

ORIGINAL ARTICLE

High throughput analysis of MHC-I and MHC-DR diversity of Brazilian cattle populations

Deepali Vasoya¹  | Priscila Silva Oliveira² | Laura Agundez Muriel³ | Thomas Tzelos¹ | Christina Vrettou¹ | W. Ivan Morrison¹ | Isabel Kinney Ferreira de Miranda Santos² | Timothy Connelley¹

¹Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, Edinburgh, UK

²Department of Biochemistry and Immunology, Ribeirão Preto Medical School, Ribeirão Preto, São Paulo, Brazil

³Translational Synthetic Biology Department, Mammalian genome editing and gene therapy, Parque de Investigación Biomédica, Carrer del Dr, Barcelona, Spain

Correspondence

Deepali Vasoya, Centre for Tropical Livestock Genetics and Health (CTLGH), The Roslin Institute, University of Edinburgh, Easter Bush Campus, Edinburgh EH25 9RG, UK.
Email: deepali.vasoya@roslin.ed.ac.uk

Funding information

Bill and Melinda Gates Foundation, Grant/Award Number: OPP1127286; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BBS/E/D/20002174; European Union's Horizon 2020 (VetBioNet), Grant/Award Number: N°731014

The major histocompatibility complex (MHC) contains many genes that play key roles in initiating and regulating immune responses. This includes the polymorphic MHCI and MHCII genes that present epitopes to CD8+ and CD4+ T-cells, respectively. Consequently, the characterisation of the repertoire of MHC genes is an important component of improving our understanding of the genetic variation that determines the outcomes of immune responses. In cattle, MHC (BoLA) research has predominantly focused on Holstein-Friesian animals (as the most economically important breed globally), although the development of high-throughput approaches has allowed the BoLA-DRB3 repertoire to be studied in a greater variety of breeds. In a previous study we reported on the development of a MiSeq-based method to enable high-throughput and high-resolution analysis of bovine MHCI repertoires. Herein, we report on the expansion of this methodology to incorporate analysis of the BoLA-DRB3 and its application to analyse MHC diversity in a large cohort of cattle from Brazil (>500 animals), including representatives from the three major *Bos indicus* breeds present in Brazil – Guzerat, Gir and Nelore. This large-scale description of paired MHCI-DRB3 repertoires in *Bos indicus* cattle has identified a small number of novel DRB3 alleles, a large number of novel MHCI alleles and haplotypes, and provided novel insights into MHCI-MHCII association - further expanding our knowledge of bovine MHC diversity.

KEYWORDS

BoLA, *Bos indicus*, cattle, MHC

1 | INTRODUCTION

The Major Histocompatibility (MHC) locus contains a large number of genes associated with antigen presentation, including the MHCI and MHCII genes, which have a primary function in presenting peptide antigens to CD8⁺

and CD4⁺ αβ T-cells respectively. Both the MHCI and MHCII genes are characterised by a high level of polymorphic diversity; for example, in humans >18,000 MHCI and > 7000 MHCII alleles have been identified (IPD-MHC database January 2020-<https://www.ebi.ac.uk/ipd/>). Most of this polymorphism is focused within the exons

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. HLA: Immune Response Genetics published by John Wiley & Sons Ltd.

encoding the domains that form the 'peptide binding groove' and so determine the motif of the peptides that can be bound and presented to T-cells by individual MHC molecules. Because of this diversity, the MHCI and MHCII genes have the capacity, at a population level, to present a diverse range of peptides and so help mitigate the threat of potential pathogens evading immune responses.

Although having analogous functions, the biology of MHCI and MHCII genes are fundamentally different in a number of ways, including both molecular structure and repertoire of allelic variants. MHCI molecules are heterodimers composed of a highly polymorphic MHCI α heavy chain and β 2-microglobulin, a monomorphic protein that interacts non-covalently with the MHCI α heavy chain and is encoded outwith the MHC locus. The MHCI α heavy chain is composed of a short intracellular domain, a transmembrane domain and 3 extra-cellular domains, of which the two most membrane-distal (α 1 and α 2, encoded by exons 2 and 3, respectively) contain prominent alpha helices that form the 'peptide binding groove.' There are two 'categories' of MHCI molecules – the 'classical' MHCI which are highly polymorphic, are expressed on all nucleated cells and are primarily involved in the presentation of peptides to TCR expressed on CD8+ T-cells and the less polymorphic 'non-classical' MHCI molecules which show a more variegated and regulated pattern of expression and have a more prominent role in regulating immune cell function through interactions with innate immune receptors (e.g., KIRs). In most mammalian species 'classical' MHCI are polygenic as well as polymorphic (for example in humans there are three classical MHCI α loci—HLA-A, -B, -C). There are multiple isotypes of MHCII molecules (e.g. in humans there are DR, DQ and DP MHCII molecules), each of which are heterodimers composed of α and β chains and function predominantly in the presentation of peptides to CD4+ T-cells. Both chains are made up of two extra-cellular domains (α 1/ α 2 and β 1/ β 2, respectively), containing alpha helices similar to those in MHCI, a transmembrane region and an intra-cytoplasmic domain. In MHCII molecules the 'peptide binding groove' is formed by the combination of the membrane-distal α 1 and β 1 domains of each chain (encoded by exon 2 of each). In humans, there is a single DRA gene, nine DRB genes (of which only three encode functional products), two copies each of DPA and DPB (although DPA-2, DPB-2 are both pseudogenes) and two copies each of DQA and DQB (although DQA-2 and DQB-2 are only expressed concomitantly in Langerhans cells²). The level of polymorphism for the MHCII genes shows marked variation, with DRB, DQB and DPB having high levels of polymorphism (3296, 1771 and 1519 alleles, respectively), DQA-1 and DPA-1 showing more limited levels of allelic

diversity (216 and 161, respectively) and DRA showing a restricted amount of variation (only 29 alleles encoding only two protein variants).

In cattle, six MHCI gene loci have been proposed based on sequences of expressed MHCI products.^{3,4} However, in contrast to humans, the number of MHCI genes expressed varies between haplotypes (ranging from 1 to 4), with varying permutations of the six loci represented in different haplotypes.⁴ Similarly, several bovine DQA and DQB genes have been identified (with up to 5 gene loci proposed for each,^{5,6} and there is variability between haplotypes in the number of DQA and DQB present, with approximately half of the haplotypes expressing only single DQA/DQB genes and the remainder expressing 2.⁷ Bovine DRA is monogenic and although there are three DRB loci, DRB1 is a pseudogene and DRB2 is not functionally transcribed⁸; consequently, only the DRB3 locus is considered to be functionally relevant. As in other species, the bovine DRA gene is essentially monomorphic,⁹ while the DRB3, DQA, DQB and MHCI genes all exhibit polymorphism (330, 65, 87 and 111 alleles respectively described in the IPD-MHC database). Due to the polygenic nature of the MHCI, DQA and DQB genes, generation of high-resolution sequence data using Sanger sequencing requires costly and laborious sub-cloning procedures, which has precluded large-scale studies of the allelic repertoire. In contrast, the essentially monogenic nature of the expressed DRB gene has permitted the development of a number of Sanger-sequencing based techniques (SBT) to directly sequence DRB3 alleles.^{10–14} These SBT methods, which focus on sequencing of the polymorphic exon 2, have made larger-scale studies feasible, and so have been used extensively to study bovine MHC diversity across the globe,^{15–20} as well as assess the association between DRB3 variants and disease and production traits.^{21–24}

The advent of next-generation sequencing (NGS) technologies, by permitting cost effective, large scale, high-resolution analysis of MHC sequences, has revolutionised HLA analysis both clinically and in academic research as described in a number of recent reviews.^{25–28} We recently published a MiSeq-based protocol for sequencing cattle MHC (BoLA) class I genes and by application to small cohorts of ~100 indigenous animals from Cameroon and Kenya demonstrated its utility in expanding our knowledge of the MHCI repertoire in breeds in which the MHCI repertoire had not been previously characterised.²⁹ *Bos indicus* breeds are adapted to tropical environments and form the dominant cattle population in large parts of the world. However, compared to European *Bos taurus* cattle, relatively little is known regarding their MHC diversity. In this study we report on the expansion of the MiSeq protocol to include analysis of DRB3 and apply this to analysis of >500 animals of

predominantly *Bos indicus* breeds of cattle in Brazil. Our results have identified a large number of new bovine MHC I alleles and haplotypes and by conducting large scale parallel sequencing of MHC I and DRB3, have provided novel insights into the associations between MHC I and MHC II genes in cattle.

2 | MATERIALS AND METHODS

2.1 | Sampling

Samples came from 331 Holstein-Friesian animals at University of Edinburgh's Langhill herd and multiple farms in Brazil as described in the results (Table 2). The work was approved by The Roslin Institute Animal Welfare and Ethical Review Panel and conducted under licence and in accordance with the UK government Animal (Scientific Procedures) Act 1986. Blood Samples were collected by jugular venepuncture into EDTA vacutainers (BD Biosciences, Oxford, UK) and erythrocytes lysed by incubation in 5× volume of RBC lysis buffer (0.144 M ammonium chloride/0.175 M Tris pH 7.4) for 5 min at room temperature. The white blood cell (WBC) pellet was washed three times in PBS, total RNA extracted using Tri-reagent (Sigma, Gillingham, UK) and cDNA synthesised using a Reverse Transcription Kit (Promega, Madison, WI, USA), both according to the manufacturer's instructions.

2.2 | PCR amplification and sample preparation

PCR amplification of MHC I genes was performed as described in Reference 29. DRB3 PCR amplification used a primer pair consisting of BoLA_DRB3for1 (TAG TGA TGC TGA TGV TGC TG) and BoLA_DRB3rev1 (GGY TGR GTC TTT GCA GGA TA), designed based on the bovine DRB3 exon1 and three nucleotide sequence data that was available in the IPD-MHC database in July 2018. Series of BoLA_DRB3for1 and BoLA_DRB3rev1 primers incorporating Illumina adaptors and multiplex identifier tags (MID) were obtained from IDT (Leuven, Belgium), permitting the generation of 192 (8 × 12) multiplexed samples. BoLA-DRB3 amplification was conducted using the Phusion High-Fidelity PCR kit (New England Biolabs, Hitchin, UK) with 50 µl reactions composed of Phusion HF amplification buffer, 3% DMSO, 0.2 mM dNTPS, 25 pmol of for and rev primers, 1 U Phusion Hot Start DNA polymerase and 1 µl of cDNA. Cycling conditions were 98°C for 30s, 30 cycles of 98°C for 10s, 61°C for 30s, 72°C for 30s and a final extension period of 72°C for 10 min. Following amplification, 5 µl of PCR product from each sample were

pooled (~192 samples in pool), a fraction of this pool (~250 µl) run on a 1% agarose gel and bands of the appropriate size (~500 bp) extracted and purified using the Qiagen Gel extraction kit (Qiagen, Manchester, UK). The extracted DNA was subsequently purified using Agencourt AMPure XP beads (Beckham Coulter, High Wycombe, UK) at a v/v ratio of 1:1 beads to PCR product and quantified using 260/280 nm absorbance readings obtained from a Nanodrop spectrophotometer (Wilmington, DE, USA). The purified DRB3 pool was mixed at a ratio of 1:2:2 with the equivalent purified products from the pooled MHC I For1/Rev3 and For2/Rev1 products to generate the library submitted for sequencing. Each library submitted for sequencing included samples from ~192 animals.

2.3 | Sequencing and bioinformatics analysis

Libraries were submitted to Edinburgh genomics for 300 bp paired-end sequencing on the Illumina MiSeq v3 platform. The raw sequencing data are submitted to ENA under accession number PRJEB44287. Sequence reads were segregated based on MID combinations ~192 datasets, the raw data assessed for quality (threshold score of >Q₂₈), and paired-end sequences were overlapped using FLASH.³⁰ Data were then processed using a bioinformatics pipeline, essentially as described in Vasoya et al. 2016 except that the filter to remove 'sequences that contained 1bp or 2bp differences from a sequence present in the same sample and present at a 30- or 50-fold higher read frequency, respectively' was replaced with a filter to remove 'all sequences that were a 1bp variant of another sequence that was present in the same sample at a > 3-fold higher frequency'. The reasons, justification and subsequent adaptations of the pipeline required to accommodate this change are described in Supplementary Data 2; retrospective application of the modified pipeline to the data published in Vasoya et al. 2016 demonstrated that the results were unaffected by the modification. MHC I and DRB3 allele sequence data was extracted from the IPD-MHC database in June 2020 and updated in May 2021.

2.4 | Nomenclature of novel MHC sequences and MHC I haplotypes

Novel MHC sequences identified in this study were named according to the IPD-MHC guidelines for nomenclature of non-human MHC generated from NGS data.³¹ All nucleotide sequences were translated to amino acid sequences and compared to the data in the IPD-MHC

database. If sequences showed ≤ 4 amino acid differences from an official designated sequence, then it was considered to be in the same allelic group as that sequence and was given a name to identify it as a novel allelic subgroup. To avoid potential confusion with official IPD-MHC naming we used alphabet characters to name novel allelic subgroups and synonymous variants (e.g., BoLA-6*040:AC). Sequences that showed >4 amino acid differences from sequences in the IPD-MHC database were considered to represent novel allelic groups. As we could not assign such MHCI sequences to specific loci, these were given the prefix BoLA-MHCI* and the allele group assigned as either gb or br, depending on origin of animal in which it was first identified (Great Britain or Brazil respectively; additionally equivalent sequences identified in a previous study²⁹ from samples obtained from Cameroon and Kenya were assigned as 'cm' and 'ke' respectively), followed by a double-digit number (e.g., BoLA-MHCI*gb20:01). To be assigned as a haplotype the combination of alleles that constituted the haplotype had to be identified in multiple animals. Novel MHCI haplotypes were given the prefix HP1 and assigned a unique combination of two numbers separated by a decimal point (e.g., HP1.24.1). Haplotypes that were composed of alleles from the same group(s) but differed at the subgroup level for one or more alleles were considered as variant haplotypes; these haplotypes shared the same primary number but were assigned different secondary numbers (e.g., HP1.73.1 and HP1.73.2). Combinations of alleles that were only observed in a single animal were considered as putative novel haplotypes and given a putative haplotype designation prefixed with unHP1.

3 | RESULTS

3.1 | Validation of a high-throughput bovine DRB3 sequencing approach

A PCR protocol that permitted the amplification of exon 2 of BoLA-DRB3 alleles using MiSeq was developed using primers that would; (i) permit amplification of all the known bovine DRB3 alleles in the IPD database (based on in silico prediction), (ii) generate amplicons that after sequencing would enable unambiguous discrimination of the different DRB3 alleles in the IPD database and (iii) generate amplicons of <500 bp and so be feasible to sequence using the MiSeq platform (size of the amplicon was ~ 324 bp; 36 bp of exon 1, 264–267 bp of exon 2 and 21 bp of exon 3). To provide preliminary validation of this protocol it was applied to the analysis of DRB3 sequences of 331 Holstein-Friesian cattle in parallel with MHCI typing using the MiSeq protocol described previously.²⁹

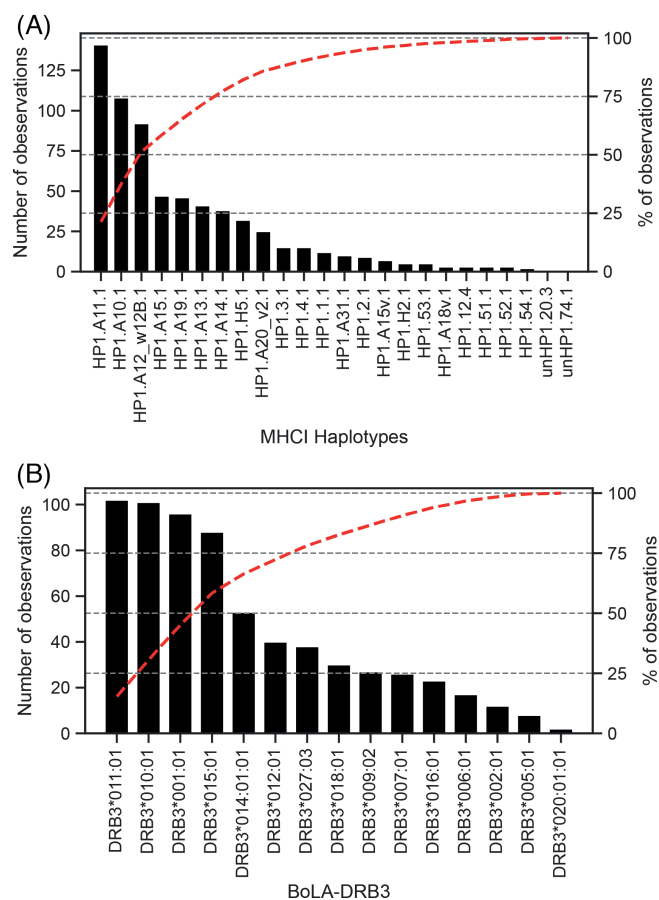


FIGURE 1 Frequency of (A) MHCI haplotypes and (B) DRB3 alleles observed in a cohort of 331 Holstein-Friesian animals. The MHCI haplotypes/DRB3 alleles are arranged in descending order of observation frequency on the horizontal axis. The number of times each haplotype was observed and cumulative percentage of observations (red line) are described in the left and right vertical axis, respectively

Analysis of the MHCI data generated from this cohort of animals identified 24 unique haplotypes, of which BoLA-A10, -A11 and -A12 were dominant, together constituting nearly $\sim 50\%$ of the haplotype observations made (Figure 1A). Seven novel haplotypes were identified: 5 were seen in multiple animals (HP1.51.1–HP1.54.1, HP1.12.4) whilst the remaining 2 (unHP1.74.1a and unHP1.20.3) were only identified in single individuals and so were assigned as putative haplotypes. All 7 of these novel haplotypes were present at low frequencies and together represented only 2.7% of the haplotype identifications (Table 1). Three of these haplotypes were constituted from previously identified alleles (HP1.53.1, HP1.54.1 and unHP1.74.1), whilst 1 was composed of only novel alleles (HP1.51.1) and the remaining 3 haplotypes included a combination of known and novel allelic sequences. Of the novel MHCI alleles, 6 were variants of

TABLE 1 Description of novel MHCI haplotypes identified from a cohort of 331 Holstein Friesian animals

Haplotype	BoLA Alleles				No. of times observed	Frequency of observation (%)
	1	2	3	4		
HP1.51.1	2*022:AA	1*074:AB			3	0.45
HP1.52.1	MHCI*gb20:01	3*004:02/3*053:01/ 3*081:01	2*016:02/2*016:03	6*091:01:AA	3	0.45
HP1.53.1	6*015:02	2*054:01			5	0.75
HP1.54.1	4*095:01	3*004:04/3*083:01			2	0.30
HP1.12.4	3*073:AA	3*004:AB:01	6*090:01		3	0.45
unHP1.74.1	3*010:01/3*010:02	3*004:02/3*053:01/ 3*081:01			1	0.15
unHP1.20.3	MHCI*cm8:03	MHCI*cm40:02			1	0.15

Note: For each haplotype the constituent alleles, the number of observations and their frequency in the cohort of animals are detailed. Novel alleles are shown in bold script. unHP1.74.1 and unHP1.20.3 are putative haplotypes that were only observed once in this cohort.

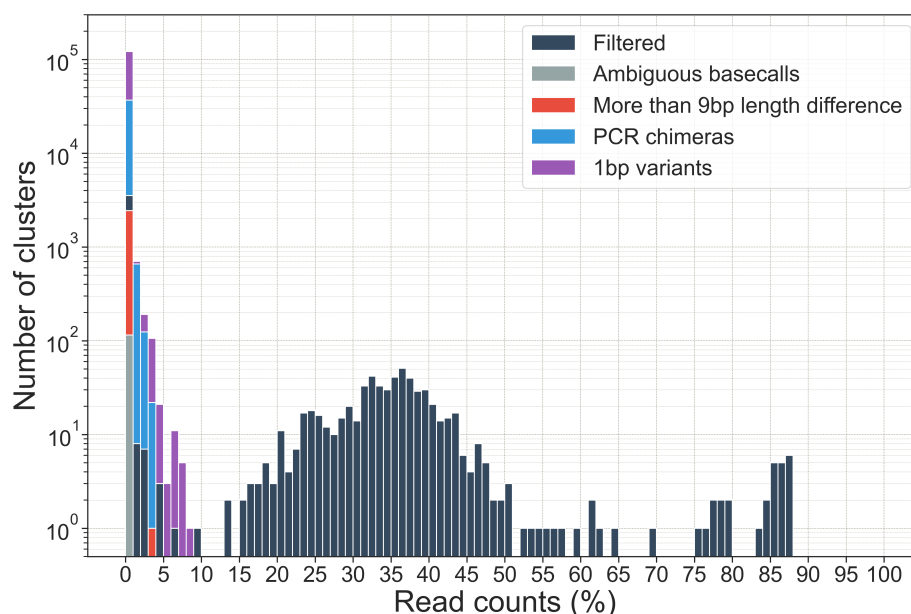


FIGURE 2 Histogram of the read frequency of unique DRB3 sequences identified from 331 Holstein-Friesian cattle. Individual DRB3 sequencing reads were clustered based on 100% identity and the percentage of reads in each cluster (with respect to the total number of reads for the individual) were plotted. Clusters that (i) had ambiguous basecalls (N) – light grey, (ii) were > 9 bp different in length from that anticipated – red, (iii) were chimeras assumed to have formed during PCR amplification – blue and (iv) contained potential PCR/sequencing errors (sequences that were 1 bp different from other sequences in the same sample present at >3× fold frequency) – purple; are removed from the dataset by filters in the pipeline. Clusters retained after this filtering (Filtered) are shown in dark grey

known alleles (2*022:AA, 1*074:AB, 6*091:01:AA, 3*073:AA, 3*004:AB:01 and MHC1*cm8:03) whilst MHC1*gb20:01 represented a novel allelic group. The sequence of each MHC1 allele is provided in Supplementary Data 1.

Analysis of the DRB3 read data was performed using a bioinformatics pipeline based on that previously developed for the MHC1 data (Supplementary Data 2b). The pipeline included filters to remove reads that (i) had ambiguous

bases (N), (ii) were > 9 bp different from that anticipated, (iii) were identified as PCR chimeras and (iv) contained potential PCR/sequencing errors (sequences that were 1 bp different from other sequences in the same sample present at >3-fold frequency). To define the frequency threshold to apply to the DRB3 data, the read frequency of each unique cluster (reads with 100% identity) in each animal was calculated, compiled and plotted on a histogram

TABLE 2 - Summary of the Brazilian cohort sample set

		Farm							Total
		AA	AB	AC	AD	AE	AF	AG	
Breed	Nelore	58		29				50	137
	Guzerat		33				100		133
	Gir				101	46			147
	Girlando				100	17			117
	Mixed		21						21
Total		58	54	29	201	63	100	50	555

Note: The number of animals in each farm (AA-AG) for each of the breeds (Nelore, Guzerat, Gir, Girlando and mixed) are detailed. Mixed animals were of various proportions of different breeds and so were excluded from downstream analyses (this group did not contain any exclusive novel alleles or haplotypes).

(Figure 2). The distribution of read frequencies revealed a distinct tri-modal pattern, with peaks evident at <7%, between 14–50%, and > 75% of read counts. It was assumed, and subsequently confirmed, that the peaks between 14–50% and > 75% corresponded to genuine DRB3 alleles expressed in heterozygous and homozygous animals respectively. The peak of clusters with a frequency of <7% were considered to represent artefacts, as signified by the dominance of clusters that were removed by the various filters applied in the bioinformatics pipeline. Based on this, a threshold between 7% and 14% was initially considered to be suitable for excluding artefacts from the DRB3 data. However, subsequent MHCI/DRB3 co-occurrence analysis (see below) suggested that both ‘filtered’ DRB3 clusters in the 5–10% range were likely to be genuine whilst those clusters present at <5% were not. Notably, the cluster present at 6.9% was a third allele in a twin (animal cLH164_gLH164 – this animal expressed 3 MHCI haplotypes as well as 3 DRB3 haplotypes, suggesting that a third MHC locus was present as a result of blood chimaerism), which may explain its abnormally low frequency. Based on these observations a threshold of 5% read frequency was incorporated into the pipeline.

Analysis of the DRB3 data for these 331 animals (Supplementary Data 3) identified 36 individuals that expressed a single DRB3 allele (of which 25 were also homozygous for the MHCI haplotype), 294 individuals that expressed 2 DRB3 alleles and a single animal that expressed three DRB3 alleles (cLH164_gLH164 – as discussed above). Fifteen different DRB3 alleles were identified in this cohort of animals (Figure 1B), with BoLA-DRB3*011:01, 010:01, 001:01 and 015:01 dominant (each at ~15%, together accounting for ~60% of the haplotype observations). All 15 DRB3 alleles found in this cohort were known alleles already present in the IPD database. The results generated from this cohort of animals provided validation for the MiSeq BoLA-DRB3 sequencing protocol.

3.2 | Analysis of MHCI and DRB3 repertoires in Brazilian cattle populations

Having validated the paired MHCI and DRB3 sequencing protocol we applied it to study a cohort of Brazilian cattle. This involved sampling a total of 555 animals from seven different farms (coded as farms AA-AG); the animals were predominantly from three different *Bos indicus* breeds—Guzerat, Nelore, Gir—and a Gir x Holstein-Friesian cross (Girlando) as summarised in Table 2.

3.2.1 | Analysis of the MHCI repertoire

Of these 555 animals 517 were heterozygous for MHCI, 7 were identified as having either 3 or 4 MHCI haplotypes (these were assumed to be twins with the number of haplotypes being accounted for by blood cell chimaerism) and 30 were homozygous for MHCI. Therefore, the total number of haplotypes observed was $n = 1119$. One animal (AE_042) expressed three alleles which could not be resolved into haplotype(s) and consequently none were assigned in this animal. A total of 93 different MHCI haplotypes were described in the dataset (Table 3). Of these, 24 had been reported previously either in the IPD-MHC database or previous work (Vasoya et al. 2016 – Supplementary Data 4), whilst the remaining 69 (74.2%) were novel. This included 57 haplotypes which were identified in multiple animals, and 12 which were designated as putative haplotypes as they were only identified in single animals. The frequency of the haplotypes showed a marked hierarchy with 2 of the novel dominant haplotypes (HP1.56.1, $n = 121$, 10.8% and HP1.84.1, $n = 90$, 8.0%) and the other haplotypes having a frequency that ranged in a continuous spectrum from 4.8% to 0.09% (Figure 3A). The data generated from the Brazilian cohort of animals are summarised in Supplementary Data 5.

TABLE 3 – Description of MHC I haplotypes identified from a cohort of 555 Brazilian cattle

Haplotype	Allele1	Allele2	Allele3	Allele4	Allele5	Number of observations	Percentage of observations
HP1.A10.1	3*002:01	2*012:01	6*092:01/6*092:02	3*086:01 N		14	1.26
HP1.A11.1	3*017:01	2*018:01/2*018:02	3*033:01 N			8	0.72
unHP1.A11.4	3*017:01	2*018:01:AA	3*033:01 N			1	0.09
HP1.A12_w12B.1	1*019:01	2*008:01				35	3.14
HP1.A13.1	1*031:01	2*032:01 N				1	0.09
unHP1.A13.2	1*031:02	2*032:01 N				1	0.09
HP1.A14.1	1*023:01	4*024:01:01:01/4*024:01:01:02	6*040:01	2*025:01		19	1.7
HP1.A15.1	1*009:01	4*024:01:01:01/4*024:01:01:02	2*025:01			13	1.17
HP1.A19.1	6*014:02	2*016:01				3	0.27
HP1.A20_v2.1	3*027:02	2*026:01/2*026:03	3*098:01 N			2	0.18
HP1.A31.1	1*021:01	2*022:01				3	0.27
HP1.BF1.1	2*054:01					1	0.09
HP1.H2.1	5*039:01					1	0.09
HP1.H5.1	3*036:01	3*037:01				15	1.35
HP1.1.1	1*028:AA	4*024:01:01:01/4*024:01:01:02				2	0.18
HP1.3.1	2*016:02/2*016:03	2*079:01	MHCI*gb5:01			9	0.81
HP1.4.1	4*076:01/4*076:02	2*054:01	3*078:01/3*078:02			19	1.7
HP1.5.4	1*029:AC	2*030:01:AA	MHCI*cm1:01			8	0.72
HP1.9.4	2*016:01:AA	MHCI*cm4:04	3*037:01			2	0.18
HP1.9.5	MHCI*cm4:05	2*016:01:AA	3*037:01			13	1.17
HP1.12.3	3*073:01	3*004:02/3*053:01/3*081:01	MHCI*cm1:02	MHCI*cm41:02		20	1.79
HP1.13.2	3*073:AF	MHCI*cm40:01	MHCI*cm1:02			3	0.27
HP1.19.1	2*089:AA	MHCI*cm7:01				9	0.81
HP1.25.2	2*022:AD	4*063:01	3*078:01/3*078:02			4	0.36
unHP1.29.2	5*064:AB					1	0.09
HP1.31.1	6*015:AA	2*012:AA				6	0.54
HP1.32.2	6*041:01	2*060:01	6*093:01			2	0.18
HP1.33.2	6*041:AC	2*016:01				5	0.45
HP1.39.3	MHCI*cm26:03	MHCI*cm31:02	MHCI*cm32:01	MHCI*cm45:01:AA		3	0.27
HP1.42.1	2*022:AA:01	3*098:ABN	3*001:01/3*001:02/3*001:03			13	1.17

(Continues)

TABLE 3 (Continued)

Haplotype	Allele1	Allele2	Allele3	Allele4	Allele5	Number of observations	Percentage of observations
HP1.43.1	3*050:AA	MHCI*ke12:01	3*004:AC			8	0.72
HP1.44.2	MHCI*ke1:02	3*004:AD	6*090:01			3	0.27
HP1.46.2	MHCI*cm4:08	2*008:AB	3*004:02/3*053:01/3*081:01			7	0.63
HP1.46.3	MHCI*cm4:09	2*008:AB	3*004:02/3*053:01/3*081:01			3	0.27
HP1.50.1	MHCI*cm10:01	3*078:01/3*078:02	MHCI*ke5:01			1	0.09
HP1.51.3	2*022:AA	1*074:AE				8	0.72
HP1.51.4	2*022:AA	1*074:AD				4	0.36
HP1.51.5	2*022:AC	1*074:AC				10	0.9
unHP1.52.2	MHCI*gb20:02	2*016:02/2*016:03	3*004:02/3*053:01/3*081:01	6*091:01		1	0.09
HP1.56.1	MHCI*br32:01	MHCI*br25:01				121	10.85
HP1.57.1	2*022:AC	MHCI*br26:01	2*062:01:AA			17	1.52
HP1.58.1	MHCI*br27:01	MHCI*br28:01	2*006:01/2*006:02/2*006:03			29	2.6
HP1.59.1	MHCI*br29:01	2*048:01	3*004:03			54	4.84
HP1.60.1	3*050:01	MHCI*br11:01	2*006:01/2*006:02/2*006:03	3*033:AHN		13	1.17
HP1.61.1	3*066:03	3*004:03	6*091:01			6	0.54
HP1.62.1	MHCI*br18:01	4*024:AA				2	0.18
HP1.63.1	6*034:01:AA	MHCI*br30:01				2	0.18
HP1.64.1	MHCI*br24:01	MHCI*br30:01				8	0.72
HP1.66.1	2*005:01	3*087:01:AA	MHCI*ke8:01			4	0.36
HP1.67.1	MHCI*ke12:01	MHCI*cm38:02	MHCI*cm41:01			5	0.45
HP1.68.1	1*074:AF	MHCI*br30:01				12	1.08
HP1.69.1	3*080:01	2*057:01	6*090:01			5	0.45
HP1.70.1	MHCI*cm25:01	2*075:AA	5*003:AB	MHCI*cm41:04	3*033:AIN	26	2.33
HP1.71.1	1*028:AB	2*032:AAN	2*025:01			17	1.52
HP1.72.1	1*074:AF	MHCI*br17:01				7	0.63
HP1.73.1	6*015:AC	2*008:AA	6*091:01			8	0.72
HP1.73.2	6*015:AB	2*008:AA	6*091:01			3	0.27
HP1.74.1	3*004:02/3*053:01/3*081:01	3*010:01/3*010:02	MHCI*br19:01			24	2.15
HP1.75.1	1*097:01:AA	2*071:01	2*025:01	6*040:01		3	0.27
HP1.76.1	3*004:02/3*053:01/3*081:01	MHCI*br10:01	MHCI*br5:01	6*092:01/6*092:02		2	0.18

TABLE 3 (Continued)

Haplotype	Allele1	Allele2	Allele3	Allele4	Allele5	Number of observations	Percentage of observations
HP1.77.1	MHCI*br20:02	MHCI*br21:01	6*090:AB	6*092:AB		13	1.17
HP1.79.1	MHCI*br7:01	MHCI*br15:01				46	4.13
HP1.80.1	MHCI*br3:01	MHCI*br13:01				3	0.27
HP1.81.1	2*022:AA	3*004:02/3*053:01/3*081:01	MHCI*br4:01			14	1.26
HP1.82.1	3*037:01	3*035:AB				32	2.87
HP1.83.1	6*013:AA	2*018:01/2*018:02				5	0.45
HP1.84.1	1*074:AG	MHCI*cm5:01	MHCI*cm1:03			90	8.07
HP1.85.1	MHCI*br24:02	2*008:01				4	0.36
HP1.86.1	MHCI*br7:01	1*074:AD	2*006:01/2*006:02/2*006:03			56	5.02
HP1.87.1	3*035:01	3*004:02/3*053:01/3*081:01	2*070:01			17	1.52
HP1.88.1	MHCI*cm41:04	MHCI*cm4:06	2*016:01:AA	3*004:02:AC		3	0.27
HP1.89.1	MHCI*br6:01	MHCI*ke12:01	3*033:01 N			2	0.18
HP1.90.1	1*021:01	2*071:AC				19	1.7
HP1.91.1	6*015:AD	2*085:AA				17	1.52
HP1.91.2	6*015:AD	2*085:AB				3	0.27
HP1.91.3	6*015:AE	2*085:AA				4	0.36
HP1.92.1	3*065:01	MHCI*br17:01	MHCI*cm41:03			14	1.26
HP1.93.1	2*047:AD	3*087:01	2*032:01 N	MHCI*ke8:03		33	2.96
unHP1.93.2	2*047:01	2*032:01 N	3*087:01	MHCI*ke8:02		1	0.09
HP1.94.1	1*074:AF	MHCI*br11:01	2*006:01/2*006:02/2*006:03			39	3.5
HP1.95.1	MHCI*br1:01	3*004:03				5	0.45
HP1.96.1	MHCI*br9:01	2*060:01	6*093:01			5	0.45
HP1.97.1	MHCI*cm9:02	1*067:AA				9	0.81
HP1.98.1	1*007:AA	2*008:01				14	1.26
HP1.99.1	2*018:AB	3*037:01	3*017:AC			7	0.63
unHP1.100.1	4*063:01	2*032:ABN	3*078:01/3*078:02			1	0.09
unHP1.101.1	2*089:01	4*094:01	3*004:04/3*083:01			1	0.09
unHP1.102.1	2*075:01	3*004:04/3*083:01	4*095:AA			1	0.09
unHP1.103.1	MHCI*br26:01	2*062:01:AA				1	0.09
unHP1.104.1	MHCI*br12:01	MHCI*cm22:01	6*040:01			1	0.09

(Continues)

TABLE 3 (Continued)

Haplotype	Allele1	Allele2	Allele3	Allele4	Allele5	Number of observations	Percentage of observations
unHP1.105.1	MHCI*cm25:01	2*022:AC	5*003:AB	1*074:AG		1	0.09
unHP1.106.1	MHCI*br23:01	3*004:02/3*053:01/3*081:01	2*047:01		1	1	0.09
HP1.195.1	MHCI*cm4:01	2*016:01:AA	3*087:01:AA		4	4	0.36

Note: For each the names of the constituent alleles, the number of observations and their frequency in the cohort of animals are detailed. Novel alleles and haplotypes are shown in bold script. Alleles shown in underlined script consistently had low read frequency in animals carrying the haplotype and were not observed in a subset of these animals (presumed to be due to being present below the threshold of 0.2% of reads). Putative haplotypes that were observed in single animals are indicated by prefix of 'un' in their names.

Consistent with data we reported previously²⁹ there was evidence of marked variation in the 'structure of transcribed MHCI haplotypes' between the individual haplotypes, with variation in both the number of alleles present (ranging from 1 to 5 alleles—Table 3) and the relative frequency at which alleles within a haplotype were expressed. For 14 of the 57 novel haplotypes identified in multiple individuals, linear regression analysis showed a high level of correlation ($R^2 > 0.90$) between the read frequency data derived from the two independent PCRs (using the For1/Rev2 and For3/Rev1 primer pairs); suggesting that for these haplotypes read frequency data was not substantively affected by PCR bias and so could be used to describe the relative expression levels of allele mRNA transcripts (e.g., Figure 4A). For an additional 21 haplotypes, the correlation was not as high but was sufficient to infer the relative expression levels of the different alleles (e.g., Figure 4B). Amongst these 35 haplotypes, a large number of different MHCI expressed 'structures' were evident (i.e., different patterns in the number and relative expression levels of alleles—Supplementary Data 6). For the remaining 22 haplotypes it was not possible to infer the structure of the expressed alleles due to substantial 'skewing' of the read frequency data as a consequence of PCR bias with the different primer pairs (e.g., Figure 4C).

A total of 176 MHCI alleles were sequenced from this cohort, which included 101 that have been previously described and 75 that were novel (of which seven were only identified in single animals). Of the 75 novel alleles 26 were members of new allele groups, 44 represented new allele subgroups (of previously described groups) and five were synonymous variants of previously described alleles. Phylogenetic analysis shows these novel alleles, including those representing novel allelic groups are distributed throughout the phylogenetic range and are thoroughly intercalated with previously identified alleles (Supplementary Data 7a). Nine (14.1%) of the novel haplotypes expressed new combinations of known alleles, 15 (21.1%) expressed only novel alleles but the majority ($n = 45\text{--}64\%$) expressed a combination of known and novel alleles.

3.2.2 | Analysis of the DRB3 repertoire

Of the 555 animals studied, 494 were heterozygous for DRB3, 52 were homozygous for DRB3 and 7 were identified with either 3 or 4 DRB3 alleles. Two animals were identified as having >4 DRB3 alleles, which we could not explain from a biological perspective; these data were excluded from subsequent analyses. Therefore, the total number of analysed DRB3 allele observations in the

dataset was 1112. A total of 45 unique DRB3 alleles were identified, 2 (4.4%) of which were novel alleles. As with the MHCI, there was a clear hierarchy in the representation of the different DRB3 alleles in the Brazilian cohort (Figure 3B), with BoLA-DRB3*036:01 ($n = 166$ –14.8%) identified approximately twice as often as the next three most frequently observed haplotypes (DRB3*015:01, DRB3*016:01 and DRB3*033:01 – all present at ~7–8%), whilst the remaining DRB3 alleles were present at frequencies ranging from 5.5% ($n = 61$) to 0.1% ($n = 1$). The combined frequency of the novel DRB3 allele observations was only 1.53%, showing these novel alleles to be relatively under-represented. As with the novel MHCI alleles, the novel DRB3 alleles are intercalated with previously identified alleles in multiple clades in the phylogenetic tree (Supplementary Data 7b).

3.2.3 | Breed-associated MHCI haplotype and DRB3 allele structures

Having described the Brazilian cohort we next evaluated the similarity of the MHCI haplotype and DRB3 allele repertoires in the Gir, Guzerat and Nelore populations. Visualisation of the MHCI haplotypes present in each population (divided by breed and farm of origin) indicated that the MHCI haplotypes in the three breeds were largely discrete, with a limited number of haplotypes being shared by multiple breeds (Figure 5A). In contrast, comparison between populations of the same breed on different farms showed a high level of overlap. Notably, HP1.56.1 (the most frequent MHCI haplotype in the cohort) was present at high frequency in both Nelore and Gir but absent from Guzerat, whilst conversely HP1.84.1 (the second most common haplotype in the cohort) was a high frequency haplotype in both the Nelore and Guzerat but absent from Gir (Figure 3A); in fact, of the 50 most frequent haplotypes only 2 were present in all 3 breeds (HP1.59.1 and HP1.3.1 – Figure 3A). To quantify the similarity between the sample populations we performed pairwise comparisons calculating the i) percentage of shared haplotypes and ii) percentage of observed MHCI in shared haplotypes between the sample populations (Table 4). For within breed comparisons, the percentage of shared haplotypes ranged from 32% to 74.5%, whereas for inter-breed comparisons the range was consistently lower – 10.3% to 30.2%; the equivalent data for the percentage of observed MHCI in shared haplotypes was 48.1%–91.2% and 10.2%–37.2%, respectively. Pairwise comparison of each of the Guzerat, Gir and Nelore populations with the Holstein-Friesian cohort showed low levels of both haplotype sharing (0% to 11.76%) and percentage of observations in shared haplotypes (0 to

25.29%). Together the data suggested that the three *Bos indicus* breeds each had a distinct MHCI repertoire but that these repertoires were more similar to each other than they were to that observed in a *Bos taurus* population.

A parallel analysis of the DRB3 (Figure 5B and Table 4) yields a similar description with intra-breed comparisons showing higher levels of similarity (52.7%–88.2% for shared DRB3 alleles and 62.3%–95.3% for observations of alleles that were shared), than in inter-breed comparisons (14.8%–57.9% and 12.7%–73.9%, respectively), which in turn was higher than that observed when the three *Bos indicus* breeds were compared to the Holstein-Friesian cohort (11.76%–37.84% and 18.14%–48.75%). However, in comparison to the MHCI repertoire data, the DRB3 data provided less resolution of the segregation of the breeds and lineages. For example, the Nelore population on farm AA had a higher percentage of shared DRB3 alleles with both the Guzerat population in farm AB and the Gir population in farm AD (55% and 57.9% respectively) than the Nelore on farm AC (53.8%). Similarly, the Gir population in farm AE and Guzerat population in farm AB both had higher levels of identity with the Holstein Friesian cohort (31.25%/48.75% and 37.84%/47.61%) than was seen in a number of the comparisons between *Bos indicus* breeds. This perhaps reflects the comparatively smaller diversity of DRB3 alleles than MHCI haplotypes. As would be anticipated, the MHCI and DRB3 repertoires in the Girlando, a Holstein-Friesian x Gir hybrid, showed high levels of similarity with both the Holstein-Friesian and Gir populations.

3.3 | Association between MHCI haplotypes and DRB3 alleles

To examine the linkage of MHCI haplotypes and DRB3 alleles we performed an analysis of co-expression of DRB3 alleles with MHCI haplotypes (Figures 6 and 7, and Supplementary Data 8 and 9). In the Holstein-Friesian cohort (Figure 6) there were clear patterns of MHCI-DRB3 association; for the 20 MHCI haplotypes that were observed in three or more individuals (the minimum number for which MHCI-DRB3 association was analysed), 9 (45%) were always associated with the same DRB3 allele and a further 7 (35%) MHCI haplotypes were associated with the same DRB3 allele in >75% of the individuals in which they were observed. In three of these haplotypes (A12, A13 and H5) secondary associations with an alternative DRB3 allele were also evident. In the other four haplotypes—A11, A31, A19 and A14—there was no single dominant DRB3 association but there

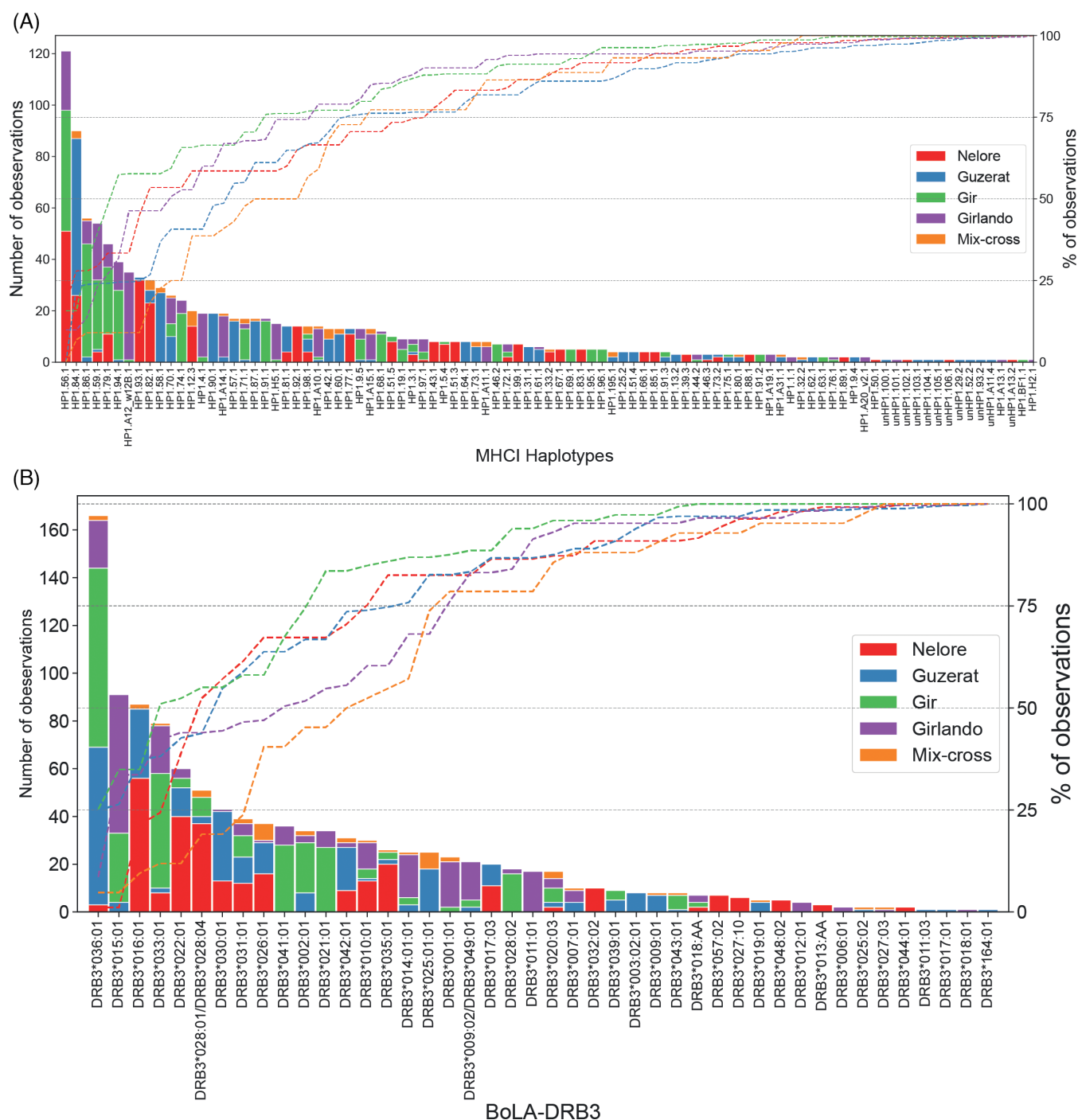
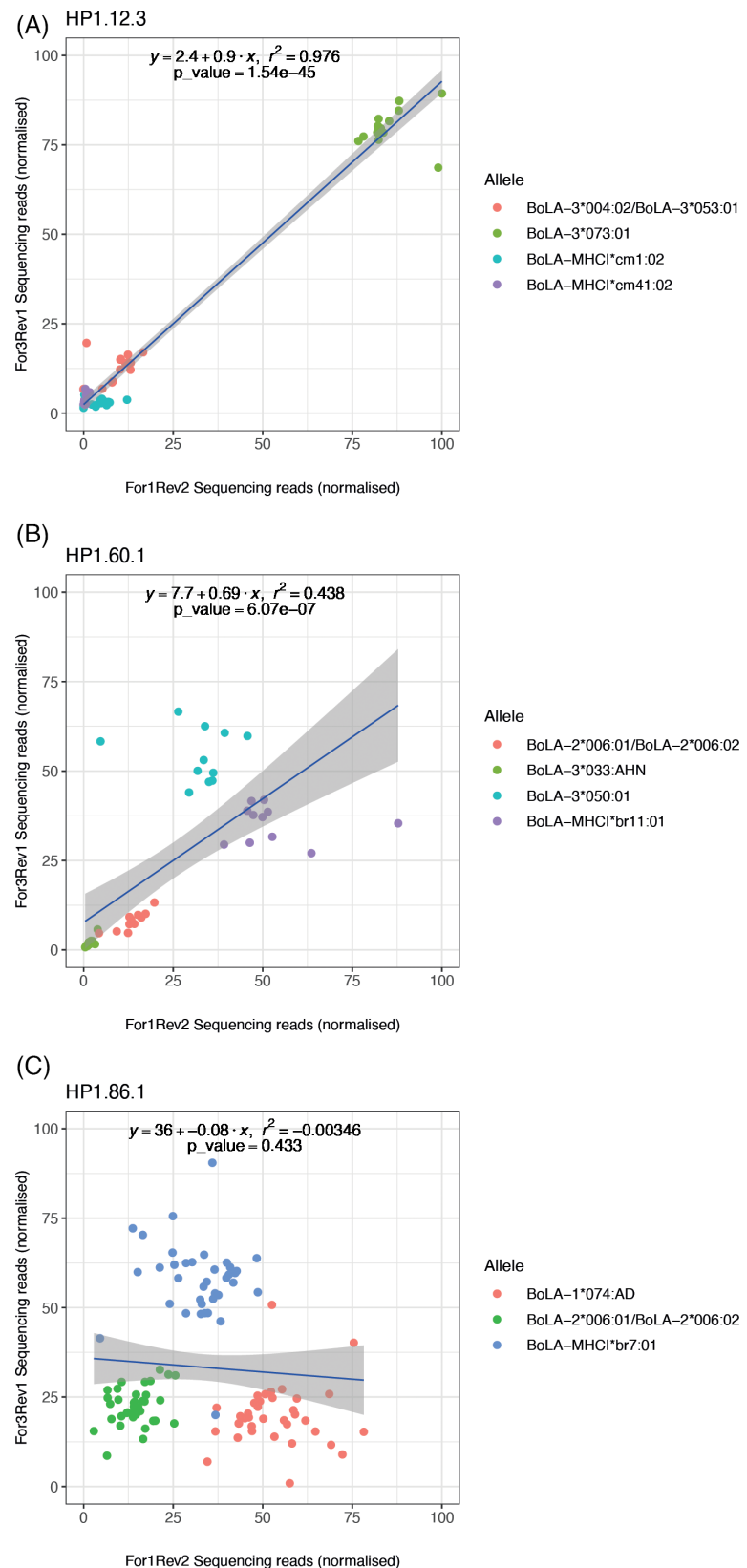


FIGURE 3 Frequency of (A) MHCI haplotypes and (B) DRB3 alleles observed in a cohort of 555 Brazilian cattle. The MHCI haplotypes/DRB3 alleles are arranged in descending order of observation frequency on the horizontal axis. The distribution of the haplotypes between the different breeds is shown by colouring as described in the legend. The number of times each haplotype was observed and cumulative percentage of observations are described in the left and right vertical axis, respectively

was association with multiple DRB3 alleles. Together, 76% of the observed MHCI-DRB3 pairings in the Holstein-Friesian population could be explained by just the primary MHCI-DRB3 associations, and 96.7% could be accounted for if secondary, tertiary and quaternary DRB3 associations were considered.

Due to the greater MHCI and DRB3 diversity in the Brazilian dataset, the results of a parallel analysis were more complex but still provided clear evidence of MHCI-DRB3 associations (Supplementary Data 8 and 9). Of the 63 MHCI haplotypes that were observed in more than three individuals, 23 (36%) were always associated with

FIGURE 4 Linear regression analysis of read frequency for alleles as derived from sequencing of amplicons generated by two independent PCRs for haplotypes (A) HP1.12.3 (B) HP1.60.1 and (C) HP1.86.1. In haplotypes for which the line of best fit had a slope of between 0.8–1.1 and had an R^2 value of >0.95 (with a $p < 0.05$) it was assumed the read frequency data was not substantively affected by PCR bias and therefore read frequency likely represented relative expression of alleles (e.g., HP1.12.3). For some haplotypes the relative expression levels of alleles could be inferred even if these criteria were not met (e.g., HP1.60.1), while for other haplotypes PCR bias led to significant skewing of the read frequency data and so relative expression of the alleles could not be inferred (e.g. HP1.86.1)



the same DRB3 allele and a further 24 haplotypes (38%) were associated with the same single DRB3 allele in $>75\%$ of the individuals in which they were observed – similar to the figures seen in the Holstein-Friesian

cohort. Of the 16 remaining MHCII haplotypes, 14 had secondary and/or tertiary and quaternary associations (as did 7 of the haplotypes with primary associations $>75\%$). Interestingly there were a number of examples

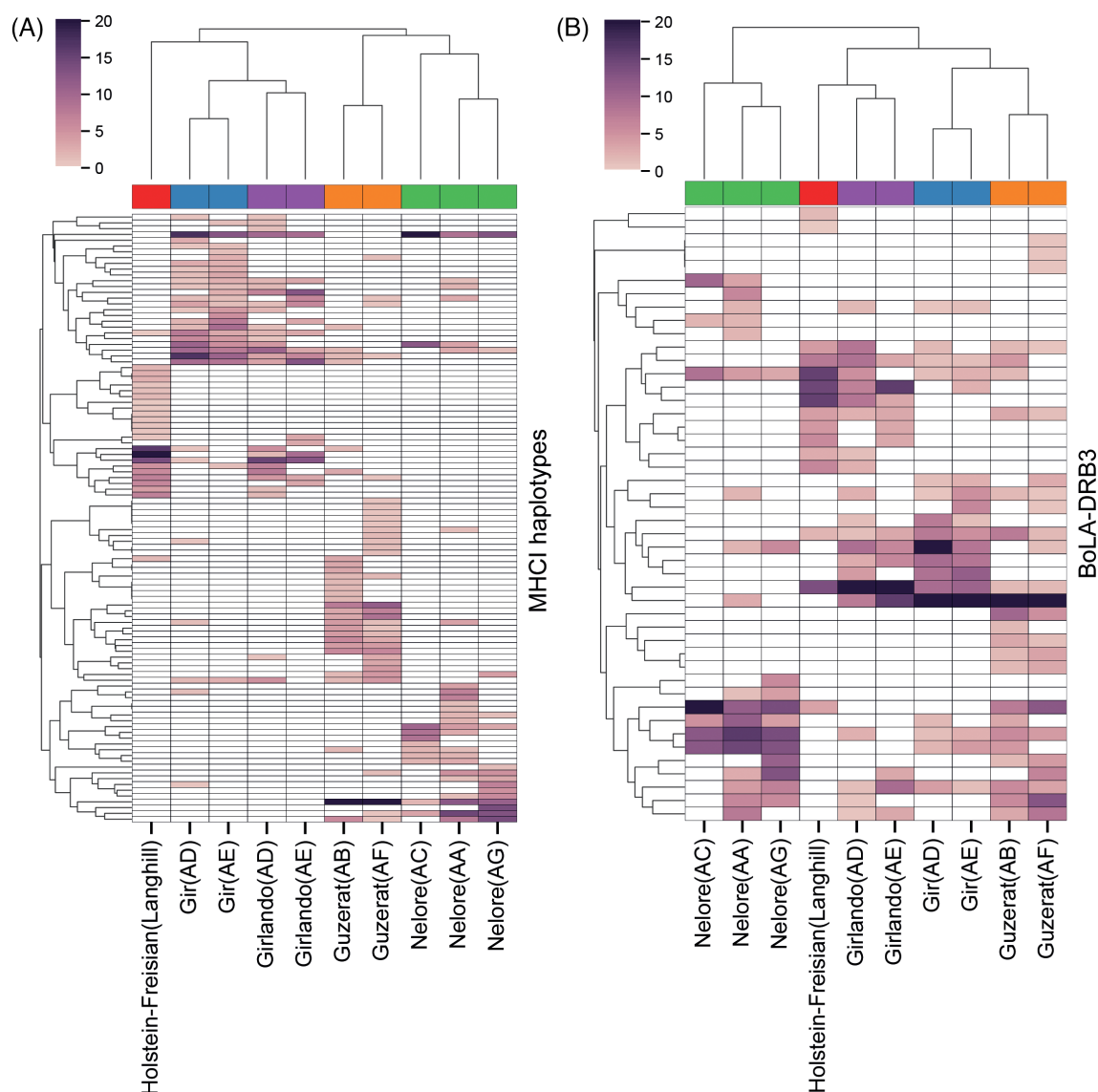


FIGURE 5 Heat-map representation of the frequency of (A) MHCI haplotypes and (B) DRB3 alleles in the Brazilian cohort grouped according to breed and farm origin. The colour of the boxes reflects the relative frequency of each MHCI haplotype as shown in the legend. The hierarchical clustering shows the relatedness between the different populations, showing that the populations cluster according to breed

where MHCI haplotypes shared between multiple breeds showed distinct breed-related DRB3 associations. One example is HP1.56.1 (Figure 7), where although association with DRB3*033:01 and 35:01 was observed in both Gir and Nelore cattle, 16:01 and 30:01 were unique to Nelore and 36:01 (and other low frequency MHCI haplotypes only observed in a single farm) was unique to Gir.

Together, 74.1% of MHCI-DRB3 co-expression in the Brazilian cohort could be explained by just the primary MHCI-DRB3 associations, and 92.9% could be accounted for if the secondary, tertiary and quaternary DRB3 associations were included; again, these figures are similar to those calculated for the Holstein-Friesian cohort. In the Brazilian cohort there was a strong correlation ($R^2 = 0.8$) between the observed frequency of DRB3 alleles and the

number of MHCI haplotypes they were associated with; for example, the most commonly observed DRB3 allele, BoLA-DRB3*036:01, was found associated with the highest number of different MHCI haplotypes (11). In contrast, no such correlation was observed in the Holstein-Friesian cohort ($R^2 = 0.006$ – data not shown).

4 | DISCUSSION

In this study, we describe the development and application of a novel high throughput sequencing (HTS) approach to characterise bovine BoLA-DRB3 alleles using the MiSeq platform. Using this in parallel with the equivalent approach for characterising bovine MHCI alleles²⁹

TABLE 4 Pairwise comparison of (A) MHCI haplotypes and (B) DRB3 alleles shared between populations included in this study

A												
			Percentage of shared MHCI haplotypes									
			Nelore			Guzerat		Gir		Girlando		Holstein-Friesian
			AA	AC	AG	AB	AF	AD	AE	AD	AE	
Percentage of observations in shared haplotypes	Nelore	AA		48.6	48.8	18.5	26.2	30.2	24	19.2	19	4
		AC	60.2		32	10.3	13	10.5	11.4	10.8	7.4	0
		AG	62.7	48.1		18.6	20	14.3	10.3	9.8	12.9	0
	Guzerat	AB	29.9	14.2	30.4		47.6	25.5	19.2	25.9	13.6	11.54
		AF	35.8	21.7	35.4	77.3		19.4	16.9	16.4	15.7	6.78
	Gir	AD	37.2	35	25.3	33.6	21.7		74.5	60.4	46.5	11.76
		AE	23.9	34	16.8	24.4	10.2	91.2		64	50	12.5
	Girlando	AD	23.8	24.7	18.2	30.4	11.7	77.4	65.9		52.4	40
		AE	16	34.4	13.4	9.9	5.2	68.4	58.3	62.7		35
	Holstein-Friesian		2.3	0	0	21.96	7.29	25.29	8.33	76.07	53.3	
B												
			Percentage of shared DRB3 alleles									
			Nelore			Guzerat		Gir		Girlando		Holstein-Friesian
			AA	AC	AG	AB	AF	AD	AE	AD	AE	
Percentage of observations of shared DRB3 alleles	Nelore	AA		53.8	71	55	47.62	50	44.4	45	29.4	11.76
		AC	74.7		52.7	35.7	14.8	33.3	25	14.3	0	18.18
		AG	82.8	62.3		48.5	40	41.4	34.5	30.3	22.2	14.81
	Guzerat	AB	73.9	52.4	56.8		72.7	57.9	47.4	57.1	44.4	37.84
		AF	69.1	26.7	53	87.2		45	45	50	47.4	26.32
	Gir	AD	55.3	12.7	35	55.6	54.6		88.2	73.7	55.2	31.25
		AE	49.1	17.5	30.5	58.5	49.5	95.3		73.7	50	31.25
	Girlando	AD	35.8	10.8	24.2	64.7	59	87.3	81.1		66.7	55.56
		AE	22.7	0	23.7	53.9	51.1	75	67.7	81.7		60
	Holstein-Friesian		18.14	21.91	18.46	47.61	24.3	37.34	48.75	81.02	71.74	

Note: The (i) percentage of haplotypes (top right) and (ii) percentage of observed haplotypes (i.e. shared haplotypes weighted by the number of times they were observed – bottom left) shared between different populations included in this study are shown. The percentage of shared haplotypes (alleles) was calculated as: Shared haplotypes (Sh_h) = (number of haplotypes in population 1 shared with population 2 (Sh_{h1}) + number of haplotypes in population 2 shared with population 1 (Sh_{h2}))*100/(total number of haplotypes in population 1 (T_{h1}) + total number of haplotypes in population 2 (T_{h2})). The percentage of shared haplotype observations as: Shared observed haplotypes (Sh_{ob}) = (number of haplotype observations in population 1 for haplotypes shared with population 2 (Sh_{ob1}) + number of haplotype observations in population 2 for haplotypes shared with population 1 (Sh_{ob2}))*100/(total number of observed haplotypes in population 1 (T_{ob1}) + total number of observed haplotypes in population 2 (T_{ob2})).

we have determined the combined MHCI/DRB3 genotypes of >800 animals, including a large cohort of animals from *Bos indicus* breeds common in Brazil.

The system for the bovine DRB3 HTS genotyping was largely based on the equivalent system we established for the bovine MHCI. However, the biology of the BoLA-DR is less complex than that of the BoLA-MHCI system, with only a single locus (DRB3) expressed and only a single

highly polymorphic exon (exon 2) contributing to the functionally important ‘peptide binding groove’ of the -BoLA-DR molecule. Consequently, the system could be simplified, with a single set of primers sufficient as i) the entirety of the polymorphic area of interest (exon 2) could be sequenced from a single amplicon and ii) the primers could be located in the highly conserved exons 1 and 3, so that ‘allele dropout’ due to reliance on a

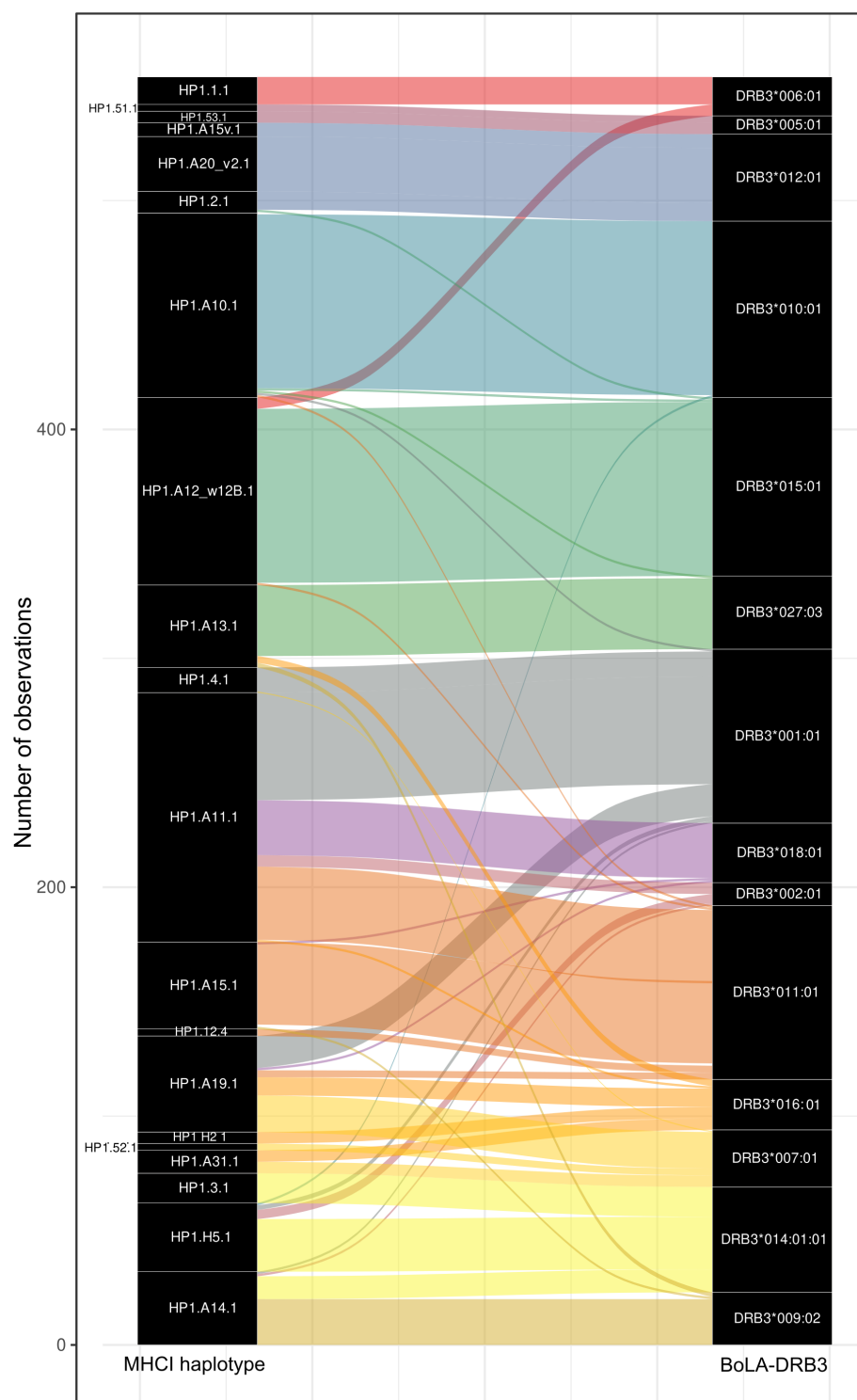
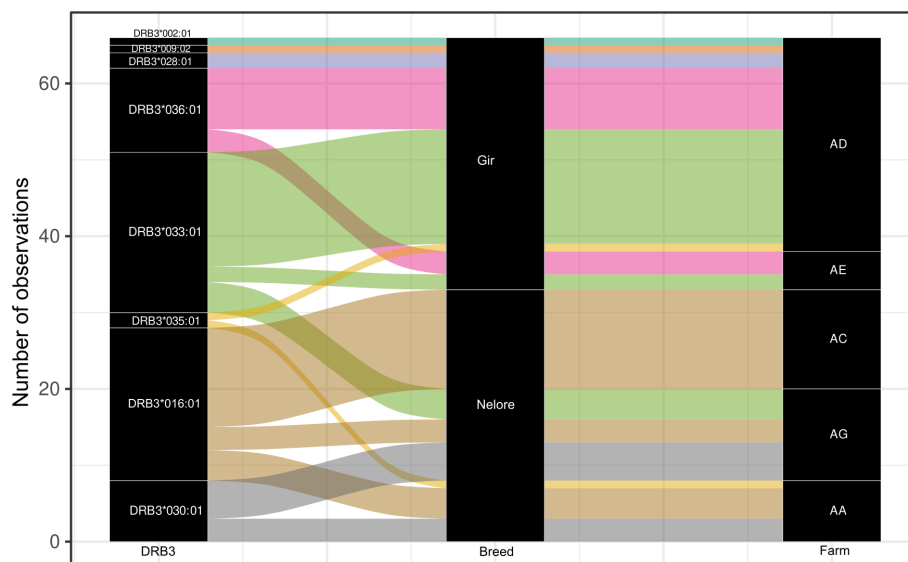


FIGURE 6 Alluvial diagrams showing the co-expression patterns for MHCII haplotypes and DRB3 alleles in the Langhill Farm Holstein-Friesian herd. Data were only presented for MHCII haplotypes that were observed at least twice in the cohort and from animals which were heterozygous for both MHCII and DRB3. For all selected animals, (i) pairwise comparison between MHCII and DRB3 was used to calculate the normalised frequency of each DRB3 for each MHCII haplotype, (ii) DRB3 and MHCII haplotypes in each animal were linked by using the best calculated normalised frequency of each DRB3 from both MHCII haplotypes, (iii) the frequency at which each MHCII-DRB3 co-occurred was calculated

single primer pair was unlikely to be a major detrimental factor. The bioinformatics pipeline used for the DRB3 system was also essentially a simplified version of that used for the MHCII, including application of many of the same filters to remove artefacts (e.g., chimera detection and removal). However, the 5% threshold for minimal read frequency was established empirically for the DRB3 pipeline using the Holstein-Friesian sample-set for

validation. We were able to set this threshold substantially higher than that used for MHCII (for which the threshold value is 0.2% of reads) due to the much lower complexity with regard to the number of gene products being amplified and sequenced, and also because the lower level of disparity between read frequencies for different alleles seen in DRB3 compared to MHCII. Bovine DRB3 typing using HTS has been attempted previously

FIGURE 7 Alluvial diagrams showing the co-expression patterns for MHC I haplotypes and DRB3 alleles in the subset of Gir and Nelore animals within the Brazilian cohort that expressed the HP1.56.1 MHC I haplotype. The data are also shown for the farm origin of the animals



using the 454 pyrosequencing platform³² but corrections of the errors in the sequence data were a major problem; the improved accuracy of MiSeq and the robust bioinformatics pipeline established herein overcome this obstacle.

For cattle the extensive application of non-HTS approaches to DRB3 genotyping of cattle populations around the world (e.g., references^{15,17,18}, has already generated a large database of known DRB3 alleles (331 and 365 listed on <https://www.ebi.ac.uk/ipd/mhc/> in June 2020 and May 2021 respectively). Consequently, we only identified a small number (2) of novel DRB3 alleles in the Brazilian cohort. This is consistent with recent studies of DRB3 diversity in both *Bos indicus* and Holstein-Friesian (and other *Bos taurus*) cattle in South America using non-HTS technologies^{16,19,20} where only three novel alleles were identified from sequences of ~1500 animals. Similarly, no novel DRB3 allele was identified in the Holstein-Friesian cohort. This summary of the data obviously carries the caveat that it assumes there is no variation outwith the amplified and sequenced portion of the DRB3 genes and so this may underestimate the number of novel alleles present in these populations (this also applies to the MHC I data and needs to be considered in how the data are interpreted). We are currently validating a PacBio sequencing approach that will enable the full length of MHC alleles to be sequenced to address this issue.

The main benefit of incorporating DRB3 analysis into the HTS platform for bovine MHC genotyping will be to provide an integrated system where MHC I and MHC II repertoires can be examined in parallel at a large scale and at high resolution. This may be especially beneficial for studies of MHC and disease association^{33–37} and may be particularly pertinent for two reasons. First, our data

has shown that in both the Holstein-Friesian and Brazilian cohorts there is a greater diversity of MHC I haplotypes than DRB3 alleles represented and that the same MHC I may be linked to multiple DRB3 (and vice versa); consequently, combined analysis of MHC I and DRB3 may give a greater granularity to any association studies. Second, in diseases where CD8+ T-cells are key mediators of immunity, for example BLV, the ability to look for associations with the MHC I as well as the MHC II repertoire may provide additional functionally relevant information.^{23,38–40}

In contrast to DRB3, a large number (75) of novel MHC I alleles were identified in the Brazilian cohort analysed in this study (and a further seven were identified in the Holstein-Friesian cohort). The disparity between the proportion of novel alleles identified for DRB3 and MHC I highlights the discrepancy in the depth to which these two sets of MHC genes have been characterised, largely attributable to the technological obstacles to high-throughput MHC I analysis prior to HTS approaches. Most of the novel alleles (73/82) were identified within the 417 pure *Bos indicus* animals included in the study. This represents the first (as far as the authors are aware) large-scale analysis of MHC I alleles from *Bos indicus* animals and provides a large quantity of data on the MHC I repertoire on these populations, which play a crucial role in the agriculture of large regions of the world.

A notable feature of the MHC I data was the presence of nearly equivalent numbers of novel alleles (82 in total) and novel haplotypes (76 in total). Most of the novel haplotypes were composed of a mix of novel and known alleles (48), with novel combinations of known alleles accounting for almost as many novel haplotypes as combinations of only novel alleles (12 vs. 16, respectively).

These data suggest that, as in humans, different permutations of alleles resulting from recombination within the MHC locus may be a major source of haplotype diversity for cattle.⁴¹ However, ~20 of the novel haplotypes were variants of previously identified haplotypes (i.e., haplotypes that contained nearly identical alleles, but where 1 or more were subgroup variants), showing that allele mutation also contributes to the diversification of the bovine MHCI haplotype repertoire.

Recombination also appears to be a major factor in diversifying the combined MHCI/DRB3 haplotypes. Although there was strong MHCI-DRB3 associations evident in both the Holstein-Friesian and Brazilian cohorts, a significant minority (~25%) of MHCI-DRB3 pairings differed from those most commonly observed, suggesting that recombination events at positions between the MHCI and MHCII loci were generating different permutations of MHCI-DRB3 and so expanding the range of extended MHCI-MHCII haplotypes. In the Brazilian cohort, although there were only 93 MHCI haplotypes and 45 DRB3 alleles identified, different permutations of the co-expressed MHCI and DRB3 alleles suggested the presence of up to ~570 unique MHCI-DRB3 haplotypes. However, analysis of a larger sample-set will be required to confirm some of these extended MHCI-DRB3 haplotypes as they were only identified in single animals. In humans, the diversity that arises from different MHCI-MHCII pairings is predominantly significant in the context of transplantation medicine and host versus donor graft responses. These are obviously not major factors in livestock; instead, understanding the extent of MHCI-MHCII linkage and the limitations of the inferences that can be made about the MHCI haplotype from DRB3 identification and vice versa in these cattle populations are of more relevance.

The derivation of our sample-set from a limited number of farms (8 in total) has a number of implications for how the data generated from the study can be interpreted. In examining the diversity of MHC present in Brazilian *Bos indicus* cattle populations a sampling strategy based on smaller sample sets from a large number of farms could have been advantageous in mitigating the potential of breeding practices (e.g., bull selection) and relatedness between individuals within herds in biasing and artificially narrowing the MHC repertoires identified. However, it was assumed, based on available data from both humans and cattle, that the MHC in cattle herds would be characterised by a predominance of a limited number of haplotypes leading to haplotype frequency distributions with long 'right-hand tails'⁴¹; as seen in the data presented herein, this assumption was correct. This has two effects that are particularly relevant to the objectives of this study. Firstly, intensive sampling of herds is

capable of capturing a large diversity of MHC haplotypes – the ratio of MHC haplotypes identified in the Brazilian cohort was 1 MHCI haplotype/6 animals (i.e., 93 MHCI haplotypes were identified in the cohort of 555 animals). Secondly, as we rely on identification of MHC allele combinations being identified in multiple individuals to confirm haplotypes, this approach (where related individuals sharing rare MHC haplotypes are more likely to be sampled) was thought to be advantageous. Even after adopting this approach, 12/69 of the novel haplotypes were designated as putative as they were combinations of alleles that were only observed in single animals; taking a more dispersed sample set would have likely led to a greater proportion of haplotypes that were only observed once. However, deriving all of the samples for a breed from a single farm would have made it impossible to discriminate between farm and breed factors influencing MHC diversity and thus preclude any assessment of breed-specific repertoires. To compensate for this, we sought to generate data from multiple farms for each breed to allow both intra-breed and inter-breed comparisons. Although the granularity of the sampling structure was insufficient to appropriately apply population genetics statistical analyses (e.g., F_{ST}) the data generated is highly supportive of the Gir, Nelore and Guzarat breeds having distinct, although not exclusive, MHCI repertoires and that the repertoires in these breeds are more similar to each other than to the MHCI repertoire in the *Bos taurus* Holstein Friesian breed. However, it should be noted the Holstein-Friesian population is anticipated to exhibit an unusually restricted MHC diversity due to high levels of selection for production traits and so is unlikely to be representative of the broader *Bos taurus* population. The equivalent data for DRB repertoires was similar but was less robust – probably reflecting the lower diversity of DRB3 alleles observed in the cohort as a whole and so a lower resolution of discrimination between the breeds. Notably, the evidence of breed-related MHCI-DRB3 association patterns identified in this study indicates that use of extended MHCI-MHCII haplotypes would reveal greater discrimination between the breeds. Although the data supported the notion of breed-specific MHC repertoires it is important to note that for both MHCI and DRB3 alleles the sequences derived from *Bos indicus* animals were intercalated with the sequences generated from other populations, suggesting that there was no distinct segregation of MHC alleles between the *Bos taurus* and *Bos indicus* lineages.

The sampling structure employed in this study may also have had an influence on the MHCI-DRB3 association data presented in the study and so needs to be considered in its interpretation. Use of limited numbers of bulls as sires in individual herds (especially for the UK

Holstein-Friesian herd) may have limited not only the diversity of MHC I and DRB3 repertoires but also the diversity of the extended MHC I-DRB3 haplotypes, leading to an over-estimation of the association between MHC I haplotypes and DRB3 alleles. Due to a lack of other datasets characterising MHC I and DRB3 association in cattle it is not possible to make comparisons to evaluate if this significantly affected our results. However, identification of the same MHC I-DRB3 associations across multiple herds, including between the Holstein-Friesian cohort in the United Kingdom and Girlando (Gir × Holstein-Friesian crosses) animals in Brazil, suggests that this was not the case. Due to the inherent complexity, a comprehensive comparative characterisation of the extended MHC I-DRB3 haplotype diversity in the cattle breeds analysed herein would require a much larger, and differently structured, study than that undertaken here.

In conclusion, we present here a new HTS system to study in parallel MHC I and DRB3 repertoires in cattle. The high number of recent publications focusing on analysis of DRB3 repertoires demonstrates the potential utility of this tool in a number of areas of research.^{40,42–47} In this study, application of this system has generated novel data on the association between MHC I haplotypes and DRB3 alleles as well as generated a substantial new dataset of MHC I allele sequences and haplotype content for three major *Bos indicus* cattle breeds of Brazil. Further expansion of the platform to incorporate analysis of the BoLA-DQ genes will enable high-throughput examination of the entire classical MHC repertoires of cattle populations. Such a tool will dramatically enhance the capacity of researchers to study the evolutionary biology of the bovine MHC system, associations between MHC and diseases resistance/susceptibility and also provide a comprehensive description of bovine MHC diversity. Combined with applications of other 'omics' technologies, such as immuno-peptidomics,⁴⁸ such a tool may also have major benefits in supporting the development of novel vaccines that aim to induce protective T-cell responses in cattle.

ACKNOWLEDGMENTS

This development of the methodology was partly funded by the BBSRC through the ISP award made to The Roslin Institute (BBS/E/D/20002174) and the European Union's Horizon 2020 (VetBioNet) research and innovation programme under grant agreement N731014. The work characterising the MHC diversity in Brazil was funded by The Bill & Melinda Gates Foundation and with UK aid from the UK Foreign, Commonwealth and Development Office (Grant Agreement OPP1127286) under the auspices of the Centre for Tropical Livestock Genetics and Health (CTLGH), established jointly by the University of Edinburgh, SRUC (Scotland's Rural College), and the

International Livestock Research Institute (the findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation nor the UK Government). The authors would like to thank the following institutions and producers, for their valuable assistance in providing samples. Nelore breed: Professor Flavio Meireles and Professor José Bento Sterman Ferraz, Experimental Farm of the School and Animal Science and Food Technology of the University of São Paulo, Pirassununga, São Paulo state; Dr. Maria Eugênia Mercadante, São Paulo State Institute of Animal Science, Sertãozinho, São Paulo state; Mr. Luciano Borges Ribeiro, Fazenda da Matinha, Uberaba, Minas Gerais state; Guzerát breed: Mr. Ricardo Junqueira Vieira, Sítio Nossa Senhora do Carmo, Heliadora, Minas Gerais state; the Tonetto family, Fazenda Perfeita União, Pirajuí, São Paulo State; Gir breed: Mr. José Coelho, Fazenda São José do Can Can, São José da Barra, Minas Gerais state; Dr. Anibal Vercesi, Fazenda Morro D'Água, Guapé, Minas Gerais state; Girolando breed: Dr. Mauricio Coelho, Fazenda Santa Luzia, Passos, Minas Gerais state. We also would like to thank the staff at the University of Edinburgh Farm: Wilson Lee, Peter Tennant, Adrian Ritchie and James Nixon in particular.

CONFLICT OF INTEREST

The authors declare no conflict of interests in the publication of this manuscript.

AUTHOR CONTRIBUTIONS

Deepali Vasoya was the lead researcher in the analysis of the data; Priscila Silva Oliveira, Laura Agundez Muriel, Thomas Tzelos and Christina Vrettou all contributed to the acquisition of the data; W. Ivan Morrison, Isabel Kinney Ferreira de Miranda Santos and Timothy Connelley contributed to the design of the study. All authors contributed to the drafting and revision of the manuscript and have approved the final version. All authors agree to be accountable for all aspects of the work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in European Nucleotide Archive at <https://www.ebi.ac.uk/ena/browser/view/PRJEB44287>, reference number PRJEB44287.

ORCID

Deepali Vasoya  <https://orcid.org/0000-0001-6764-5418>

REFERENCES

1. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res.* 2020;48(D1):D948–D955.

2. Lenormand C, Bausinger H, Gross F, et al. HLA-DQA2 and HLA-DQB2 genes are specifically expressed in human Langerhans cells and encode a new HLA class II molecule. *J Immunol.* 2012;188(8):3903-3911.
3. Hammond JA, Marsh SG, Robinson J, Davies CJ, Stear MJ, Ellis SA. Cattle MHC nomenclature: is it possible to assign sequences to discrete class I genes? *Immunogenetics.* 2012;64(6):475-480.
4. Holmes EC, Roberts AF, Staines KA, Ellis SA. Evolution of major histocompatibility complex class I genes in Cetartiodactyls. *Immunogenetics.* 2003;55(4):193-202.
5. Gelhaus A, Forster B, Horstmann RD. Evidence for an additional cattle DQB locus. *Immunogenetics.* 1999;49(10):879-885.
6. Gelhaus A, Forster B, Wipperfurth C, Horstmann RD. Evidence for an additional cattle DQA locus, BoLA-DQA5. *Immunogenetics.* 1999;49(4):321-327.
7. Andersson L, Rask L. Characterization of the MHC class II region in cattle. The number of DQ genes varies between haplotypes. *Immunogenetics.* 1988;27(2):110-120.
8. Burke MG, Stone RT, Muggli-Cockett NE. Nucleotide sequence and northern analysis of a bovine major histocompatibility class II DR beta-like cDNA. *Anim Genet.* 1991;22(4):343-352.
9. Zhou H, Hickford JG, Fang Q, Byun SO. Short communication: identification of allelic variation at the bovine DRA locus by polymerase chain reaction-single strand conformational polymorphism. *J Dairy Sci.* 2007;90(4):1943-1946.
10. Takeshima SN, Matsumoto Y, Miyasaka T, et al. A new method for typing bovine major histocompatibility complex class II DRB3 alleles by combining two established PCR sequence-based techniques. *Tissue Antigens.* 2011;78(3):208-213.
11. Takeshima SN, Matsumoto Y, Aida Y. Short communication: establishment of a new polymerase chain reaction-sequence-based typing method for genotyping cattle major histocompatibility complex class II DRB3. *J Dairy Sci.* 2009;92(6):2965-2970.
12. Takeshima S, Ikegami M, Morita M, Nakai Y, Aida Y. Identification of new cattle BoLA-DRB3 alleles by sequence-based typing. *Immunogenetics.* 2001;53(1):74-81.
13. Baxter R, Hastings N, Law A, Glass EJ. A rapid and robust sequence-based genotyping method for BoLA-DRB3 alleles in large numbers of heterozygous cattle. *Anim Genet.* 2008;39(5):561-563.
14. Miltiadou D, Law AS, Russell GC. Establishment of a sequence-based typing system for BoLA-DRB3 exon 2. *Tissue Antigens.* 2003;62(1):55-65.
15. Miyasaka T, Takeshima SN, Matsumoto Y, et al. The diversity of bovine MHC class II DRB3 and DQA1 alleles in different herds of Japanese black and Holstein cattle in Japan. *Gene.* 2011;472(1-2):42-49.
16. Giovambattista G, Takeshima SN, Ripoli MV, et al. Characterization of bovine MHC DRB3 diversity in Latin American creole cattle breeds. *Gene.* 2013;519(1):150-158.
17. Takeshima SN, Miyasaka T, Polat M, et al. The great diversity of major histocompatibility complex class II genes in Philippine native cattle. *Meta Gene.* 2014;2:176-190.
18. Takeshima SN, Miyasaka T, Matsumoto Y, et al. Assessment of biodiversity in Chilean cattle using the distribution of major histocompatibility complex class II BoLA-DRB3 allele. *Tissue Antigens.* 2015;85(1):35-44.
19. Takeshima SN, Giovambattista G, Okimoto N, et al. Characterization of bovine MHC class II DRB3 diversity in south American Holstein cattle populations. *Tissue Antigens.* 2015;86(6):419-430.
20. Takeshima SN, Corbi-Botto C, Giovambattista G, Aida Y. Genetic diversity of BoLA-DRB3 in south American zebu cattle populations. *BMC Genet.* 2018;19(1):33.
21. Baltian LR, Ripoli MV, Sanfilippo S, Takeshima SN, Aida Y, Giovambattista G. Association between BoLA-DRB3 and somatic cell count in Holstein cattle from Argentina. *Mol Biol Rep.* 2012;39(7):7215-7220.
22. Miyasaka T, Takeshima SN, Jimba M, et al. Identification of bovine leukocyte antigen class II haplotypes associated with variations in bovine leukemia virus proviral load in Japanese black cattle. *Tissue Antigens.* 2013;81(2):72-82.
23. Takeshima SN, Sasaki S, Meripet P, Sugimoto Y, Aida Y. Single nucleotide polymorphisms in the bovine MHC region of Japanese black cattle are associated with bovine leukemia virus proviral load. *Retrovirology.* 2017;14(1):24.
24. Yoshida T, Mukoyama H, Furuta H, et al. Association of the amino acid motifs of BoLA-DRB3 alleles with mastitis pathogens in Japanese Holstein cows. *Anim Sci J.* 2009;80(5):510-519.
25. Carapito R, Radosavljevic M, Bahram S. Next-generation sequencing of the HLA locus: methods and impacts on HLA typing, population genetics and disease association studies. *Hum Immunol.* 2016;77(11):1016-1023.
26. Hosomichi K, Shiina T, Tajima A, Inoue I. The impact of next-generation sequencing technologies on HLA research. *J Hum Genet.* 2015;60(11):665-673.
27. Sanchez-Mazas A, Nunes JM. Does NGS typing highlight our understanding of HLA population diversity?: some good reasons to say yes and a few to say be careful. *Hum Immunol.* 2019;80(1):62-66.
28. Petersdorf EW, O'Huigin C. The MHC in the era of next-generation sequencing: implications for bridging structure with function. *Hum Immunol.* 2019;80(1):67-78.
29. Vasoya D, Law A, Motta P, et al. Rapid identification of bovine MHCI haplotypes in genetically divergent cattle populations using next-generation sequencing. *Immunogenetics.* 2016;68(10):765-781.
30. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27(21):2957-2963.
31. Maccari G, Robinson J, Bontrop RE, et al. IPD-MHC: nomenclature requirements for the non-human major histocompatibility complex in the next-generation sequencing era. *Immunogenetics.* 2018;70(10):619-623.
32. Lee BY, Hur TY, Jung YH, Kim H. Identification of BoLA-DRB3.2 alleles in Korean native cattle (Hanwoo) and Holstein populations using a next generation sequencer. *Anim Genet.* 2012;43(4):438-441.
33. Derakhshani H, Plaizier JC, De Buck J, Barkema HW, Khafipour E. Association of bovine major histocompatibility complex (BoLA) gene polymorphism with colostrum and milk microbiota of dairy cows during the first week of lactation. *Microbiome.* 2018;6(1):203.
34. Oprzadek J, Sender G, Pawlik A, Lukaszewicz M. Locus BoLA-DRB3 is just an ordinary site of the polygene when explaining genetic variance of somatic cell count and milk yield. *J Dairy Res.* 2015;82(4):449-452.

35. Sun L, Song Y, Riaz H, Yang L. Effect of BoLA-DRB3 exon2 polymorphisms on lameness of Chinese Holstein cows. *Mol Biol Rep*. 2013;40(2):1081-1086.
36. Duangjinda M, Buayai D, Pattarajinda V, et al. Detection of bovine leukocyte antigen DRB3 alleles as candidate markers for clinical mastitis resistance in Holstein x zebu. *J Anim Sci*. 2009;87(2):469-476.
37. Duangjinda M, Jindatajak Y, Tipvong W, et al. Association of BoLA-DRB3 alleles with tick-borne disease tolerance in dairy cattle in a tropical environment. *Vet Parasitol*. 2013;196(3-4):314-320.
38. Takeshima SN, Ohno A, Aida Y. Bovine leukemia virus proviral load is more strongly associated with bovine major histocompatibility complex class II DRB3 polymorphism than with DQA1 polymorphism in Holstein cow in Japan. *Retrovirology*. 2019;16(1):14.
39. Bai L, Takeshima SN, Isogai E, Kohara J, Aida Y. Novel CD8 (+) cytotoxic T cell epitopes in bovine leukemia virus with cattle. *Vaccine*. 2015;33(51):7194-7202.
40. Forletti A, Lutzelschwab CM, Cepeda R, Esteban EN, Gutierrez SE. Early events following bovine leukaemia virus infection in calves with different alleles of the major histocompatibility complex DRB3 gene. *Vet Res*. 2020;51(1):4.
41. Slater N, Louzoun Y, Gragert L, Maier M, Chatterjee A, Albrecht M. Power laws for heavy-tailed distributions: modeling allele and haplotype diversity for the national marrow donor program. *PLoS Comput Biol*. 2015;11(4):e1004204.
42. Giovambattista G, Moe KK, Polat M, et al. Characterization of bovine MHC DRB3 diversity in global cattle breeds, with a focus on cattle in Myanmar. *BMC Genet*. 2020;21(1):95.
43. Giovambattista G, Takeshima SN, Moe KK, et al. BoLA-DRB3 genetic diversity in Highland creole cattle from Bolivia. *HLA*. 2020;96(6):688-696.
44. Eirin M, Carignano H, Shimizu E, et al. BoLA-DRB3 exon2 polymorphisms among tuberculous cattle: nucleotide and functional variability and their association with bovine tuberculosis pathology. *Res Vet Sci*. 2020;130:118-125.
45. Lazebnaya IV, Perchun AV, Lazebny OE. Intrabreed and interbreed variation of the BOLA-DRB3.2 gene in the Kostroma and Yaroslavl indigenous Russian cattle breeds. *Immunogenetics*. 2020;72(6-7):355-366.
46. Lo CW, Borjigin L, Saito S, et al. BoLA-DRB3 Polymorphism is associated with differential susceptibility to bovine leukemia virus-induced lymphoma and proviral load. *Viruses*. 2020;12(3):352. <http://doi.org/10.3390/v12030352>.
47. Bohorquez MD, Ordonez D, Suarez CF, et al. Major histocompatibility complex class II (DRB3) genetic diversity in Spanish Morucha and Colombian Normande cattle compared to Taurine and zebu populations. *Front Genet*. 2019;10:1293.
48. Nielsen M, Connelley T, Ternette N. Improved prediction of bovine leucocyte antigens (BoLA) presented ligands by use of -mass-spectrometry-determined ligand and in vitro binding data. *J Proteome Res*. 2018;17(1):559-567.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Vasoya D, Oliveira PS, Muriel LA, et al. High throughput analysis of MHC-I and MHC-DR diversity of Brazilian cattle populations. *HLA*. 2021;98(2):93-113. <https://doi.org/10.1111/tan.14339>