Discussion
DISCUSSION

The BF gene produces class I molecules. The Class I molecules consists of MHC-encoded a chain or heavy polypeptide chain and non-MHC encoded β polypeptide chain i.e. β2 microglobulin or β2m. The β2 microglobulin is attached by a number of non-covalent bonds to the heavy chain. The a chain is comprised of five main regions or domains i.e. three extracellular domain i.e. N or α1, C1 or α2 and C2 or α3; a trans-membrane domain (polypeptide chain passes through the plasma membrane) and a cytoplasmic domain (within the cytoplasm of the cell). A groove, known as peptide binding region (PBR) is formed by the interaction of the α1 and α2 domains. This peptide binding region (PBR) has a floor of 8 stranded β pleated sheets with two opposite walls made by parallel strands of an α helix. The α1 contribute 4 strands of β pleated sheet and one α helix, while α2 contribute another 4 strands of β pleated sheet and one α helix. The greatest variability in amino acids occurs in the α1 and α2 sequences that form the groove that interacts with amino acids in the peptide fragments. This polymorphism among the class I MHC products creates variation in the chemical surface of the peptide binding groove. For any given MHC molecules, binding of a peptide usually requires the peptide to have one or more specific amino acids at a fixed position, frequently the terminal or penultimate amino acid of the peptide. Binding of the specific amino acid in the groove of the MHC molecule occurs in what is termed as anchor site (s). The other amino acids can be variable so that each MHC molecule can bind many different peptides. Other polymorphic residues of the MHC molecule are those in contact with the T cell receptor (TCR), which interacts with both peptide and the MHC molecules itself. The α3 is highly conserved and is homologous to
Ig constant domains and non-covalently bound β2m, an invariant molecule, also homologous to Ig constant domains. These two interacts with α1 and α2 domains to maintain their proper confirmation. The importance of the highly conserved region of α3 domain is that CD8, a molecule expressed on cytolytic T cells that recognizes class I MHC molecules, binding to this region.

5.1. Complete CDS of BF2 region in red jungle fowl

The complete CDS comprising of 1033 bp was obtained from the annotated sequence of 1235 bp. This 1235 bp annotated sequence was obtained by their joining the sequence of 673 bp fragment and 593 bp fragment amplified in RJF. The 1033 bp CDS was composed 64 bp of exon-1, 264 bp exon-2, 273 bp exon-3, 273 bp exon-4, 108 bp exon-5, 33 bp exon-6 and 18 bp exon-8. In chicken the complete CDS of BF2 gene was reported to be 1066 bp (Hunt and Fulton, 1998, Wallny et al., 2006 and Shaw et al., 2007). The CDS of RJF same to 33 bp was shorted then chicken and the deference was due to absence of exon-7. While was 33 bp in size except the absence of exon-7 CDS absence in RJF. All other exons showed no size variation with those of chicken.

In RJF the signal peptide was comprised of 21 amino acids, while the mature peptide was comprised of 323 amino acids. In chicken the signal peptide of 21 bp, the mature peptide of 334 amino acids was reported (Livant et al., 2004, Yan et al., 2005a). The mature peptide of RJF seems to be 11 amino acids shorter than the chicken and this difference is due to absence of 11 amino acids of cytoplasmic domain II. The 323 amino acids mature peptide was composed of 3 extra cellular domain i.e. α1 domain, α2 domain and α3 domain, transmembrane domain, cytoplasmic domain I and cytoplasmic domain III.
5.2. Polymorphism in BF2 between red jungle fowl and other chicken haplotypes

5.2.1. Nucleotide sequence variation

The nucleotide sequences of BF2 exons in different chicken B haplotypes (Table 4.1) were retrieved from database and studied for sequence homology within them selves as well as with the RJF. With in the chicken B haplotypes, exon-2 and exon-3 were found showed high polymorphism, showing 14.65-14.78 % polymorphism, where as the % polymorphism ranged from 3.03-6.48 % in other exons. Hunt and Fulton (1998) analyzed the polymorphic nature of eleven alleles expressed by the major class I locus (B-FIV) in chickens and reported high variability occur in exons-2 and 3 encoding the α1 and α2 domains. Singh (2009) reported comparatively lower percent nucleotide variation within the chicken sequences i.e. 7.85 in α1 domain and 9.16 % in α2 domain. The possible reason for this comparatively lower polymorphism within chicken sequences might be the lesser number of chicken sequences taken in his study i.e. only 6 against the 23 sequences taken in present study. Within chicken B haplotypes, the ratio of transition to transversion nucleotide substitutions was 1:1 in exon-2 and exon-3 but in exon-4 and exon-5, a bias for transition was observed as the ratio of transition to transversion was 5:1. Similarly, an upward bias for non-synonymous nucleotide substitutions was observed in exon-1, exon-2, exon-3 and exon-4 as the ratio of synonymous and Non-synonymous substitutions ranged from 1:2 in exon-4 to 1:9 in exon-3. In other exons, the ratio of synonymous and Non-synonymous nucleotide substitutions was close to 1:1.

Between RJF and chicken B haplotypes the percent polymorphism was maximum in exon-2 and exon-3 (15.15 and 15.38 % respectively) followed by exon-1 (9.4%). In other exons, the % polymorphism ranged from 3.03 to 6.48 %. While the ratio of transition to transversion nucleotide substitutions was 1:1 in exon-2 and exon-3, an upward bias for transition was observed in exon-4 and
exon-5 as the ratio of transition to transversion was ~ 6:1. Similarly, a highly upward bias for non-synonymous nucleotide substitutions was observed in exon-2 and exon-3 (1:5 to 1:9), while in other exons, this ratio ranged from (1.2:1 to 1:2).

5.2.2. Amino acid sequence variation

Within chicken B haplotypes, the percent amino acids polymorphism was maximum in a1 domain and a2 domain (27.27% and 26.37% respectively) followed by signal peptide (14.28%) and trans-membrane domain (11.11%). The a3 domain showed comparatively much lower percent amino acids variability i.e. 7.69 % as compared to a1 domain and a2 domain. While the cytoplasmic domain I was completely conserved, 16.66% polymorphism was observed in cytoplasmic domain III. Singh (2009) reported 15.91 % percent aa variability in a1 domain and 15.38 % percent aa variability in a2 domain. Lima-Rosa et al., (2004) reported high amino acid variability of 37 % in a1 as well as in a2 domain in chicken sequences.

Between RKF and chicken B haplotypes, a1 domain and a2 domain showed maximum percent amino acids polymorphism i.e. 27.27% and 26.37%, respectively, followed by 14.28 % in signal peptide and 11.11 in trans-membrane domain (11.11%). The a3 domain showed comparatively much lower percent amino acids variability i.e. 8.79 % as compared to a1 domain and a2 domain. Cytoplasmic domain I showed complete conservation, but 16.66 % polymorphism was observed in cytoplasmic domain III.

5.3. Polymorphism in BF2 between red jungle fowl and other poultry species

5.3.1. Nucleotide sequence variation

The exon-2 and exon-3 were found to be comprised of 264 nt and 273 nt, respectively in red jungle fowl. In chicken, the exon-2 and 3 were of the same size i.e. 264 nt and 273 nt, however other poultry species showed size variation in the exon-2 and exon-3. The respective
sizes of exon-2 and exon-3 were 270 and 273 in quail, 264 and 276 in duck and goose; and 270 and 269 (partial) in guinea fowl.

While red jungle fowl showed least sequence variability of 3.40 % in exon-2 and 5.49 % in exon-3, with chicken, much higher variability was observed between RJF and poultry species other than chicken i.e. 19.26 % to 27.65 % and 11.23 % to 25.72 % in exon-2 and exon-3, respectively. The ratio of transition to transversion nucleotide substitution was close to 1:1 between red jungle fowl and poultry species including chicken, however an upward biasness was observed for non-synonymous between the red jungle fowl and poultry species other than chicken as the ratio of synonymous to non-synonymous nucleotide substitution 1:2 to 1:4 in exon-2 and exon-3. Singh (2009) also reported high nucleotide variability between the poultry species i.e. 29.26 % to 43.70 % in exon-2 and (with duck) 17.47 % to 37.92 % in exon-3. He further reported an upward bias for non-synonymous substitution in exon-2 as well as in exon-3 between the poultry species.

5.3.2. Amino acid variability

The $\alpha_1$ domain and $\alpha_2$ domain were found to be comprised of 88 aa and 91 aa, respectively in red jungle fowl. In chicken, $\alpha_1$ domain and $\alpha_2$ domain were of the same size i.e. 88 aa and 91 aa, however other poultry species showed size variation in $\alpha_1$ domain and $\alpha_2$ domain. The respective sizes of $\alpha_1$ domain and $\alpha_2$ domain were 90 and 91 in quail, 88 and 92 in duck and goose; and 90 and 87 (partial) in guinea fowl. Earlier workers also reported 88 aa in $\alpha_1$ and 91 aa in $\alpha_2$ in chicken BF2 gene (Livant et al., 2004, Yan et al., 2005a). Xia et al., (2004) reported that the $\alpha_1$ and $\alpha_2$ domain of Anpl-UDA01 was made up of 88 and 92 aa respectively.

While red jungle fowl showed least sequence variability i.e. 6.81 % and 8.79 % in $\alpha_1$ domain and $\alpha_2$ domain, much higher variability was observed between RJF and poultry species other than chicken in these regions (30.00 % to 48.86 % and 17.39 % to 34.78 %,
respectively). Singh (2009) also reported high sequence variability ranging from 41.11 to 56.67 % in α1 domain and from 24.72 % to 53.93 % in α2 domain between the poultry species. Lima-Rosa et al., (2004) reported high amino acid variability of 37 % in α1 as well as in α2 domain in chicken sequences. Moon et al., (2005) compared the five MHC class I genes in ducks and found that majority of polymorphic amino acid residues were within these two domains with 42 polymorphic residues in α1 and 30 polymorphic residues within α2. Xia et al., (2004) reported much lower homology between the duck and chicken (55.2 – 64.6 %). Moon et al., (2005) reported the identity of the amino acid sequences ranges from 81.9% - 87.4% between different BF I genes in duck.

5.4. Structural and functional homology of RJF BF2 gene

5.4.1. With Chicken B haplotypes

RJF α1, α2 and α3 domains showed the overall conservation of structure of the PBR region. Two disulphide binding cysteines i.e. C99 and C161 in α2 domain and two disulphide binding cysteines i.e. C109 and C255 in α3 domain; a potential N-glycosylation site i.e. N85 in α1 domain were conserved. All the conserved residues interacting with the amino terminus of the bonds peptide in the HLA/H2 PBS i.e. Y7, Y59, Y159 and Y171 were conserved in RJF also i.e. Y7, Y58, Y156 and Y168. Three of the four conserved residues, which interacts with the carboxyl terminus of the peptides in HLA/H2 i.e. T143, K146 and W146 were also conserved in RJF i.e. T140, K143 and W144. However, the residue Y84 was not conserved and replaced by R83 in red jungle fowl and chicken.

Between RJF and chicken B haplotypes, majority of polymorphism residue (~ 70%) was observed in and around the region that from a helix as compared to 20 % polymorphism in β strand of α1 domain, but in α2 domain β strand region showed comparatively more polymorphism 60% in comparison as compared to that observed in a helix region (40%). Lima-Rosa et al., (2004) also reported that the
majority of the polymorphic residues (73%) in $\alpha_1$ domain occurred in or around the region that form $\alpha$ helix, while in $\alpha_2$ domain, 65% of the polymorphic residues were in or around the region that forms the four $\beta$ strands.

Among the presumed sites interacting with peptide, 45% were polymorphic in $\alpha_1$ domain, while in $\alpha_2$ domain, only 21.25% were polymorphic. Among the presumed sites coming in contact with $\beta$ micoglobulin, only 10% of sites were polymorphic in each domain. However for the sites interacting with T cell receptor, 40% were polymorphic in $\alpha_2$ domain, while in $\alpha_1$ and $\alpha_3$ domains, only 16.66% were polymorphic. Lima-Rosa et al., (2004) reported that approximately 33% polymorphic sites in $\alpha_1$ domain and 32% in $\alpha_2$ domain were peptide binding sites.

5.4.2. With different poultry species

Red jungle fowl BF2 gene showed the overall conservation of $\alpha_1$ and $\alpha_2$ domains. Two disulphide binding cysteines i.e. C$^{99}$ and C$^{161}$ in $\alpha_2$ domain and a potential N-glycosylation site i.e. N$^{85}$ in $\alpha_1$ domain were conserved. Sequence alignment with mammalian MHC showed that many of the conserved features of mammalian MHC class I glycoprotein was conserved in guinea fowl. Out of the four conserved residues interacting with the amino terminus of the bonds peptide in the HLA/H2 PBS i.e. Y$^7$, Y$^{59}$, Y$^{159}$ and Y$^{171}$ three were conserved in red jungle fowl and different poultry species i.e. Y$^7$, Y$^{58}$ and Y$^{156}$, while Y$^{168}$ was replaced by H in duck. Three of the four conserved residues, which interacts with the carboxyl terminus of the peptides in HLA/H2 i.e. T$^{143}$, K$^{146}$ and W$^{146}$ were conserved in red jungle fowl and different poultry species i.e. i.e. T$^{140}$, K$^{143}$ and W$^{144}$, except in quail, where K143 has been replaced by A. Residue Y$^{84}$ was not conserved and replaced by R$^{85}$ in red jungle fowl and different poultry species except in quail, where it has been replaced by L.

In chicken, the conserved pattern in $\alpha_1$ and $\alpha_2$ domain with their mammalian counter parts have been reported by various
workers. Kaufman et al. (1992) reported that both chicken $\alpha_1$ and $\alpha_2$ domains resembles mammalian class I molecules. Some conserved residues with important functions were seven residues that binds the ends of the peptide, two residues that bind CD8 and three residues that are phosphorylated. Hunt and Fulton (1998) reported many conserved features of mammalian class I gene in chicken BFIV. These included the residues interacting with amino terminal of the bound peptide ($Y^7$, $Y^{58}$, $Y^{156}$ and $Y^{168}$) and three conserved residues, which interact with the carboxyl terminal of the peptide ($T^{140}$, $K^{143}$ and $W^{144}$). The $R^{83}$ residue is not conserved in chicken.

Yan et al., (2005a,b) reported the conserved cysteins at 99th and 161st positions involved in intra chain disulfide bond and a potential N-glycosylation site at 85-87 position in $\alpha_2$ domain. The overall conservation of BF2 could be observed within the sequences, which also displayed relative conservation involved in PBD, CD8+ interaction sites and $\beta_2m$ contact sites. These included $Y^7$, $Y^{58}$, $Y^{156}$ and $Y^{168}$, which interacts with amino terminal of the bound peptide and $T^{140}$, $K^{143}$ and $W^{144}$, which interact with the carboxyl terminal of the peptide. The $R^{83}$ residue is not conserved in chicken.

Xia et al., (2004) showed that in duck MHC class I $\alpha_1$ and $\alpha_2$ domains retained the various conserved features such as cysteins at 99th and 162nd positions and potential N-glycosylation site at positions 85-87. These residues which interacts with amino terminal of the bound peptide i.e. $Y^7$, $Y^{58}$, $Y^{156}$ and $Y^{168}$ and the residues interacting with the carboxyl terminal of the peptide i.e. $T^{140}$, $K^{143}$ and $W^{144}$ were found to be conserved in duck. The $R^{83}$ residue is not conserved in duck as it was also not conserved in chicken also.

Moon et al., (2005) compared the MHC class I sequences from duck with the amino acid sequences for hallmarks of classical MHC class I genes. Among residues predicted to be involved in peptide anchoring ($Y^7$, $Y^{59}$, $Y^{84}$ (R$^{84}$ in non-mammalian vertebrates), $Y^{123}$, $T^{143}$, $K^{146}$, $W^{147}$, $Y^{159}$, and $Y^{171}$), all were conserved in most loci. The $Y^{123}$ sequence is replaced by phenylalanine or leucine in all five duck
genes, but is also seen in other avian sequences and many other species (45). The UDA did not have the conserved Y\textsuperscript{171}; however, histidine at this residue (H\textsuperscript{169}) was seen in other mammalian classical MHC class I sequences. All sequences have the site for N-linked glycosylation at N\textsuperscript{85}.

Shiina et al., (1999) reported that MHC class I amino acid sequences i.e. Coja-C and Coja-D in quail retained characteristic features required for functioning as antigen presenting glycoproteins such as cysteine residues at positions 101\textsuperscript{st} and 163\textsuperscript{rd} for interdomain disulfide bond formation, a single conserved N-linked glycosylation site at N\textsuperscript{87}, G\textsuperscript{88}, S\textsuperscript{89}, three potential sites for phosphorylation at Y\textsuperscript{315}, S\textsuperscript{325} and S\textsuperscript{328}, conserved residues at 229, 230 and 232-236 for β\textsubscript{2}m contact sites and at positions 180, 191 and 207-209 for α\textsubscript{1}, α\textsubscript{2} and α\textsubscript{3} domain interaction sites and at positions 221-227 for CD8 binding sites. At least 35 amino acid residues were predicted to be involved in formation of putative A-F pockets at the peptide binding region sites in the α helix and β sheet of the α\textsubscript{1} and α\textsubscript{2} domains of the Coja class I molecules. Among them, the eight invariant residues (Y\textsuperscript{7}, Y\textsuperscript{59}, Y\textsuperscript{84} or R\textsuperscript{84}, T\textsuperscript{143}, K\textsuperscript{146}, W\textsuperscript{147}, Y\textsuperscript{159} and Y\textsuperscript{171}) that anchor the NH\textsubscript{2}- and COOH-termini of the antigenic and form pockets A and F are conserved in classical class I molecules of chicken, mouse and human and quail.

Xia et al., (2005) cloned goose MHC class I cDNA (Ancy-MHC I) and found that the mature peptides of Ancy-MHC I cDNA encoded 333 amino acids. The genomic organization is composed of eight exons and seven introns. Based on the genetic distance, six Ancy-MHC I genes from six individuals can be classified into four lineages. A total of nineteen amino acid positions in peptide-binding domain showed high scores by Wu-kabat index analysis. The Ancy-MHC I amino acid sequence displayed seven critical HLA-A2 amino acids that bind with antigen polypeptides and have an 85.4-98.9% amino acid homology with each genes and a 59.8-66.0% amino acid homology with chicken MHC class I.
5.5. Genetic distances

5.5.1. Genetic distances between red jungle fowl and chicken B haplotypes

Between the red jungle fowl and different chicken B haplotypes, genetic distances were very low and ranged from 0.046 to 0.087. Similarly, the genetic distances (Poisson correction) were estimated using the cumulative amino acid variability in \( \alpha_1 \) domain and \( \alpha_2 \) domain between the red jungle fowl and different chicken B haplotypes were very low and ranged from 0.081 to 0.157. Using the pair wise genetic distances based on nucleotide as well as amino acid variability in \( \alpha_1 \) and \( \alpha_2 \) domains between red jungle fowl and chicken B haplotypes, the phylogenetic tree was constructed. Both the phylogenetic trees revealed that RJF cluster along with the chicken B21 haplotype. The very high closeness between RJF and B21 haplotype is extremely important as the B21 haplotype is resistant to M.D. Singh (2005) reported RJF sequence showed maximum similarity with BW1 haplotype of RJF, followed by B21 haplotype, irrespective of its genetic background i.e. either it is from White Leghorn or Thai native chicken based on sequence homology of \( \beta_1 \) domain of BLB2 gene. Allele B21 is strongly associated with resistance to MD, whereas B1 exhibited poor fitness (Kaffmann et al., 1995). Higher similarity of BLB2 gene from RJF with the similar sequence from B21 haplotype suggest the higher disease resistance ability of RJF.

5.5.2. Genetic distances between red jungle fowl and different poultry species

Red jungle fowl showed very low genetic distances with chicken based on either cumulative nucleotide sequence variability in \( \alpha_1 \) domain and \( \alpha_2 \) domain (0.047) or based on cumulative amino acid variability in \( \alpha_1 \) domain and \( \alpha_2 \) domain (0.546). However, between RJF and poultry species other than chicken, the estimates ranged from 0.206 to 0.345, when based on nucleotide variability and from 0.318 to 0.546, when based on amino acid variability. Based on nucleotide
sequence variability of β1 domain of BLB2 gene, Singh et al., (2005a) reported showed maximum sequence homology between RJF and chicken, followed by that with pheasant, while the duck was most distant from RJF.

Between the poultry species, estimates ranged from 0.200 between duck and goose to 0.404 between quail and goose, when based on nucleotide variability and from 0.273 between duck and goose to 0.596 between quail and goose as well as between quail and duck, when based on amino acid variability. Singh (2009) also reported the genetic distances between the poultry species ranging from 0.186 between duck and goose to 0.401 between quail and duck using nucleotide sequence variability in α₁ and α₂ domains, while the estimates ranged from 0.307 between chicken and quail to 0.602 between quail and duck, when based on cumulative amino acid variability in α₁ domain and α₂ domain.

Phylogenetic tree constructed by using pair wise genetic distances based on nucleotide variability as well as on amino acid variability revealed two major clusters, comprising of guinea fowl, quail, chicken and RJF in one, while duck and goose in other. In first cluster, RJF grouped with chicken. Singh (2009) also revealed the clustering of duck and goose separately from other poultry species. In first cluster, guinea fowl make separate branch, while chicken and quails are clustered together. Shiina et al., (1999) reported the separate clustering of MHC class I sequences from quail and chicken. Xia et al., (2004) found that duck MHC class I clusters quite distantly from chicken MHC.

5.6. PCR RFLP studies

The primers designed were successful to amplify the 599 bp fragment comprising most of exon-2, intron-2 and half of the exon-3 in red jungle fowl from genomic DNA. While the PCR RFLP of 599 bp BF2 fragment with Hae II revealed the polymorphic profile in red jungle fowl, the PCR RFLP with Xho I could not detect polymorphism.
The genotypic frequencies of AA, BB and AB genotypes observed in PCR RFLP with *Hae* II were 0.50, 0.17 and 0.33, respectively, showing low amount of heterozygocity. Reports on PCR RFLP of BF2 gene in red jungle fowl are lacking in literature, but the PCR RFLP of 277 bp of β1 domain of BLB2 in RJF was reported by several workers. Singh (2005) used PCR RFLP to detect polymorphism in β1 domain of BLB2 gene using *Hae* III, *Rsa* I and *Taq* I. While, *Hae* III could not detect any polymorphism in RJF population, *Rsa* I and *Taq* I showed polymorphic RE profile. PCR-RFLP profile with *Rsa* I identified two genotypes i.e. AA and AB in RJF population with the frequency of 0.66 and 0.33, respectively, while the frequencies of the genotypes identified with PR-RFLP with *Taq* I i.e. AA and AB was 0.83 and 0.17, respectively. It showed the low amount of heterozygocity in RJF population. Singh *et al.*, (2005b) also studied PCR-RFLP in exon-2 of BLB2 gene with *Taq* I and *Pst* I restriction enzymes in RJF and white Leghorn. No PCR-RFLP polymorphism was observed with *Taq* I, however *Pst* I detected the polymorphism. They reported the presence of high hetrozygocity in White Leghorn, which might be due to increased variability during the process of diversification.

Most of the studies on PCR RFLP involved BLB2 gene and are in chicken, however, Ewald *et al.*, (2007) done PCR RFLP in BF2 gene and reported significant BF2 associations with a subset of traits were observed in two commercial broiler lines. The BF2*21* allele was positively associated with antibody titre to infectious bursal disease virus in both lines. Other associations were line-specific. Using BLB2 gene in PCR RFLP studies in chicken, Ahmed (2001) could not get the polymorphism with, using *Taq* I and *HaeIII* restriction enzymes, between the divergent lines for antibody response to SRBC as well as CMI to mitogens. Similarly, Shivakumar (2003) studied the polymorphism in the β1 exon of the BLB II gene in the divergent lines for response to sheep red blood cell in white leghorn chicken. However, using *Taq* I restriction enzyme, he could not get the polymorphism between high titer line and low titer line in IWG as well.
as in IWJ lines of White Leghorn. Similarly, Muthukumar (2003) studied the 2nd generation of HSRBC, LSRBC, HCM and LCM lines in broiler. He studied polymorphism using PCR-RFLP between these lines, however using Taq I, he also could not get polymorphism between these lines.