600 alleles and 200 haplotypes identified for the chicken BF-BL region

600 alleles and 200 haplotypes identified for the chicken BF-BL region

Clive A. Tregaskes[1,2], Rebecca J. Martin[1], Lan Huynh[3], Nicolas Rocos[3] and Jim Kaufman[1,2,3]

1. Department of Pathology, University of Cambridge, UK
2. Department of Veterinary Medicine, University of Cambridge, UK
3. Institute for Immunology and Infection Research, University of Edinburgh, UK

The major histocompatibility complex (MHC) of mammals is a large and complex region, with hundreds of genes and much recombination, but encodes a few highly polymorphic classical class I and class II molecules that have central roles in immune responses (Kaufman 2016). The functional equivalent of the mammalian MHC in chickens is the BF-BL region, which is remarkably simple and compact with few genes, most of which are critical to the function of classical MHC molecules, so that this region was originally dubbed a "minimal essential MHC". Moreover, recombination within the BF-BL region was considered to be rare, so that this region could exist as relatively stable haplotypes, with co-evolution between these closely-linked genes leading to functional consequences (Kaufman et al., 1999; Kaufman, 2018; Tregaskes and Kaufman, 2021). However, most of these ideas arose from the analyses of a few "standard haplotypes" dating back to the original descriptions by Briles and co-workers (Miller et al., 2004; Miller and Taylor, 2016; Afrache et al., 2020).

We set out to understand more about the diversity of MHC alleles and haplotypes in different chicken populations, starting with reference strand-mediated conformational analyses (RSCA) followed by cloning and sequencing (Potts et al., 2019). As the need for higher through-put became clear, we developed a polymerase chain reaction-next generation sequencing (PCR-NGS) system to type the classical class II B genes *BLB1* and *BLB2*, and the classical class I genes *BF1* and *BF2*. Taking advantage of the compact nature of chicken MHC genes, we amplified exon 2 through the intron to the end of exon 3 (roughly 750 nucleotides) from genomic DNA, and used the Illumina MiSeq to paired-end sequence both exons from each gene (Fig. 1). Coupled with DNA isolation using relatively cheap reagents (which worked for most samples) and a double-barcoding system (12 pairs of barcoded primers each for BF and BL, and 96 barcoded Illumina adaptors), we were able to analyse up to 1152 samples in a run. We developed a bespoke bioinformatics pipeline that automatically generated sequences for all the alleles present, compared them to known alleles and then assembled them into known haplotypes, leaving the unknown sequences to be analysed by inspection (Martin, 2021; C. A. Tregaskes, R. J. Martin and J. Kaufman, unpublished).

The initial samples included DNA, blood cells or tissues primarily from experimental lines, red jungle fowl, commercial flocks, fancy birds and African chickens provided by many collaborators. Altogether, 22 different MiSeq runs were performed covering roughly 20,000 samples. For some populations, we used the microsatellite LEI0258 (Fulton et al., 2006) and BF2-specific PCRs to confirm and extend the assignments (Bertzbach et al., 2022; L. Huynh, C. A. Tregaskes and J. Kaufman, unpublished). For some samples of blood cells, flow cytometry was performed to determine the expression level of the class I (BF) molecules on erythrocytes that is known to correlate inversely with peptide repertoire (Chappell et al., 2015), which in some cases was determined by immunopeptidomics. Almost all the experimental work is now complete, except for some PacBio sequencing that has become necessary to assign BLB sequences to the appropriate loci (N. Rocos, C. A. Tregaskes and J. Kaufman, unpublished). We identified roughly 600 alleles and found over 200 haplotypes, but there is much analysis to complete, so only an initial overview of some preliminary results will be summarised in this report, with some details expected to change as the analyses are refined.

For understanding alleles and haplotypes, we began by assembling the data known from the scientific literature as well as from nucleotide databases (such as NCBI/GenBank) (Afrache et al., 2020). From the literature, sixteen complete BF-BL haplotypes were known from sequencing cosmids, bacterial artificial chromosomes (BACs) and long-range PCR products, and other haplotypes could be assembled from complete or partial (exon 2-exon 3 for BF, exon 2 and exon 2-exon 3 for BLB) genomic sequences, as well as complete or partial cDNA sequences. However, many gene and cDNA sequences in the databases had to be ignored even if published, since they were deposited from single studies and differed by only one or two nucleotides, with the associated papers revealing that only one amplification had been carried out, so most of those sequences were not separate alleles but the result of nucleotide misincorporation. Altogether, 17 standard haplotypes seemed secure, and an additional 16 haplotypes were suggested (Afrache et al., 2020). Even these data must be treated with some caution; for example, the B6 and B15 haplotypes reported in the largest haplotype-sequencing project (Hosomichi et al., 2008) have not yet been found in any sample examined by PCR-NGS.

It was very easy to assign the 339 class I sequences found by PCR-NGS (after 22 MiSeq runs) to the *BF1* and *BF2* loci (Martin, 2021) since they were found almost exclusively in different clades by neighbour-joining, maximum likelihood and minimum evolution tree building algorithms (Fig. 2). The exceptions include nine alleles related to BF1*0201 and BF1*0901 which cluster together, as well as a couple of other sequences present in other clades, all in the BF2 part of the tree. The *BF2* locus was more polymorphic, with 247 *BF2* alleles compared to 92 *BF1* alleles. Despite largely being in one clade, *BF1* alleles had much sequence diversity, with deep branches in the BF1 clade tree. However, much of this variation was not obviously in the peptide-binding site, with 74% of the BF1 sequences having His9 and Asp24, which may mean a wide peptide repertoire (as recently found from the structure of BF1*1901). Moreover, 89% of the BF1 sequences had an identical or near-identical sequence in the region of the C1/C2 epitope on the α1 helix, consistent with the suggested role of BF1 molecules as natural killer (NK) ligands (Ewald and Livant, 2004; Kim et al., 2018).

In contrast to class I loci, it became very difficult to assign all 259 BLB sequences to the *BLB1* and *BLB2* loci (Martin, 2021), largely due to finding haplotypes with two new class II B sequences, both of which were most closely-related to known BLB1 sequences. The known BLB1 and BLB2 sequences (as well as many identified by PCR-NGS) were mixed in phylogenetic trees (Fig. 2). Ongoing experiments using “between gene” primers and PacBio sequencing to assign sequences to the *BLB1* locus adjacent to *Blec* and to the *BLB2* locus adjacent to *BRD2* have resolved the ambiguities in 26 of 47 unclear BLB haplotypes (N. Rocos, C. A. Tregaskes and J. Kaufman, unpublished). These experiments have revealed the same sequence in one locus in one haplotype and the other locus in a second haplotype, as well as one haplotype with the same sequence in both loci. It seems likely that most haplotypes will have to be checked to ensure that new sequences most closely-related to known BLB2 sequences are actually located in the *BLB2* (that is *BRD2*-adjacent) locus. Given the facts that the *BLB1* and *BLB2* genes are in opposite transcriptional orientation and that most of the gene sequences are nearly identical, one possibility is that recombination between homologous sequences in the *BLB1* and *BLB2* loci leads to inversion; in the PacBio run done thus far, there has been no convincing evidence for such inversions.

The optimal choice of nomenclature for BF and BLB alleles continues to be unclear. The old accepted nomenclature was based on haplotypes, so the same sequence in two haplotypes would have different names (Miller et al., 2004). Based on the nomenclature system originally described for human MHC alleles and widely used for other species (Ballingall et al., 2018; Robinson et al., 2020; Afrache et al., 2020), the gene designation would be separated from the allele designations by a star or asterisk, with distantly-related alleles of a single locus differing in the first two numbers (eg., BLB1*02 versus BLB1*04), and with closely-related alleles of a single locus having the same first two

numbers and differing in the next two numbers after a colon (eg., BLB1*02:01 and BLB2*02:02). Haplotypes would then be constructed by strings of alleles (eg., BLB1*02:01-BLB2*02:02-BF1*02:04-BF2*02:05 or in short 2-2:02-2:04-2:05). This elegant solution ran into trouble from the criteria for close relationship, in that the number of sequence differences within clade of closely-related sequences could exceed the number of sequence differences between two sequences from different clades. Moreover, the same BLB sequence has now been found experimentally in both the *BLB1* and *BLB2* loci, so how should it be named? At the moment, designations for many sequences are simply *ad hoc*, as we struggle to develop a consistent approach.

Of the "standard haplotypes", seven were exhaustively analysed over some years in our lab, five of which were included among the 14 subsequently analysed by another lab in a single sequencing paper (Wallny et al., 2006; Shaw et al., 2007; Hosomichi et al., 2008). Of these 16 haplotypes, all had different *BF2* alleles except for two B15 haplotypes which differed in *BF1*. The original B15 haplotype described (and almost all subsequently) had no expressed *BF1* allele (as did the B14 haplotype), but the B15 haplotype from a chicken line in Japan had a *BF1* allele present. Among the 242 haplotypes identified by PCR-NGS (although there are a few from published data that we have not found), 27 (11%) have no *BF1* allele amplified by the primers used. Originally Southern blots suggested an insertion in the BF1 loci that were not expressed (Wallny et al., 2006; Shaw et al., 2007), but the latest experiments with primers outside the gene have amplified this region in those haplotypes, and identified a deletion of the whole *BF1* gene which is the result of two short direct (but imperfect) repeats (N. Rocos, F. J. Coulter and J. Kaufman, unpublished).

Of the "standard haplotypes", B19 was identified as a recombinant of B12 and B15 haplotypes, and three haplotypes (B5, B8 and B11) were also found to be recombinants, although by "gene conversion" of long stretches of DNA (Wallny et al., 2006; Shaw et al., 2007; Hosomichi et al., 2008). Among the 242 haplotypes, 128 (53%) could have arisen by recombination between *BLB2* and *BF1* (some with subsequent mutation to produced closely-related alleles); 22 BLB haplotypes are in combination with 96 BF haplotypes (with closely-related *BF* alleles combined, since they might have arisen from mutation subsequent to recombination). The most extreme is the BLB haplotype 5-5, which is associated with 25 different BF haplotypes (with closely-related BF alleles combined). There is also apparent recombination between *BF1* and *BF2*, with 15 different *BLB1-BLB2-BF1* haplotypes in association with 37 *BF2* alleles. As an example, 4-8-4 is found with BF2*24:01, 53:01 and four closely-related 43 alleles.

The first analyses by RSCA were performed with high level (elite, greatgrandfather) lines of commercial breeders, and we were shocked at the low diversity of these populations; some had only a single BF-BL haplotype. In order to better understand the commercial chickens that are actually in the field, we obtained farm-level samples from our collaborators, examining six broiler lines, 15 egg-layer lines and one dual purpose line. The take-home message is that there are typically very few haplotypes, mostly four or five haplotypes above 1% genotype frequency, usually with one haplotype by far the majority. In particular, a haplotype not described before (made up of alleles previously described, provisionally called B31) is present in over 33-64% genotypes of the six farm-level broiler flocks. Similarly, a previously undescribed haplotype provisionally called B9:02 dominates nearly all the brown egg layer flocks. If these numbers are representative, then there are billions of chickens in the field that are MHC homozygotes.

How do these commercial chickens survive with such low MHC diversity? Part of the answer may be that most of the high frequency haplotypes are those with BF2 alleles that have low cell surface expression and promiscuous peptide binding (for those with known peptide motifs). Such so-called "promiscuous haplotypes" are known to protect against a variety of economically-important infectious diseases in chickens and have been suggested to act as generalists, in contrast to

"fastidious haplotypes" which may act as specialists (Chappel et al., 2015; Kaufman, 2018; Tregaskes and Kaufman, 2021). A few haplotypes with high-expressing *BF2* alleles are found in some populations; these may function as specialists or have some other useful attribute(s).

A wealth of information has already emerged from this PCR-NGS typing, but there is much more to be learned by finishing the detailed analysis of commercial, fancy and African chickens. Moreover, there are many chickens worldwide that have not been examined by this kind of analysis, particularly in South, Southeast and East Asia. As the typing methods benefit from longer reads that cover more genes, many interesting attributes of the chicken MHC are likely to be revealed.

Figure legends

Fig. 1. The basis for the PCR-NGS typing of the chicken MHC. Organisation of the BF-BL region and gene names from Kaufman et al 1999 (*RING3* is now known as *BRD2*); primers are designated by lab names. Figure from Martin 2021.

Fig. 2. The chicken classical class I sequences mostly separate into two large clades, while classical class II B sequences are all mixed together in phylogenetic trees. BLB* indicates class II B sequences that could not easily be assigned to the *BLB1* or *BLB2* loci in the haplotypes examined. Figure from Martin 2021, derived from data available in 2020.

Acknowledgements
We thanks Dr. Thomas Tan for setting up most of the first PacBio run, and many technicians, students and collaborators that contributed work and samples to this effort over the last seven years. We thank Dr. Jacqueline Smith and Dr. Magda Migalska for critical reading.

Statement of Ethics
No animals were used by the authors for the work described in this report.

Conflict of Interest Statement
The authors declare no conflicts of interest.

Funding Sources
The authors thank the Wellcome Trust (Investigator Award 110106/A/15/Z to JK), the BBSRC (DTP PhD in Biological Sciences at the University of Cambridge to RJM), and the University of Edinburgh for support.

Author Contributions
CAT, BJM and JK were involved in conceptualisation, all authors in experimentation, data analysis and interpretation, JK wrote the first draft and all authors had the opportunity for editorial comment.

Data Availability Statement
Published data is available in the cited publications. Unpublished data awaits final processing and interpretation before it is submitted for deposition in GenBank and the ImmunoPolymorphism database and for publication in scientific journals.

References

Afrache H, Tregaskes CA, Kaufman J. A potential nomenclature for the Immuno Polymorphism Database (IPD) of chicken MHC genes: progress and problems. Immunogenetics. 2020 Feb;72(1-2):9-24. doi: 10.1007/s00251-019-01145-6. PMID: 31741010.

Ballingall KT, Bontrop RE, Ellis SA, Grimholt U, Hammond JA, Ho CS, et al. Comparative MHC nomenclature: report from the ISAG/IUIS-VIC committee 2018. Immunogenetics. 2018 Nov;70(10):625-632. doi: 10.1007/s00251-018-1073-3. PMID: 30039257.

Bertzbach LD, Tregaskes CA, Martin RJ, Deumer US, Huynh L, Kheimar AM, et al. The Diverse Major Histocompatibility Complex Haplotypes of a Common Commercial Chicken Line and Their Effect on Marek's Disease Virus Pathogenesis and Tumorigenesis. Front Immunol. 2022 May 27;13:908305. doi: 10.3389/fimmu.2022.908305. PMID: 35693787.

Chappell P, Meziane el K, Harrison M, Magiera Ł, Hermann C, Mears L, et al. Expression levels of MHC class I molecules are inversely correlated with promiscuity of peptide binding. Elife. 2015 Apr 10;4:e05345. doi: 10.7554/eLife.05345. PMID: 25860507.

Ewald SJ, Livant EJ. Distinctive polymorphism of chicken B-FI (major histocompatibility complex class I) molecules. Poult Sci. 2004 Apr;83(4):600-5. doi: 10.1093/ps/83.4.600. PMID: 15109057.

Fulton JE, Juul-Madsen HR, Ashwell CM, McCarron AM, Arthur JA, O'Sullivan NP, Taylor RL Jr. Molecular genotype identification of the Gallus gallus major histocompatibility complex. Immunogenetics. 2006 Jun;58(5-6):407-21. doi: 10.1007/s00251-006-0119-0. PMID: 16738938.

Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, et al. Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol. 2008 Sep 1;181(5):3393-9. doi: 10.4049/jimmunol.181.5.3393. Erratum in: J Immunol. 2010 May 1;184(9):5415. PMID: 18714011.

Kaufman, J. (2016) “Genetics and genomic organisation of the major histocompatibility complex (MHC)”. In: Encyclopedia of Immunobiology, vol. 2, pp 166-173. M. J. H. Ratcliffe, editor in chief. Academic Press, Oxford.

Kaufman J. Generalists and Specialists: A New View of How MHC Class I Molecules Fight Infectious Pathogens. Trends Immunol. 2018 May;39(5):367-379. doi: 10.1016/j.it.2018.01.001. PMID: 29396014.

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken B locus is a minimal essential major histocompatibility complex. Nature. 1999 Oct 28;401(6756):923-5. doi: 10.1038/44856. PMID: 10553909.

Kim T, Hunt HD, Parcells MS, van Santen V, Ewald SJ. Two class I genes of the chicken MHC have different functions: BF1 is recognized by NK cells while BF2 is recognized by CTLs. Immunogenetics. 2018 Sep;70(9):599-611. doi: 10.1007/s00251-018-1066-2. PMID: 29947944.

Martin, RJ. Diversity and Evolution in the Avian Major Histocompatibility Complex. PhD thesis, University of Cambridge. 2021.

Miller MM, Taylor RL Jr. Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. Poult Sci. 2016 Feb;95(2):375-92. doi: 10.3382/ps/pev379. PMID: 26740135.

Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, et al. 2004 Nomenclature for the chicken major histocompatibility (B and Y) complex. Immunogenetics. 2004 Jul;56(4):261-79. doi: 10.1007/s00251-004-0682-1. PMID: 15257423.

Potts ND, Bichet C, Merat L, Guitton E, Krupa AP, Burke TA, et al. Development and optimization of a hybridization technique to type the classical class I and class II B genes of the chicken MHC. Immunogenetics. 2019 Nov;71(10):647-663. doi: 10.1007/s00251-019-01149-2. PMID: 31761978.

Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Res. 2020 Jan 8;48(D1):D948-D955. doi: 10.1093/nar/gkz950. PMID: 31667505.

Shaw I, Powell TJ, Marston DA, Baker K, van Hateren A, Riegert P, et al. Different evolutionary histories of the two classical class I genes BF1 and BF2 illustrate drift and selection within the stable MHC haplotypes of chickens. J Immunol. 2007 May 1;178(9):5744-52. doi: 10.4049/jimmunol.178.9.5744. PMID: 17442958.

Tregaskes CA, Kaufman J. Chickens as a simple system for scientific discovery: The example of the MHC. Mol Immunol. 2021 Jul;135:12-20. doi: 10.1016/j.molimm.2021.03.019. PMID: 33845329.

Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A. 2006 Jan 31;103(5):1434-9. doi: 10.1073/pnas.0507386103. PMID: 16432226.

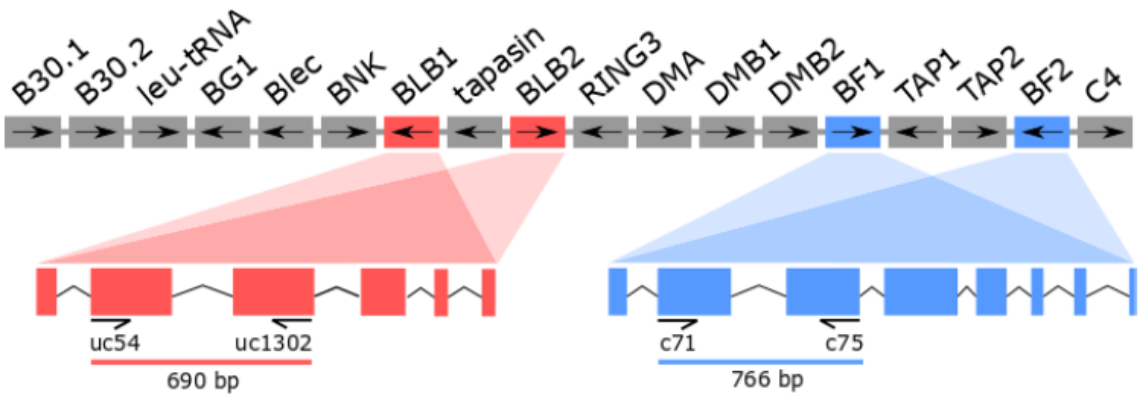

Fig. 1. The basis for the PCR-NGS typing of the chicken MHC. Organisation of the BF-BL region and gene names from Kaufman et al 1999 (RING3 is now known as BRD2) with genes and exons of class I in blue and class II B in red; primers are designated by lab names. Figure from Martin 2021.

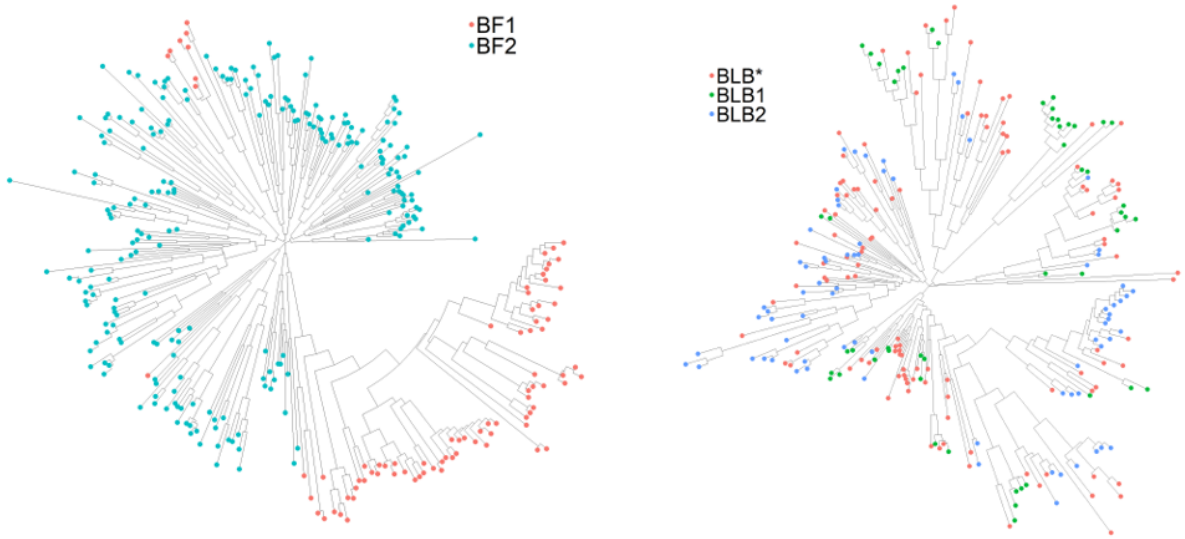

Fig. 2. The chicken classical class I sequences mostly separate into two large clades, while classical class II B sequences are all mixed together in phylogenetic trees. BLB* indicates class II B sequences that could not easily be assigned to the BLB1 or BLB2 loci in the haplotypes examined. Figure from Martin 2021, using data available in 2020.