Vol. 212, No. 1, 1995 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 6, 1995 Pages 110-117

Nucleotide sequence and the molecular evolution of a new *A2* gene in the *DQ* subregion of the bovine major histocompatibility complex+

Akira Morooka[1,2], Masatoshi Asahina[1,3], Chikara Kohda[1,4], Shigeru Tajima[1,5], Masashi Niimi[1,4], Yoshii Nishino[1], Masahiro Sugiyama[2] and Yoko Aida[1]*

[1]Laboratory of Gene Technology and Safety, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan

[2]Nippon Veterinary and Animal Science University, Tokyo 180, Japan

[3]Faculty of Agriculture, Iwate University, Morioka 020, Japan

[4]Faculty of Agriculture, Tohoku University, Sendai 981, Japan

[5]Medicine Research Division, Tokyo Medical and Dental University, Tokyo 113, Japan

Received June 4, 1995

SUMMARY: cDNA clones encoding the bovine major histocompatibility complex (MHC) class II DQα chain were isolated. One clone, MQ9, encoded a primary translated product of 255 amino acids, with a signal peptide of 23 amino acids and a mature polypeptide of 232 amino acids. A new *A2* gene in the *DQ* subregion of the bovine genome was identified from a comparison of amino acid sequences encoded by class II *A* genes among several species and the construction of a phylogenetic tree. It was revealed that MQ9 is most closely related to the ovine *DQA2* genes among sequences from various mammalian species. By contrast, the *BoLA-DQA* genes previously isolated are more closely related to ovine *DQA1* than to the *BoLA-DQA2* gene, and they represent *BoLA-DQA1* genes. Thus, the presence of two *BoLA A* genes, which may be expressed and functional in the bovine, as well as in sheep was confirmed. A large number of amino acids unique to products of *DQA2* genes of bovine and ovine origin were identified when the predicted amino acid sequences for both species were compared, and most of the *DQA2*-specific residues were located in the α1 domain and were conserved with respect to products of *DQA1* genes of ruminants. Thus, several characteristics of the bovine *DQA* genes were found to differ from those of human and rodent genes, despite similarities in gene structure and in nucleotide sequence.

The class II regions of the bovine lymphocyte antigen (*BoLA*) system, the major histocompatibility complex (MHC) of cattle, have been under intensive investigation because of their assumed role in the genetic control of immune responses to pathogen-derived peptide antigens. Analyses of restriction fragment length polymorphism (RFLP), and the characterization of cloned

+Accession numbers for MQ9, AQ16-1 and KQ217 are D50045, D50046 and D50047, respectively.

*To whom correspondence should be addressed. Fax: +81-298-36-9050.

0006-291X/95 $12.00

bovine class II genes has provided evidence for the existence of one *DRA* gene and at least three *DRB* genes, one or two *DQA* and *DQB* genes, one *DOA* and one *DNA* gene, and three novel genes, namely, *DYA*, *DYB* and *DIB* (1-8). In the rabbit, rat and mouse, the *DQA*-equivalent gene is a transcriptionally active single-copy gene (9-11). In humans, there are two *DQA* loci. However, only *DQA1* is expressed (12). By contrast, sheep appear to have two loci, *DQA1* and *DQA2*, that are both transcriptionally active (13). To date, the structural organization of the *BoLA-DQA* genes has not been characterized in detail. The number of *DQA* genes varies between haplotypes in the bovine, as is also the case for human *DRB* genes (14), ovine *DQA* genes (13) and bovine *DQB* genes (1). Cloning and sequencing of a genomic clone, W1, and of two cDNA clones, NQ1 and α5, which correspond to the *BoLA-DQA* gene and are more closely related to ovine *DQA1* genes than to ovine *DQA2* genes, as well as cloning and sequencing of the *DQA* genes of various mammalian species, have confirmed the presence of a *DQA1* gene in the bovine (5,15,16). By contrast, there is no evidence, as yet, for a *DQA2* gene in the bovine. In present study, we isolated and characterized cDNA clones that represent a novel *BoLA-DQA2* gene, which differs from the *BoLA-DQA1* gene, and we studied phylogenetic relationships to other genes on the basis of nucleotide sequences.

MATERIALS AND METHODS

Preparation of cDNA library. A cDNA library from the bovine lymphoid cell line BLSC-KU-1 was constructed in the mammalian expression vector pCDM8, as described previously (17,18), and it was introduced into *E. coli* Mc1061/P3.

Isolation and characterization of BoLA-DQA2 cDNAs. To screen the cDNA library, Hybond-N nylon filters (Amersham, UK) carrying approximately 2 x 10^6 colonies of the cDNA library, were screened with a 1.1-kilobase pair (kbp) *Pst* I fragment of the *HLA-DPA* cDNA clone, pDAα1–4 (19), by colony hybridization under low-stringency conditions. Plasmid DNAs were isolated and subcloned into pBluescript II SK (+) (Stratagene, Heidelberg, Germany), and the nucleotide sequences of both strands were determined by the dideoxy chain-termination method (20) with a *Bca*BEST sequencing kit (Takara Shuzo Co., Shiga, Japan). Computer analyses were performed with the program package from the Wisconsin Genetics Computer Group, which included the FASTA and TFASTA programs (21).

Construction of a dendrogram. A phylogenetic tree showing evolutionary relationships was constructed from the matrix of sequences similarities calculated with the UPGMA program (22).

RESULTS AND DISCUSSION

A partial *BoLA-DQA* cDNA clone, KQ217, with a 1,478-bp insert, was isolated from the BLSC-KU-1 cDNA library by colony hybridization under low-stringency conditions, with a cDNA for human *DPA* as probe. The nucleotide sequence of this clone was similar to the sequences of exons 2, 3, and 4 of a genomic *BoLA-DQA* clone, W1, reported by van der Poel *et al.* (5). The cDNA did not, however, contain the sequence of part of the signal peptide, suggesting that the 5' end of the

Fig. 1. Complete nucleotide and predicted amino acid sequences of the *BoLA-DQA* cDNA clone MQ9. Arrows designate the putative α-chain domains; 5'UT, 5' untranslated region; SP, signal peptide; α1, first domain; α2, second domain; CP; connecting peptide; TM, transmembrane region; CY, cytoplasmic tail and 3'UT, 3' untranslated region. Two conserved cysteine residues involved in disulfide bridges and two putative sites of *N*-linked glycosylation are underlined and the nucleotides forming a stop codon are indicated by asterisks.

gene in the KQ217 clone had been deleted. To obtain longer cDNA clones, the library (approximately 2 x 10^6 colonies) was again screened with a 1.5-kb *Xba* I fragment of KQ217 as probe and two clones were isolated: they were designated MQ9 (1,192 bp) and AQ16-1 (925 bp). Clone MQ9 was a cDNA that contained the full-length coding region of a bovine *DQA* gene.

The cDNA sequences and the deduced amino acid sequence of MQ9 revealed the general features of a functional class II gene (Figs. 1 and 2). The MQ9 clone is 1,192 nucleotides (nt) long with a single open reading frame (ORF) that begins with an ATG initiation codon (nt 39 to 41) and ends with a TGA stop codon (nt 804 to 806). The ORF encodes a protein of 255 amino acids. The *N*-terminal region of the putative protein begins at nucleotide position 108, being preceded by a signal sequences of 23 predominantly non-polar amino acids and followed by a mature polypeptide of 232 amino acids with α1 (positions 1 to 87), α2 (positions 88 to 181) and connecting-peptide/transmembrane/cytoplasmic (positions 182 to 232) domains. There is a 3'-untranslated region of 355 nucleotides, with a typical polyadenylation signal sequence (AATAAA; nt 1,142-1,147) before a poly(A) tail. Thus, MQ9 encodes the complete precursor molecule encoded by *BoLA-DQA*, containing the signal sequence. Two cysteine residues that form the intramolecular disulfide bridge of

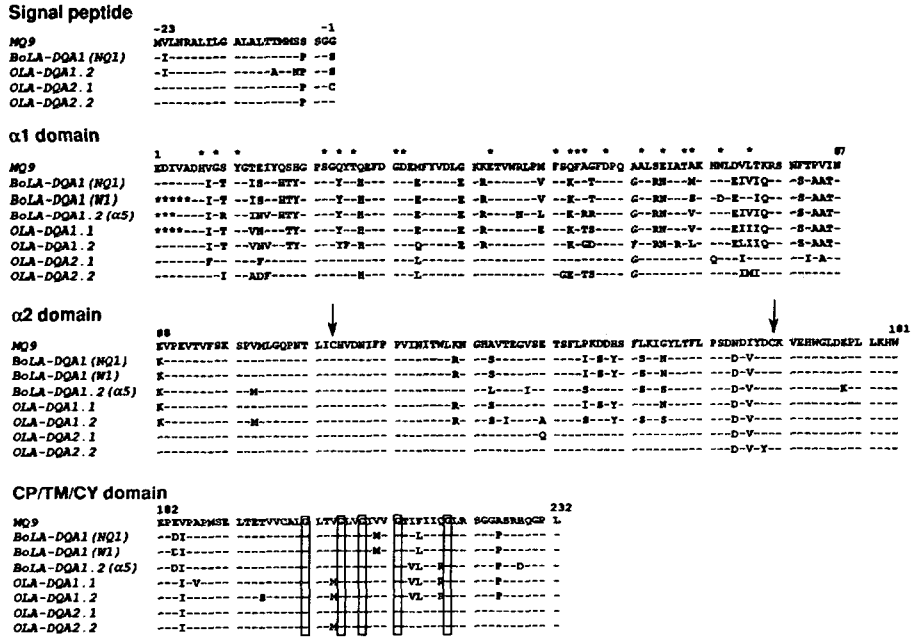

Fig. 2. Alignment of the amino acid sequence predicted from the *BoLA-DQA* clone MQ9 with those predicted from genes for the DQ α-chains of bovine (*BoLA*) (5,13,16) and sheep (*OLA*) (30). Dashed lines indicate sequence identity with respect to MQ9. The numbers denote positions of amino acids in the mature protein. Two putative sites of *N*-linked glycosylation and highly conserved residues in the immunoglobulin-like region are indicated by shaded boxes; the residues that participate in interactions with the α-chain are boxed; and two conserved cysteines are indicated by arrows. A dot indicates a residue in the antigen-recognition site (ARS) according to the proposed model of the class II molecule (28).

the α2 domain (positions 110 and 166) and two putative sites for glycosylation (positions 81 to 83 and 121 to 123) are all found at conserved positions in products of MQ9 and of *DQA* genes of other species, as previously observed by Figueroa and Klein (23). Moreover, all of ten residues that are highly conserved in the immunoglobulin chain (positions 145, 148, 154, 164, 166, 168, 170, 173, 176 and 181) are present in the protein encoded by MQ9, as indicated by shaded boxes in Figure 2 (24,25). Another conserved feature of DQA molecules found in the product of MQ9 is a set of five glycine residues (positions 201, 205, 208, 212 and 219) that are believed to participate in the interaction with the β-chain (26).

A comparison of the predicted amino acid sequences of each domain encoded by the MQ9 clone with those of the MHC class II α-chains of bovine, ovine, swine, rabbit, equine, mouse, rat and human origin, and a phylogenetic tree constructed from these data, are shown in Table 1 and Figure 3. The phylogenetic tree clearly shows that the *DRA/DPA* genes and the *DQA/DYA/DNA* genes form two major separate branches, and the MQ9 clone is on the *DQA* rather than on the *DYA* and *DNA* of the *DQA/DYA/DNA* branch. The MQ9-encoded protein exhibits a high degrees of overall identity (81-82%) when compared with sequences encoded by two *BoLA-DQA* cDNA clones, NQ1 and α5, and the partial genomic *BoLA-DQA* clone, W1, but only 67% and 55% identity to those encoded by

Table 1. Identities in terms of predicted amino acid sequences of various domains# among MQ9 cDNA and genes for class II α-chains from several species*, expressed as percentages

	SP	α1	α2	CP/TM/CY	Total
BoLA-DQA1 (NQ1)	87.0	67.8	88.3	90.2	82.0
BoLA-DQA1 (W1)	-	67.1	88.3	90.2	81.9
*BoLA-DQA1**2 (α5)	-	63.1	89.4	86.3	81.2
OLA-DQA1.1	-	65.1	90.4	86.3	80.7
OLA-DQA1.2	78.3	65.5	87.2	86.3	89.3
OLA-DQA2.1	91.3	90.8	96.8	98.0	94.5
OLA-DQA2.2	95.7	83.9	96.8	96.1	92.2
SLA-DQA (c)	65.2	67.8	86.2	90.2	79.1
SLA-DQA (d)	60.9	72.4	89.4	90.2	81.2
RLA-DQA	-	70.7	79.8	90.2	78.9
ELA-DQA	82.6	69.0	83.0	76.5	76.9
HLA-DQA1 (1)	73.9	67.4	87.2	90.2	79.6
HLA-DQA1 (2)	73.9	66.7	86.2	88.2	78.8
HLA-DQA1 (3)	-	70.9	81.9	88.2	79.2
HLA-DQA1 (4)	-	65.1	85.1	84.3	77.6
HLA-DQA1 (5)	-	66.7	84.0	88.2	78.4
HLA-DQA2	65.2	66.7	86.2	84.3	77.7
H2-IAα	-	69.0	73.4	92.2	76.0
RT1.Bα	69.2	65.5	75.5	88.2	74.1
HLA-DPA	-	50.5	70.2	56.9	52.6
HLA-DNA	60.9	77.1	78.5	72.7	74.0
BoLA-DRA (MR1)	21.7	47.1	62.8	56.9	60.0
BoLA-DYA (R2)	-	62.2	73.4	62.8	67.0

#The regions indicated in the Table are defined in the legend to Figure 1.
*Amino acid sequences deduced from genes for class II α-chains were taken from the references listed in the legend to Figure 2.

bovine *DYA* and *DRA* genes, respectively. In addition, the extent of the identity between the putative MQ9 protein and putative ovine, human, swine, rabbit, equine, mouse and rat proteins is rather high, ranging from 74% to 95%. Thus, the MQ9 gene is more closely related to the *DQA* gene of other mammalian species than to the *DRA* and *DYA* genes of bovine origin, and the *DPA* and *DNA* genes of human origin. The phylogenetic tree also indicates that the sequences of various mammalian *DQA* genes form four separate major branches, a rodent branch, a human branch, an artiodactyla branch, and a rabbit/equine branch, and MQ9 is located on the artiodactyla branch. It appears, therefore, that the *BoLA-DQA* genes are more similar to their counterparts in ovine and swine than to the *DQA* genes of rodent, rabbit, equine and human. Furthermore, sequences from artiodactyla species, such as bovine, ovine and swine, are separated into three additional clusters. One cluster contains MQ9 and ovine *DQA2* genes, such as *OLA-DQA2.1* and *OLA-DQA2.2*, suggesting that MQ9 is closest to the ovine *DQA2* genes. Moreover, MQ9 and *OLA-DQA2.1* are included in one additional cluster. Indeed, *OLA-DQA2.1* shares the greatest overall identity, in terms of predicted amino acid sequences, to MQ9 (94.5%) among orthologous comparisons, followed by *OLA-DQA2.2* (92.2%). By contrast, the *DQA1* genes of bovine, such as NQ1, α5, W1 and two *OLA-DQA1* genes, are separated from this cluster and their products exhibit approximately 83.2% identity to the protein encoded to by the MQ9 clone. A third cluster appears to be restricted to the swine *DQA* genes, with 80% identity to the MQ9 protein. Thus, the presence of the two *BoLA A* genes in the *DQ* subregion

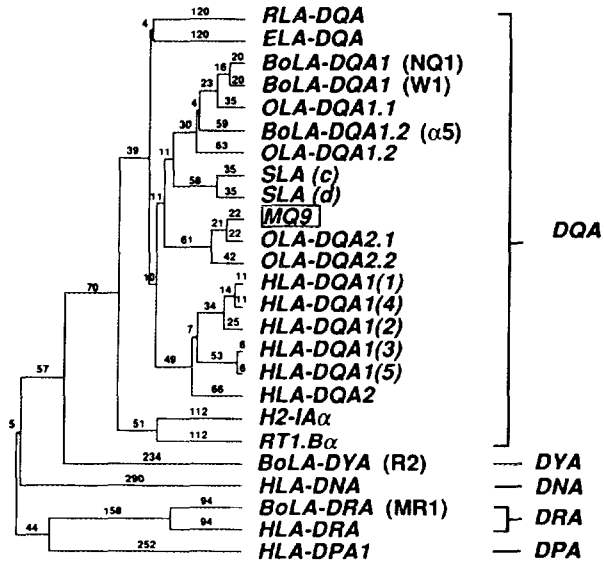

Fig. 3. Phylogenetic tree of MHC class II A genes of mammalian species, including bovine (*BoLA*) (5,13,16,17), ovine (*OLA*) (30), human (*HLA*) (31-38), swine (*SLA*) (39), dog (*DLA*) (40), rabbit (*RLA*) (10), rat (RT) (41), and mouse (H2) (42). The tree was constructed from the predicted amino acid sequences of the α1, α2, CP, TM and CY domains. Numbers above the horizontal lines indicate the evolutional distance (10^{-3}) between one protein and another.

is confirmed. The two genes represented by NQ1, W1 and α5 together and by MQ9 are designated *BoLA-DQA1* and *BoLA-DQA2*, respectively, and we have designated the gene that corresponds to the MQ9 clone *DQA2.1*. The evolutionary relationships indicate that, following the separation of the artiodactyla and human lineages after the bifurcation of the rodent and rabbit/equine/ardiodactyla/human lineages, a second gene duplication occurred that led to the ruminant *DQA1* gene on one branch and the ruminant *DQA2* gene on the other. This duplication was apparently independent of that leading to the presence of two *DQA* genes in human, which probably occurred about 15-20 million years ago (27).

The amino acid sequences predicted from for *DQA* genes from ruminants, including bovine and sheep, are aligned in Figure 2. The 19 amino acid unique to products of *DQA2* genes that differ from most of products of *DQA1* genes are identified. The substitutions at positions 18, 19, 24, 40, 42, 64, 65, 74, 78, 82, 84, 86, 88, 215, 225 are conservative with respect to *DQA1* genes, whereas substitutions at positions 69, 130, 142 and 152 are not. The α1 domain contains most of the *DQA2*-specific residues, and the three amino acids at positions 24, 69 and 84 are located at positions that participate in the putaitve antigen-recognition site (ARS) proposed in the model of Brown and co-workers (28). In addition, 15 amino acid differences were identified between MQ9 and *OLA-DQA2.1*. The α1 domain contains most of these substitutions, whereas the remaining domains appear to be relatively well conserved.

In this study, we identified a new *BoLA-DQA2* gene. A bovine cDNA clone that was most similar to the ovine *DQA2* gene, among various orthologous comparisons, was cloned. The absence

of variations in our clone at residues of which amino acids are highly conserved among products of mammalian class II *A* genes supports the notion that this gene might be expressed. Co-transfection with the *DQA2* cDNA MQ9 described here and the *DQB2* cDNA NB17 isolated previously (29) of COS cells identified two pairs of *A* and *B* genes, which were capable of pairing and subsequent expression of products at the cell surface. By PCR using *DQA2*-specific primers, the second exon of transcribed cattle *DQA2* gene was amplified from the cDNA library of bovine lymphoid cell line. These results clearly suggest that the *BoLA-DQA2* gene might be expressed and functional. The sequence information obtained in this study also provides additional information about the previously isolated NQ1, W1 and α5 clones, which represent the bovine *DQA1* gene. Thus, in contrast to human and rodent loci, two *BoLA A* loci that appear to be transcriptionally active in the *DQ* subregion were confirmed in the bovine, as well as in the sheep. The functional significance of this variation is not understood. Perhaps the history of species-specific pathogens influences the number and expression of MHC genes, or perhaps the these species have taken different evolutionary tracks to arrive at the same functional endpoint.

ACKNOWLEDGMENTS: The authors thank Dr. H. Amanuma for encouragement and Drs. H. Inoko and A. Ando for their gift of pDAα1–4. This study was supported by Special Coordination Funds from the Science and Technology Agency of the Japanese Government.

REFERENCES

1. Andersson, L., and Rask, L. (1988) *Immunogenetics* 27, 110-120.
2. Muggli-Cockett, N. E., and Stone, R. T. (1988) *Anim. Genet.* 19, 213-225.
3. Muggli-Cockett, N. E., and Stone, R. T. (1989) *Anim. Genet.* 20, 361-369.
4. Groenen, M. A. M., van der Poel, J. J., Dijkhof, R. J. M., and Giphart, M. J. (1990) *Immunogenetics* 31, 37-44.
5. Van der Poel, J. J., Groenen, M. A. M., Dijkhof, R. J. M., Ruyter, D., and Giphart, M. J. (1990) *Immunogenetics* 31, 29-36.
6. Stone, R. T., and Muggli-Cockett, N. E. (1990) *Anim. Genet.* 21, 353-360.
7. Sigurdardottir, S., Borsch, C., Gustafsson, K., and Andersson, L. (1991) *Anim. Genet.* 22, 199-209.
8. Sigurdardottir, S., Lunden, A. and Andersson, L. (1988) *Anim. Genet.* 19, 133-150.
9. Flavell, R. A., Allen, H., Huber, B., Wake, C., and Widera, G. (1985) *Immunol. Rev.* 84, 29-50.
10. LeGuern, C., Weissman, J. D., Marche, P. N., Jouvin-Marche, E., Laverriere, A., Bagnato, M. R., and Kindt, T. J. (1987) *Immunogenetics* 25, 104-109.
11. Watters, J. W. F., Locker, J. D., Kunz, H. W., and Gill, T. J. (1987) *Immunogenetics* 26, 220-229.
12. Auffray, C., Lillie, J. W., Korman, A. J., Boss, J. M., Frechin, N., Guillemot, F., Cooper, J., Mulligan, R. C., and Strominger, J. L. (1987) *Immunogenetics* 26, 63-73.
13. Scott, P. C., Maddox, J. F., Gogolin-Ewens, K. J., and Brandon, M. R. (1991) *Immunogenetics* 34, 80-87.
14. Andersson, G., Larhammar, D., Widmark, E., Servenius, B., Peterson, P. A., and Rask, L. (1987) *J. Biol. Chem.* 262, 8748-8758.
15. Xu, A., Mckenna, K., and Lewin, H. A. (1993) *Immunogenetics* 37, 231-234.
16. Nishino, Y., Tajima, S., and Aida, Y. *Immunogenetics*, in press.
17. Aida, Y., Kohda, C., Morooka, A., Nakai, Y., Ogimoto, K., Urao, T., and Asahina, M. (1994) *Biochem. Biophys. Res. Commun.* 204, 195-202.
18. Aida, Y., Niimi, M., Asahina, M., Okada, K., Nakai, Y., and Ogimoto, K. (1995) *Biochem. Biophys. Res. Commun.* 209, 981-988.
19. Inoko, H., Ando, A., Kimura, M., and Tsuji, K. (1985) *Present and Future Development of Transplantation.*, pp. 188-220, Tokai University Press, Tokyo.

20. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
21. Pearson, W. R., and Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444-2448.
22. Saitou, N., and Nei, M. (1987) *Mol. Biol. Evol.* 4, 406-425.
23. Figueroa, F., and Klein, J. (1986) *Immunol Today* 7, 78-81.
24. Korman, A. J., Auffray, C., Schamboeck, A., and Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6013-6017.
25. Orr, H. T., Lancet, D., Robb, R. J., Lopez de Castro, J. A., and Strominger, J. L. (1979) *Nature* 282, 266-270.
26. Cosson, P., and Bonifacino, J. S. (1992) *Science* 258, 659-662.
27. Trowsdale, J., Young, J. A. T, Kelley, A. P., Austin, P. J., Carson, S., Meunier, H., So, A., Erlich, H. A., Spielman, R. S., Bodmer, J., and Bodmer, W. F. (1985) *Immunol. Rev.* 85, 5-43.
28. Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L., and Wiley, D. C. (1993) *Nature* 364, 33-39.
29. Dikiniene, N., and Aida, Y. (1995) *Immunogenetics*, 42, 75.
30. Fabb, S. A., Maddox, J. F., Gogolin-Ewens, K. J., Baker, L., Wu, M-J., and Brandon, M. R. (1993) *Anim. Genet.* 24, 249-255.
31. Auffray, C., Korman. A. J., Roux-Dosseto, M., Bono, R., and Strominger, J. L. (1982) *Proc. Natl. Acad. Sci. USA* 79, 6337-6341.
32. Auffray, C., Lillie. J. W., Arnot, D., Grossberger, D., Kappes, D., and Strominger, J. L. (1984) *Nature* 308, 327-333.
33. Chang, H. C., Moriuchi, T., and Silver, J. (1983) *Nature* 305, 813-815.
34. Schenning, L., Larhammar, D., Bill, P., Wiman, K., Jonsson, A. K., Rask, L., and Peterson, P. A. (1984) *EMBO J.* 3, 447-452.
35. Moriuchi, J., Moriuchi, T., and Silver, J. (1985) *Proc. Natl. Acad. Sci. USA* 82, 3420-3424.
36. Marsh, S. G. E., and Bodmer, J. G. (1993) *Immunogenetics* 37, 79-94.
37. Jonsson, A.-K., and Rask, L. (1989) *Immunogenetics* 29, 411-413.
38. Koppelman, B., and Cresswell, P. (1990) *J. Immunol.* 145, 2730-2736.
39. Hirsch, F., Sachs, D. H., Gustafsson, K., Pratt, K., Germana, S., and LeGuern, C. (1990) *Immunogenetics* 31, 52-56.
40. Sarmiento, U. M., DeRose, S., Sarmiento, J. I., and Storb, R. (1992) *Immunogenetics* 35, 416-420.
41. Barran, P. A., and McMaster, W. R. (1987) *Immunogenetics* 26, 56-62.
42. Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams, V. E., and McDevitt, H. O. (1983) *Cell* 34, 169-177.