Class II MHC typing for a Holstein/Boran population for which p67C vaccination outcome data is available

Anna Lacasta[1], Sonal Henson[1] and Isaiah Obara[2]
[1]International Livestock Research Institute; [2]Freie Universität Berlin

Overview

The ILVAC *T. parva* subunit vaccine team previously observed a huge variability in antibody titres and epitope recognition to p67C after vaccination. Because of the pivotal role of the MHC in immunosurveillance and responses to vaccines, it is hypothesized that modulation of $CD4^+$ T-cell responses to p67C by class II MHC type may influence the antibody response. The proposed programme of research sought to exploit advantages afforded by the availability of a Holstein/Boran cattle population for which antibody titres and epitope recognition to p67C have been evaluated. The key data will be a comprehensive catalogue of genetic polymorphisms within the cattle class II MHC DRB3, DQA and DQB. The extent to which allele specific information from individual cattle and similarities in machine learning predicted peptide motifs correlates with differential antibody titres and vaccination outcomes will be assessed. The project has three distinct milestones that show clear continuity, and the purpose of this report is to outline the discrete steps that have been undertaken towards successful completion of the milestones.

Milestone 1– Synthesis of bovine cDNA

Activity 1: Defining the sample set

A rationally chosen set of 62 samples from an inventory of cryopreserved PBMCs were selected for characterization of expressed class II MHC sequences. The considerations for selection of samples included: (i) the immunogen used for vaccination, (ii) the absolute antibody titers at day of challenge, (iii) lethal dose used, (iv) ECF score and (v) if the animal was considered immune or susceptible.

Table 1. Samples selected for MHC typing

Animal ID	Immunogen	Lethal dose	ECF score	Immunity (1=immune, 0=susceptible)	DOB	Breed
BK007	3x450ug s-p67C	LD73	4.48	1	2013-07-07	Boran
BK010	3x450ug s-p67C	LD73	7.68	0	2013-07-17	Boran
BK011	3x450ug s-p67C	LD73	3.83	1	2013-08-19	Boran
BK018	3x450ug s-p67C	LD73	4.67	1	2013-08-08	Boran
BK019	3x450ug s-p67C	LD73	1.87	1	2013-08-08	Boran
BK020	3x450ug s-p67C	LD73	7.23	0	2013-09-08	Boran
BK021	3x450ug s-p67C	LD73	3.09	1	2013-09-08	Boran
BK028	3x450ug s-p67C	LD73	2.12	1	2013-07-30	Boran
BK029	3x450ug s-p67C	LD73	5.13	1	2013-08-17	Boran
BK036	3x450ug s-p67C	LD73	8.47	0	2013-09-01	Boran
BK039	3x450ug s-p67C	LD73	5.55	1	2013-08-27	Boran
BM172	3x200ug s-p67C	LD56	8.15	0	2016-04-20	Friesians
BN077	HBcAg-p67C + SV-p67C	LD93	4.42	1	2017-03-17	Friesians
BN085	HBcAg-p67C + SV-p67C	LD93	0.25	1	2017-05-20	Friesians
BN086	HBcAg-p67C + SV-p67C	LD93	6.92	0	2017-06-10	Friesians

BN092	HBcAg-p67C + SV-p67C	LD93	7.23	0	2017-06-10	Friesians
BN093	HBcAg-p67C + SV-p67C	LD93	1.13	1	2017-07-27	Friesians
BN094	HBcAg-p67C + SV-p67C	LD93	0.4	1	2017-07-05	Friesians
BN097	HBcAg-p67C + SV-p67C	LD93	7.79	0	2017-05-10	Friesians
BN099	HBcAg-p67C + SV-p67C	LD93	0.17	1	2017-06-15	Ayshire
BN100	HBcAg-p67C + SV-p67C	LD93	7.06	0	2017-05-15	Ayshire
BN103	HBcAg-p67C + SV-p67C	LD93	5.27	1	2017-05-05	Friesians
BN104	HBcAg-p67C + SV-p67C	LD93	5.71	1	2017-04-10	Friesians
BN107	HBcAg-p67C + SV-p67C	LD93	6.79	0	2017-06-01	Ayshire
BN108	HBcAg-p67C + SV-p67C	LD93	8.23	0	2017-06-10	Friesians
BN113	HBcAg-p67C + SV-p67C	LD93	1.3	1	2017-06-25	Friesians
BN115	HBcAg-p67C + SV-p67C	LD93	1.28	1	2017-05-15	Friesians
BP008	p67C-I53-50A	LD60	0.52	1	2017-10-10	Friesians
BP017	p67C-I53-50A	LD60	6.72	0	2017-11-10	Friesians
BP021	p67C-I53-50A	LD60	5.44	1	2017-12-20	Friesians
BP028	p67C-I53-50A	LD60	8.08	0	2018-01-20	Friesians
BP034	p67C-I53-50A	LD60	3.57	1	2017-12-25	Friesians
BP036	p67C-I53-50A	LD60	1.53	1	2017-11-20	Friesians
BP038	p67C-I53-50A	LD60	0.44	1	2017-09-10	Friesians
BP040	p67C-I53-50A	LD60	0.17	1	2017-12-10	Friesians
BP042	p67C-I53-50A	LD60	0.92	1	2017-11-15	Friesians
BP010	p67C-I53-50A + SV-p67C	LD60	0.85	1	2017-10-17	Friesians
BP013	p67C-I53-50A + SV-p67C	LD60	1.15	1	2017-11-10	Friesians
BP014	p67C-I53-50A + SV-p67C	LD60	0.17	1	2017-09-25	Friesians
BP018	p67C-I53-50A + SV-p67C	LD60	0.24	1	2017-12-10	Friesians
BP019	p67C-I53-50A + SV-p67C	LD60	7.92	0	2017-11-17	Friesians
BP025	p67C-I53-50A + SV-p67C	LD60	1.41	1	2018-01-17	Friesians
BP026	p67C-I53-50A + SV-p67C	LD60	0.46	1	2017-11-17	Friesians
BP031	p67C-I53-50A + SV-p67C	LD60	7.87	0	2017-12-28	Friesians
BP033	p67C-I53-50A + SV-p67C	LD60	0.82	1	2017-11-20	Friesians
BP084	p67C-I53-50A	LD67	3.68	1	2018-02-05	Friesians
BP085	p67C-I53-50A	LD67	5.71	1	2018-01-10	Friesians
BP086	p67C-I53-50A	LD67	7.45	0	2018-01-15	Friesians
BP087	p67C-I53-50A	LD67	0.60	1	2018-11-30	Friesians
BP088	p67C-I53-50A	LD67	7.49	0	2018-10-20	Friesians
BP089	p67C-I53-50A	LD67	1.08	1	2018-03-28	Friesians
BP090	p67C-I53-50A	LD67	3.78	1	2018-03-14	Friesians
BP091	p67C-I53-50A	LD67	1.75	1	2018-03-20	Friesians
BP092	p67C-I53-50A	LD67	5.60	1	2018-04-24	Friesians
BP094	p67C-I53-50A	LD67	1.32	1	2018-03-30	Ayshire
BP095	p67C-I53-50A	LD67	4.81	1	2018-03-16	Friesians
BP097	p67C-I53-50A	LD67	7.68	0	2018-06-10	Friesians
BP098	p67C-I53-50A	LD67	2.90	1	2018-03-10	Friesians
BP099	p67C-I53-50A	LD67	0.47	1	2018-04-10	Friesians
BP119	p67C-I53-50A	LD67	0.60	1	2018-03-20	Friesians

Activity 2: Total RNA isolation

Cellular RNA was extracted from PBMCs using Trizol (Invitrogen) according to the manufacturer's instructions but with a few modifications. The optimised protocol is provided below:

Homogenization

- Thaw cells quickly in a water bath at 37 degrees, and transfer to a 50mL falcon tube.
- Add 10mL warm PBS, dropwise while swirling the tube to dilute out the DMSO
- Centrifuge cells for 10min at 1200rpm. Discard the supernatant
- Lyse cells by adding 400uL of TRIzol reagent to the cell pellet.
- Pipette lysate up and down to homogenise and transfer to a 2mL Eppendorf tube.

Phase separation

- Add 200uL of Chloroform. Securely cap the tube and thoroughly mix by shaking vigorously.
- Incubate tubes for 3 min at room temperature
- Centrifuge sample for 15min at 12,000xg at 4oC. Transfer upper aqueous phase to a new 2mL eppendorf tube by angling the tube

RNA precipitation

- Add 500uL of very cold Isopropanol to the aqueous phase to precipitate the RNA
- Incubate for 10min at 4oC.
- Centrifuge for 15min at 12,000xg at 4oC.

RNA wash

- Add 1mL of cold 75% ethanol
- Vortex the sample briefly and centrifuge for 5min at 7500xg at 4oC. Repeat this step twice
- Discard the supernatant

Redissolving RNA

- Air-dry the RNA pellet for 5-10min at room temperature
- Resuspend RNA pellet in 30uL of RNase-free water by pipetting up and down.
- Incubate in heat-block at 56oC for 10min, to completely dissolve the RNA.
- Aliquot 10uL of RNA for quality checks and cDNA synthesis.
- Store the rest of the RNA at -80oC

Total RNA was treated with Dnase, to remove contaminating genomic DNA. The RNA was quantified both using the Qubit Fluorometer and by measuring Abs 260/280 with a nanodrop UV spectrophotometer. The total RNA yield for all the samples was well above the minimum 1 ng required to synthesize first-strand cDNA using the SuperScript® First-Strand Synthesis System.

Activity 3: Oligo dt-primed cDNA synthesis

The total RNA was used as template for oligo dt-primed cDNA synthesis using the Invitrogen’s SuperScript IV First-Strand Synthesis System with ezDNase Enzyme according to the manufacturer's instructions.

Activity 4: Quality assessment of purified cDNA

This assessment was accomplished by using aliquots of the cDNA as template for amplification of the bovine DRB3, DQA and DQB genes. Since MHC variants are continually being discovered through sequence-based typing and deposited in public repositories, we first verified that the primers used for

cDNA quality control were conserved in MHC II sequences available in BoLA databases as of June 2021 and are furthermore conserved between cattle and relatively distantly related species like the buffalo.

Only the polymorphic peptide binding regions of the class II transcripts were amplified, and we obtained the expected product sizes as shown in the table below.

Primer	Sequence (5'-3')	Annealing temp	Amplicon size (bp)
DRB3-F	TAG TGA TGC TGA TGV TGC TG	61	364
DRB3-R	GGY TGR GTC TTT GCA GGA TA		
DQA-F	STG GRR GTG AAG ACA TYG TG	61	301
DQA-R	GAY TTG GRA AAC ACA GYC AC		
DQB-F	GGR CYG AGG GCA GAG ACT	61	319
DQB-R	GGR GAG ATG GTC ACT GTA GG		
PCR cycling conditions			
98 for 30s, (98 for 10s, 61 for 30s, 72 for 45s) x 30 cycles, 72 for 10 min.			

Activity 5. Ascertaining that the primers amplify the correct loci

We sought further confirmation that the primers were amplifying the target loci. This was achieved by sanger sequencing of all the three amplicons from a subset of the samples. We used the raw sequence data produced from Sanger sequencing, assembled the chromatograms for the bidirectional data in Geneious Prime. Each consensus sequence was searched against a database of known cattle class II MHC sequences using BLASTN. From our results, it is clear that the primers are amplifying the correct loci.

In sum, we isolated total cellular RNA from a rationally chosen set of samples, performed Dnase treatment to remove contaminating genomic DNA and synthesized cDNA using Invitrogen's SuperScript reverse transcription system and oligo(dT) primers. We further ascertained that the cDNA was of the desired quality and the primers were amplifying the correct target loci. The cNDA was shipped to LGC Berlin for typing of the polymorphic regions of cattle class II MHC transcripts (target amplification, library preparation, sample indexing and Illumina MiSeq run).

Milestone 2– Generation of raw Illumina reads

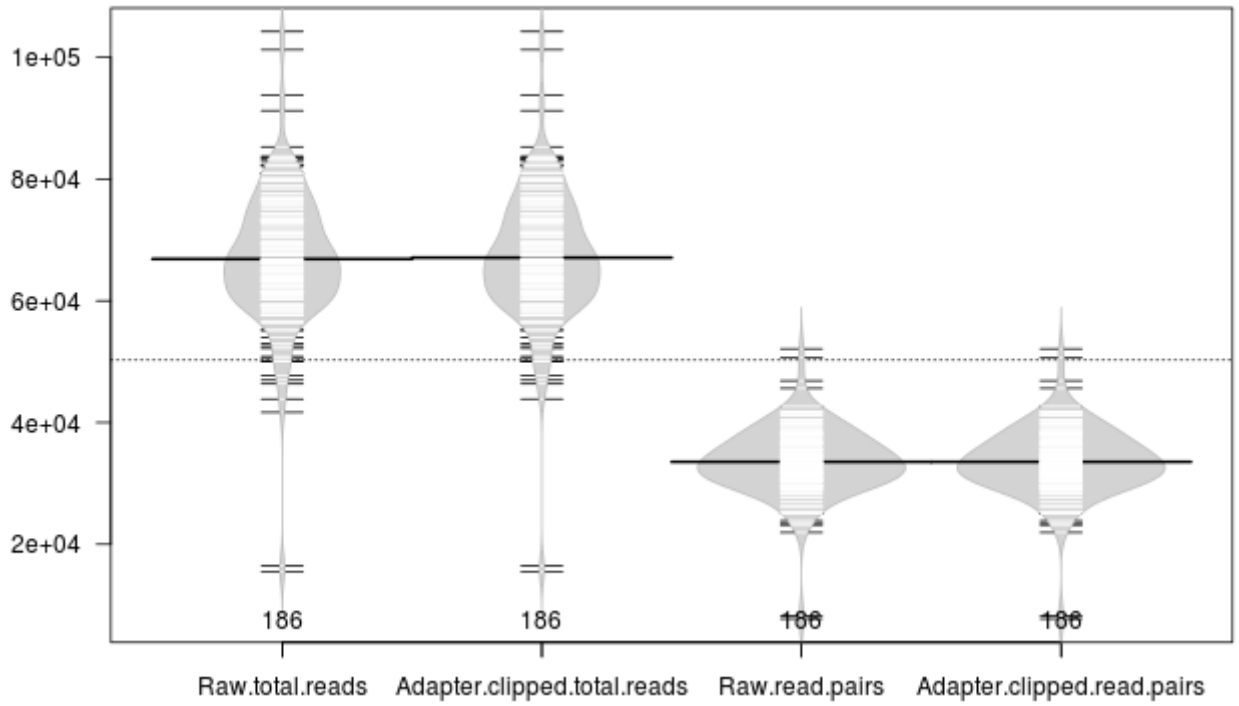
The library preparation, sample indexing and Illumina MiSeq run (300 bp paired-end strand-specific sequencing using Illumina MiSeq V3) was done at LGC Genomics, Berlin (https://shop.lgcgenomics.com/). The list of the 62 samples typed for the polymorphic regions of the bovine DQA, DQB and DRB 3 class II MHC transcripts is provided below

```
BP086 BP087 BP088 BP089 BP090 BP091 BP092 BP094 BP095 BP097 BP098
BP099 BP119 BN077 BN085 BN092 BN093 BN094 BN097 BN099 BN100 BN103
BN104 BN107 BN108 BN113 BN115 BP040 BP036 BP042 BP038 BM172 BK007
BK010 BK011 BK018 BK019 BK020 BK021 BK028 BK029 BK036 BK039 BP008
BP017 BP021 BP028 BP034 BN086 BP013 BP014 BP018 BP019 BP025 BP026
BP031 BP033 BP084 BP085 BL093 BL101
```

The raw sequences after base calling were initially processed as follows:

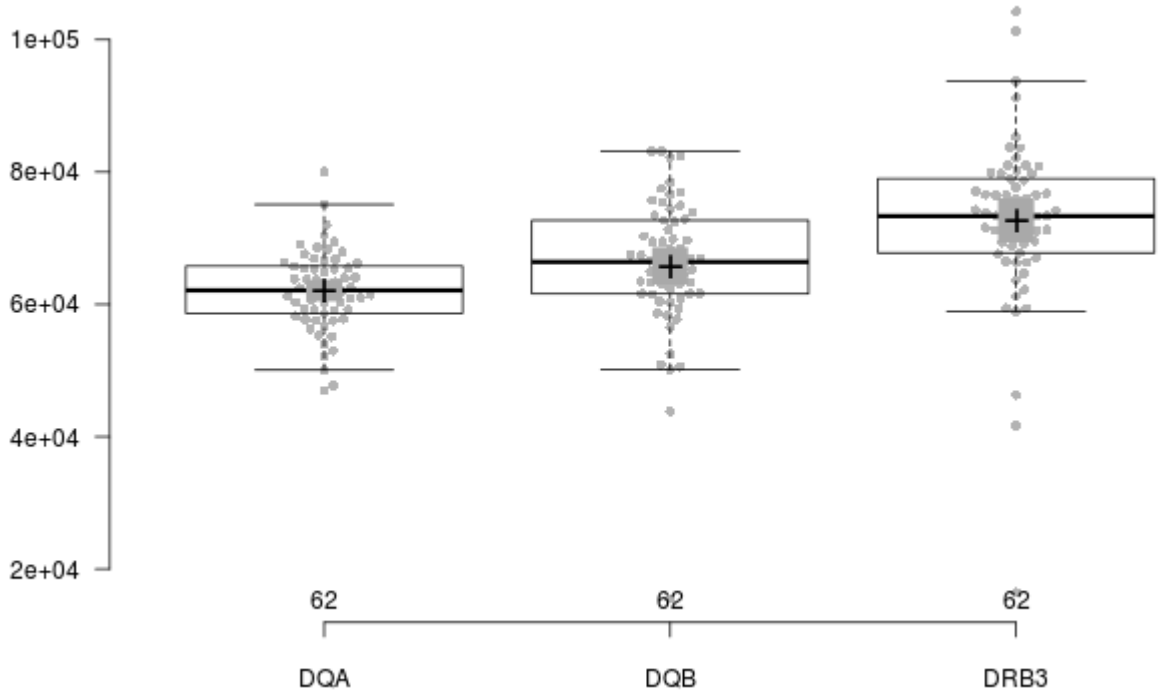
- Demultiplexing of all libraries for each sequencing lane using the Illumina bcl2fastq v2.20 software [1]. Here 1 or 2 mismatches or Ns were allowed in the barcode read when the barcode distances between all libraries on the lane allowed for it
- Clipping of sequencing adapter remnants from all reads

This analysis yielded a total of 12,480,784 adapter clipped raw reads and 6,240,392 adapters clipped read pairs. The total read count distribution is shown in the figure below.



Black lines show the means; white lines represent individual data points; polygons represent data density. n= 186 samples.

The mean per-gene raw read count was: DQA - total raw reads (3,849,696), mean raw read count (62091 ± 6048 SD); DQB - total raw reads (4,081,844), mean raw read count (65836 ± 10576 SD) and DRB - total raw reads (4,507,580), mean raw read count (72702 ± 12639 SD). The per-gene total read count distribution is shown in the figure below.

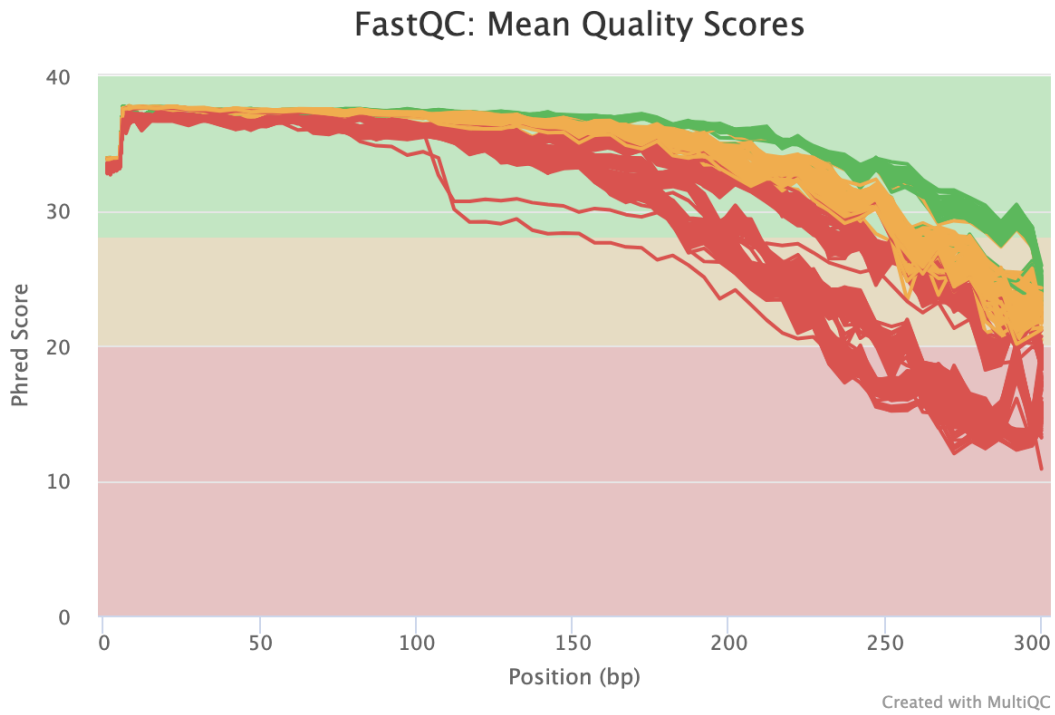


Centre lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means; bars indicate 95% confidence intervals of the means; data points are plotted as open circles. n = 62 sample points.

Milestone 3– Disaggregating variants from artifacts and predicting peptide binding specificities

Activity 1. Initial quality control

Before starting the downstream analysis, the paired end sequencing data was quality controlled using FastQC (version 0.11.9) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [2]. FastQC checks the raw sequence data for per base sequence quality, per tile sequence quality, per sequence quality scores, per base sequence content, per sequence GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences, and adapter content. The results illustrate that a majority of the outputs were of high quality. The per-sequence quality scores were above Q20 for R1 of all samples across the length of the reads, whereas the often-observed decline in average read quality for R2 was also observed in our samples towards the end of the sequencing cycle. The plot in the figure below shows the distribution of quality scores at each position in the read across all reads for all samples.



Distribution of quality scores at each position in the read across all reads for all samples.

Activity 2. Disaggregating variants from artifacts

An eventual aim of this part of the analysis is to filter the reads to disaggregate true alleles from artifactual and chimeric sequences. The reads are filtered following a stepwise stringent criterion originally described for class I MHC characterization in a cohort of Holstein-Friesian and African cattle breeds [3]. The allele calling workflow was subsequently adapted to class II MHC characterization [4]. Chimeric sequences generated in amplicons were identified by assessing if a sequence read could be a combination of two different, but more frequent read clusters. Further downstream analysis ultimately resulted in:

- a comprehensive catalogue of genetic polymorphisms within the DQA, DQB and DRB3 loci. As already mentioned, application of high throughput next generation sequencing technologies in combination with stringent artifact and chimera filtering algorithms not only shows the most prevalent alleles in this population, but also the presence of minor/novel alleles.
- an assessment of the extent at which expression of specific alleles correlated with differential antibody responses following immunisation with the *T. parva* p67 sporozoite antigen
- a distinction of the alleles regarding their peptide binding specificities and functional distance between the alleles, towards the goal of evaluating if functionally divergent clusters (supertypes) correlate to antibody response categories.
- an assessment of whether the class II MHC molecules expressed by animals with high antibody titres show elevated posterior probabilities of positive selection, based on dN/dS ratios under a Bayesian population genetics framework.

References

1. Illumina. bcl2fastq2 Conversion Software. URL : https://support.illumina.com/sequencing/sequencingsoftware/bcl2fastq-conversionsoftware.html.
2. Simon Andrews. FastQC A Quality Control tool for High Throughput Sequence Data. URL : http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
3. Vasoya D, Law A, Motta P, Yu M, Muwonge A, Cook E, Li X, Bryson K, MacCallam A, Sitt T, Toye P, Bronsvoort B, Watson M, Morrison WI, Connelley T. Rapid identification of bovine MHCI haplotypes in genetically divergent cattle populations using next-generation sequencing. Immunogenetics. 2016 Nov;68(10):765-781. doi: 10.1007/s00251-016-0945-7. Epub 2016 Aug 11. PMID: 27516207; PMCID: PMC5056950.
4. Vasoya D, Oliveira PS, Muriel LA, Tzelos T, Vrettou C, Morrison WI, de Miranda Santos IKF, Connelley T. High throughput analysis of MHC-I and MHC-DR diversity of Brazilian cattle populations. HLA. 2021 Aug;98(2):93-113. doi: 10.1111/tan.14339. Epub 2021 Jun 17. PMID: 34102036.

Funding for this work was provided by the CGIAR Research Program on Livestock.
December 2021

Recommended citation: Lacasta, A., Henson, S. and Obara, I. 2021. *Class II MHC typing for a Holstein/Boran population for which p67C vaccination outcome data is available.* Nairobi, Kenya: ILRI.