Allelic polymorphism in the ovine DQA1 gene

H. Zhou and J. G. H. Hickford

Animal and Food Sciences Division, Lincoln University, Canterbury, New Zealand

ABSTRACT: Variation in the ovine DQA1 gene was investigated by amplification of exon 2 using PCR, followed by single-strand conformational polymorphism (SSCP) analysis, cloning, and DNA sequencing. Fourteen novel SSCP patterns, representing 14 different sequences, were identified. Eight of these 14 sequences were identical to published DQA1 sequences from sheep, whereas the remaining six were novel but similar to the published DQA1 sequences from sheep and cattle. These six new sequences exhibited conserved region and variable region patterns similar to the published sheep DQA1 sequences, but were different than the published DQA2 sequences from sheep. All of these 14 putative sheep DQA1 sequences fulfilled the criteria used by the established bovine leukocyte antigens major histocompatibility complex nomenclature committee for assignment as new alleles. Comparison of the available DQA1 sequences from sheep and cattle revealed several clusters of ovine DQA1 sequences, and some sheep alleles were more similar to cattle alleles than other sheep alleles. The occurrence of trans-species polymorphism suggests the action of balancing selection at the DQA1 locus. Twenty-four percent of the nucleotide positions showed variation within exon 2, and this variation seems to have arisen largely by point mutation and gene conversion. The nonsynonymous and synonymous substitution rates were similar in both the putative antigen-binding site codons and the putative nonantigen-binding site codons. The extensive polymorphism reported in this article is consistent with polymorphism reported at the bovine DQA1 locus.

Key Words: DQA1, Major Histocompatibility Complex, Polymorphism, Sheep, Single-Strand Conformational Polymorphism

©2004 American Society of Animal Science. All rights reserved. J. Anim. Sci. 2004. 82:8–16

Introduction

The class II molecules of the major histocompatibility complex (MHC) present fragments of predominantly exogenous antigens (Ag) to CD4+ T-lymphocytes (Germain and Margulies, 1993). The genes encoding these molecules are highly polymorphic. Variation in these genes may impact on immune responses to pathogens, which may lead to variation in disease susceptibility (Racioppi et al., 1991).

The class II MHC gene region of sheep has an organization similar to that of humans (Scott et al., 1987). Within this gene region, two subregions, DR and DQ, exhibit high levels of polymorphism (Amills et al., 1998). In contrast to humans, high levels of polymorphism are observed in the DQ subregion (Escayg et al., 1996). This, together with the absence of a functional DP subregion (Scott et al., 1987), suggests that DQ is of primary importance for antigen presentation. In cattle, it has been shown that DQ molecules present antigens to CD4+ T-cells and the interhaplotype pairing of DQα and DQβ molecules form functional restriction elements (Glass et al., 2000).

Two DQA genes, DQA1 and DQA2, have been identified in sheep (Scott et al., 1991). Both genes are polymorphic. Restriction fragment length polymorphism analyses have identified seven alleles plus a null allele at the DQA1 locus and 16 alleles at the DQA2 locus (Wright and Ballingall, 1994; Escayg et al., 1996). Sequence analysis of some of these RFLP alleles and others has revealed nine DQA1 and 10 DQA2 sequences (Snibson et al., 1998; Zhou and Hickford, 2001). This suggests that variation at the DQA loci has not yet been well characterized. Lower levels of sequence variation may make some alleles difficult to differentiate, especially by RFLP analysis; hence, sequence polymorphism in the DQA genes may have been underestimated.

In this study, variation in the ovine DQA1 gene was investigated in a large number of sheep of different breeds using single-strand conformational polymorphism (SSCP) analysis, cloning, and sequencing.
Table 1. Primers for amplifying exon 2 of the ovine DQA1 gene

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence (5’ → 3’)</th>
<th>Locationb</th>
<th>Amplimer size</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1-up</td>
<td>ACCTGACTCAcCTGACCACAa</td>
<td>348 to 367</td>
<td>263 bp</td>
</tr>
<tr>
<td>DQA1-dn</td>
<td>AACACATACTGTGTTAGCAGCA</td>
<td>594 to 616</td>
<td></td>
</tr>
</tbody>
</table>

aA mismatched nucleotide (lower case) was introduced to decrease self-complementarity, and a mismatched nucleotide in bold was introduced to increase the likelihood of 3’ T overhangs on the PCR amplimers for cloning.
bNucleotide positions refer to the ovine DQA1 sequence reported in Scott et al. (1991).

Materials and Methods

Sheep and DNA Sources

Three hundred sheep from six different breeds (Merino, Corriedale, Borderdale, Romney, Awassi, and Finnish Landrace), sired by unrelated rams, were used in this study. Genomic DNA used in PCR amplification was isolated from blood using a high-salt procedure as described by Montgomery and Sise (1990) or extracted from blood collected on FTA Classic cards (Whatman BioScience, Newton, MA), following the manufacturer’s protocol.

PCR Amplification

Specific PCR primers DQA1-up and DQA1-dn (Table 1) used to amplify the second exon of the ovine DQA1 gene were designed based on a published ovine DQA1 sequence (M33304, Scott et al., 1991) and with comparison to DQA2 sequences (M33305, Scott et al., 1991; Z28421, Wright and Ballingall, 1994). The amplimers contain the whole of exon 2 and the flanking intron sequences from the DQA1 gene. Primers were synthesized by Proligo (Proligo LLC, Boulder, CO).

Initial PCR amplifications for allele detection used a nonproof-reading Taq DNA polymerase (Qiagen, Hilden, Germany). In later amplifications for allele sequencing, a proof-reading enzyme, ProofStart DNA polymerase (Qiagen), was used in order to reduce PCR-associated nucleotide substitutions. Each PCR was performed in a 20-μL reaction volume containing 50 ng of genomic template DNA or genomic DNA on one 1.2-mm punch of an FTA card, 0.25 μM of each primer, 150 μM of dNTP (ABgene, Surrey, U.K.), 1 U of DNA polymerase, and 1× reaction buffer (supplied containing 1.5 mM MgCl2). Amplification was carried out in an iCycler (Bio-Rad, Hercules, CA) and consisted of denaturation at 94°C for 2 min, followed by 32 cycles of 94°C for 30 s, 61°C for 30 s, and 72°C for 30 s, with a final extension step at 72°C for 5 min. Amplimers were visualized by electrophoresis in 1% Seakem LE agarose (BioWhittaker Molecular Applications, Rockland, ME) gels using 1× TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM Na2EDTA) containing 200 ng/mL of ethidium bromide.

Single-Strand Conformational Polymorphism Analysis

A 0.7-μL aliquot of each amplimer was mixed with 7 μL of loading dye (98% formamide, 10 mM EDTA, 0.025% bromophenol blue, 0.025% xylene-cyanol). After denaturation at 95°C for 5 min, samples were cooled rapidly on ice and then loaded on 16 cm × 18 cm, 16% acrylamide: bisacrylamide (37.5:1) (Bio-Rad) gels. Electrophoresis was performed using Protean II xi cells (Bio-Rad), at 320 V for 20 h at 5°C in 0.5× TBE buffer. Gels were silver stained according to the method of Bassam et al. (1991).

Cloning of PCR Amplimers and Screening of Clones

Amplimers representative of each unique SSCP banding pattern were ligated with the pCR 4 Blunt-TOPO vector (Invitrogen, San Diego, CA) according to the manufacturer’s instructions. A 2-μL aliquot of the ligation mixture was used to transform competent Escherichia coli cells (One Shot INVαF’, Invitrogen), following the protocol recommend by the manufacturer. Between 10 and 15 insert-positive colonies for each transformation were picked and incubated overnight in Terrific broth (Invitrogen) at 37°C, in a shaking rotary incubator (225 rpm).

Plasmids containing inserts were recovered from the transformants by boiling for 10 min in 0.8% Triton X-100 solution. After centrifugation at 13,000 × g for 2 min, 1 μL of the supernatant was amplified by PCR using Taq DNA polymerase (Qiagen). Amplimers from clone inserts were run alongside amplimers from the corresponding genomic DNA for comparison of the banding patterns. Only those clones which presented identical SSCP patterns to that of the corresponding genomic DNA were selected for subsequent DNA sequencing.

DNA Sequencing

Plasmid from selected clones was extracted using a QIAprep Spin Miniprep kit (Qiagen). It was sequenced in both directions using M13 forward and reverse primers at the Waikato DNA Sequencing Facility, University of Waikato, Hamilton, New Zealand. Identical sequences obtained from at least three clones from different sheep, or two independent PCR amplifications from
the same sheep, were subjected to further sequence analysis.

Sequence Analysis

Sequence alignments, translations, and comparisons were carried out using DNAMAN (Version 4.0, Lynnon BioSoft, Vaudreuil, Canada). The BLAST algorithm was used to search the NCBI GenBank databases (http://www.ncbi.nlm.nih.gov/) for homologous sequences.

Nucleotides believed to be involved in antigen presentation were identified by comparison with the models produced by Brown et al. (1993) and Paliakasis et al. (1996). Genetic distances for nonsynonymous (dN) and synonymous (dS) substitutions were calculated according to the method of Nei and Gojobori (1986) using the Jukes and Cantor (1969) correction. All means of dN and dS and standard errors were estimated using the MEGA program (Version 2.1; http://www.megasoftware.net/).

A neighbor-joining tree (Saitou and Nei, 1987) was constructed on the basis of genetic distances, estimated by the Kimura (1980) two-parameter method, using the MEGA program. The following cattle DQA1 sequences, together with the sheep DQA1 sequences listed in Table 4, were used to construct the tree: BoLA-0101 (Z48185), BoLA-0102 (Z48194), BoLA-0103 (Z48187), BoLA-0202 (Z48190), BoLA-0204 (Z48188), BoLA-0301 (Z48195), BoLA-0401 (Z48196), BoLA-0801 (Z48186), BoLA-1001 (Z48191), BoLA-1201 (Z48193), BoLA-1301 (Z48192), and BoLA-1401 (Z48189) from Gelhaus et al. (1995); BoLA-1202 (D50454) from Nishino et al. (1995); BoLA-1203 (M30117) from van der Poel et al. (1990); and BoLA-1302 (Z79507) from Ballingall et al. (1997). GenBank sequences were trimmed to the length corresponding to the PCR amplimers before generating the neighbor-joining tree. The tree was rooted with the human DQA1 sequence HLA-DQA1*0101 (L34082; Yasunaga et al., 1996) as an outgroup.

Nomenclature

Because no workshop for the nomenclature of sheep MHC has been held, the criteria for assigning new alleles and the rules for naming alleles for the ovine DQA1 gene were based on those proposed for the cattle MHC (Davies et al., 1997; http://www.projects.roslin.ac.uk/bola/bolahome.html). The criteria are that when a sequence is based on clones derived from PCR amplification, there have to be at least three identical clone sequences. Names are based on the predicted AA sequences and consist of four or five digits, where the first two digits indicate the major type, the third and fourth digits indicate the subtype, and the fifth digit (if present) indicates unexpressed variation (silent substitutions). Alleles that differ by less than five AA in the first domain are assigned as subtypes within a single major type.

Results

Amplification of the Ovine DQA1 Exon 2 and SSCP Analysis

Amplimers of 269 bp were obtained with genomic DNA from most sheep using primers DQA1-up and DQA1-dn. These amplimers exhibited polymorphism upon SSCP analysis. Under the established conditions, 14 unique SSCP patterns could be detected (Figure 1). These primers did not amplify DNA from a small portion of animals, including those that had previously been typed as DQA2-F or -G homozygous by Southern hybridization and hence were probably DQA1-null, as reported previously by Hickford et al. (2000).

Sequence Polymorphism of the Ovine DQA1 Exon 2

Cloning of PCR amplimers representative of the 14 SSCP patterns followed by DNA sequencing revealed 14 different sequences (Figure 2). No more than two sequences were isolated from clones derived from each animal. All sequences were confirmed by comparing the SSCP patterns of the cloned PCR amplimers with those of the genomic PCR amplimer from which the sequence was isolated. Additional confirmation was provided by identifying each sequence in more than one animal.

Of the 14 sequences identified, eight were identical to previously published ovine DQA1 sequences, whereas the remaining six sequences were unique, but shared close homology to previously published DQA1 sequences derived from sheep and cattle. Between one and 11 nucleotide differences were observed between new sequences described in this study were submitted to the NCBI GenBank and assigned the accession numbers AF317616, AF317617, AY229894, and AY230208 to AY230210.
Figure 2. Alignment of the ovine DQA1 nucleotide sequences. Sequences detected in this study are named based on the bovine major histocompatibility complex (MHC) nomenclature (see Table 4), whereas the previously reported sequences that were not detected in this study are labeled with a GenBank accession number in parentheses. Newly identified sequences are in bold. A dash indicates identity with the top sequence. The primer binding sequences and exon 2 are indicated. Nucleotides in the sequences that were not detected in this study and that are different to the closest sequences detected here are marked with boxes.

these six new sequences and the published ovine DQA1 sequences (Figure 2).

Twenty-four percent of nucleotide positions were polymorphic in the 14 DQA1 sequences (Figure 2). In the putative antigen-binding site (ABS) codons, the frequency of nonsynonymous substitutions was slightly higher than that of synonymous substitutions. A similar pattern was observed in the codons of putative non-ABS. Nonsynonymous and synonymous substitutions both occurred more frequently in the ABS codons than in the non-ABS codons (Table 2).

The predicted AA sequences of the ovine DQA1 alleles detected in this study are shown aligned with the predicted amino sequences of the published ovine DQA2 alleles (Figure 3).

Relationship Between the Ovine DQA1 Alleles

Of all the available ovine DQA1 sequences, one published sequence (M93430; Fabb et al., 1993) and one unpublished sequence listed in the NCBI GenBank

Table 2. Relative frequencies of nonsynonymous (dN) and synonymous (dS) substitutions in the antigen-binding sites (ABS) and non-ABS for the 14 ovine DQA1 alleles

<table>
<thead>
<tr>
<th>Binding sites</th>
<th>dN, %</th>
<th>dS, %</th>
<th>dN/dS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>17.0 ± 5.0</td>
<td>11.5 ± 6.9</td>
<td>1.5</td>
</tr>
<tr>
<td>Non-ABS</td>
<td>6.4 ± 2.0</td>
<td>6.0 ± 1.8</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Comparison of exon 2 AA sequences of ovine DQA1 and ovine DQA2 alleles (excluding the primer binding regions). The names of the ovine DQA1 alleles refer to Table 4, and those of the ovine DQA2 alleles are labeled with the species designation and a GenBank accession number in parentheses. Amino acids are presented as one-letter codes and the numbering above the aligned sequences refers to the human DQα chain (Paliakasis et al., 1996). Newly identified ovine DQA1 alleles are shown in bold. A dash represents identity with the OLA-DQA1*0101 sequence and a cross indicates putative sites involved in peptide binding as proposed for the human DR1 (Brown et al., 1993) and human DQα molecules (Paliakasis et al. 1996).

(L49464), were not detected in this study. Because insufficient information is available to assign these two sequence names using the recognized criteria, their assignment as allelic sequences awaits further confirmation and therefore they were not analyzed here.

The ovine DQA1 alleles analyzed shared 84.5 to 99.6% nucleotide homology in the 226-bp exon 2 sequence (excluding the primer binding sequences). This corresponded to the 73 to 98.6% similarity in the predicted AA sequences (Table 3). These alleles were named based on the identities and similarities between the AA sequences and this, along with their historical names, is summarized in Table 4.

A neighbor-joining phylogenetic tree was constructed from these ovine DQA1 sequences and 23 published bovine DQA1 sequences, together with a reference sequence of human DQA1. This tree revealed several clusters of ovine DQA1 sequences, and some of the ovine DQA1 sequences were more similar to cattle sequences than to ovine sequences (Figure 4).

Table 3. Comparisons of ovine DQA1 exon 2 sequences

<table>
<thead>
<tr>
<th>Allele</th>
<th>*0101</th>
<th>*0102</th>
<th>*0103</th>
<th>*0104</th>
<th>*0201</th>
<th>*0401</th>
<th>*0402</th>
<th>*0302</th>
<th>*0301</th>
<th>*0701</th>
<th>*0501</th>
<th>*0601</th>
<th>*0801</th>
<th>*0901</th>
</tr>
</thead>
<tbody>
<tr>
<td>*0101</td>
<td>99.6</td>
<td>98.2</td>
<td>97.8</td>
<td>97.3</td>
<td>94.7</td>
<td>89.9</td>
<td>91.6</td>
<td>91.6</td>
<td>91.2</td>
<td>91.2</td>
<td>93.8</td>
<td>94.2</td>
<td>92.5</td>
<td></td>
</tr>
<tr>
<td>*0102</td>
<td>98.6</td>
<td>97.8</td>
<td>97.3</td>
<td>94.7</td>
<td>89.9</td>
<td>91.6</td>
<td>91.6</td>
<td>91.2</td>
<td>90.7</td>
<td>91.7</td>
<td>93.4</td>
<td>93.8</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td>*0103</td>
<td>93.6</td>
<td>93.2</td>
<td>93.2</td>
<td>96.6</td>
<td>97.3</td>
<td>98.9</td>
<td>92.5</td>
<td>92.5</td>
<td>92.0</td>
<td>91.6</td>
<td>94.7</td>
<td>94.2</td>
<td>92.9</td>
<td></td>
</tr>
<tr>
<td>*0104</td>
<td>93.2</td>
<td>93.2</td>
<td>98.6</td>
<td>94.7</td>
<td>97.3</td>
<td>98.9</td>
<td>92.5</td>
<td>92.0</td>
<td>91.6</td>
<td>90.3</td>
<td>91.2</td>
<td>93.4</td>
<td>92.5</td>
<td></td>
</tr>
<tr>
<td>*0201</td>
<td>87.8</td>
<td>87.8</td>
<td>89.2</td>
<td>87.8</td>
<td>87.6</td>
<td>90.3</td>
<td>90.3</td>
<td>90.7</td>
<td>88.9</td>
<td>93.8</td>
<td>91.6</td>
<td>91.6</td>
<td>92.0</td>
<td></td>
</tr>
<tr>
<td>*0401</td>
<td>81.1</td>
<td>81.1</td>
<td>82.4</td>
<td>81.1</td>
<td>81.8</td>
<td>87.4</td>
<td>89.7</td>
<td>89.3</td>
<td>94.7</td>
<td>95.1</td>
<td>88.1</td>
<td>84.5</td>
<td>85.4</td>
<td></td>
</tr>
<tr>
<td>*0402</td>
<td>85.1</td>
<td>85.1</td>
<td>86.5</td>
<td>85.1</td>
<td>82.4</td>
<td>95.9</td>
<td>94.7</td>
<td>92.0</td>
<td>92.4</td>
<td>87.6</td>
<td>88.1</td>
<td>90.7</td>
<td>97.2</td>
<td></td>
</tr>
<tr>
<td>*0302</td>
<td>86.5</td>
<td>86.5</td>
<td>87.8</td>
<td>86.5</td>
<td>83.8</td>
<td>90.5</td>
<td>89.1</td>
<td>99.6</td>
<td>88.5</td>
<td>87.6</td>
<td>91.2</td>
<td>90.3</td>
<td>88.1</td>
<td></td>
</tr>
<tr>
<td>*0301</td>
<td>85.1</td>
<td>85.1</td>
<td>86.5</td>
<td>85.1</td>
<td>83.8</td>
<td>89.2</td>
<td>89.5</td>
<td>88.1</td>
<td>87.4</td>
<td>92.0</td>
<td>89.8</td>
<td>87.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*0701</td>
<td>83.8</td>
<td>82.4</td>
<td>81.9</td>
<td>79.7</td>
<td>78.4</td>
<td>74.3</td>
<td>78.4</td>
<td>79.7</td>
<td>78.4</td>
<td>89.8</td>
<td>93.4</td>
<td>92.9</td>
<td>88.9</td>
<td></td>
</tr>
<tr>
<td>*0501</td>
<td>81.1</td>
<td>81.1</td>
<td>81.9</td>
<td>81.7</td>
<td>87.4</td>
<td>73.0</td>
<td>77.0</td>
<td>77.0</td>
<td>77.0</td>
<td>85.7</td>
<td>85.0</td>
<td>85.1</td>
<td>84.5</td>
<td></td>
</tr>
<tr>
<td>*0601</td>
<td>83.8</td>
<td>83.8</td>
<td>85.1</td>
<td>83.8</td>
<td>81.1</td>
<td>75.7</td>
<td>79.7</td>
<td>79.7</td>
<td>78.4</td>
<td>90.5</td>
<td>84.5</td>
<td>85.1</td>
<td>89.5</td>
<td></td>
</tr>
<tr>
<td>*0801</td>
<td>86.5</td>
<td>85.1</td>
<td>85.1</td>
<td>83.8</td>
<td>82.4</td>
<td>78.4</td>
<td>82.4</td>
<td>82.4</td>
<td>81.1</td>
<td>86.5</td>
<td>88.5</td>
<td>89.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*0901</td>
<td>81.1</td>
<td>81.1</td>
<td>81.7</td>
<td>79.7</td>
<td>73.0</td>
<td>74.3</td>
<td>77.0</td>
<td>77.0</td>
<td>75.7</td>
<td>86.5</td>
<td>85.1</td>
<td>84.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Nucleotide sequence identities are shown above the diagonal and predicted AA identities are shown below the diagonal. All sequences compared exclude the PCR primer binding regions, and the similarities are shown in percentages.
Table 4. Proposed nomenclature of the ovine DQA1 alleles

<table>
<thead>
<tr>
<th>Recommended allele name&lt;sup&gt;a&lt;/sup&gt;</th>
<th>GenBank accession number</th>
<th>Previous equivalent name</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQA1*0101</td>
<td>Z28418</td>
<td>DQA1</td>
<td>Wright and Ballygall (1994)</td>
</tr>
<tr>
<td>DQA1*0102</td>
<td>AY230210</td>
<td>DQA1</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0103</td>
<td>M33304</td>
<td>DQA1</td>
<td>Scott et al. (1991)</td>
</tr>
<tr>
<td>DQA1*0104</td>
<td>AY230209</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0201</td>
<td>AF317617</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0301</td>
<td>AY230268</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0302</td>
<td>AP276955</td>
<td>DQA1-I</td>
<td>Zhou and Hickford (2001)</td>
</tr>
<tr>
<td>DQA1*0401</td>
<td>L49463&lt;sup&gt;b&lt;/sup&gt;</td>
<td>DQA1.3</td>
<td>Snibson et al (1998)</td>
</tr>
<tr>
<td>DQA1*0402</td>
<td>AY265308</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0501</td>
<td>Z28420</td>
<td>DQA1</td>
<td>Wright and Ballygall (1994)</td>
</tr>
<tr>
<td>DQA1*0601</td>
<td>Z28518</td>
<td>DQA1</td>
<td>Wright and Ballygall (1994)</td>
</tr>
<tr>
<td>DQA1*0701</td>
<td>AP276954</td>
<td>DQA1-H</td>
<td>Zhou and Hickford (2001)</td>
</tr>
<tr>
<td>DQA1*0801</td>
<td>AY229894</td>
<td>This study</td>
<td>This study</td>
</tr>
<tr>
<td>DQA1*0901</td>
<td>AP276956</td>
<td>DQA1-J</td>
<td>Zhou and Hickford (2001)</td>
</tr>
<tr>
<td>Sequences requiring confirmation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M93430&lt;sup&gt;c&lt;/sup&gt;</td>
<td>L49464&lt;sup&gt;bd&lt;/sup&gt;</td>
<td>DQA1</td>
<td>Fabb et al. (1993)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DQA1.5</td>
<td>Snibson et al. (1998)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Based on the proposed nomenclature for the bovine major histocompatibility complex (www.projects.roslin.ac.uk/bola/bolahome.html).

<sup>b</sup>Partial exon 2 sequence.

<sup>c</sup>Messenger RNA-derived sequence.

<sup>d</sup>Unpublished sequence retrieved from the NCBI GenBank (http://www.ncbi.nlm.nih.gov/).

Discussion

Sequence analysis of exon 2 of the ovine DQA1 genes investigated in this study revealed 14 sequences at this locus. Six of these were newly identified sequences. As more animals from more breeds are screened, more novel sequences may also be detected. In addition, analysis of parts of the gene, outside of exon 2 may reveal even further genetic variation. This confirms that the ovine MHC DQA1 is a highly polymorphic locus. The level of diversity detected for the ovine DQA1 locus appears to be lower than that for the human DQA1 (22 alleles; Marsh et al., 2002), but is similar with that for DQA1 in cattle (15 alleles; Ballygall et al., 1998).

Sheep that are typed by Southern hybridization to be DQA2-F or -G homozygotes contain two DQA2-like genes but lack a DQA1 gene (i.e., they are DQA1-null; Hickford et al., 2000). No PCR amplifiers were detected with these sheep, confirming this finding and also confirming the specificity of the PCR amplification reaction utilized in this study. This is also supported by the sequencing data that reveals that all sequences isolated were either identical or showed closest sequence homology to previously published DQA1 sequences from sheep and cattle. In addition, the six new sequences identified in this study exhibited conserved region and variable region sequence patterns similar to those of the published ovine DQA1 sequences, but different from ovine DQA2 sequences (Figure 3). This strongly supports the assertion that they are new ovine DQA1 allelic sequences.

Caution was exercised in assigning the new sequences identified in this study since they were based on clones derived from PCR amplification. Misincorporation of nucleotides by DNA polymerases, even those with 3′ → 5′ exonuclease activity (such as ProofStart DNA polymerase; Eckert and Kunkel, 1991), and in vitro recombination (Bradley and Hillis, 1997) may occur during PCR amplification. This has been acknowledged with the finding that some published MHC allelic sequences from a variety of species are most likely to be the result of PCR artifacts (Ennis et al., 1990; Kennedy et al., 2002).

All the sequences reported here were confirmed to be unique by comparison of the patterns generated by the SSCP analysis of amplifiers derived from genomic DNA with the patterns generated from clones. In addition, the amplifiers of sheep from which sequences with a single nucleotide difference were isolated, generated distinct SSCP patterns, and these matched with the patterns obtained from corresponding clones that were sequenced as well. These checks ensured that the new sequences represented genuine ovine DQA1 alleles, and not PCR or sequencing artifacts.

A close examination of two previously reported ovine DQA1 sequences (M93430 and L49464) that were not detected in this study revealed that they had two and three single nucleotide differences compared with the sequences detected here (*0301 and *0601, respectively). These nucleotide differences occurred at positions that are conserved in all of the other 14 sequences analyzed here, raising the possibility that they either represent rare undefined DQA1 alleles that are not present in the sheep examined in this study, or are analysis or sequencing artifacts. The sequence of M93430 is from a clone derived from a complementary DNA sequence, while the sequence of L49464 is from a clone derived from a cDNA sequence. This suggests that the nucleotide differences in these sequences may be the result of PCR artifacts rather than genuine allelic differences.

Finally, it is important to note that the proposed nomenclature for the ovine DQA1 alleles is based on the proposed nomenclature for the bovine MHC (www.projects.roslin.ac.uk/bola/bolahome.html). This nomenclature is designed to provide a consistent and standardized way of assigning names to alleles at the DQA1 locus, and it is expected that this nomenclature will be adopted by the international community of researchers working with ovine MHC genetics.
Figure 4. Neighbor-joining tree (Saitou and Nei, 1987) of sheep and cattle DQA1 exon 2 sequences. The tree was constructed using the nucleotide sequences of sheep and cattle listed in Table 4 and from the NCBI GenBank based on 500 bootstrap replications. Sheep sequences are in bold and newly identified alleles are boxed. The numbers at the forks indicate the bootstrap confidence values (Felsenstein, 1985) and branch lengths are proportional to genetic distance.

DNA library (Fabb et al., 1993), whereas that of L49464 is from a sequence that was amplified from genomic DNA (Snibson et al., 1998). Considering that no confirmation of the authenticity of these two sequences is available, they were therefore not assigned allele names in this study.

It is difficult to determine whether the inability to produce a PCR amplimer from some sheep is due to the absence of a DQA1 gene or to the presence of alleles whose sequences are too different to be amplified by the primers used. However, many of the DNA samples that did not amplify with the DQA1 primers were genotyped using Southern hybridization at the DQA2 locus (as per Hickford et al., 2000), and they were either DQA2-F, -G, -I, and -J homozygous or heterozygous for variations of these four DQA2 alleles (data not shown). These DQA2 alleles are probably associated with the DQA1-null allele in sheep (Hickford et al., 2000). However, because the potential also exists for there to be new DQA1 alleles with novel sequences in the primers-binding regions used in this study, further analyses of these sheep with other PCR primers and other techniques such as RFLP-Southern hybridization are required to confirm this.

Although the sheep DQA1 alleles shared close sequence similarity to each other (Figures 2 and 3), at times they clustered into groups with cattle DQA1 sequences (Figure 4). These clusters of sheep and cattle alleles may be derived from primordial sequences that were present in a common ancestor and have persisted in the sheep and cattle populations since their divergence. This has been referred to as the trans-species hypothesis (Klein, 1987). This pattern of evolution suggests the action of natural selection on the DQA1 gene, since neutral polymorphism is not expected to persist very long in a population. Pathogen recognition may provide selection pressure to maintain particular MHC sequences, and the observation that sheep and cattle share similar allelic sequences may be evidence of the need for a specific immune response to a common pathogen.

Most of the ovine DQA1 alleles reported here possess different sequence motifs at the positions that are considered to be important for peptide binding and pocket formation in the human DQA1 molecules (Figure 5; Brown et al., 1993; Paliakasis et al., 1996). In the human DQα chain, residues at α72, α75, α76, and α79 are invariable and deemed crucial for forming hydrogen bonds to the backbone of a foreign peptide (Paliakasis et al., 1996). However, residues at two of these three positions (α72 and α79) were variable in the sheep DQA1 molecules.

Sequence variation was also observed at positions that are predicted to form the first and the fifth pockets at the ABS; with the latter, variation was observed at four positions (α72, α75, α76, and α79). Substitutions at these positions may impact the specificity of the ABS and therefore affect peptide binding ability. The importance of specific residues within the ABS and how peptide binding can be altered with only one or two amino acid changes in some alleles, has been clearly illustrated in humans (Seidl et al., 1997; Toussirot et al., 1999).

There were two pairs of ovine DQA1 alleles (*0101 and *0102; *0103 and *0104), which only had one AA difference between alleles within the pair. These AA substitutions were non-conservative, but occurred at the positions that are considered to be conserved non-ABS sites in human DR and DQ molecules (Brown et al., 1993; Paliakasis et al., 1996). Because the x-ray crystal structure has not been determined for sheep DQ or for any closely related species, the effect of these substitutions on antigen presentation is speculative.
It is generally accepted that the polymorphism of MHC genes is driven by a form of natural selection called balancing selection (Hughes, 1999). The number of nonsynonymous substitutions per nonsynonymous site (dN) is expected to exceed the number of synonymous substitutions per synonymous site (dS) in the codons encoding the ABS if balancing selection is, or recently has been, acting on these positions. However, with these ovine DQA1 sequences, nonsynonymous substitutions occurred at about the same rate as synonymous substitutions in either the codons of the ABS or non-ABS codons (Table 2). The low ratio of dN:dS seen here may be due to several reasons. 1) Balancing selection has a weaker effect at this locus in sheep. In common chimpanzee, common baboon, and hamadryas baboon, dN was not significantly greater than dS in the ABS codons at the DQA locus, but a significant difference was found in the human (Bergstrom and Gyllensten, 1995). 2) Although an important selected site may lie within the DQA1 locus, sufficient interlocus or interallele recombination (gene conversion) exists to disrupt the effects of balancing selection. The presence of clustered sequence variation between alleles (Figure 2) suggests that gene conversion may have occurred for the ovine DQA1 gene, but because gene conversion is expected to be an essentially random process, it does not fit the very specific pattern of dN exceeding dS in the ABS codons that characterize MHC loci (Hughes and Nei, 1988, 1989). 3) Ovine DQA1 molecules may possess ABS different to those defined for the human DR1 (Brown et al., 1993). As discussed above, AA substitutions at the positions important for forming hydrogen bonds to antigenic peptide backbones, especially the nonconservative substitution at α79, may result in different ABS. 4) These alleles identified so far at DQA1 may only represent those that are closely related in sequence. Other alleles with less similarity to these identified alleles may exist, and further sequencing may reveal that dN exceeding dS in the ABS codons. It has been documented that some sheep contain a DQA1 and a DQA2 gene, whereas others do not possess a sequence homologous to DQA1, instead they have a sequence that is more homologous to the DQA2 gene (called DQA2-like) (Snibson et al., 1998). Considering the possible function of DQA1 molecules in immune response, the presence of a DQA2-like sequence instead of a DQA1 sequence in some sheep suggests that these DQA2-like sequences may act functionally as DQA1 alleles and contribute to diversity at this locus.

The comparison of sequences between these ovine DQA1 alleles suggests that point mutation and gene conversion have a major role in the generation of polymorphism at the DQA1 locus. While point mutation in MHC is positively selected, this is a slow process, compared to gene conversion. The combination of these mechanisms may hasten the evolution of new alleles to maximize pathogen recognition.

Implications

The polymorphism we have detected in the exon 2 region of the ovine DQA1 gene makes the gene a potential genetic marker for the study of disease susceptibility and resistance. This would require an accurate typing mechanism for differentiating different alleles of the DQA1 gene in sheep. The single-strand conformational polymorphism exhibited in the exon 2 region could be used to develop such a typing system.

Literature Cited


van der Poel, J. J., M. A. Groenen, R. J. Dijkhof, D. Ruyter, and M. J. Giphart. 1990. The nucleotide sequence of the bovine MHC class II alpha genes: \( DRA, DOA, \) and \( DYA \). Immunogenetics 31:29–36.

