association, Johne's disease, SNP chip.

Introduction

Bovine Paratuberculosis, commonly referred to as Johne's disease, is a contagious bacterial disease estimated to be present in 68.1% of US dairy herds (USDA – Animal and Plant Health Inspection Service (APHIS) 2008) and which results in annual losses exceeding US \$200 million (Ott *et al.* 1999). The bacterium *Mycobacterium avium* ssp. *paratuberculosis* (*Map*) is responsible for Johne's disease and

causes reduced milk production, reproductive failure, weight loss and eventual death. Johne's disease is not treatable, and vaccine efficacy against *Map* is controversial.

Current management practices include culling cows that test positive for *Map* or exhibit clinical signs of the disease. However, clinical signs of *Map* infection may be delayed 5 years or more years after the initial exposure (Stabel 2006). Current diagnostic testing has a limited sensitivity for detecting *Map* infection in pre-clinical animals, at which time they are likely to be spreading *Map* to other animals in the herd through faecal contamination of the environment, food and water (Whitlock *et al.* 2000). The low overall sensitivity of enzyme-linked immunosorbent assay (7–35%) and faecal (38–65%) diagnostic tests and the long incubation period of the disease present major roadblocks to the

Address for correspondence

H. Neibergs, Department of Animal Sciences, Washington State University, Pullman, WA 99164, USA.
E-mail: neibergs@wsu.edu

Accepted for publication 15 February 2009

control of Johne's disease (Whitlock *et al.* 2000; McKenna *et al.* 2005; Collins *et al.* 2006). The identification of animals with genetic susceptibility to infection by *Map* would provide a mechanism to reduce the incidence of Johne's disease. Further, a reduction of *Map* in the environment may also be beneficial to humans, as the presence of *Map* has also been associated with Crohn's disease (Bentley *et al.* 2008).

Resistance to *Map* infection has been found to be heritable (Koets *et al.* 2000; Mortensen *et al.* 2004; Gonda *et al.* 2006; Hinger *et al.* 2008), with heritability estimates ranging from 0.06 to 0.102. However, attempts to locate loci associated with resistance to paratuberculosis have had limited success. Gonda *et al.* (2007) found evidence for a quantitative trait locus on *Bos taurus* chromosome 20 (BTA20) associated with paratuberculosis susceptibility. Hinger *et al.* (2007) investigated the association with paratuberculosis of eight microsatellites located in or near *Map* susceptibility candidate genes in 1179 (594 positive) German Holstein cows. However, none showed any significant associations. While *Map* susceptibility genes have yet to be identified in the bovine, Reddacliff *et al.* (2005) found an association of one microsatellite allele in *SLC11A1* (formerly *NRAMP1*) with *Map* resistance in sheep.

The purpose of this study was to identify loci associated with *Map* infection status using the Illumina BovineSNP50 BeadChip whole genome SNP assay. The population of animals used in this study was Holstein cows from four geographically distinct herds culled for any reason, including having a *Map* positive diagnostic test. *Map* infection status was evaluated from culture diagnostic testing of faecal samples and intestinal tract tissues. We sought to identify loci associated with four different case/control classifications: the presence of *Map* in the tissue, presence of *Map* in faeces, presence of *Map* in both tissue and faeces and presence of *Map* in tissue but not faeces. The first classification tested if there were loci associated with a tissue positive result for *Map*, regardless of faecal status. These loci may be associated with the bacterium's ability to infect the host's cells. The second classification tested for loci associated with faecal shedding that may be independent of those associated with *Map* tissue infection; these loci may correspond to regions responsible for a cow's ability to transmit *Map* through its faeces. Further, this analysis is comparable to studies that inferred disease status by the presence of faecal *Map*, in the absence of tissue data. When considering both faecal and tissue results to classify animals, four possible subgroups arise: *Map* negative controls (faecal-, tissue-), faecal shedding but tissue negative (faecal+, tissue-), tissue-infected but not faecal shedding (faecal-, tissue+), and tissue-infected and faecal shedding 'shedding' (faecal+, tissue+). The third classification tests for loci associated with tissue-infected but not faecal shedding animals (cases) relative to negative animals (controls) and the fourth classification tests 'shedding'

animals (cases) relative to negative animals (controls). These two classifications were expected to identify loci associated with tissue-infected animals that were not faecal shedding as compared with 'shedding' animals and thus to allow determination of whether any loci are shared. The identification of loci associated with each of these phenotypes may be used to develop a marker-assisted selection programme to aid in the control of Johne's disease.

Materials and methods

Study population and phenotypes

Two hundred forty-five Holstein cows from three herds comprising the RDQMA study (Herd A (New York), B (Pennsylvania) and C (Vermont); Pradhan *et al.* 2009) and a fourth herd (Herd D) from Pennsylvania were followed to culling between January 1999 and November 2007 and subsequently were assessed for the presence of *Map* in both faecal and necropsy tissue. Tissue samples from the ileum, ileo-caecal valve and two adjacent ileo-caecal lymph nodes were cultured for the presence of *Map* following the protocols previously described in Whitlock *et al.* (2000). In addition, faecal samples were taken at necropsy and were cultured for the presence of *Map* using the procedure described in Whitlock *et al.* (1996). Each animal was considered tissue infected if any cultured sample from any of the four tissues had >0 CFU/g of tissue. Faecal culture status was classified in a similar manner. Ninety-five animals were classified as tissue positive and 45 animals were classified as faecal positive. Within the tissue positive samples, faecal culture detected the presence of *Map* in only 40% of the samples (38 of 94, when both tissue and faecal results were present, one tissue positive sample had a missing faecal sample). Six animals tested positive in faecal culture but not in tissue culture. The faecal CFU/g values in these animals were less than five and may represent what many consider to be faecal 'pass-through'. A faecal 'pass-through' is when an animal has a faecal positive test because of the consumption of *MAP*, but is not *Map* infected. The distributions of the faecal and tissue results by herd are shown in Table S1. Each sample was further characterized for potential confounding variables such as age at culling (median 56.5 months of age, quantiles 1 and 3 equal to 44.3 and 72.25 months respectively) and herd of origin.

Genotyping

DNA was extracted from the tissue of each animal using the Puregene DNA extraction kit as per manufacturer's instructions (Gentra). Sample DNA was quantified and genotyped using the Illumina BovineSNP50 BeadChip (unpublished data). The Illumina BovineSNP50 BeadChip assay contains 55 074 SNPs with a mean spacing of one SNP every 49.4 kb (median spacing of 37 kb; quartiles 1

and 3 equal to 27.6 and 54 kb respectively; and a maximum distance of 1.45 Mb) based on the *BTAU4.0* assembly (<ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/>). The BovineSNP50 BeadChip also contains an additional 1828 SNPs, which are located on unassigned contigs (Chr. Un) in the *BTAU4.0* assembly. All samples were genotyped using *BEADSTUDIO* (Illumina) software and a custom cluster file developed from more than 7000 samples and multiple *B. taurus* cattle breeds at the University of Missouri.

Genotype quality assurance

Seven animals (one from Herd A, five from Herd B and one from Herd D) were excluded from the analysis for quality because of >10% missing genotypes. An excess no-call rate is an indicator of low quality DNA. To assess technical variation, one animal genotyped twice resulted in a >99% identity of called genotypes (two mismatches). Multi-Dimension Scaling (MDS) analysis identified 15 animals that were clearly distinct from the majority of animals, two animals from Herd A, five animals from Herd B, one from Herd C and seven animals from Herd D (see Appendix S1). The projection of the data onto the first two MDS axes is shown in Fig. 1. These animals were removed from further analysis, resulting in a substantial reduction of the genomic inflation factor (based on median chi-squared, data not shown). After removing animals for genotype quality and excessive genomic variance, we compared the age ranges of

the remaining control animals (tissue negative) with those for tissue *Map*-positive animals. The mean age of control animals was 58.1 months, while the mean age of the tissue *Map*-positive animals was 60.5 months of age. Five control animals were significantly younger (approximately 12 months of age when culled) than the youngest of the tissue *Map*-positive animals (approximately 23 months of age when culled). These animals were excluded from the study as their youth may have affected our ability to detect *Map* in their tissues and/or faeces. The resulting mean age at culling for control animals (tissue negative) was 60 months with a range of 21.8–144, months and the mean age at culling was 60.5 months for *Map* tissue positive animals with a range of 22.9–135 months. After removing 27 samples for quality, genetic variability and age, 218 animals remained in the study with an average genotype call rate of 98.9% (mean of 54 475 SNPs). Of these, 90 animals tested positive for *Map* in at least one tissue sample (tissue tissue-positive) and 41 animals tested positive for *Map* in faecal samples (faecal positive). Thirty-five animals tested positive in both faecal and tissue samples ('shedding'), six animals tested positive in faecal culture but not tissue ('pass-through'), 54 animals tested positive in at least one tissue sample but not faecal sample ('infected'), 112 samples tested negative for *Map* in both faecal and tissue samples ('negative'), 11 samples were untested for either faecal or tissue and seven of these animals were untested for both faecal and tissue samples. These animals were left in the study to contribute to SNP level quality assurance but did not contribute to any association test statistic. The distribution of animal numbers by herd is shown in Table S2.

We excluded 1276 SNPs with >10% genotype no-call rate and 8317 with a minor allele frequency < 0.01, of which 6356 were monomorphic. Genome-wide, 45 683 SNPs (82.9%) passed these quality control filters.

Statistical analysis

Standard one-degree of freedom (d.f.) allelic, 1-d.f. dominance, 1-d.f. recessive and 2-d.f. (genotypic) tests of association with genotype between cases and controls were calculated. When evidence for stratification occurred, a within-herd Cochran-Mantel-Haenszel (CMH) test $2 \times 2 \times K$ ($K = 4$ herds) of association with genotype was performed. All calculations and plots were performed using the R statistical environment and PLINK (Purcell *et al.* 2007; Version 1.04). For genome-wide association (GWA), uncorrected P -values $< 5 \times 10^{-7}$ provided strong evidence of association, and uncorrected P -values between 5×10^{-5} and 5×10^{-7} were considered to provide moderate evidence (Wellcome Trust Case Control Consortium 2007). Physical positions and alleles are expressed in terms of the forward strand of the reference genome (*BTAU4.0*, <ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/>).

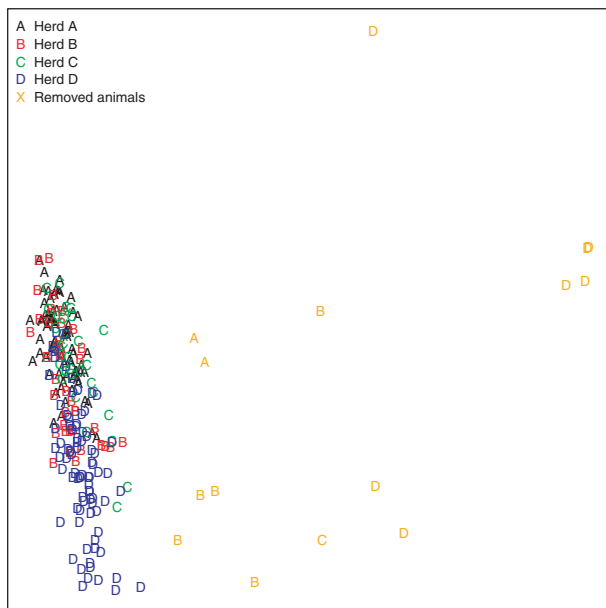


Figure 1 Multidimensional scaling (MDS) plot. An MDS plot provides a spatial representation of data that can facilitate interpretation and reveal structural relationships in the data. In this experiment, the MDS plot identified 15 animals (grey) whose multilocus genotypes differed significantly from the other animals. These animals were removed from further analysis.

The resulting sample size of 218 animals used in this study is lower when compared with many human case-control studies, which frequently reach to thousands of individuals. The primary factors in this study limiting the sample size are the difficulty in attaining a high number of samples and the significant cost associated with determining phenotype; as such, the power to detect causative alleles is expected to be lower than in comparable human studies.

Results

A case-control GWA study was conducted to test four different classifications of *Map* infection status defined by results from *Map* culture from both tissue and faecal samples. Single SNP analysis was conducted to test the association of loci to each of the four classifications of *Map* infection status.

Association of loci with *Map*-infected tissue (tissue positive vs. tissue negative)

Cases in this analysis were defined as animals with a *Map*-positive tissue result ($n = 90$) and controls were animals with a *Map*-negative tissue result ($n = 119$) regardless of faecal shedding status. We found a strong association with the ss86341066 SNP located on BTA3 (111 682 510 bp) using the basic allelic model ($P = 3 \times 10^{-7}$; $P = 0.014$ after

Bonferonni correction). Seven SNPs were found to be moderately associated ($P < 5 \times 10^{-5}$) with *Map*-infected tissue and were located on chromosomes 1 and 21 for the basic allele frequency difference model and on chromosomes 5 and 16 using the dominance model (Table 1). There was no evidence for population substructure on the quantile-quantile (Q-Q) plot (see Appendix S1, Fig. S1), or based on the genomic inflation factor (based on median chi-square value, $\lambda_{gc} = 1$).

Association of loci with *Map* faecal positive (faecal positive vs. faecal negative)

Cases in this analysis were defined as animals with a positive faecal culture result ($n = 41$) and controls were defined as animals with a negative faecal culture result ($n = 168$), regardless of culture of tissue status. The standard allelic, dominance, recessive and genotypic models showed evidence that the results were influenced by population substructure (identified by Q-Q plot, Fig. S2 and genomic inflation factor, $\lambda_{gc} = 1.10$). This was likely due to Herds A, B and C containing fewer faecal positive animals relative to Herd D. To account for population substructure, we performed stratified analysis (CMH test) by herd. Investigation of the resulting Q-Q plot (Fig. S3) and the genomic inflation factor ($\lambda_{gc} = 1.0$) shows that this test successfully accounted for population substructure. CMH test results showed no strong associations

Table 1 Genomic regions associated with *Map*.

SNP	Chr	Pos (bp)	Phenotype	Test	Minor allele (freq)	Odds ratio (minor allele)	P-value	RefSeq genes (1 Mb)
rs29012843	1	3 083 368	Tissue	Allelic	A (0.43)	2.31	3.264e-05	<i>SOD1</i>
rs29012842	1	3 083 498	Tissue	Allelic	A (0.43)	2.31	3.264e-05	<i>SOD1</i>
rs41924236	1	17 106 221	Infected	Allelic	G (0.31)	0.31	4.269e-05	None
ss86341066	3	111 682 510	Tissue	Allelic	A (0.36)	0.33	3.062e-07 *(0.014)	<i>FOXJ3, EDN2, CTPS,</i>
			Infected	Allelic		0.33	3.290e-05	<i>CITED4, NFYC</i>
ss61502325	5	73 634 207	Tissue	Dom	A (0.34)	0.30	3.708e-05	<i>TRRAP, ALDH1L2</i>
ss61567534	5	106 209 963	Faecal	CMH	G (0.01)	91.2	4.163e-05	<i>TAS2R42, MAGOHB,</i>
								<i>KLRA1, KLRJ1</i>
ss86329857	7	47 688 319	Shedding	CMH	G (0.46)	3.689	3.723e-05	<i>SPOCK1</i>
ss86299036	8	74 335 842	Faecal	CMH	A (0.46)	0.30	3.389e-05	<i>STC1</i>
rs43679450	9	647 609	Shedding	CMH	A (0.01)	41.50	2.511e-05	None
rs43679467	9	698 262	Shedding	CMH	A (0.01)	41.50	2.511e-05	None
ss117969611	9	786 610	Shedding	CMH	C (0.01)	41.50	2.511e-05	None
rs43070062	9	813 310	Shedding	CMH	C (0.01)	NA	1.029e-07 *(0.005)	None
			Faecal	CMH		NA	1.661e-05	
rs42396640	16	27 186 964	Tissue	Dom	G (0.41)	3.71	4.632e-05	<i>PARP1, POLR1D, PSEN2,</i>
								<i>CABC1, SCCPDH</i>
rs42397377	16	27 237 455	Tissue	Dom	G (0.41)	3.97	2.579e-05	<i>PARP1, POLR1D, PSEN2,</i>
								<i>CABC1, SCCPDH</i>
rs29025761	21	26 660 590	Infected	Allelic	C (0.09)	5.29	8.115e-06	<i>BCL2A1, ZFAND6, MESDC2,</i>
			Tissue	Allelic		4.18	3.453e-05	<i>IL16, MCEE</i>
ss46526325	23	48 404 721	Faecal	CMH	G (0.35)	3.07	4.392e-05	<i>EEF1E1</i>

MAP, *Mycobacterium avium* subspecies *paratuberculosis*; CMH, Cochran-Mantel-Haenszel test.

*Bonferonni significance, when applicable, in brackets.

with a P -value of $<5 \times 10^{-7}$; however, four SNPs were found to be moderately significant ($P < 5 \times 10^{-5}$), and were located on chromosomes 5, 8, 9 and 23 (Table 1).

Association of loci with *Map* tissue infection but faecal culture negative (infected)

Cases in this analysis were defined as animals with a positive culture of tissue result and a negative faecal culture ($n = 54$) and controls were defined as animals with a negative culture of tissue result and a negative faecal culture ($n = 112$). No strong association with any SNPs was identified using the dominant, recessive, or genotypic model, but three SNPs were identified with moderate P -values using the allelic model. Two SNPs located on chromosomes 3 and 21 were the same as those found associated with tissue positive status. The third SNP was a newly identified locus found on chromosome 1, 14 Mb downstream of the SNPs found with a moderate association with a tissue positive result (Table 1). Analysis of 'infected' animals did not show evidence for population substructure in the Q-Q plot (Fig. S4) or genomic inflation factor ($\lambda_{gc} = 1$).

Association of loci with the class *Map* tissue infected, faecal positive (shedding)

Cases in this analysis were defined as animals with a positive tissue and a positive faecal culture result ($n = 25$), while controls were defined as animals with a negative tissue and a negative faecal culture result ($n = 112$). The standard allelic, dominance, recessive and genotypic models showed evidence that results were influenced by population substructure (identified by Q-Q plot, Fig. S5 and genomic inflation factor, $\lambda_{gc} = 1.06$). To account for population substructure, we performed a stratified analysis (CMH test) by herd. Investigation of the resulting Q-Q plot (Fig. S6) and the genomic inflation factor ($\lambda_{gc} = 1.0$) shows that this test

successfully accounted for population substructure. One strong association was found with SNP rs43070062 located on BTA9 using the CMH test stratifying by herd ($P = 1 \times 10^{-7}$; $P = 0.004$ after Bonferroni correction). Four other SNPs, three located on chromosome 9 adjacent to rs43070062, showed moderate association ($P = 2.5 \times 10^{-5}$), as did one on chromosome 7 ($P = 3.72 \times 10^{-5}$) (Table 1).

Discussion

A GWA study was conducted to identify loci associated with *Map* infection status using four different case/control classifications of the samples: presence of *Map* in tissue; presence of *Map* in faeces; presence of *Map* in tissue but not faeces, and the presence of *Map* in both tissue and faeces. Results from the first classification, presence of *Map* in tissue, produced a single strong association at ss86341066 (odds ratio of the minor allele = 0.33, Bonferroni $P = 0.014$) with resistance to *Map* penetration of the tissue, and four moderate associations (see Fig. 2). RefSeq genes located within 1 Mb identified the gene *endothelin 2* (*EDN2*) located 31 kb downstream of ss86341066. *EDN2* has been found to be highly expressed in the gastrointestinal tract of the mouse. Takizawa *et al.* (2005) found that, in intestinal epithelial cells, *EDN2* could be secreted into the lamina propria and the dome region in Peyer's patch and that it might modulate immune cells for mucosal defence. Further, McCartney *et al.* (2002) found *EDN2* to be highly expressed in mucosal biopsies of humans, including samples taken from patients with inflammatory bowel disease (IBD). However, no associations were found between human IBD and mRNA expression.

Association with the presence of *Map* in faeces, however, did not yield any strong associations and only four moderate associations, three of which were on different chromosomes to those identified as being associated with *Map*

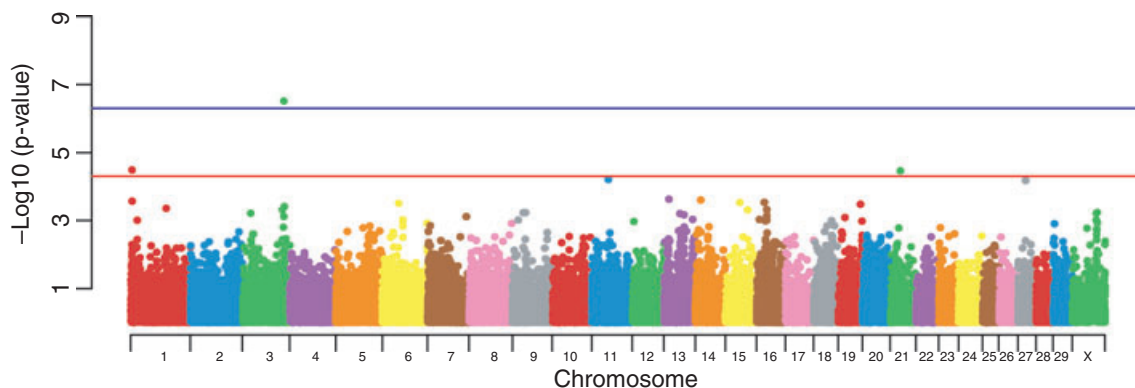


Figure 2 Genome-wide plot of $-\log_{10}(P\text{-values})$ for an association of loci with *Mycobacterium avium* subspecies *paratuberculosis* (*Map*)-infected tissue (tissue positive vs. tissue negative). Chromosomes 1–29 and Chromosome X are shown separated by colour. The horizontal red line is drawn at $-\log_{10}(5 \times 10^{-5})$ and the horizontal blue line is drawn at $-\log_{10}(5 \times 10^{-7})$ to show those significant at the moderate and strong levels of significance respectively.

infection of tissue. The fourth locus, on chromosome 5, was 30 Mb from the SNP found to be moderately associated with *Map* infection of tissue. Comparing the rank order of SNPs by significance for the tissue positive results to the faecal positive results showed the two to have a different significance ordering (data not shown). It is well established that the sensitivity of *Map* faecal culture for detecting infected animals in the early stages of infection is low. In this study, faecal culture only detected the presence of *Map* infection 40% of the time in the presence of tissue infection. The differences in sensitivities between the diagnostic tests and the resulting definition of phenotype may then significantly impact on the loci detected to be associated with *Map*, and on their corresponding interpretation. This observation led us to question whether there was a difference in loci associated with a tissue positive and faecal negative classification and a tissue positive and faecal positive classification.

If the inclusion of faecal culture results in the analysis was to provide no additional information, one would expect the rank order of SNPs, by significance, to be similar in these two analyses. However, this is not what was found. Results from tissue infection with no faecal shedding did not identify any strongly associated regions but did identify three moderately associated regions. One SNP found to have a strong association (ss86341066) in the tissue positive culture test had a moderate association in this test. Results from the analysis of the tissue positive infection in conjunction with faecal shedding phenotype produced one strong association on chromosome 9 (rs43070062, odds ratio of the minor allele = 45.1, Bonferonni $P = 0.005$) and four moderate associations, three of which were neighbours of rs43070062 (see Fig. 3). There are no Bovine RefSeq genes within 1 Mb of rs43070062. The identification of different loci associated with tissue infection, compared with tissue infected and faecal shedding, suggests that there may be different loci important to different stages of the disease.

The infection of cells by *Map* is the first step towards Johne's disease, but it may not be a guarantee of disease. Similarly, faecal shedding of *Map* may occur because of environmental exposure, but without tissue infection, and the animal may not become diseased. Within hours of *Map* ingestion, the bacterium attaches to the host's intestinal mucosa and penetrates the mucosa/M cells overlying Peyer's patches (Momotani *et al.* 1988; Wu *et al.* 2007). *Map* that survives phagocytosis by macrophages can then acquire nutrients for growth and replication. *Map* is then later disseminated in macrophages and is associated with early paucibacillary lesions. This cell-mediated immunological response restricts the expansion of *Map* (Clarke 1997; Waters *et al.* 1999). At this stage, and depending on other factors (such as the dose of *Map*, the genetics of the host, local tissue cytokine concentrations, stress and hormone levels of the cow) the animal may either clear *Map* or may develop a persistent *Map* infection that culminates in Johne's disease (Harris & Barletta 2001). For those animals that progress to Johne's disease, humoral-mediated responses replace cell-mediated responses. These animals exhibit faecal shedding, positive serological tests, greater *Map* replication rates, multibacillary lesions and clinical disease (Whittington & Sergeant 2001). The very different processes involved in early disease and late disease may also be reflected in the different loci found to be associated with the classification of tissue infection as compared with the 'shedding' classification. Because we do not fully understand the pathogenesis of the disease, we do not refer to loci associated with tissue infection or the 'shedding' classification as Johne's susceptibility or resistance loci, because these loci may have different roles in the expression of Johne's disease at different stages. Therefore, the use of both tissue and faecal diagnostic testing provides a more complete profile of the health status of the cow.

While others have found evidence of the association with Johne's disease on BTA20, our study shows no evidence for

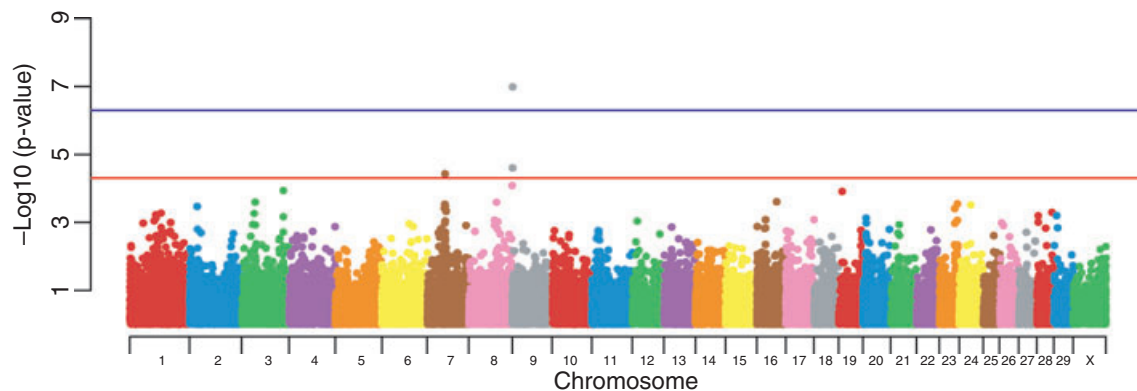


Figure 3 Genome-wide plot of $-\log_{10}(P\text{-values})$ for an association of loci with *Map* tissue-infected and faecal-positive phenotype ('shedding'). Chromosomes 1–29 and Chromosome X are shown separated by colour. The horizontal red line is drawn at $-\log_{10}(5 \times 10^{-5})$ and the horizontal blue line is drawn at $-\log_{10}(5 \times 10^{-7})$ to show those significant at the moderate and strong levels of significance respectively.

a locus associated with Johne's disease on BTA20 in any of the four phenotype classifications analysed here (Gonda *et al.* 2007). Likewise, no significant loci were found within 1 Mb of *SLC11A1*, a candidate gene found to be associated with *Map* in sheep. Additionally, in the Crohn's disease literature, two genes consistently have the most significant association with Crohn's: *IL23R* and *NOD2* (also known as *CARD15*). While others have shown *NOD2* to have 'no association' with Johne's disease (Taylor *et al.* 2006), no one has yet looked at the *IL23R* ortholog on BTA3 (located at 83 Mb). Focusing on all SNPs within 1 Mb of both *NOD2* and *IL23R*, we found no loci in the vicinity of these genes to be associated with any *Map* infection phenotype (data not shown).

To our knowledge, this is the first GWA study of Johne's disease in cattle using the recently released Illumina BovineSNP50 BeadChip. Future work may include replication of this study with a larger number of animals and fine mapping of candidate regions.

Acknowledgements

MS was supported by the Center for Reproductive Biology, Washington State University. JFT, SDK and RDS were supported by National Research Initiative grants 2005-35205-15448, 2005-35604-15615, 2006-35205-16701 and 2006-35616-16697 from the USDA Cooperative State Research, Education and Extension Service. TF, RW, YS, JSVK, JSK, EH and JMS were supported in part by the USDA-Agricultural Research Service (Agreements. 58-1265-3-155, 58-1265-3-156, 58-1265-3-158 and 58-1265-4-020) for the Regional Dairy Quality Management Alliance (RDQMA) and the Johne's Disease Integrated Program (JDIP, USDA contract 45105). The authors would also like to thank Jan Vierck (Washington State University), Terry Fyock (University of Pennsylvania), Ernest Hovingh (Penn State University) and Jeff Karns (ARS-USDA) for their contributions with sample collection and preparation.

References

- Bentley R.W., Keenan J.I., Geary R.B., Kennedy M.A., Barclay M.L. & Roberts R.L. (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**, 1168–72.
- Clarke C.J. (1997) The pathology and pathogenesis of paratuberculosis in ruminants and other species. *Journal of Comparative Pathology* **116**, 217–61.
- Collins M.T., Gardner I.A., Garry F.B., Roussel A.J. & Wells S.J. (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**, 1912–9.
- Gonda M.G., Chang Y.M., Shook G.E., Collins M.T. & Kirkpatrick B.W. (2006) Genetic variation of *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. *Journal of Dairy Science* **89**, 1804–12.
- Gonda M.G., Kirkpatrick B.W., Shook G.E. & Collins M.T. (2007) Identification of a QTL on BTA20 affecting susceptibility to *Mycobacterium avium* ssp. *paratuberculosis* infection in US Holsteins. *Animal Genetics* **38**, 389–96.
- Harris N.B. & Barletta R.G. (2001) *Mycobacterium avium* subsp. *paratuberculosis* in veterinary medicine. *Clinical Microbiology Review* **14**, 489–512.
- Hinger M., Brandt H., Horner S. & Erhardt G. (2007) Short communication: association analysis of microsatellites and *Mycobacterium avium* subspecies *paratuberculosis* antibody response in German Holsteins. *Journal of Dairy Science* **90**, 1957–61.
- Hinger M., Brandt H. & Erhardt G. (2008) Heritability estimates for antibody response to *Mycobacterium avium* subspecies *paratuberculosis* in German Holstein cattle. *Journal of Dairy Science* **91**, 3237–44.
- Koets A.P., Adugna G., Janss L.L.G., van Weering H.J., Kalis C.H.J., Wentink G.H., Rutten V.P.M.G. & Schukken Y.H. (2000) Genetic variation of susceptibility to *Mycobacterium avium* subsp. *paratuberculosis* infection in dairy cattle. *Journal of Dairy Science* **83**, 2702–8.
- McCartney S.A., Ballinger A.B., Vojnovic I., Farthing M.J.G. & Warner T.D. (2002) Endothelin in human inflammatory bowel disease: comparison to rat trinitrobenzenesulphonic acid-induced colitis. *Life Sciences* **71**, 1893–904.
- McKenna S.L.B., Keefe G.P., Barkema H.W. & Sockett D.C. (2005) Evaluation of three ELISAs for *Mycobacterium avium* subsp. *paratuberculosis* using tissue and fecal culture as comparison standards. *Veterinary Microbiology* **110**, 105–11.
- Momotani E., Whipple D.L., Thiermann A.B. & Cheville N.F. (1988) Role of M cells and macrophages in the entrance of *Mycobacterium paratuberculosis* into domes of ileal peyer's patches in calves. *Veterinary Pathology* **25**, 131–7.
- Mortensen H., Nielsen S.S. & Berg P. (2004) Genetic variation and heritability of the antibody response to *Mycobacterium avium* subspecies *paratuberculosis* in Danish Holstein cows. *Journal of Dairy Science* **87**, 2108–13.
- Ott S.L., Wells S.J. & Wagner B.A. (1999) Herd-level economic losses associated with Johne's disease on US dairy operations. *Preventative Veterinary Medicine* **40**, 179–92.
- Pradhan A., Van Kessel J.S., Karns J.S. *et al.* (2009) Dynamics of endemic infectious diseases of animal and human importance on three dairy herds in the northeastern US. *Journal of Dairy Science*, **92**, 1811–25.
- Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**, 559–75.
- Reddacliff L.A., Beh K., McGregor H. & Whittington R.J. (2005) A preliminary study of possible genetic influences on the susceptibility of sheep to Johne's disease. *Australian Veterinary Journal* **83**, 435–41.
- Stabel J.R. (2006) Host Responses to *Mycobacterium avium* supsp. *Paratuberculosis*: a complex arsenal. *Animal Health Research Reviews* **7**, 61–70.
- Takizawa S., Uchida T., Adur J., Kozakai T., Kotake-Nara E., Quan J. & Saida K. (2005) Differential expression of *endothelin-2* along the

- mouse intestinal tract. *Journal of Molecular Endocrinology* **35**, 201–9.
- Taylor K.H., Taylor J.F., White S.N. & Womack J.E. (2006) Identification of genetic variation and putative regulatory regions in bovine *CARD15*. *Mammalian Genome* **17**, 892–901.
- USDA – Animal and Plant Health Inspection Service (APHIS) (2008) *Johne's Disease in U.S. Dairies, 1991–2007*. USDA – APHIS, Washington DC, *USDA Info Sheet*, April, http://nahms.aphis.usda.gov/dairy/dairy07/Dairy2007_Johnes.pdf.
- Waters W.R., Stabel J.R., Sacco R.E., Harp J.A., Pesch B.A. & Wannemuehler M.J. (1999) Antigen-specific B-cell unresponsiveness induced by chronic *Mycobacterium avium* subsp. *paratuberculosis* infection of cattle. *Infection and Immunity* **67**, 1593–8.
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–78.
- Whitlock R.H., Rosenberger A.E., Sweeney R.W. & Spencer P.A. (1996) Distribution of *M. paratuberculosis* in tissues of cattle from herds infected with Johne's disease. In: *Proceedings of the Fifth International Colloquium on Paratuberculosis*, 29th September–4th October 1996 (Ed. by R.J. Chiodini, M.E. Hines & M.T. Collins), pp. 168–74. International Association for Paratuberculosis, Rehoboth, MA.
- Whitlock R.H., Wells S.J., Sweeney R.W. & Van Tiem J. (2000) ELISA and fecal culture for paratuberculosis (Johne's disease): sensitivity and specificity of each method. *Veterinary Microbiology* **77**, 387–98.
- Whittington R.J. & Sergeant E.S.G. (2001) Progress towards understanding the spread, detection and control of *Mycobacterium avium* subsp. *paratuberculosis* in animal populations. *Australian Veterinary Journal* **79**, 267–78.
- Wu C.-W., Livesey M., Schmoller S.K., Manning E.J.B., Steinberg H., Davis W.C., Hamilton M.J. & Talaat A.M. (2007) Invasion and persistence of *Mycobacterium avium* subsp. *paratuberculosis* during early stages of Johne's disease in calves. *Infection and Immunity* **75**, 2110–9.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Quantile–quantile (Q–Q) plot for *P*-values from a 1-d.f. test for association (allelic) vs. those expected for an association of loci with *Mycobacterium avium* subspecies *paratuberculosis* (*Map*)-infected tissue (tissue positive vs. tissue negative). The Q–Q plot shows little to no evidence of a deviation from the expected null distribution of *P*-values and therefore shows no evidence of population substructure.

Figure S2 Q–Q plot for *P*-values from a 1-d.f. test for association (allelic) vs. those expected for an association of loci with *Map* faecal-positive (faecal positive vs. faecal negative). The Q–Q plot shows evidence of a deviation from

the expected null distribution of *P*-values and therefore evidence of population substructure.

Figure S3 Q–Q plot for stratified test of association ($2 \times 2 \times 4$ CMH test) within-herd vs. expected for an association of loci with *Map* faecal-positive (faecal positive vs. faecal negative). The Q–Q plot shows little to no evidence of a deviation from the expected null distribution of *P*-values and therefore no evidence of population substructure.

Figure S4 Q–Q plot for 1-d.f. test for association (allelic) vs. expected for an association of loci with *Map* tissue-infected, faecal-negative (infected) cases vs. controls. The Q–Q plot shows little to no evidence of a deviation from the expected null distribution of *P*-values and therefore no evidence of population substructure.

Figure S5 Q–Q plot for *P*-values from a 1-d.f. test for association (allelic) vs. expected for an association of loci with the *Map* tissue-infected, faecal-positive ('shedding') cases vs. controls. The Q–Q plot shows some evidence of a deviation from the expected null distribution of *P*-values and therefore evidence of population substructure.

Figure S6 Q–Q plot for the stratified test of association ($2 \times 2 \times 4$ CMH test) within-herd vs. expected for an association of loci with the *Map* tissue-infected, faecal-positive ('shedding') cases vs. controls. The Q–Q plot shows little to no evidence of a deviation from the expected null distribution of *P*-values and therefore no evidence of population substructure.

Table S1 (a) *Map* tissue culture status by herd before quality control filtering of the data. Tissue-positive status was determined as at least one sample from any of four tissues yielding at least one colony forming unit of *Map* (peak CFU/g > 0). (b) *Map* faecal culture status by herd before quality control filtering of the data. Faecal-positive status was determined as at least one faecal sample containing at least one colony forming unit of *Map* (peak CFU/g > 0).

Table S2 (a) *Map* culture of tissue status by herd after quality control filtering of the data. Tissue-positive status was determined as at least one sample in any of four tissues containing at least one colony forming unit of *Map* (peak CFU/g > 0). (b) *Map* faecal culture status by herd after quality control filtering of the data. Faecal-positive status was determined as at least one faecal sample containing at least one colony forming unit of *Map* (peak CFU/g > 0).

Appendix S1 Methods utilized for quantile–quantile plots and MDS.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors.