Characterization of the Bovine Major Histocompatibility Complex

By
Monica McArthur

Submitted to the
Office of Honors Programs and Academic Scholarships
Texas A&M University
In partial fulfillment of the requirements for

1998-99 University Undergraduate Research Fellows Program

April 15, 1999

Approved as of style and content by:

Loren Skow (faculty advisor)

Susanna Finnell, Executive Director
Honors Programs and Academic Scholarships

Abstract

Characterization of the Bovine Major Histocompatibility Complex

Monica McArthur, (Loren Skow) University Undergraduate Fellow, 1998-1999,
Texas A&M University, Department of Veterinary Anatomy and Public Health,

Genes within the major histocompatibility complex (MHC) are involved in immune response to infectious agents, tumor metastasis, stress response, gametogenesis, and development, including embryogenesis. Therefore, characterization of the organization and content of this region will aid in the identification and isolation of genes that can be used to increase disease resistance and improve health and productivity in cattle.

In both mice and humans, the organization of the MHC is conserved and found in a small, continuous region of about 300Mb of DNA. However, genes of the bovine MHC are found in two clusters on chromosome 23 separated by approximately 20Mb of DNA. The disruption of the organization of the bovine MHC was caused by a large inversion that apparently is common to all artiodactyl species.

To better define the breakpoints of the inversion, we screened a bacterial artificial chromosome (BAC) library and isolated a clone (BAC 27) that contains both the centromeric gene glutamyl cystein synthetase (GCS) as well as an MHC gene, DYA. Therefore, the centromeric breakpoint for the inversion is contained in BAC 27 between GCS and DYA.

Using sequences obtained from BAC 27 subclones several bovine MHC genes involved in the inversion can be identified. These genes may play an important role in immune response and productivity in cattle and will increase our understanding of the organization of the bovine MHC. Comparing the organization of the bovine MHC with the organization found in mice, humans, and other animals, will help determine the evolutionary history of this important region of the genome.

Introduction

The major histocompatibility complex (MHC) is the genetic region that contains genes involved in immune response to infectious agents, the discrimination between self and non-self, tumor metastasis, response to stress, gametogenesis, and development, including embryogenesis. Most MHC genes function in the immune response. The genes of the mammalian MHC are classified into three groups designated class I genes, class II genes, and class III genes (Figure 1). The class I and class II genes of the MHC encode antigen-presenting proteins involved in the immune response. The class I and class II proteins are anchored to the cellular membrane and present foreign material, usually peptides, for recognition by the T-cells. The class I genes are also known as the transplantation antigens. These gene products are found on almost all cells, and are required for cell-mediated immunity by their interaction with T-cell receptors on killer T-lymphocytes. Like the class I gene products, the class II gene products are anchored to the cellular membrane and present foreign material for recognition by T-cells. However, the class II MHC molecules are located only on the lymphocytes and macrophages (cells of the immune system). MHC class II molecules are responsible for antigen presentation in the humoral immune response of B-lymphocytes that secrete immunoglobulins. Also included among the class II genes are antigen-processing genes such as *LMP-2* and *LMP-7* (low molecular weight protease), proteases involved in breaking down intracellular proteins for presentation by MHC molecules. Also included in the third class of MHC genes, class III, are the genes encoding components of the complement cascade. The complement proteins interact with antigen-antibody complexes to initiate cell lysis. Additionally, there are several other genes that appear unrelated to the immune response

that are also located within the MHC region. These include genes for heat shock proteins, tumor necrosis factors, and hormone metabolism.

Because of the biological potency of the genes in the MHC region, it is important to develop an understanding of its organization and contents. Genes isolated from this region have been found to increase immune response to infectious agents, improve productivity, and reproduction in chicken, pigs, humans, mice, and cattle.

Although somewhat different from the MHC in mammals, the MHC in chickens also contains genes responsible for immune response to a variety of foreign agents. The MHC in chickens, B, is composed of several clusters of highly polymorphhic genes. The protein structure of the class I and class II antigens is similar to the overall structure of these proteins in mammals. However, the MHC in chickens also includes a class IV region where the B blood group antigens are located. (Miller et al)

The porcine MHC, SLA (swine leukocyte antigen), is located on pig chromosome 7, but class II SLA genes are located on the opposite side of the centromere from the tightly linked class I and class III groups. Many genes have been isolated from the class I and III regions; however, there is still little known about the contents and organization of the class II region. (Vaiman et al)

The MHC is well characterized in humans and mice. The human MHC region (HLA or human leukocyte antigen) is located on the short arm of chromosome 6. The groups of genes are located in the order II, III, I, and cover approximately 4Mb of DNA. Most of the genes of the HLA have been identified, mapped, and sequenced within each group of genes. The murine MHC region (H-2) is also well characterized, and displays a very similar organization to that of the HLA. H-2 is located on mouse chromosome 17.

Like the HLA, the H-2 region is confined to a approximately 4Mb of DNA. The order of the groups of genes in mice is I, II, III, I, I. Unlike the human MHC, H-2 contains three distinct groups of class I genes (Figure 1).

The MHC region in cattle (BoLA or bovine leukocyte antigen) is not as well characterized as that of mice or humans, but shows some similarity in its organization. The BoLA region is found on bovine chromosome 23 in two distinct clusters separated by approximately 20 Mb of DNA (figure 1). This is unlike the murine and human MHC regions that are tightly linked into a relatively small region of DNA. The basic order of the gene groups II, III, I is conserved. However, the BoLA class II region has been disrupted so that an additional cluster of class II genes (the class IIb genes) are found near the centromere, 20 Mb away from the other MHC genes. Mapping of bovine chromosome 23 indicates that the disruption of the organization of the bovine MHC was caused by a large inversion that apparently is common to all artiodactyl species.

To better define the breakpoints of the inversion, we screened a bacterial artificial chromosome (BAC) library and isolated a clone (BAC27) that contains both the centromeric gene glutamyl cysteine synthetase, GCS (not an MHC gene) and the MHC class IIb gene DYA. In all taxa other than ruminants, GCS maps distant from MHC genes or on a different chromosome from MHC genes. Therefore, the centromeric breakpoint for the inversion is most likely contained in BAC27 (Figure 2).

The functions of many of the genes located in the MHC regions of other species have been identified and found to be important in disease resistance, health and productivity. Identification and isolation of genes in cattle that can be used to increase disease resistance, health and productivity will be highly beneficial. We also hope to

further characterize the centromeric breakpoint of the BoLA inversion to help understand the functional correlates of MHC organization and the evolutionary events that have shaped the contemporary MHC of mammals. In order to achieve these goals, we have employed a variety of methods including exon trapping, nested deletion, shotgun cloning, hybridization, and sub-library formation.

Methods

Exon Trapping

In order to identify functional genes homologous to those found in HLA and H-2, we first tried the exon trapping system. The purpose of this system is to selectively identify, isolate, and amplify exons from complex DNA. The exon trapping system uses a modified plasmid vector that selects for functional exon splice sites (Figure 3). We screened a BAC library and identified 15 clones that contain BoLA sequences. These BACs were digested with *Eco* RI, *Bam* HI, and *Sau* 3A restriction endonucleases to create a variety of fragments of different lengths. The digested BAC DNA was then ligated into the HIV1-tat gene flanked by viral splice sites and exons of the pSP3 vector. The vector requires a eukaryotic host in order to process the RNA and remove introns. The cell line used for these experiments was the COS 7 African monkey liver cell line. The cells were grown at 35°C in a 5 % CO2 incubator in Dulbecco's modified Eagle medium (DMEM) supplemented with 10 % fetal bovine serum, glutamine, nonessential amino acids, and penicillin-streptomycin. Sixty-percent confluency is required for transfection using either electroporation or cationic lipid LipofectACE (Life Technologies/GIBCO-BRL). After transfection, if an entire exon and flanking intron sequence is included in a fragment, the exon will be found in the mature polyA+ cytoplasmic RNA (Figure 4). Isolation of the mRNA from these cells allows PCR amplification of the exons for subcloning and sequencing.

The plasmid vector, pSP3, contains modifications that allow for the preferential incorporation of exons from complex DNA sources (BACs) and elimination of vector only clones and false positives. (Nisson et al) The HIV1-tat gene has been introduced

into the plasmid and selects for functional splice sites in BAC's. Vectors that do not incorporate exons from the BACs need to be eliminated. This is achieved by the introduction of *Bst* XI half-sites (Figure 5). *Bst* XI has an interrupted palindromic recognition site that allows the conservation of splice site integrity while introducing a selective feature at the same location. If an insert is spliced into the HIV-tat exon, the *Bst* XI recognition site is interrupted in the processed RNA. However, failure of the plasmid to incorporate an exon will lead to an intact *Bst* XI site. Processed RNA can be isolated and converted to double stranded complimentary DNA (cDNA) by reverse transcriptase polymerase chain reaction (RT-PCR). Restriction digest of this cDNA with *Bst* XI will cleave only vector copies with an uninterrupted *Bst* XI site (those that do not contain insert DNA) eliminating the most predominant form of false positives. PCR is then used to amplify the vector + exon samples before cloning. To facilitate cloning and add directionality, restriction sites were incorporated into the vector that will be present in the final PCR product.

Nested Deletions

DNA isolated from BACs known to contain BoLA sequence was digested with *Eco* RI. The resulting fragments were ligated into pUC19 (Figure 6) for amplification. PCR amplification was performed using universal primers based on abundant restriction enzyme sequences. The PCR products were then partially digested with *Sau* 3A. Each *Sau* 3A partial digest was divided into two aliquots and digested to completion with the rare cutters *Xho* I or *Not* I. These double digests were directionally cloned into pBS

(Figure 7). These pBS ligations were transformed into NovaBlue Competent Cells. DNA was isolated from 10 of the transformed colonies for secondary PCR amplification.

Shotgun Cloning

DNA was isolated from the BAC clones identified in the library screen for BoLA sequence. Originally, an alkaline lysis mini-prep (5-ml liquid culture) was used for BAC DNA isolation; however, the DNA yield was insufficient for further experiments. Therefore, the QIAGEN Midi Kit Protocol was used to isolate higher concentrations of DNA from a subset of the 15 BACs containing BoLA sequence. The yield from this protocol was approximately 20 ug. BAC 27, known to contain the centromeric breakpoint of the BoLA class IIb inversion was partially digested with *Sau* 3A for 5 minutes at room temperature. The fragments generated by this digestion were ligated with Ready-To-Go pUC18 *Bam* HI/BAP + ligase (Pharmacia Biotech) and transformed into NovaBlue Competent Cells (Novagen). DNA was isolated from 100 of the transformed colonies using the alkaline lysis plasmid mini-prep. Approximately 20 of the DNA samples, representing BAC 27, were sequenced using the M13 reverse primer. The sequence from these sub-clones was then entered into Gene-Bank for comparison with known sequences.

Hybridizations

In addition to isolating DNA from the transformed colonies, the colonies were transferred to nitrocellulose filters, lysed with an alkaline pH, and baked onto the filters. These filters were then hybridized with ^{32}P radioisotope probes of GCS, TAP1, and

LMP2 sequences. The probes were made using the Random Prime Labeling Kit (Boehinger Manneheim). The filters were probed overnight at 37°C. Low stringency wash conditions were used to remove unincorporated probe, and the filters were exposed on film to detect hybridization with the probe DNA.

BAC 27 Sub-library

A sub-library was created with Bam HI digests of DNA isolated from BAC 27 by the QUIAGEN midi-prep protocol. The individual fragments were cut out of a 1 % low gelling temperature agarose gel and ligated with Ready-To-Go pUC18 Bam HI/IAP (Pharmacia Biotech) (Figure 6). Each ligation was then transformed into NovaBlue Competent Cells. Transformed colonies were selected for DNA isolation and sequencing. In order to keep the concentration of salt low, N-laurylsarcosamine was substituted for SDS (sodium dodecyl sulfate) during the lysis stage of the plasmid mini-prep protocol. This method gave very clean (low salt) DNA; however, the DNA pellet was very light and difficult to see and was often lost during precipitation. Therefore, it took several attempts to actually isolate the DNA.

Results and Discussion

Exon Trapping

The digested BAC DNA was ligated successfully with pSP3. However, unfortunately, the growth rate of the COS 7 cells was substantially slower than originally expected. This method was originally chosen because sequencing only the expressed sequences was thought to be more efficient than trying to sequence through large segments of unexpressed sequence. However, the pSP3 vector can take inserts of only 2-3 kb with a relatively high efficiency. This limited and possibly selected cloning of inserts makes this system substantially less efficient and necessitated the use of digestions with multiple restriction endonucleases in order to insure that all of the sequences would be included in the mRNA population isolated from the COS 7 cells. Such a large number of clones required an even larger number of cells for transfection. The growth rate of the COS 7 cells was not what we had expected, and after 5 months of trying to grow the cells to a high enough density to transfect, we decided to try another method.

Shot-gun Cloning

The sequences obtained from the 20 analyzed sub-clones showed sequence homology to *E. coli* and small repetitive bovine sequences. This was due to preferential ligation between the vector and the smallest fragments of DNA included in the ligation. Apparently there were many small fragments of *E. coli* genomic DNA that contaminated the BAC digest. The clones that did not contain genomic DNA fragments, contained small fragments of repetitive bovine sequences. In order to reduce the amount of small fragments in the ligations we made a sub-library using *Bam* HI restriction fragments.

Hybridization

Development of the films exposed to the BAC 27 sub-clone filters after hybridization with TAP 1 and LMP 7 showed no positive colonies. This is consistent with the current model of the region surrounding the hypothesized centromeric breakpoint and the location of the human homolog to DYA.

Sub-library

DNA was isolated from the colonies transformed with pUC18/BAC 27 *Bam* HI digest. Several attempts were made to sequence this DNA with the M13 forward primer, but no sequence was obtained. Most likely, the DNA concentration was not high enough due to the use of N-laurylsarcosamine during the lysis stage of the plasmid mini-prep protocol.

Future Direction

Since we began this project, a new BAC DNA isolation protocol has been developed. DNA isolated by this protocol is sufficiently pure to allow direct cycle sequencing of the BAC DNA. Therefore, "primer walking" along the sequence of a BAC is now possible without subcloning. Each sequencing reaction provides approximately 400 base pairs of sequence from each end of the BAC insert for a total of 800 base pairs per sequencing reaction. Primers can be designed from the end of the newest sequence, and another set of sequencing reactions performed (Figure 8a). The insert in BAC 27 is approximately 130,000 base pairs (130kb) in length. At a rate of 800bp per sequencing reaction, it will take about 200 sets of primers and sequencing reactions. Synthesis and shipping time for the primers will be the rate-limiting step in this method. However, if primers are designed from known internal sequences such as the genes GCS and DYA, then six primers can be used simultaneously to increase the speed of the "primer walking" to 2400 bp per set of sequencing reactions (Figure 8b). Although the number of primers and sequencing reactions required to sequence the entire BAC 27 insert remain the same, by sequencing from multiple sites, the time to completion of the BAC 27 sequences will be decreased dramatically.

Figure 1 Comparative maps of the HLA, H-2, and BoLA regions showing the grouping of the MHC genes into class I, class II, and slass III. The hypothesized inversion of the BoLA class IIb region is also shown.

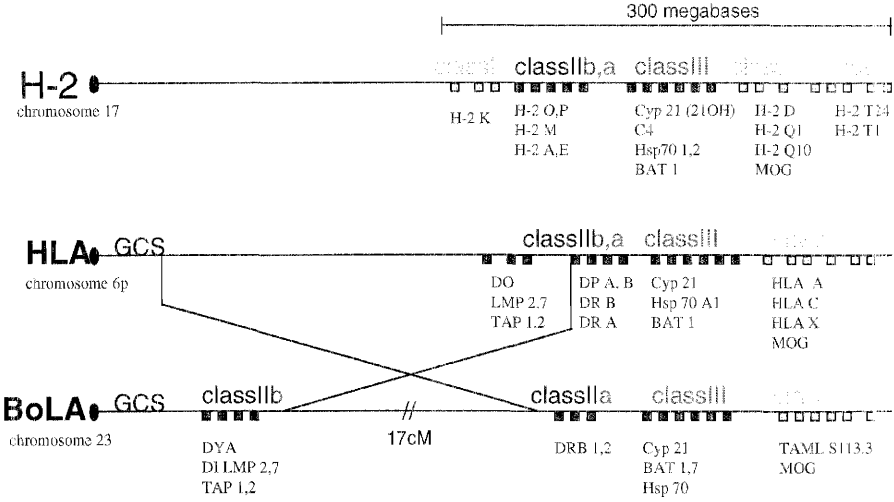
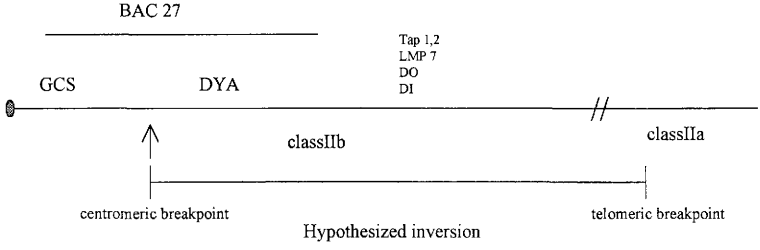


Figure 2 Hypothesized inversion of the BoLA class IIb region. The centromeric breakpoint for this inversion is found in BAC 27.



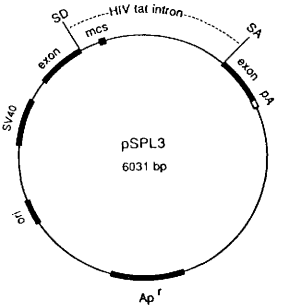


Figure 3 Map of pSPL3. The vector contains sequences that enable replication in *E. coli* and COS-7 cell hosts (bacterial and SV40 origins of replication are present). An ampicillin resistance marker allows for selection of subclones. An MCS (multiple cloning site) that interrupts the HIV-tat intron provides several restriction sites for the subcloning of genomic DNA. After transfection of COS-7 cells, transcription occurs at high levels facilitated by the SV40 promoter (SV40). Processing of the transcript results in removal of the HIV-tat intron via splicing in which the vector exons are combined at the splice donor (SD) and splice acceptor (SA) sites. Cytoplasmic RNA is polyadenylated [pA=SV40 poly (A) addition recognition sequence]. Exons are trapped from genomic DNA cloned into pSPL3 as a result of interaction of the vector splice sites (derived from HIV-tat) with splice sites flanking exons contained in genomic DNA. (Nisson et al)

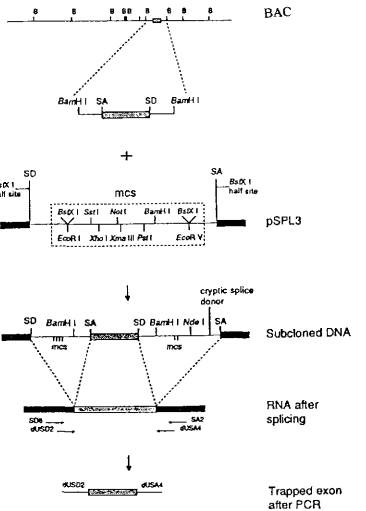


Figure 4 pSPL3 exon trapping of BAC DNA. BAC DNA is digested with a restriction enzyme. In this example, a single exon (shaded box) is contained with in a Bam HI fragment. The exon is flanked by splice acceptor (SA) and splice donor (SD) sites. The pSPL3 vector is prepared for subcloning at the same restriction site. The MCS of pSPL3 contains several restriction enzyme sites. The solid boxes represent exon sequences with in the vector. After subcloning genomic DNA into pSPL3, DNA is isolated and transfected into COS-7 cells. Total RNA is isolated for RNA-based analysis. (Nisson et al)

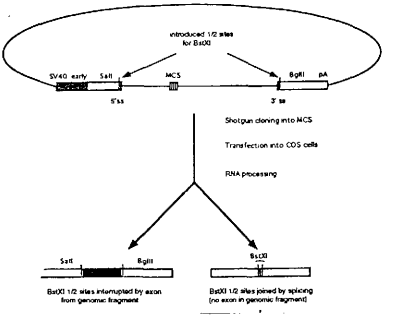


Figure 5 Schematic representation of pSPL3 and Bst XI restriction site generation as a result of RNA processing. Genomic fragments are cloned into the MCS located in the vector intron. Transient transfection into COS-7 cells allows for the production of RNA driven by upstream SV40 promoter sequences. Half restriction sites are located on opposite sides of the intron and are joined by RNA processing. RNA is converted to double-stranded cDNA followed by Bst XI digestion. (Church et al)

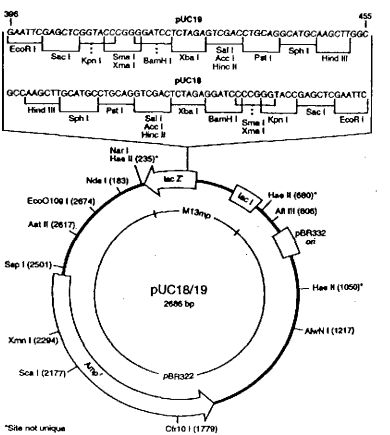
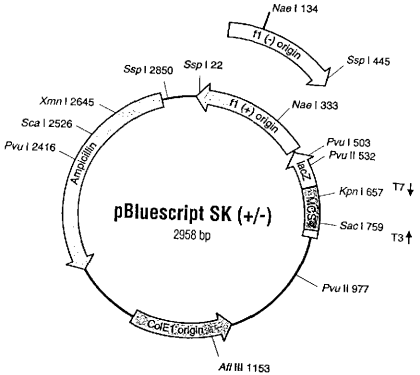


Figure 6 Map of pUC19 and pUC18 including multiple cloning site.



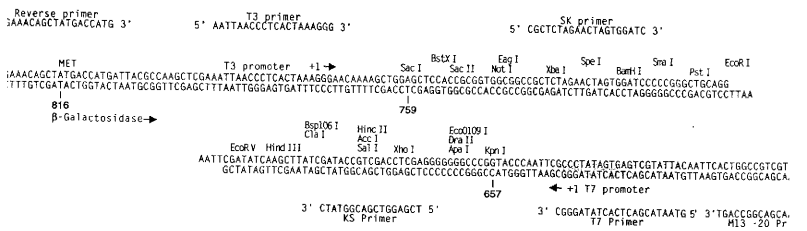


Figure 7 Map of pBS including the multiple cloning site sequence. Not I / Sau 3A and Xho I / Sau 3A double digests were used to directionally clone fragments into the pBS vector.

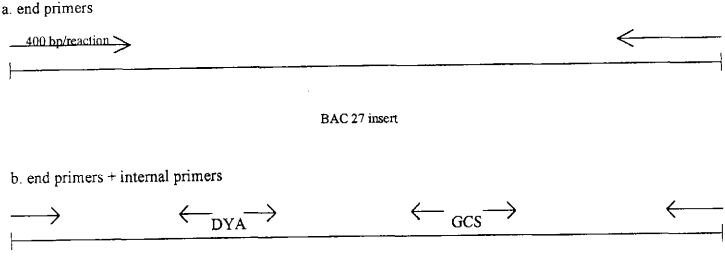


Figure 8 Direct sequencing of BAC 27 by "primer walking". (a) using only end primers to sequence allows two sequencing reactions before new primers must be designed (800 bp of sequence) (b) using two additional internal sequences (DYA and GCS) for sequences allows six sequencing reactions before new primers must be designed increasing the number of base pairs generatedfrom 800 to 2400 in the same amount of time.

References

Band M, Larson JH, Womack JE, Lewin HA (1995) A radiation hybrid map of BTA23: Identification of a chromosomal rearrangement leading to seperation of the cattle MHC class II subregions. *Genomics* 53: 269-275.

Bronson SK, Pei J, Taillon-Miller P, Chorney MJ, Geraghty DE, Chaplin DD (1991) Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. *Proc Natl Acad Sci* USA 88: 1676-1680.

Church DM, Stotler CJ, Rutter JL, Murrell JR, Trofatter JA, Buckler AJ. (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genetics* Jan 6: 98-105

Jarell VL, Lewin HA, Da Y, Wheeler MB (1995) Gene-centromere mapping of bovine DYA, DRB3, and PRL using secondary oocytes and first polar bodies: evidence for four-strand double crossovers between DYA and DRB3. *Genomics* May 1; 27 (1): 33-9.

Kasahara M (1997) New insights into the genomic organization and origin of the major histocompatibility complex: Role of chromosomal (genome) duplication in the emergence of the adaptive immune system. *Hereditas* 127: 59-65

Lamont, SJ (1998) The chicken major histocompatibility complex and disease. *Rev Sci Tech* Apr;17 (1): 128-42.

Lewin B, Genes VI. New York:Oxford University Press, 1997.

Lodish H, Baltimore D, Berk A, Zipursky SL, Matudaire P, Darnell J. Molecular Cell Biology. New York:Scientific American Books, 1995.

Miller MM, Goto RM, Taylor RL, Zoorob R, Auffray C, Briles RW, Briles WE, Bloom SE. (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex (MHC) NOR microchromosome and evidence for high-frequency recombination associated with the nucleolar organizer region *Proc Natl Acad Sci USA* Apr 30 93(9): 3958-3962

Muggli-Cockett NE, Stone RT (1991) Restriction fragment length polymorphisms in bovine major histocompatibility complex class II B-chain genes using bovine exon-containing hybridization probes. *Animal Genetics* 22: 123-136

Nisson PE, Ally A, Watkins PC. (1994) Protocols for trapping internal and 3'-terminal exons. *PCR Methods and Applications* Aug; 4 (1): S24-39

Stone RT, Muggli-Cockett NE (1993)BoLA-DIB: species distibution, linkage with DOB, and Northern analysis. *Animal Genetics* 24: 41-45.

Vaiman M, Chardon P, Rothschild MF (1998) Porcine major histocompatibility complex. *Rev Sci Tech* Apr;17 (1): 95-107.

van der Poel JJ, Groenen MAM, Dikhof RJM, Ruyter D, Giphart MJ (1990) The nucleotide sequencee of the bovine MHC class II alpha genes: DRA, DQA, and DYA. *Immunogenetics* 31: 29-36

Wolf, Stephen L. Introduction to Cell and Molecular Biology. Belmont:Wadsworth Publishing Company,1995.