bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 **Variations in MHC-DRB1 exon2 and associations with Brucellosis susceptibility**
2 **in Chinese Merino sheep**

3 Yue'e Chen[1, #], Wanyun Xu[1, #], Chuangfu Chen[1], Hugh T Blair[2], Jianfeng Gao[1]*

4 1 School of Life Sciences, Shihezi University, Xinjiang, 832003, PR China

5 2 Institute of Veterinary, Animal and Biomedical Sciences (IVABS), Massey University, New
6 Zealand

7 # These authors contributed equally to this work.

8 * **Address correspondence to:** Jianfeng Gao

9 School of Life Sciences, Shihezi University, North road 4, Shihezi, 832003, Xinjiang, PR
10 China. Tel: +86-0993-2031130, E-mail: jianfengg1689@163.com.

11 **Running title:** The relationship between MHC-DRB1 and Brucellosis susceptibility

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12 ## Abstract

13 The aim of this study was to investigated the association between MHC-DRB1 exon2 and
14 Brucellosis susceptibility in Chinese Merino sheep. MHC-DRB1 exon2 was amplified by
15 polymerase chain reaction (PCR) from 126 healthy and 67 Brucellosis-infected Chinese
16 Merino sheep. PCR products were analyzed using the SSCP technique, and then cloned to
17 allow sequencing of the different alleles. For each SNP, allelic and genotypic frequencies
18 were compared between case and control samples, in addition the association with Brucellosis
19 susceptibility was determined. Haplotypes and their frequencies were established and
20 analyzed by SHEsis online software. There were forty-one single nucleotide polymorphisms
21 (SNPs) in the 270 bp DNA sequence. The distribution of C>T alleles at locus 109 was
22 significantly different between case and control samples. The linkage disequilibrium (LD)
23 analysis showed that there were nine LD blocks in MHC-DRB1 exon2 and strong LD
24 between SNPs existed in every Block. Haplotype analysis identified nine haplotypes with
25 strong LD, but only Hap8 and Hap9 in case-control groups were significantly different
26 ($P<0.05$); neither haplotype contained the C>T allele at locus 109. In conclusion, genetic
27 variants of MHC-DRB1 gene exon2 demonstrated associations with Brucellosis susceptibility,
28 indicating that further research is warranted.

29 **Keywords:** MHC-DRB1, Chinese Merino, SNPs, Brucellosis, Susceptibility, Haplotypes

30 **Abbreviations:** MHC: Major Histocompatibility Complex; OLA: Ovine Lymphocyte
31 Antigen; SNP: Single Nucleotide Polymorphism

32

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

33 **Introduction**

34 Brucellosis, caused by the *Brucella* genus, is a widespread, chronic, zoonotic disease of
35 animals and humans (Corbel, 1997). Xinjiang Province in Western China is an important
36 source of production animals, including sheep, cattle, goats and deer. Chinese Merino sheep
37 are one of the most populous breeds in Xinjiang. The presence of *Brucella* is serious, not only
38 because of the harmful effect on human health, but also because infected sheep have lower
39 production. Consequently, there is interest in finding effective means to both prevent *Brucella*
40 infections and to reduce the presence of the *Brucella* organism in Xinjiang. *Brucella* is
41 considered a major health problem requiring urgent action.

42 The Major Histocompatibility Complex (MHC) is a genetic region comprised of clustered
43 genes which control the regulation of immune response (Ji et al., 2010). Two groups of
44 cell-surface glycoproteins, termed Class I and Class II molecules, are the major
45 MHC-encoded effector molecules and they are intimately involved in T- and
46 B-lymphocyte-mediated immune reactions. The ovine MHC, or Ovine Lymphocyte Antigen
47 (OLA), harbors clusters of immunological genes involved in overall resistance/susceptibility
48 of animals to infectious diseases (Danchin et al., 2004; Flajnik and Kasahara, 2001; Kaufman,
49 2002). The ovine DRB1 locus is located in the MHC Class II region, and its function is to
50 present extracellular-derived peptides to the immune system. To date, associations between
51 ovine Class II genetic markers and disease susceptibility have been limited to the MHC Class
52 II *Ovar*-DRB1 locus. Reported relationships in sheep include bovine leukemia virus disease,
53 Maedi-Visna and hydatidosis (Larruskain et al., 2010; Li et al., 2011). The majority of

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

54 nucleotide polymorphisms in Class II loci locate to exon2, consequently, this region has been
55 the principal target in genotyping studies designed to associate MHC genetic diversity with
56 susceptibility to disease (Fallin et al., 2001; Konnai et al., 2003).

57 Single Nucleotide Polymorphisms (SNPs) belong to the third generation molecular marker
58 technology. Each non-synonymous SNP (nsSNP) changes one amino acid in the gene product
59 causing Single Amino-acid Polymorphisms (SAP) (Schaefer et al., 2012). SNPs are useful
60 genetic markers and can assist in searching for genetic risk factors associated with complex
61 diseases. The application of SNPs in the recognition and identification of disease
62 susceptibility genes has become a key research area. Following completion of the Human
63 Genome Project in 2003 and the International Human Genome Haplotype Map (HapMap) in
64 2005, the HapMap offered an important additional tool to discover genetic variants associated
65 with diseases. Research using SNPs and haplotypes will play an important role in exploring
66 genetic and pathogenic mechanisms of complex diseases.

67 Individual SNPs play only a minor role in explaining the genetic variation in complex
68 diseases. However, the genetic information provided by haplotypes is more useful in
69 describing the polygenic nature of genetic diseases (Zaykin et al., 2002). As a consequence,
70 haplotype analysis is increasingly becoming the preferred method for the study of complex
71 diseases (Dukkipati et al., 2006; Sayers et al., 2005). If a specific haplotype can be shown to
72 be distributed differently between case and control samples, this provides strong evidence that
73 the haplotype is associated with the disease, and provides a shortcut for polygenic disease
74 research. While haplotype analysis has been applied to study of the human MHC (Yang et al.,

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

75 2004; Zhang et al., 2003), this research technique has not been widely applied to investigate
76 the association between animal MHC haplotypes and disease susceptibility.

77 The purpose of this study was to investigate associations between SNPs and haplotypes of the
78 MHC-DRB1 exon2 in Chinese Merino sheep and Brucellosis susceptibility.

79 **Results**

80 The Rose Bengal Plate Agglutination Test (RBPT) was used to detect Chinese Merino sheep
81 that expressed antibodies to *Brucella* and were thus considered positive for *Brucella* infection
82 (Figure 1). Out of 193 sheep tested, 126 (65%) were negative for *Brucella*, and 67 (35%)
83 tested positive.

84 PCR amplification products of MHC-DRB1 exon2 were examined via SSCP electrophoresis
85 (Figure 2). Due to the PCR-SSCP method having high sensitivity, many SNP genotypes were
86 detected. Some genotypes were only detected in one individual, and it would be useful to
87 expand the number of sheep sampled to be confident these low frequency SNPs are real.

88 In this study, the PCR-SSCP technique and sequence alignment results demonstrated that
89 MHC-DRB1 gene exon2 of Chinese Merino sheep was richly polymorphic. Sequence
90 alignments were analyzed by GeneDoc software, and the results of alignment are shown in
91 Figure 3. A total of forty-one SNPs were identified in the 270bp length. SNP allelic and
92 genotypic frequencies for case and control groups are shown in Table 1 and Table 2,
93 respectively. The Hardy-Weinberg equilibrium tests showed no significant differences
94 (P>0.001).

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

95 Insertion and deletion sites did not exist in these SNPs. There were nine sites belonging to the
96 PIP and thirty-two belonging to the SP. There were 18 SP conversion sites , including 7 A/G
97 and 11 T/C, and 14 transversion sites, including 1 C/A, 6 G/T, 4 C/G and 3 A/T. The
98 conversion rate was higher than the transversion rate because the cytosine residue CpG
99 dinucleotide is the most frequently mutated sites in the genome. Most of them are methylated,
100 and can spontaneously deaminate resulting in the formation of thymine, which is consistent
101 with the reported results.

102 The sequence length of MHC-DRB1 exon2 is 270bp, and it can be translated into 89 amino
103 acids. Amino acid sequence analysis showed there were twenty-three mutations causing
104 amino acid changes (Table 3), eight nonsense-mutations and three sense-mutations. For
105 Chinese Merino sheep MHC-DRB1 exon2 of allelic gene sequences of amino acid homology
106 analysis, the homology was more than 81%. The results reflected the consistent with
107 MHC-DRB1 nucleotide polymorphism loci.

108 The MHC-DRB1 exon2 sequences of different sheep breeds were aligned through GenDoc
109 software. The results (Table 4) demonstrated that in different studies, the number of SNPs and
110 SAPs were different. However, in the majority of studies between 50% and 70% of SNPs
111 caused amino acid changed. The current study was at the upper limit of this range whereby
112 30 out of 41 (73%) SNPs resulted in amino acid changes.

113 In this study, the polymorphism of MHC-DRB1 gene exon2 associated with Brucellosis
114 susceptibility, the result of analysis found that MHC-DRB1 was likely to be one of the genes
115 associated with Brucellosis susceptibility, C>T alleles at the 109 locus in the case-control

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

116 samples distribution existed significant difference (*P*<0.05), and preliminary analysis
117 suggested that MHC-DRB1 exon2 109 C>T associated with Brucellosis susceptibility.
118 Association analysis was conducted for each genotype of the gene polymorphisms, showing
119 that the site of DD/Dd/dd genotype in the case-control samples distribution had no significant
120 difference (*P*>0.05).

121 Haplotypes generally had more information content than individual SNPs. Therefore, we
122 performed linkage disequilibrium(LD) and haplotype analysis for the SNPs with MAFs<5%
123 and the genotype distributions were in Hardy-Weinberg equilibrium (*P*<0.001) in Chinese
124 Merino sheep. The standardized measure of LD denoted as D' were calculated for all pairs of
125 SNPs. Among forty-one SNPs in MHC-DRB1 exon2, only twenty-nine SNPs were eligible,
126 and were used to analyze LD in both case and control, the haplotypes LD map as shown in
127 Figure 4, indicated the LD of the SNPs, the number in box such as 99 is 0.99, the greater D'
128 value means the stronger LD degree between each two SNPs, D'>0.9 is usually considered
129 highly LD, and D'>0.7 is deemed to the two SNPs located in same block, which can perform
130 haplotypes analysis. The block diagram of color from light to deep (white and red), said the
131 LD degree from low to high, deep red means completely linkage. LD analysis found that the
132 MHC-DRB1 exon2 had nine LD Blocks (Table 5), and each two SNPs had strong LD in
133 every Block.

134 Due to a series of genetic diseases are often not caused by a single SNP loci, but by the
135 combination of SNPs on several sites, therefore, in the study of disease association analysis,
136 based on the multiple sites of SNPs studies tend to have more strength and more convincing

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

137 than a single SNP loci. Because haplotype frequency cannot less than 0.03 in a population, the
138 haplotype combination which frequency yet reached 0.03 should be ignored during statistics,
139 so that only nine haplotype combinations were analyzed in the MHC-DRB1 exon2 (Table 6),
140 the result of analysis found that Hap8 and Hap9 these two haplotypes frequency in the case
141 group was 12.5% and 15% respectively, was significantly higher than control group,
142 haplotype frequency distribution difference was statistically significant (P<0.05), initially
143 speculated that these two haplotypes may associated with Brucellosis susceptibility.

144 **Discussion**

145 A high degree of polymorphism and base mutation are a very prominent feature of vertebrate
146 MHC genes, having become a hot research topic in livestock resistance breeding. Forty-one
147 SNPs in MHC-DRB1 gene exon2 of Chinese Merino sheep observed in this study are newly
148 identified and have not been reported previously. These results confirmed that PCR-SSCP is a
149 useful tool for easy and efficient identification of DNA polymorphisms and can be employed
150 for evaluating genetic variability in large livestock populations. Furthermore, according to the
151 previously study about SNPs and SAPs, the results fully demonstrated that MHC-DRB1 gene
152 of Chinese Merino sheep is really highly polymorphic.

153 Compared with other varieties sheep polymorphic in MHC-DRB1 gene exon2, there were a
154 lot of differences in different breeds of sheep. The discrepancy in different populations was
155 probably caused by the following reasons: one the one hand, some polymorphisms can only
156 be preserved and inherited in ancient and special animals. On the other hand, different
157 selection purpose and selection history lead to the diversity.

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

158 The main function of MHC is antigen presentation, playing a very important role in the
159 immune system of animal body, and it has a close relationship with the livestock disease
160 susceptibility. Many disease susceptibility are associated with MHC, as a candidate gene for
161 disease susceptibility has become one of the research hotspots in modern molecular immune
162 genetic. Through the study of the association between SNPs and the Brucellosis susceptibility,
163 the 109 C>T was found to be significantly associated with Brucellosis susceptibility, which
164 provide scientific basis for choice Brucellosis susceptibility molecular marker-assisted,
165 genetic loci for further search with susceptibility and lay the foundation for genetic breeding,
166 but about whether the 109 C>T as genetic markers of Brucellosis, also need to attack toxic
167 experiment will be carried out to further verify. After further verification, this SNP could be a
168 useful molecular marker for use in poultry breeding.

169 In case and control association studies, when LD exists in each two SNP, haplotypes
170 generally had more information content than individual SNPs. This study calculated the
171 degree of LD and inferred haplotype using SHEsis online software. Because of LD of SNPs in
172 very close genetic distance, but in order to successfully identify may lead to changes in the
173 disease, many relational analysis method using multiple high density SNPs sites to
174 study(Epstein and Satten, 2003), on the other hand, haplotype is just a collection of closely
175 linked SNPs on the same chromosome allele, it contains a number of pairwise LD information,
176 so the gene location problem, based on haplotype association are more powerful than based
177 on single SNP loci analysis (Zaykin et al., 2002).

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

178 The haplotype analysis found that nine haplotypes were constructed by using twenty -nine
179 SNPs, if LD did not exist between each two SNPs, twenty-nine SNPs should produce 2^{29}
180 haplotypes, but only nine haplotypes were detected among these SNPs in this study because
181 of strong LD. In addition, we also investigated haplotype frequencies between case and
182 control in Chinese Merino sheep, and speculated that Hap8 and Hap9 may associated with the
183 Brucellosis susceptibility which may relate to immune traits association studies and lay the
184 foundation for molecular marker assisted selection. Further studies as well as functional
185 analysis are required to fully elucidate how these interesting gene polymorphisms may affect
186 MHC-DRB1 gene exon2 activity and/or probably act as candidate markers associated with
187 Brucellosis susceptibility in sheep.

188 Materials and Methods

189 Sample collection and DNA isolation

190 Blood samples were obtained from 193 Chinese Merino sheep farmed by the Xinjiang
191 Production and Construction Corps Agricultural 9th Division 170 Regiment. Genomic DNA
192 was isolated from whole blood samples with the use of phenol chloroform extraction, and
193 stored at -20°C for later use.

194 Detection of *Brucella*

195 The Rose Bengal Plate Agglutination Test (RBPT) was used to detect *Brucella* antibodies. A
196 serum sample of 30μl was evenly mixed with 30 μl antigen at room temperature, and reaction
197 results were recorded after 4 to 10 min, results were immediately compared to positive serum.

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

198 Serum agglutination with any degree of granularity or flocculation was designated as positive,

199 while no agglutination was designated as negative.

200 **Amplification of OLA-DRB1 exon2**

201 Exon2 of OLA-DRB1 (Accession Number: FR848372) was amplified using PCR with the

202 following oligo nucleotide primers: DRB1-1: 5' TAT CCC GTC TCT GCA GCA CAT TTC

203 3' and DRB1-2: 5' CTC GCC GCT GCA CAC TGA AAC TCT 3'. PCR was performed in a

204 reaction volume of 20 µl containing 50 ng/µl ± 4 ng/ul genomic DNA, 10 µl 2×PCR Master

205 Mix, 1 µl of each primer (10 umol/L) and 7 µl ddH_2O. The PCR was undertaken in a

206 Mastercycler gradient thermocycler (Eppendorf China Limited) under the following

207 conditions: denaturation at 94°C for 5 min, followed by 94°C for 30 s, 63°C for 1 min, and

208 72°C for 1 min for 30 cycles, and a final extension at 72°C for 10 min. The amplified

209 fragment of 286 bp consisted of 16 bp of intron1 and the entire exon2 of 270 bp. An aliquot of

210 5 µl reaction product was used to check the concentration and quality of the PCR products by

211 agarose gel electrophoresis.

212 **SSCP (Single-strand conformational polymorphism) Analysis**

213 Each PCR product (4 µl) was mixed well with 9 µl of denaturing solution (95% formamide,

214 20 mM EDTA, 0.05% bromophenol blue and 0.05% xylene-cyanole). Mixtures were

215 submitted to a denaturation step at 95°C for 5 min, and then rapidly chilled on ice. Ten µl of

216 the denatured mixture was directly electrophoresed on precast 10% (Acr : Bis=39:1)

217 non-denaturing polyacrylamide gels for SSCP. After electrophoresis, one glass plate was

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

218 removed and the gel on the second glass plate was stained in 0.1% $AgNO_3$ for 8 min, and then
219 briefly rinsed with distilled water two times, for 20 s each time. Gels were developed in 400
220 ml containing 1.2% (w/v) NaOH and 0.4% formaldehyde. When the desired band intensity
221 was achieved, development was stopped. Stained gels were rinsed with distilled water and
222 gels were scanned using an X-ray lamp.

223 **Cloning and sequencing**

224 Following PCR-SSCP analysis, PCR amplification products were recycled and purified by an
225 agarose gel purification kit to enable identification of different genotypes for each individual.
226 Recycled DNA fragments were attached to the pMD19-T vector, and transformed into TOP10
227 E.coli competent cells on LB plates coated with Amp. Clones were inoculated onto the culture
228 medium containing Amp and allowed to culture. Thus, the bacteria were used as a template
229 for PCR amplification to identify recombinant clones. If they were amplified completely
230 consistently with the objective fragment, they were identified as positive clones. Bacteria
231 containing positive clones were sent to BGI, Beijing, for sequencing. Sequence results were
232 analyzed using DNASTAR and GeneDoc software.

233 **Data processing and statistical analysis**

234 The genotypic frequency of SNPs from the MHC-DRB1 gene exon2 were analyzed for
235 Hardy-Weinberg equilibrium (HWE) in case and control Chinese Merino sheep using
236 chi-squared with a significance level of P=0.001. Comparisons of allelic and genotypic

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

237 frequencies in case-control groups and association analysis between the polymorphisms and
238 Brucellosis susceptibility were undertaken using SHEsis software.

239 **Linkage disequilibrium and Haplotype analysis**

240 SNPs whose minor allele frequencies (MAFs) were <5%, failed genotyping due to technical
241 errors, or failed to meet Hardy-Weinberg equilibrium (P<0.001) were removed from the
242 haplotype structure. The pattern of Linkage Disequilibrium (LD) between the SNPs was
243 measured by the LD coefficient D'=D/Dmax., The magnitude of LD between matching sites
244 was used to indicate whether or not LD existed. The value of D' was calculated by the online
245 genetics software SHEsis. This software tests the HWE of the sample population, analyses
246 LD of large samples while testing for multiple SNP loci at the same time and constructs
247 haplotype associations (Li et al., 2009; Yong and Lin, 2005).

248 **Acknowledgments**

249 We thank everyone from the Xinjiang Production and Construction Corps Agricultural 9th
250 Division, 170 Regiment, who provided samples and their participation and support for this
251 work. We are also warmly appreciative to Ma Runlin Professor in the Institute of Genetics
252 and Developmental Biology, Chinese Academy of Sciences, for his helpful advice and thank
253 BGI, Beijing, for providing the sequencing platform.

254 **Conflict of interest:** There are not any conflicts of interests.

255

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

256 **References**

257 Corbel, M.J. 1997. Brucellosis: an overview. *Emerg Infect Dis*. 3:213.

258 Danchin, E., V. Vitiello, A. Vienne, O. Richard, P. Gouret, M.F. McDermott, and P.
259 Pontarotti. 2004. The major histocompatibility complex origin. *Immunol Rev*.
260 198:216-232.

261 Dukkipati, V., H. Blair, D. Garrick, and A. Murray. 2006. 'Ovar-Mhc '—Ovine major
262 histocompatibility complex: Role in genetic resistance to diseases. *New Zeal Vet J*.
263 54:153-160.

264 Epstein, M.P., and G.A. Satten. 2003. Inference on haplotype effects in case-control studies
265 using unphased genotype data. *Am J Hum Genet*. 73:1316-1329.

266 Fallin, D., A. Cohen, L. Essioux, I. Chumakov, M. Blumenfeld, D. Cohen, and N.J. Schork.
267 2001. Genetic analysis of case/control data using estimated haplotype frequencies:
268 application to APOE locus variation and Alzheimer's disease. *Genome Res*.
269 11:143-151.

270 Flajnik, M.F., and M. Kasahara. 2001. Comparative genomics of the MHC: glimpses into the
271 evolution of the adaptive immune system. *Immunity*. 15:351-362.

272 Kaufman, J. 2002. The origins of the adaptive immune system: whatever next? *Nature*
273 *Immunol*. 3:1124-1125.

274 Konnai, S., Y. Nagaoka, S. Takeshima, M. Onuma, and Y. Aida. 2003. Sequences and
275 diversity of 17 new Ovar-DRB1 alleles from three breeds of sheep. *Eur J*
276 *Immunogenet*. 30:275-282.

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

277 Larruskain, A., E. Minguijón, K. García-Etxebarria, B. Moreno, I. Arostegui, R.A. Juste, and
278 B.M. Jugo. 2010. MHC class II DRB1 gene polymorphism in the pathogenesis of
279 Maedi–Visna and pulmonary adenocarcinoma viral diseases in sheep.
280 *Immunogenetics*. 62:75-83.
281 Li, R., W. Hui, B. Jia, G. Shi, Z. Zhao, H. Shen, Q. Peng, L. Lv, Q. Zhou, and H. Li. 2011.
282 The relationship between MHC-DRB1 gene second exon polymorphism and
283 hydatidosis resistance of Chinese merino (Sinkiang Junken type), Kazakh and
284 Duolang sheep. *Parasite: journal de la Société Française de Parasitologie*. 18:163.
285 Li, Z., Z. Zhang, Z. He, W. Tang, T. Li, Z. Zeng, L. He, and Y. Shi. 2009. A
286 partition-ligation-combination-subdivision EM algorithm for haplotype inference
287 with multiallelic markers: update of the SHEsis (http://analysis. bio-x. cn). *Cell Res*.
288 19:519-523.
289 Sayers, G., B. Good, J. Hanrahan, M. Ryan, J. Angles, and T. Sweeney. 2005. Major
290 histocompatibility complex DRB1 gene: its role in nematode resistance in Suffolk
291 and Texel sheep breeds. *Parasitology*. 131:403-409.
292 Schaefer, C., A. Meier, B. Rost, and Y. Bromberg. 2012. SNPdbe: constructing an nsSNP
293 functional impacts database. *Bioinformatics*. 28:601-602.
294 Yang, S., H.S. Ge, A.P. Zhang, S.C. Wei, M. Gao, H.Y. Wang, J.J. Chen, M. Li, Y.H. Liang,
295 and P.P. He. 2004. Haplotype associations of the MHC with psoriasis vulgaris in
296 Chinese Hans. *Clin Exp Dermatol*. 29:399-405.

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

297 Yong, Y., and H. Lin. 2005. SHEsis, a powerful software platform for analyses of linkage
298 disequilibrium, haplotype construction, and genetic association at polymorphism loci.
299 *Cell Res*. 15:97-98.
300 Zaykin, D.V., P.H. Westfall, S.S. Young, M.A. Karnoub, M.J. Wagner, and M.G. Ehm. 2002.
301 Testing association of statistically inferred haplotypes with discrete and continuous
302 traits in samples of unrelated individuals. *Hum Hered*. 53:79-91.
303 Zhang, X.-J., A.-P. Zhang, S. Yang, M. Gao, S.-C. Wei, P.-P. He, H.-Y. Wang, Y.-X. Song, Y.
304 Cui, and J.-J. Chen. 2003. Association of HLA class I alleles with psoriasis vulgaris
305 in southeastern Chinese Hans. J Dermatol Sci. 33:1-6.

306

307

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

308 ## Figure Legends

309 **Figure 1.** Detection results of Brucella. 1-3: The positive Brucella. 0: Positive control.

310 **Figure 2.** PCR-SSCP detection results of MHC-DRB1 exon2. M: DNA Marker DL2000.
311 1-17: Different individuals DRB1 exon2 of genotype.

312 **Figure 3.** Nucleotide sequence alignment of MHC-DRB1 gene exon2 alleles from Chinese
313 Merino Sheep. Note: The corresponding blank figures represent no polymorphic locus and
314 omit the sequence in the figure.

315 **Figure 4.** The figure of linkage disequilibrium Haplotypes based case-control sample. Note:
316 The first line number indicates the location of SNPs; and the second line of figures indicate
317 the number of SNPs, a total of 29. The number in graph box is the value of linkage
318 disequilibrium(D'), D' values greater the darker, which means that the higher the degree of
319 linkage disequilibrium. Conversely represents the lower the degree of linkage disequilibrium.

320

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

321 # Tables

322 **Table 1.** Allele frequencies of PCR-SSCP products about MHC-DRB1 exon2

Loci	SNP type	Case Allele Frequency	Control Allele Frequency	*P* value
15	G/C	0.375/0.625	0.500/0.500	0.259777
27	A/C	0.775/0.225	0.700/0.300	0.445885
28	A/C/T	0.350/0.175/0.475	0.375/0.250/0.375	0.596163
62	C/A/G	0.650/0.300/0.050	0.775/0.175/0.050	0.415949
63	T/G	0.725.0.275	0.725/0.275	1.000000
64	C/T	0.725/0.275	0.825/0.175	0.284171
67	C/T	0.475/0.525	0.375/0.625	0.365639
70	G/T	0.775/0.225	0.725/0.275	0.605600
72	C/T/G	0.500/0.350/0.150	0.675/0.200/0.125	0.250346
73	T/C	0.950/0.050	0.975/0.025	0.556225
74	C/G	0.700/0.300	0.675/0.325	0.809387
75	C/T	0.900/0.100	0.875/0.125	0.723478
76	G/T	0.675/0.325	0.725/0.275	0.625608
85	T/A/G	0.325/0.475/0.200	0.325/0.325/0.350	0.251407
86	C/G/A	0.425/0.450/0.125	0.650/0.250/0.100	0.117623
104	G/A	0.925/0.075	0.950/0.050	0.644189
106	A/G	0.975/0.025	0.900/0.100	0.165840
109	C/T	0.325/0.675	0.575/0.425	0.024641*
114	C/T	0.175/0.825	0.350/0.650	0.075293
115	C/T	0.750/0.250	0.625/0.375	0.227779
117	G/C/T	0.775/0.200/0.025	0.650/0.300/0.050	0.455680
140	C/T	0.375/0.625	0.425/0.575	0.648098

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

144	A/T	0.425/0.575	0.475/0.525	0.653116
159	C/T	0.900/0.100	0.875/0.125	0.723478
171	A/G	0.950/0.050	0.900/0.100	0.395907
172	G/C	0.050/0.950	0.025/0.975	0.556225
174	A/G/T	0.075/0.050/0.875	0.125/0.050/0.825	0.756228
175	A/T	0.950/0.050	0.975/0.025	0.556225
190	A/G	0.825/0.175	0.775/0.225	0.576173
200	G/T	0.675/0.325	0.850/0.150	0.065913
206	G/C	0.800/0.200	0.800/0.200	1.000000
207	A/T	0.825/0.175	0.800/0.200	0.774538
227	A/G	0.075/0.925	0.025/0.975	0.304887
231	A/G	0.725/0.275	0.575/0.425	0.159583
237	C/T	0.575/0.425	0.475/0.525	0.370490
245	G/T	0.700/0.300	0.650/0.350	0.633092
246	C/T	0.650/0.350	0.600/0.400	0.644189
249	G/T	0.225/0.775	0.225/0.775	1.000000
252	G/T/C	0.700/0.300/0.000	0.525/0.450/0.025	0.201897
253	A/G/T/C	0.225/0.350/0.275/0.150	0.275/0.450/0.225/0.050	0.407304
262	A/G	0.925/0.075	0.925/0.075	1.000000

323 Note: the case group and normal control group of Chinese Merino sheep MHC-DRB1 alleles,

324 *$P<0.05$

325 **Table 2.** Genotype frequencies of PCR-SSCP products about MHC-DRB1 exon2

Loci	AH	AL	Genotype frequency in case	Genotype frequency in control	*P* value
G15C	C	G	0.550/0.150/0.300	0.350/0.300/0.350	0.374 222
A27C	A	C	0.700/0.150/0.150	0.500/0.400/0.100	0.208 108

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

T63G	T	G	0.550/0.350/0.100	0.450/0.550/0.000	0.213 430
C64T	C	T	0.550/0.350/0.100	0.650/0.350/0.000	0.338 465
C67T	T	C	0.300/0.450/0.250	0.450/0.350/0.200	0.618 440
G70T	G	T	0.550/0.450/0.000	0.450/0.550/0.000	0.527 109
T73C	T	C	0.900/0.100/0.000	0.950/0.050/0.000	0.548 327
C74G	C	G	0.500/0.400/0.100	0.450/0.450/0.100	0.945 797
C75T	C	T	0.800/0.200/0.000	0.800/0.150/0.050	0.564 718
G76T	G	T	0.500/0.350/0.150	0.550/0.350/0.100	0.883 548
G104A	G	A	0.850/0.150/0.000	0.900/0.100/0.000	0.632 607
A106G	A	G	0.950/0.050/0.000	0.850/0.100/0.050	0.485 672
C109T	T	C	0.500/0.350/0.150	0.400/0.350/0.250	0.139 499
T114C	T	C	0.700/0.250/0.050	0.450/0.150/0.400	0.249 167
C115T	C	T	0.600/0.300/0.100	0.450/0.350/0.200	0.556 504
C140T	T	C	0.350/0.550/0.100	0.350/0.450/0.200	0.648 344
T144A	T	A	0.350/0.450/0.200	0.250/0.550/0.200	0.765 928
C159T	C	T	0.800/0.200/0.000	0.750/0.250/0.000	0.704 969
A171G	A	G	0.950/0.000/0.050	0.850/0.100/0.050	0.347 999
C172G	C	G	0.900/0.100/0.000	0.950/0.050/0.000	0.548 327
A175T	A	T	0.900/0.100/0.000	0.950/0.050/0.000	0.548 327
A190G	A	G	0.700/0.250/0.050	0.650/0.250/0.100	0.830 950
G200T	G	T	0.550/0.250/0.200	0.800/0.100/0.100	0.237 128
G206C	G	C	0.600/0.400/0.000	0.650/0.300/0.050	0.515 377
A207T	A	T	0.700/0.250/0.050	0.650/0.300/0.050	0.938 030
G227A	G	A	0.900/0.050/0.050	0.950/0.050/0.000	0.598 389
A231G	A	G	0.600/0.250/0.150	0.400/0.350/0.250	0.441 902
C237T	C	T	0.350/0.450/0.200	0.350/0.250/0.400	0.289 936

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

G245T	G	T	0.600/0.200/0.200	0.500/0.300/0.200	0.747 584
C246T	C	T	0.600/0.100/0.300	0.450/0.300/0.250	0.283 728
T249G	T	G	0.700/0.150/0.150	0.650/0.250/0.100	0.691 758
A262G	A	G	0.900/0.050/0.050	0.850/0.150/0.000	0.362 661

326 Note: the genotype representation: DD/Dd/dd (D for high frequency allele; D for high

327 frequency allele). AL: low frequency allele name; AH: high frequency allele name

328 **Table 3.** Variable sites and amino acid changes of Chinese Merino sheep MHC-DRB1 exon2

Number	SNP type	AA change	Number	SNP type	AA change
1	15G>C	His/Gln	22	140C>T	Leu/Pro
2	27A>C	His/Gln	23	144A>T	Gly
3	28A>C>T	Arg/-	24	159C>T	Arg
4	62C>A>G	Pro/His	25	171A>G	Ala
5	63T>G	Pro/His	26	172G>C	Arg
6	64C>T	Arg/Trp	27	174A>G>T	Arg
7	67C>T	Pro/Ser	28	175A>T	Thr/-
8	70G>T	Ala/Ser	29	190A>G	Arg/Gly
9	72C>T>G	Ala/Ser	30	200G>T	Cys/Phe
10	73T>C	Cys/Ser	31	206G>C	Gly/Ala
11	74C>G	Cys/Ser	32	207A>T	Gly/Ala
12	75C>T	Cys/Ser	33	227A>G	Arg/His
13	76G>T	Ala/Ser	34	231A>G	Tyr/-
14	85T>A>G	Ser/Ala	35	237C>T	Asn
15	86C>A>G	Ser/Ala	36	245G>T	Leu/Arg
16	104G>A	Ser/Asn	37	246C>T	Leu/Arg
17	106A>G	Thr/Ala	38	249G>T	Ser

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

18	109C>T	Trp/Arg	39	252G>T>C	Phe/Arg
19	114C>T	Arg	40	253A>G>T>C	Asp/Tyr/Asn
20	115C>T	Pro/Ser	41	262A>G	Gly/Arg
21	117G>C>T	Pro/Ser			

329 “ - ”：means termination codon

330 **Table 4.** The number of SNP and amino acid mutation loci in different sheep breeds

Breeds	Authors	Submit date	SNPs No.	SAPs No.	SAPs/SNPs
Arta	Theodorou,G., Spetsarias,S.	2012/6/25	34	21	0.618
Bergamasca	Ballingall,K.T., Steele,P.J.,	2010/4/28	23	16	0.696
Blue Faced Leicester	Stear,M.J.,Stirling,D.	2008/9/16	25	17	0.680
Kozani	Theodorou,G., Spetsarias,S.	2012/6/25	17	12	0.706
Lori Bakhtiari	Nikbakht Brujeni,G.,Rezaii,H.	2010/9/2	25	15	0.600
Prealpe	Ballingall,K.T., Tassi,R.,	2009/5/12	34	19	0.559
Red Maasai	Ballingall,K.T., Steele,P.	2010/8/25	43	26	0.605
Scottish Blackface	Stirling,D., Stear,M.J.	2008/9/16	58	31	0.534
Sopravvisana	Ballingall,K.T., Steele,P.J.,	2010/4/28	19	10	0.526
Zandi	Nikbakht Brujeni,G., Rezaii,H.	2010/9/2	19	7	0.368

331 **Table 5.** The composition of every LD Block with different SNPs

LD Block	SNPs
Block1	T63G、C64T
Block2	C64T、C67T、G70T
Block3	G70T、C74G
Block4	G76T、G104A
Block5	C109T、C114T、C115T

bioRxiv preprint doi: https://doi.org/10.1101/038539; this version posted February 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Block6	C140T、A144T、C159T
Block7	A171G、G172C、A190G、G200T
Block8	A207T、A227G、A231G
Block9	G245T、C246T、G249T

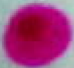
332 **Table 6.** The Haplotypes frequencies of SNPs in case-control MHC-DRB1 exon2

	Haplotypes	Case(freq.)	Control(freq.)	Chi2	*P* value
Hap1	CATCCGCCGGCCTCACACAGGAGATGCTA	2.00(0.050)	1.00(0.025)	0.139	0.708833
Hap2	CATCTGCCGGCCTCACACAGGAGATGCTA	0.00(0.000)	2.00(0.050)	2.822	0.093002
Hap3	CATCTGCCGGTTCTTCGCGGGAGGCTTTA	2.00(0.050)	0.50(0.013)	0.379	0.538296
Hap4	CATTCGCCGGTTCTTCACAGGTGGCGCTA	0.00(0.000)	2.00(0.050)	2.822	0.093002
Hap5	CCTTCGCCGGTTCTTCACAGGAGACTTTA	0.00(0.000)	2.00(0.050)	2.822	0.093002
Hap6	GATCTGCCGGTTCTTCACAGGAGATGCGA	0.00(0.000)	2.00(0.050)	2.822	0.093002
Hap7	GATCTGGCTGCCTCACACAGGAGATTTTA	0.00(0.000)	2.00(0.050)	2.822	0.093002
Hap8	CATCTGCCGGTTCTTCACAGGAGATGCGA	5.00(0.125)	0.00(0.000)	4.725	0.029752
Hap9	CATTCGCCGGTTCTTCACATGAGACGCTA	6.00(0.150)	0.00(0.000)	5.946	0.014769

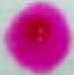
333

0

1



2



3

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

2000bp

1000bp

750bp

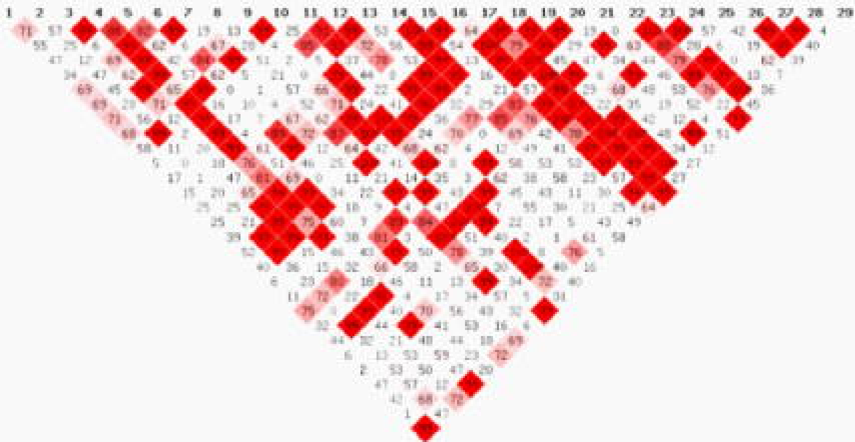
500bp

250bp

1-25 30-60 88-102 112-133 146-188 209-225 263-270

80 140 200 240 260

7(2) 1 270
7(5) 1 270
15(1) 1 270
15(2) 1 270
15(5) 1 270
21(2) 1 270
21(5) 1 270
31(1) 1 270
31(3) 1 270
49(1) 1 270
49(2) 1 270
49(3) 1 270
57(1) 1 270
57(2) 1 270
112(1) 1 270
112(2) 1 270
112(3) 1 270
113(1) 1 270
113(2) 1 270
113(3) 1 270
124(1) 1 270
124(3) 1 270
200(1) 1 270
200(2) 1 270
228(4) 1 270
228(4) 1 270
204(1) 1 270
204(2) 1 270


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29