WILEY

HLA
Immune Response Genetics

ORIGINAL ARTICLE OPEN ACCESS

# Four Target Resequencing for the Bovine Major Histocompatibility Complex Region. Proof of Concept

Guillermo Giovambattista[1,2] | Arisa Kawamura[3] | Akane Ishida[3] | Yukine Murakawa[3] | Kazuyoshi Hosomichi[4] | Fumihiro Nagata[1] | Yoko Aida[1] | Shin-nosuke Takeshima[3]

[1]Laboratory of Global Infectious Diseases Control Science, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan | [2]Institute of Veterinary Genetics (IGEVET, UNLP-CONICET), Faculty of Veterinary Sciences, National University of La Plata, La Plata, Argentina | [3]Department of Food and Nutrition, Jumonji University, Saitama, Japan | [4]Laboratory of Computational Genomics, School of Life Science, Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan

**Correspondence:** Yoko Aida (yoko-aida@g.ecc.u-tokyo.ac.jp) | Shin-nosuke Takeshima (takesima@jumonji-u.ac.jp)

**Received:** 1 July 2024 | **Revised:** 6 January 2025 | **Accepted:** 26 January 2025

**Funding:** This work was supported by the Livestock Promotional Subsidy of the Japan Racing Association (JRA).

**Keywords:** bovine leukocyte antigen (BoLA) | genetic diversity | Holstein | Japanese black (wagyu) | next-generation sequencing | target resequencing

## ABSTRACT

The bovine leukocyte antigen (BoLA) comprises four regions that contain a high density of polymorphic genes and frequently show gene copy number variations (CNV). Therefore, genotyping *BoLA* using genome-wide resequencing is difficult. This study aimed to develop four probe sets for resequencing of the *BoLA* region using a hybridization capture target next–generation sequencing (NGS) method. This proof of concept showed and discussed the several applications of the used strategy. DNAs from nine Japanese Black cows and one Holstein cow were genotyped for *BoLA-DRB3* using PCR sequence-based typing (SBT). DNA libraries were constructed using the KAPA HyperPlus Kit, and *BoLA* DNA sequences were enriched using the SeqCap EZ kit and four custom-made probes. Based on preliminary results, the probe set BoLA2 was selected for further analysis. This analysis resulted in a mean coverage of 90.8% with an average depth of 108 reads. A total of 113,646 SNPs and 17,995 indels were detected, several of which have previously been described in the dbSNP database. This allowed the genotyping of class II genes, including *BoLA-DRB3*. A comparison between target resequencing and PCR-SBT assays did not show conflicts between the *BoLA-DRB3* genotyping results. CNV analysis based on read number inferred that the *BoLA-DQA1*, *BoLA-DQA2*, *BoLA-DQA5*, and *BoLA-DQB* genes would be present in the homozygous or heterozygous states or absent, allowing for the definition of four class II and three class I haplotypes. In addition, CNV of non-classical class I genes were also observed. In conclusion, results show that approach used in this study is a cost–effective strategy for sequencing large samples for many research purposes.

## 1 | Introduction

The major histocompatibility complex (MHC) is one of the most polymorphic regions in the vertebrate genome and plays a major role in host immune responses against pathogens [1]. The bovine MHC, named bovine leukocyte antigen (BoLA), is located on chromosome 23 and comprises: class I, which expands into a region of approximately 3 Mb and includes genes involved in primarily presenting antigen peptides to $CD8^+$ T cells; class III, which extends into a region of approximately 1.92 Mb and includes genes related to immune response such as complement, cytokines, and heat shock proteins; and class II, which was found within two regions, designated IIa and IIb [2]. Despite general structure being maintained during the evolution, from fish to mammals, a major rearrangement within the class II region has led to the division of the BoLA region into two subregions, class IIa and class IIb, that expand throughout 0.26 and 0.60 Mb, respectively. The class IIa subregion contains the functionally

expressed *DR* and *DQ* genes, whose gene products contain processed peptides derived from extracellular antigens to $CD4^+$ T-helper cells. In contrast, the class IIb molecules are involved in the processing and transport of antigen peptides within the cell [2, 3]. Other relevant characteristics of this complex are the extensive polymorphisms present in several of its genes, which have been hypothesized to be maintained by balancing or overdominance selection [4, 5], and the gene copy number variations between *BoLA* haplotypes [6–10]. In the Class II region, genes like *BoLA-DRB3*, *BoLA-DRA*, and *BoLA-DQB1* are consistently found across all haplotypes reported to date. However, other genes such as *BoLA-DQB*, *BoLA-DQA2*, and *BoLA-DQA5* exhibit variability, being present in some haplotypes and absent in others. Similarly, in the Class I region, there is considerable variation in the number and combination of classical and nonclassical Class I genes among different haplotypes [6].

In particular, *BoLA-DRB3* alleles were associated with resistance/susceptibility to infectious diseases, such as virus-induced lymphoma and proviral load in bovine leukaemia virus (BLV) infections, somatic cell counts in milk in mastitis, endo and ectoparasites, dermatophilosis, immune response traits, responses to vaccination, and production traits (e.g., milk yield). Furthermore, *BoLA-DQA1* has been associated with the proviral load in BLV infection and mastitis [11].

Owing to the main features described above, the study of MHC has generated growing interest during the last decades for evolutionary biologists and researchers interested in population genetics, animal health, and production. Genotyping of MHC genes is of great interest for many purposes such as understanding the response to infectious diseases as well as population and evolutionary studies. However, genotyping efforts have mainly focused on the analysis of a few class II (*BoLA-DRB3*, *BoLA-DQA*, and *BoLA-DQB*) and Class I (*BoLA1*, *BoLA2*, *BoLA3*, and *BoLA-NC1*) genes [12] because the complicated genomic structure of the MHC regions is difficult to analyse [2, 13–18]. During the last decades, three types of methods have been used to analyse at the molecular level the BoLA class II variability [18], based on polymerase chain reaction (PCR)-restriction fragment length polymorphism (RFLP) [19], PCR-sequence-based typing (PCR-SBT) [13–16], and amplicon next-generation sequencing (NGS) [17] based assays. The PCR-RFLP was popular due to its simplicity and cost–effectiveness. However, its accuracy was limited because it couldn't distinguish between closely related alleles at the nucleotide or amino acid level. The PCR-SBT assay improved accuracy by directly sequencing the PCR products, allowing for precise detection of alleles based on their nucleotide sequence, and the identification of amino acid motifs. In this method, the alleles were determined by comparison with reported allele databases. However, the PCR-SBT has the disadvantage that unexplained results and the presence of putative new variants have to be confirmed by cloning and Sanger sequencing. The benefits of using amplicon NGS for analysing BoLA class II genes. This approach is that allows the determination of the phase of multiple polymorphic sites within the second exon. This is crucial for increasing the accuracy of allele detection, without the need for validation by cloning and sequencing.

New NGS methods allow for the simultaneous analysis of millions of DNA or cDNA fragments from the whole genome to several selected genes, specific gene regions, or a bulk transcriptome in a single test [20]. In 2009, the first draft version of the bovine genome was published [21], and for each year thereafter whole-genome NGS resequencing was performed every year [22–27]. Compared to whole-genome sequencing (WGS), targeted NGS aims to achieve 'targeted enrichment' of genome subregions to significantly reduce the sequencing of genomic loci of interest (sequencing coverage), increase the number of reads by genomic position (sequencing depth), and reduce costs and effort [28]. Increasing the depth of sequencing is crucial for guaranteeing the accuracy and precision of variant (single nucleotide polymorphisms (SNPs) and indels) detection, avoiding false positives, and providing higher sensitivity to call rare variants. Highly consistent detection of informative genetic markers is critical for the detection of genetic traits and disorders. Applications of targeted NGS methods include exome sequencing, genotyping for parentage or trait analysis, inherited disease research, identification of rare variants, microbiome and environmental DNA sequencing, and methylation analyses. Currently, different enrichment assays are available, such as restriction site-associated DNA sequencing (RADSeq) assays that use a non-targeted genotyping by sequencing (GBS) approach that sequences the DNA adjacent to restriction enzyme sites, GBS methods that are based on multiple amplicon enrichment, and hybridization captures target regions that sequence region capture using a set of probes [29, 30].

Thus, progress in genotyping using SNP arrays and WGS has successfully accelerated the implementation of genomic selection to identify genetically superior cattle at an earlier age [31, 32]. However, the score for the current genomic selection may not reflect most *BoLA* alleles, although various economic traits in cattle are associated with these alleles [11]. Because the SNP array cannot set probes in complicated genomic regions, including the coding regions of *BoLA* genes, to determine *BoLA* alleles, the objective of this study was to develop and validate four sets of probes, designed based on the bovine genome versions of ARS-UCD1.2 and UMD3.1, for resequencing of the *BoLA* region using a hybridization capture target NGS method. In this proof of concept, the several applications of this type of assays were discussed, showing that it could be a useful tool for deep analysis of the entire genetic diversity and structure of *BoLA*, a crucial region involved in the immune response, at the population level, to identify copy gene number variations (CNV) in the different haplotypes, as well as for further detailed association analysis of regions found in genome wide association study (GWAS) studies and direct identification of the genes responsible for resistance/susceptibility to infectious and autoimmune diseases.

## 2 | Methods

### 2.1 | Animals

At the beginning of the study, nine Japanese Black cows and one Holstein cow with known *BoLA-DRB3* genotypes were selected from a large sample size ($N_{\text{Japanese Black}} = 507$, and $N_{\text{Holstein}} = 143$) that was genotyped for the *BoLA-DRB3* gene using the PCR-SBT in a previous work of our research

**TABLE 1** | Details of analysed animals and the genotypes obtained using PCR sequence-based typing (SBT) and hybridization capture target NGS method.

| | | | | *BoLA-DRB3* alleles | |
|---|---|---|---|---|---|
| **Animal ID** | **Breed**[a] | **Age**[b] | **Sex** | **Determined by PCR-SBT** | **Determined by target NGS** |
| # 1 | H | 2 | ♀ | **012:01/*014:01* | **012:01/*014:01* |
| # 2 | J | 5 | ♀ | **014:01/*016:01* | **014:01/*016:01* |
| # 3 | J | 5 | ♀ | **016:01/*016:01* | **016:01/*016:01* |
| # 4 | J | 5 | ♂ | **005:03/*012:01* | **005:03/*012:01* |
| # 5 | J | 1 | ♂ | **002:01/*015:01* | **002:01/*015:01* |
| # 6 | J | 14 | ♀ | **007:01/*010:01* | **007:01/*010:01* |
| # 7 | J | 14 | ♀ | **007:01/*009:02* | **007:01/*009:02* |
| # 8 | J | 13 | ♀ | **007:01/*014:01* | **007:01/*014:01* |
| # 9 | J | 10 | ♀ | **011:01/*012:01* | **011:01/*012:01* |
| # 10 | J | 16 | ♀ | **010:01/*011:01* | **010:01/*011:01* |

[a]Nine Japanese Black (J) cattle and one Holstein (H) cattle were used in this study.
[b]Age (year) at the time of the sample collection.

group [8] (Table 1). This study was approved by the Animal Ethical Committee, and the Animal Care and Use RIKEN Animal Experiments Committee (approval number H29-2-104). Genomic DNA was extracted from whole blood using Wizard Genomic DNA Purification Kits (Promega KK, Tokyo, Japan), according to the manufacturer's recommendations. For *BoLA-DRB3* genotyping, the *BoLA-DRB3* second exon was amplified by PCR using primers DRB3FRW (5′-CGCTCCTGTGAYCAGATCTATCC-3′) and DRB3REV (5′-CACCCCCGCGCTCACC-3′), and purified PCR fragments were sequenced using the ABI PRISM BigDye1.1 Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems; Foster City, CA, USA). Raw sequence data were analysed using the Assign 400ATF ver. 1.0.2.41 software (Conexio Genomics, Fremantle, Australia) to determine the *BoLA-DRB3* genotype. This strategy allowed the inclusion of cows that contained 10 of the most abundant *BoLA-DRB3* alleles in Japanese Black and Japanese Holstein [8] (Table 1).

## 2.2 | Assay Design and NGS Sequencing

Four sets of probes (Set1, Set4, BoLA1, and BoLA2) were developed to resequence the *BoLA* region (Table 2). The Set1 and Set4 were designed based on the UMD3.1 bovine genome ensembles, where the BoLA1 and BoLA2 sets were constructed using the ARS-UCD1.2 cattle genome version. Set1 and BoLA1 cover only the selected genes located within a 6.2 Mb and 4.2 Mb of the *BoLA* region, respectively, while the target of the Set4 and BoLA2 sets include the whole 6.2 Mb and 4.2 Mb sequences of the *BoLA* region, respectively. The difference in coverage (size of region analysed) between the developed probe sets is a consequence of the covered regions, the choice of genes included in the probe set (selected genes), and the number and placement of probes within each gene. While BoLA2 and Set 4 probe sets capture exonic, intronic and intergenic, BoLA1 and Set 1 probe sets capture only the gene region (exonic and intronic). The information about target regions included in each probe set is detailed in the Table S1.

DNA libraries were constructed from each DNA sample using the KAPA HyperPlus Kit (Roche Molecular Systems Inc., San Francisco, CA, USA), and *BoLA* DNA sequences were enriched using the SeqCap EZ Kit (Roche Molecular Systems Inc.) and a custom-made probe set. These technologies are based on the hybridization capture of target regions of genomes specific to the researchers' objectives using five biotinylated probes. After isolating the probes annealed to complementary fragments with streptavidin-coated beads, library fragments were sequenced [29, 30]. The resulting data focused on the targeted regions, providing a greater read depth for variant identification.

## 2.3 | DNA-Seq Data Analysis

Quality control of the DNA libraries was performed using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), which produced high–quality libraries suitable for sequencing. Libraries were sequenced using a 300 bp paired-end method on an Illumina Miseq system (Illumina Inc., San Diego, CA, USA) at the Department of Bioinformatics and Genomics Laboratory, Kanazawa University (Kanazawa, Japan). The quality of the obtained 'DNA fastq sequences was analysed using the FastQC software (FastQC A Quality Control tool for High Throughput Sequence Data). Fastq files were processed using the GATK Best Practices of the Germline short variant discovery (SNPs + Indels) workflow version 4.4.0 (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels). Briefly, Illumina adapters from the reads were trimmed using the Trimmomatic version 0.39 software [33], and thereafter the processed fastq

**TABLE 2** | Main features of the four sets of probes designed based on ARS-UCD1.2 and UMD3.1 bovine genome versions for re-sequencing of bovine leukocyte antigen (BoLA) region by a hybridization capture target NGS sequencing method.

| Probe set | Genome reference | Genome region | Target genes[a] | Target region[b] (bp) | Target length Mb[c] (average coverage) | Coverage of BoLA region[d] | Average depth[e] | Variant pass rate[f] |
|---|---|---|---|---|---|---|---|---|
| Set 1 | bosTau8 (UMD3.1) | chr23: 6,975,319—7,527,615 | *DSB-KIFC1* | 1,246,135 | 1.08 (87%) | 17% | 34 | 61% |
| — | | chr23: 25,351,486 – 31,044,230 | *DQA2-POM212L2* | — | — | — | — | — |
| Set 4 | bosTau8 (UMD3.1) | chr23: 6,975,319 – 7,527,615 | *DSB-KIFC1* | 6,245,040 | 4.43 (71%) | 71% | 131 | 75% |
| — | — | chr23: 25,351,486 – 31,044,230 | *DQA2-POM212L2* | — | — | — | — | — |
| BoLA1 | bosTau9 (ARS-UCD1.2) | chr23: 6,982,199 – 7,585,473 | *GCLC-KIFC1* | 492,593 | 0.57 (75%) | 14% | 218 | 70% |
| — | | chr23: 25,400,732 – 28,956,144 | *ELOVL5-MOG* | — | — | — | — | — |
| BoLA2 | bosTau9 (ARS-UCD1.2) | chr23: 6,982,199 – 7,585,473 | *GCLC-KIFC1* | 4,158,686 | 3.97 (90.8%) | 73% | 108 | 83% |
| — | — | chr23: 25,400,732 – 28,956,144 | *ELOVL5-MOG* | — | — | — | — | — |

[a]Target genes = fist and last genes include in the target region.
[b]Target region = size in base pair (bp) of the region encompass from the first to the last probe.
[c]Target length = the obtained average coverage in Megabases (Mb) of the target region by each probe set.
[d]Coverage of the BoLA region = percentage of the whole BoLA region coverage by each probe set.
[e]Average depth = average number of reads per nucleotide position site.
[f]Variant pass rate = percentage of variants that passed the filters.

files were aligned to the ARS-UCD1.2 reference genome ensemble using the BWA software [34]. SAM and BAM files were processed using the SAMtools v1.7 software [35]. The next step was to call variants (SNPs and indels) per sample using the HaplotypeCaller software to produce GVCFs files for SNPs and indels separately, which were consolidated into a GenomicsDB database using the GenomicsDBImport tool. Joint genotyping was performed using the GenotypeGVCFs tool, the recommended VQSR filtering was applied, and the variants were recalibrated using the ApplyVQSR tool to produce the final multisample callset in GVCFs for SNPs and indels separately for further analysis. Finally, the SNPeff software version 5.1 (https://pcingola.github.io/SnpEff) was used to annotate the SNPs and indels, detect novel polymorphisms, and evaluate their effects on the targeted regions. Polymorphisms were visualised using the IGV software [36]. The phases of the multiple polymorphic sites present simultaneously in the reads mapped on the second exon of *BoLA* genes were determined, and the sequences obtained were then compared against the alleles reported in the IPD-MHC database. This allow to identify which alleles are present in each studied animal.

### 2.4 | Data Analysis

To evaluate the genetic diversity along the entire *BoLA* region, Minor allele frequencies (MAF) and heterozygosity ($h_e$) for each SNP were calculated from VCF files using Plink software 1.9 [37]. Linkage disequilibrium (LD) was estimated within each *BoLA* class region using the command --r2 --ld-window 1000 --ld-window-kb 1000 --ld-window-r2 0 implemented in PLINK 1.9, and visualised using the ggplot R package. The average and standard deviation of the estimated $r^2$ values were calculated for each class region of *BoLA* using R functions. LD analysis linkage blocks among exon 2 of the *BoLA* class II genes were constructed and visualised using the four-gamete block (GAM) and solid spine block (SPI) methods implemented in Haploview version 4.2 [38].

### 2.5 | CNV Identification

The copy number variations (CNVs) in the BoLA genes were inferred by comparing the number of mapped reads in the genome loci of these genes within and among animals, considering that the number of reads is correlated with two, one or zero copies of a specific gene. The analysis includes class IIa genes (e.g., B*oLA-DQA*, *BoLA-DQB*) and class I genes (classic and non-classic). The number of reads mapped to these genes is compared within an animal and among different animals to estimate the average depth. Furthermore, comparing among animals helps confirm the probe's performance in capturing the locus. Absence of a gene copy is indicated by no mapped reads in the locus, while reads are present in upstream and downstream sequences. The loci with a single gene copy in one animal was inferred because have about half the reads compared to upstream and downstream sequences. In addition, known single-copy loci in the genome (e.g., *BoLA-DRB3* and *BoLA-DRA*) are also used as internal controls for each animal.

## 3 | Results and Discussion

### 3.1 | Optimal Probe for *BoLA* Region Re-Sequencing by a Hybridization Capture Target NGS Sequencing Method

DNA libraries were constructed for 10 cattle samples, as shown in Table 1, and then enriched using the four probe sets, Set1, Set4, BoLA1, and BoLA2, and finally sequenced. The processed results showed that the depth varied from 108 in BoLA2 to 218 in BoLA1, whereas the number of SNPs that passed after the data was filtered and recalibrated ranged from 61% in Set1 to 83% in BoLA2 (Table 2). The results obtained from each probe set were compared to evaluate the advantages and disadvantages of each probe set for genetic variant discovery studies (reads aligned to a genome and depth and uniformity of coverage). In addition, the BoLA2 probe set has the advantage that it was developed using the ARS-UCD2.1 genome assembly which in general has several advantages (more continuity, depth, accuracy, completeness, and annotated genes) over the UMD3.1 version [39]. Although BoLA1 (which was also developed using the ARS-UCD2.1) presented greater depth than BoLA2, the latter presented a higher coverage of the *BoLA* region (73% vs. 14%) and a higher percentage of SNPs that passed the filter (83% vs. 70%) (Table 2). In conclusion, BoLA2 was the optimal probe set for BoLA genotyping and was selected for further analyses, as discussed in detail in the following sections. In this proof of concept, the several applications of this type of assays were discussed, showing that it could be a useful tool for deep analysis of the entire genetic diversity and structure of *BoLA*.

### 3.2 | Performance of BoLA2 Targeted-NGS Assay

The quantitative and qualitative performance of the BoLA2 targeted-NGS assay for *BoLA* region resequencing was evaluated for 10 cattle samples, as shown in Table 1, using the FastQC software (output files were uploaded at https://osf.io/kbavc/files/osfstorage). These analyses showed a total of 25.9 million reads, with an average of 2.59 million reads per sample, ranging between 1.80–3.26 million (Table S2). A low percentage of overrepresented sequences, which corresponded to Illumina universal adapters, were detected, and 83%–89% of the sequences remained if they were deduplicated. Most of the sequences had a length between 250 and 350 bp, maintaining average Phred values per base higher than 28 until nucleotide position 150–210. In conclusion, these results demonstrated the high performance of the BoLA2 targeted-NGS assay.

Of the 3.97 Mb of the target region captured by the BoLA2 targeted-NGS assay, a mean of 90.8% was successfully mapped to the ARS-UCD1.2 genome, ranging from 89.9 to 91.9 among samples, whereas the average depth of 108 after deduplicated sequences, ranging from 68 to 143 among samples (Table 3). These values were higher than the threshold (10–20×) used by some laboratories for the detection of germline heterozygous variants, and similar to the recommended minimum read depth (100×) for detecting variants in exon sequencing to reach a reliable variant call [40, 41]. As expected, the sequencing depth obtained in the present work were higher than the mean values reached in whole bovine genome resequencing studies

**TABLE 3** | Total number of SNPs and INDELs detected in the ten cows in the whole bovine leukocyte antigen (BoLA) region within different types of genomic regions using the BoLA2 targeted NGS assay.

| Type of genomic region | Number of SNPs (%) | Number of INDELs (%) |
|---|---|---|
| Downstream | 30,445 (14.97) | 4,680 (15.07) |
| Exon | 7,840 (3.86) | 545 (1.75) |
| Intergenic | 73,500 (36.15) | 11,429 (36,79) |
| Intron | 58,455 (28.75) | 9,078 (29.22) |
| Splice site acceptor | 18 (0.009) | 3 (0.01) |
| Splice site donor | 29 (0.01) | 6 (0.02) |
| Splice site region | 601 (0.30) | 72 (0.23) |
| Upstream | 29,985 (14.75) | 4,830 (15.55) |
| 3′ UTR | 1,715 (0.84) | 325 (1.05) |
| 5′ UTR | 736 (0.36) | 91 (0.29) |

[25, 42–46], but was in the same range as other previously reported studies based on different target NGS assays [47–50].

### 3.3 | *BoLA* Genetic Diversity

The BoLA2 probes designed for the MHC region allowed us to capture and resequence a 4,158,686 bp target region that included the entire *BoLA* region: class I, class III, class IIa, and class IIb (Table 4). This analysis identified, after applying filters and recalibrating the data according to the GATK Best Practices, a total of 113,646 SNPs in the analysed sample, 45,896 (40.38%) of which have previously been described in the dbSNP database (Table S3). This resulted in an average rate of one variant every 35 bp, ranging from 4.59 for class IIa to 184 in the class III region. Most of these polymorphic sites were biallelic, with only 5145 multiallelic variants (4.34%) (Table S3). The number of variants detected may depend on different factors such as genetic background and the number of animals analysed. However, to contextualise the results obtained in this work, they are compared below with values previously published by other authors. As of 2018, more than 82 million SNPs have been identified through whole-genome sequencing of more than 2700 bulls from 121 breeds that cover a reasonable proportion of the world's cattle diversity [26]; several of these SNPs are breed-specific [51]. If the bovine genome is considered to have a size of 2.7 Gb (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002263795.1/), the number of SNPs reported until 2018 by the 1000 Bull Genomes Project represents one single polymorphism variant every 33 bp, similar to the rate observed in the present study, despite only 10 animals from two breeds being analysed, as shown in Table 1. This could be evidence of a higher level of polymorphism in the *BoLA* region, mainly in some functionally specific sequences, compared to the average of the whole bovine genome. Furthermore, more missense changes than silent ones were observed in the functional coding sites (57.40% vs. 41.34%), revealing other main characteristics of several *BoLA* genes that may be essential for the recognition of antigens and the initiation of the immune response. The number of heterozygous SNP in each animal varied from 19,976 (19.49%) to 36,308 (35.07%) (Table S4). These values were lower than those previously reported for Danish Holstein and Hanwoo cattle using whole-genome resequencing [24, 25].

Multisample indel calling resulted in the identification of 17,995 processed variants (8045 insertions and 9950 deletions) in comparison to the reference genome ARS-UCD1.2.105, 82,823 (50.99%) of which correspond to known variants because of entries in the dbSNP database. The number of heterozygous indel in each animal varied from 2821 (15.68%) to 4976 (27.65%) (Table S3).

The SNP/indel rate of 6.58 SNPs was within the range reported in previous studies (e.g., 3.36 in Hayes and Daetwyler [26]; 7.96 in Das et al. [25]; 9.03 in Zhang et al. [27]; 10.02 in Kawahara-Miki et al. [22]; 10.65 in Choi et al. [23]; 11.75 in Lee et al. [24]). Tables 2 and S2 summarise the results obtained for the entire *BoLA* and each class region using the SNPeff software.

### 3.4 | Linkage Disequilibrium (LD)

The LD within each class was estimated using the $r^2$ method implemented in Plink 1.9 (Figure 1). The average $r^2$ was $0.183 \pm 0.231$, $0.229 \pm 0.283$, $0.226 \pm 0.288$, and $0.204 \pm 0.286$ for classes I, IIa, IIb, and III, respectively. LD has been extensively studied at the whole-genome level in several breeds of cattle, including Holstein and Japanese Black, using SNP microarrays and NGS data [52, 53]. These studies showed that LD tends to decay when genetic distance increases, following an inverse exponential curve. The LD results obtained in the present work for classes I, IIb, and III exhibited an expected LD distribution with $r^2$ values; however, class II showed an atypical trend with low LD until a distance of 60 kb and two peaks of $r^2$ higher than 0.7 between 60 and 100 kb (Figure 1A–D). The LD among the second exons of *BoLA* class II genes is shown in Figure 2. These results indicate the presence of long linkage blocks in this *BoLA* class, but with low $r^2$ values at short distances because of the high variability within these blocks. Similar results were reported by Fukunaga et al. [9] who found a strong LD in the absence of *BoLA-DQB2 (DQB2*deletion*)* and *BoLA-DQA5* (*DQA5*deletion*) and *the presence of the BoLA-DRB3*27:03* allele. Analysis of a large sample size is necessary to confirm the observed LD pattern for the class IIa region. Elucidating this point is important for understanding the structure and evolution of *BoLA* and determining its effect on GWAS studies.

### 3.5 | Copy Gene Number Variation

As mentioned above, one of the main characteristics of the *BoLA* complex is the CNVs between *BoLA* haplotypes within and among breeds [2, 8, 9]. CNVs were detected based on the difference between the probe signals obtained using microarrays or the TaqMan Copy Number Assay; however, the development of NGS allowed them to detect CNVs by counting the mapped reads from two individuals and reporting regions that exhibited statistically significant read depth differences [54]. Compared to previous methods based on SNP chips and aCGH for detecting

**TABLE 4** | Genotypes detected for bovine leukocyte antigen (BoLA) Class II genes using the developed hybridization capture target NGS re-sequencing method. Percentage of identity was indicated between brackets with the closer allele.

| Animal ID | DQA2 | DQB | DQA5 | DQB1 | DQA1 | DRB3 | DRB2 | DRA |
|---|---|---|---|---|---|---|---|---|
| # 1 | DQA*022:08/ DQA*022:08 | del | del | DQB*012:01/ DQB*010:05 (95%) | DQA*014:01/ DQA*012:05 | DRB3*012:01/DRB3*014:01 | 1/1 | 1/1 |
| # 2 | DQA*022:08/ DQA*022:08 | del | del | DQB*010:01[a]/ DQB*013:01 (99%) | DQA*010:04/ DQA*010:04 | DRB3*014:01/DRB3*016:01 | 1/1 | 1/1 |
| # 3 | DQA*022:08/ DQA*022:08 | del | del | DQB*010:01[a]/ DQB*013:01 (99%) | DQA*010:04/ DQA*012:06 | DRB3*016:01/DRB3*016:01 | 1/1 | 1/1 |
| # 4 | DQA*034:01/ DQA*034:01 | del | del | DQB*010:01[a]/ DQB*014:02 | DQA*010:03/ DQA*001:01 | DRB3*005:03/DRB3*012:01 | 1/1 | 1/1 |
| # 5 | DQA*022:09/ DQA*022:09 | del | del | DQB*010:03 (99%)/ DQB*002:01 (99%) | DQA*002:03/ DQA*010:03 | DRB3*002:01/DRB3*015:01 | 1/2 | 1/1 |
| # 6 | DQA*022:03:03/ DQA*022:03:03 | DQB*020:02/DQB*020:02 | DQA*028:02/DQA*028:02 | DQB*001:05/ DQB*001:05 | DQA*001:01/ DQA*001:01 | DRB3*007:01/DRB3*010:01 | 1/1 | 1/1 |
| # 7 | DQA*022:03:03/ DQA*022:03:03 | DQB*020:02/DQB*020:02 | DQA*028:02/DQA*028:02 | DQB*018:07[b]/ DQB*018:09 (98%) | DQA*001:001/ DQA*002:07 | DRB3*007:01/DRB3*009:01 | 1/1 | 1/2 |
| # 8 | DQA*022:09/ DQA*022:03:03 | DQB*020:02/DQB*020:02 | DQA*028:02/DQA*028:02 | DQB*010:02:01/ DQB*018:05 (97%) | DQA*001:01/ DQA*010:04 | DRB3*007:01/DRB3*014:01 | 1/1 | 1/2 |
| # 9 | DQA*034:01/ DQA*022:08 | del | del | DQB*018:10 (97%)/ DQB*001:02 (98%) | DQA*012:05/ DQA*010:03 (99%) | DRB3*011:01/DRB3*012:01 | 1/1 | 1/3 |
| # 10 | DQA*034:01/ DQA*034:01 | del | del | DQB*010:01[b]/ DQB*014:02 | DQA*010:04/ DQA*010:03 | DRB3*010:01/DRB3*011:01 | 1/1 | 1/3 |

[a]Alternative allele could be *BoLA-DQB*010:02:01*.
[b]Alternative allele could be *BoLA-DQB*018:03*, *BoLA-DQB*018:03*, and *BoLA-DQB*018:08*.

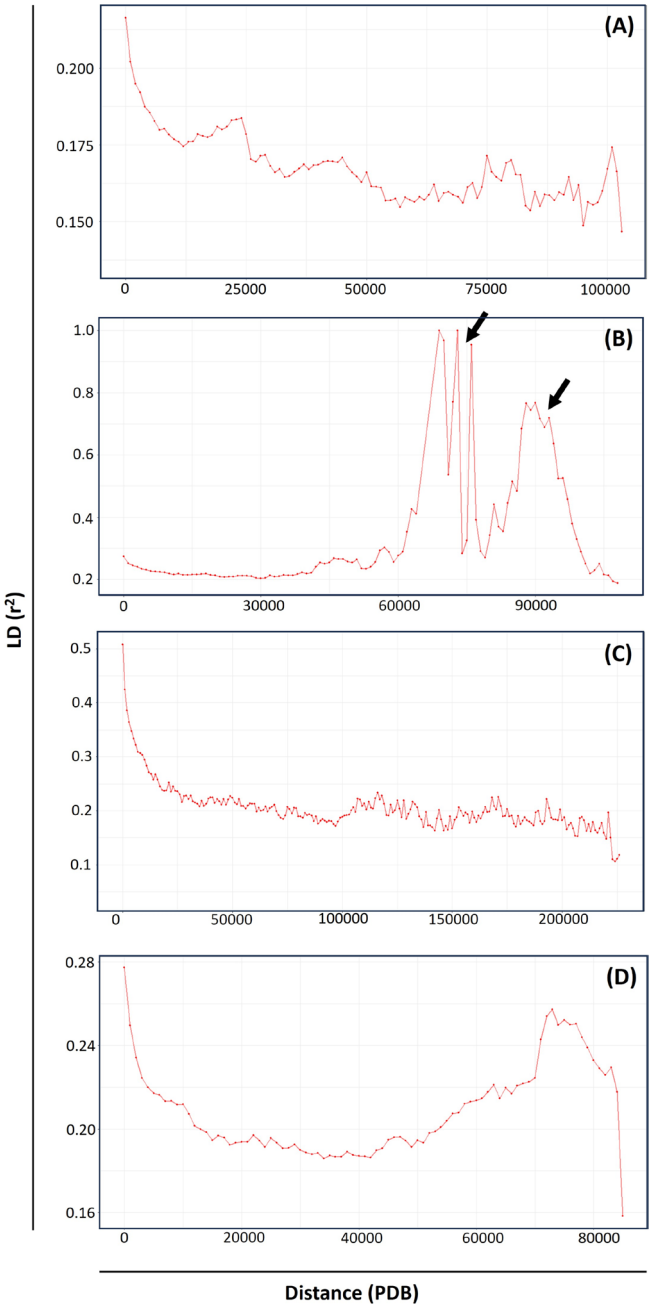


**FIGURE 1** | Trend of the linkage of disequilibrium (LD), measured through the $r^2$ index, between single nucleotide polymorphisms (SNPs) within Bovine leukocyte antigen (BoLA). (A) class I, (B) class IIa, (C) class IIb, and (D) class III. Arrows indicate the two peaks of $r^2$ higher than 0.7 between 60 and 100 kb observed in the class IIa LD analysis.

CNV, NGS has many advantages in terms of both CNV numbers and sizes because the sequencing approach overcomes the sensitivity limits of previous methods and can more precisely identify CNV boundaries [55].

Different class IIa genes have been annotated in the ARS-UCD1.2 genome version, such as *BoLA-DRB3* (Chr23: 25,723,691—25,734,819), *BoLA-DRB2* (Chr23: 25,795,927—25,828,806), LOC100848815 (Also known as *BoLA-DQA1*; Chr23: 25,691,259—25,695,296), *BoLA-DQA2* (Chr23: 25,582,895—25,589,220), *BoLA-DQA5* (Chr23: 25,631,993—25,643,886), *BoLA-DQB* (Chr23: 25,607,491—25,620,857), *BLA-DQB* (Also known as *BoLA-DQB1*; Chr23: 25,672,687—25,680,507), *BoLA-DRA* (Chr23: 25,838,391—25,843,092), and *BoLA -DRB (BoLA-DRB1* pseudogene; Chr23: 25,768,250—25,780,395). Furthermore, genes such as *BoLA-NC2*, *BoLA-NC4*, the *BoLA* pseudogene, *BoLA-NC3*, *BoLA3*, *BoLA-NC10*, *BoLA-NC1*, and *BoLA2* were mapped to the class I region.

The number of reads mapped for each class II and I gene was compared among the 10 analysed cattle to identify putative gene copy variations. This analysis revealed that *BoLA-DQA2*, *BoLA-DQB*, *BoLA-DQA5*, and *BoLA-DQA1* class II genes were present in two gene copies, in hemizygous state (only one copy), or completely absent, depending on the individual samples, which allowed us to define four class II haplotypes (Figure 3). The IGV image captures presented in Figure 4 illustrate this analysis, showing the regions belonging to class II, including BoLA-*DQB—BoLA-DQA5* (Figure 4A), and class I, comprising the pseudogene *BL3-7* (Figure 4B) in two samples, with zero and two copies of these genes. These results agree with the data recently published by Fukunaga et al. [9]. These authors used the TaqMan Copy Number Assay to evaluate the copy number within the *BoLA* region based on the analyses of only six class II genes and reported three class II haplotypes in Japanese Holstein cattle according to the presence or absence of the *BoLA-DQB*, *BoLA-DQA2*, and *BoLA-DQA5* genes. In addition, analyses of the class I region demonstrated that *BoLA-NC4*, *BoLA pseudogene*, and *BoLA-NC3* class I genes were present in homozygous or heterozygous states or absent, depending on the individual samples, indicating the presence of three class I haplotypes (Figure 3). Furthermore, combining the results of the copy numbers and alleles obtained by targeted NGS based on multiplex PCR, they identified nine *BoLA* haplotypes in class I and class II genes among the 10 analysed cattle (Figure 3).

*BoLA* alleles could not be distinguished between *BoLA-DQA* and *BoLA-DQB* using the IPD-MHC Database. Previous studies inferred the number of alleles and assigned the variants to putative copies by phylogenetic tree analysis [8, 56, 57]. In contrast, using NGS, alleles can be identified and directly assigned to a specific gene copy by analysing the mapped reads without inferences. This will contribute to determining the *BoLA* haplotypes and examining these variations in resistance to infectious diseases.

The class I region extension is considerably larger and more complex than that of class II. To date, the bovine section of the IPD-MHC contains six classical and 10 non-classical genes (updated on May 22, 2024). However, the number of copies and their combinations varied according to the class I haplotypes [6]. Moreover, these genes were not completely annotated in the latest version of the cattle genome. To solve this, the number of count reads mapped in different classical and non-classical class II genes were compared among samples to determine the presence or absence of each gene, identifying three putative haplotypes; nine haplotypes were determined when class II and class I information were combined (Figure 3). In addition, polymorphic phases were determined and the allelic sequences of these genes were assigned to a particular gene using the BLAST tool implemented on the IPD-MHC webpage. Thus, the genome positions of these genes are still unclear, and the developed assay

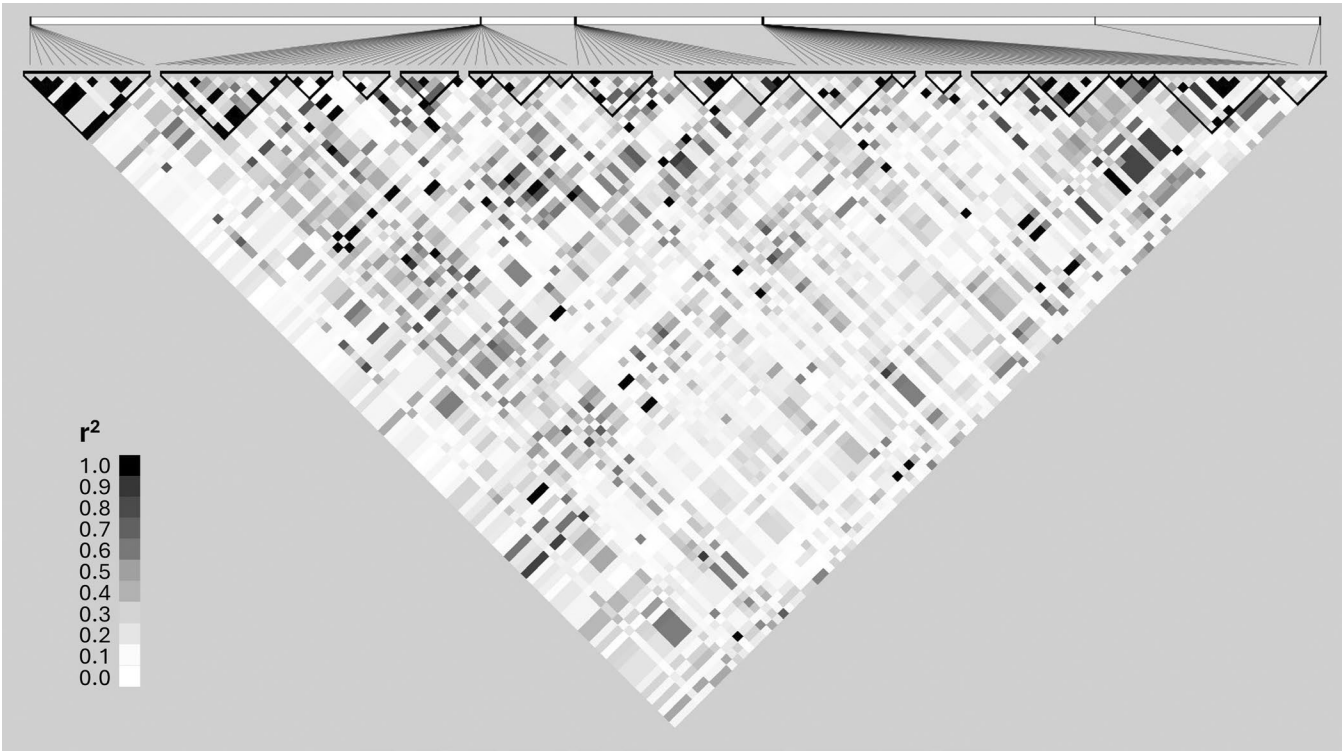


**FIGURE 2** | Linkage of disequilibrium (LD) analysis performed between the single nucleotide polymorphisms (SNPs) located within the second exon of the *BoLA* class II genes using the four-gamete rule implemented in the Haploview software. The scale of grey indicates the magnitude of $r^2$ values.

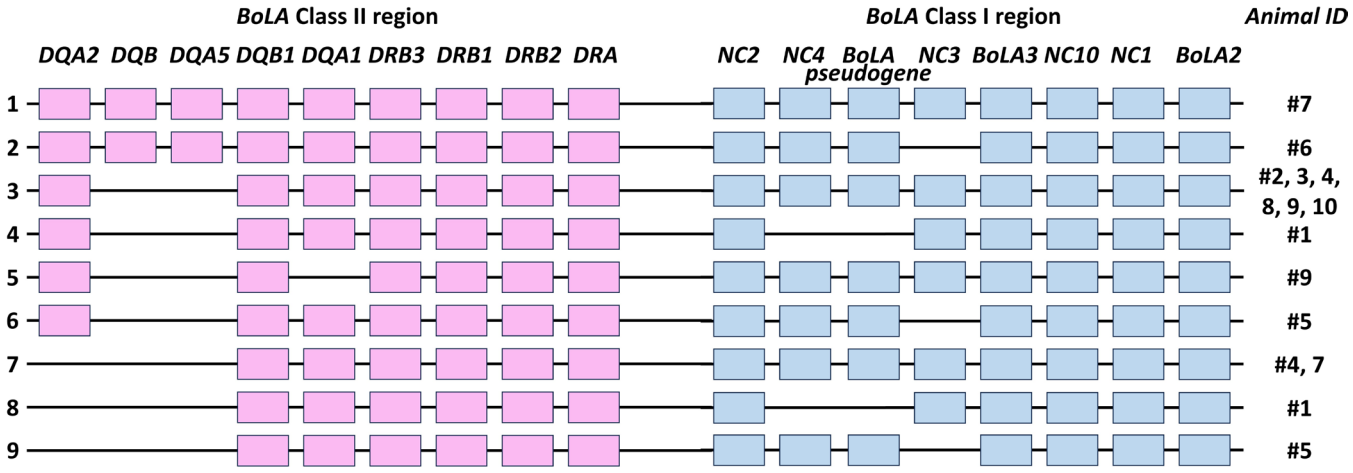


**FIGURE 3** | Schematic diagram of the inferred *BoLA class II* and *class I* haplotypes in the studied sample based on the read count mapped to the reference genome. On the right side, the animal IDs are indicated.

would contribute to increasing our knowledge about the structure of class I. Further analysis including a large sample size is necessary to elucidate this issue within different cattle breeds.

### 3.6 | Gene Genotyping

In the present study, the processed Fastq files obtained using the BoLA2 targeted-NGS assay were aligned to the ARS-UCD1.2 reference genome ensemble, and the SAM/BAM files were performed. By inspecting overlapping forward and reverse reads, the phases of the multiple polymorphic sites present simultaneously in the second exon of these *BoLA* genes were determine (Figure 5). The consensus sequences obtained from the second exon are then compared against the alleles reported in the IPD-MHC database. This step helps in identifying which specific alleles are present.

The 10 cattle included in the present study were first genotyped for the *BoLA-DRB3* gene using PCR-SBT which included 10 alleles in total (Table 1). A comparison between the two methods showed no conflicts between their results, inferring the same *BoLA-DRB3* genotypes for each animal (Table 1). To date more than 386 *BoLA-DRB3* alleles have been reported in the


**FIGURE 4** | IGV software capture showing the phase of different polymorphic sites.

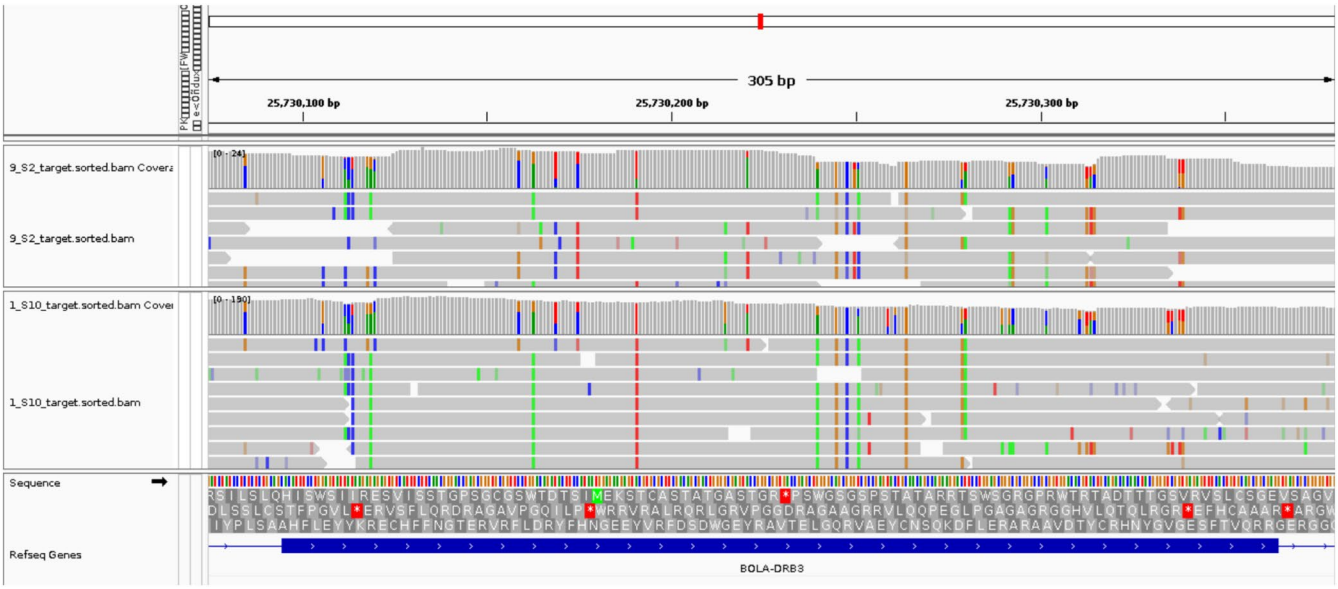


**FIGURE 5** | IGV software capture showing the haplotype phases of the *BoLA-DRB3* second exon.

IPD-MHC database (accessed on January 5th, 2025). The main challenge to the development of a genotyping assay for this and other *BoLA* gene is the ability to identify all reported alleles. In the PCR-SBT methods, this genotyping capacity depends on the primers design, since mutations in the primer sequences would make amplification of some alleles difficult. This has been extensively discussed in previous works [13–16]. By contrast, the assay development in the present work employs probes to enrich genomic DNA for target regions and then the DNA is sequenced using the adapters added during library construction. As described in Table S1, target sequences are much longer than primer sequences, which increases the probability of being able to capture all target regions included in the design, regardless the alleles present in *BoLA* genes.

As mentioned above, *BoLA* alleles were not distinguished among the *BoLA-DQA* and *BoLA-DQB* genes in the IPD-MHC Database; however, the developed targeted NGS assay allowed simultaneous determination of the genotypes of other *BoLA* genes (Table 4). In the present study, four *BoLA-DQA2*, one *BoLA-DQA5*, one *BoLA-DQB*, 14 *BoLA-DQB1* (including nine novel alleles that must be confirmed), eight *BoLA-DRB1* *pseudogene*, and two *BoLA-DRB2* and three *BoLA-DRA* genes were identified and assigned to a specific gene copy. Regarding class I genes, the allele diversity varied from one in *BoLA-NC3* and *BoLA-NC10* to several in *BoLA2* in the analysed sample (Figure 3). Sequencing many animals using the BoLA2 targeted-NGS assay would allow the estimation of linkage disequilibrium between different genes and the determination of haplotypes present in a particular breed.

The alleles of *BoLA* class I and class II genes are defined by the extensive polymorphisms present in their second exon, which encodes antigen binding sites (ABS). However, the targeted NGS method developed in the present study allows the re-sequencing of the entire region and analysis of polymorphisms present in other regions of interest for their biological functions. Recently, the annotation of functional regulatory elements in cattle was reported by Goszczynski et al. [58] in the context of the International Functional Annotation of Animal Genomes Consortium (FAANG; www.faang.org). This work confirmed the presence of the upstream regulatory region (URR) of the *BoLA-DRB3* gene that is located upstream of the initiation of transcription and is composed of highly conserved sequence

motifs that include from 5′ to 3′ the W, X, Y, CCAAT, and TATA boxes. Polymorphisms in URR can affect the expression levels of *this* genes. Analysis of the URR of the *BoLA-DRB3* gene revealed the presence of six SNPs (four of which are known polymorphisms) [59, 60], that result in 11 URR sequences. Simultaneous genotyping of regulatory regions and ABS could be relevant for studying the linkage disequilibrium and coevolution of these important functional regions.

### 3.7 | Balance Among Cost, Coverage, Depth and Accuracy of Polymorphism Detection

The output of NGS sequencer depend of the used instruments. High-end NGS instruments can produce up to 8 terabytes per flow cell. For genome assembly like ARS-UCD2.0 (2.8 gigabytes), over 0.2 terabytes are need to sequence to get the recommended depth (100×) for accurate detecting variants [40, 41]. Considering the coverage of the BoLA2 probe set (4.2 Mb), this output is enough to sequence more than 450 samples with a 100× depth, optimising cost (more than 1000 dollar vs. less than 150 dollar per sample) and reducing computing and storage needs.

In conclusion, the results of the present work showed that the BoLA2 target NGS assay based on probe capture allowed re-sequencing of the entire *BoLA* region with high quality, more than 89% of coverage, and enough depth (higher than 68) to detect confidence polymorphisms. This approach is a cost-effective strategy for sequencing many samples for different research purposes, such as estimating the rates of genetic diversity, including the detection of novel polymorphisms, alleles, and haplotypes; the level of LD; gene copy variation; runs of homozygosity; detection of selection signatures; and performing GWAS studies.

#### Author Contributions

S.T. and Y.A. conceived and designed the study. S.T. and K.H. developed the methodology. G.G., A.K., A.I., Y.M., F.N., and S.T. were acquired and analysed. G.G., S.T., and Y.A. interpreted data. G.G. and Y.A. drafted and revised the manuscript. All the authors have read and agreed to the published version of the manuscript.

#### Acknowledgements

This study was supported by the Livestock Promotional Subsidy of the Japan Racing Association (JRA). The authors thank members of the Department of Food and Nutrition, Jumonji University, and the Laboratory of Global Infectious Diseases Control Science, Graduate School of Agricultural and Life Sciences, University of Tokyo, for their technical assistance, help, and suggestions. We would like to thank Kaltech Co. Ltd. (https://kaltec.co.jp/en/, accessed on August 22, 2022) for helping with the organisation of the laboratory.

#### Conflicts of Interest

The authors declare no conflicts of interest.

#### Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### References

1. J. Klein, *Natural History of the Major Histocompatibility Complex*, 99th ed. (John Wiley Sons Inc, 1986).

2. S.-N. Takeshima and Y. Aida, "Structure, Function and Disease Susceptibility of the Bovine Major Histocompatibility Complex," *Animal Science Journal* 77 (2008): 138–150, https://doi.org/10.1111/j.1740-0929.2006.00332.x.

3. Y. Aida, "Characterization and Expression of Bovine MHC Class II Genes," *Bulletin de la Societe Franco-Japonaise des Sciences Veterinaires* 6 (1995): 17–24.

4. P. W. Hedrick, T. S. Whittam, and P. Parham, "Heterozygosity at Individual Amino Acid Sites: Extremely High Levels for HLA-A and -B Genes," *Proceedings of the National Academy of Sciences of the United States of America* 88 (1991): 5897–5901, https://doi.org/10.1073/pnas.88.13.5897.

5. A. L. Hughes and M. Nei, "Nucleotide Substitution at Major Histocompatibility Complex Class II Loci: Evidence for Overdominant Selection," *Proceedings of the National Academy of Sciences of the United States of America* 86 (1989): 958–962, https://doi.org/10.1073/pnas.86.3.958.

6. Y. Aida, S. Takeshima, C. L. Baldwin, and A. K. Kaushik, "Bovine immunogenetics," in *The Genetics of Cattle*, vol. 2015, ed. D. J. Garrick and A. Ruvinsky (CABI, 2015), 153–191.

7. E. J. Glass, R. A. Oliver, and G. C. Russell, "Duplicated DQ Haplotypes Increase the Complexity of Restriction Element Usage in Cattle," *Journal of Immunology* 165, no. 1 (2000): 134–138, https://doi.org/10.4049/jimmunol.165.1.134.

8. T. Miyasaka, S. N. Takeshima, H. Sentsui, and Y. Aida, "Identification and Diversity of Bovine Major Histocompatibility Complex Class II Haplotypes in Japanese Black and Holstein Cattle in Japan," *Journal of Dairy Science* 95, no. 1 (2012): 420–431, https://doi.org/10.3168/jds.2011-4621.

9. K. Fukunaga, Y. Yamashita, and T. Yagisawa, "Copy Number Variations in BOLA-DQA2, BOLA-DQB, and BOLA-DQA5 Show the Genomic Architecture and Haplotype Frequency of Major Histocompatibility Complex Class II Genes in Holstein Cows," *HLA* 96 (2020): 601–609, https://doi.org/10.1111/tan.14086.

10. E. Peripolli, N. B. Stafuzza, M. A. Machado, et al., "Assessment of Copy Number Variants in Three Brazilian Locally Adapted Cattle Breeds Using Whole-Genome Re-Sequencing Data," *Animal Genetics* 54 (2023): 254–270, https://doi.org/10.1111/age.13298.

11. A. Ruvinsky and A. Teale, "Genetics of Disease Resistance," in *The Genetics of Cattle*, vol. 2015, ed. D. J. Garrick and A. Ruvinsky (CABI, 2015), 192–215.

12. G. Maccari, J. Robinson, K. Ballingall, et al., "IPD-MHC 2.0: An Improved Inter-Species Database for the Study of the Major Histocompatibility Complex," *Nucleic Acids Research* 45 (2017): D860–D864.

13. S.-N. Takeshima, M. Ikegami, M. Morita, Y. Nakai, and Y. Aida, "Identification of New Cattle BoLA-DRB3 Alleles by Sequence-Based Typing," *Immunogenetics* 53 (2001): 74–81.

14. S.-N. Takeshima, A. Miki, M. Kado, and Y. Aida, "Establishment of a Sequence-Based Typing System for BoLA-DQA1 Exon 2," *Tissue Antigens* 69, no. 2 (2007): 189–199, https://doi.org/10.1111/j.1399-0039.2006.00751.x.

15. S.-N. Takeshima, Y. Matsumoto, and Y. Aida, "Short Communication: Establishment of a New Polymerase Chain Reaction–Sequence-Based Typing Method for Genotyping Cattle Major Histocompatibility Complex Class II DRB3," *Journal of Dairy Science* 92, no. 6 (2009): 2965–2970, https://doi.org/10.3168/jds.2008-1999.

16. S.-N. Takeshima, Y. Matsumoto, T. Miyasaka, 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* 78, no. 3 (2011): 208–213, https://doi.org/10.1111/j.1399-0039.2011.01708.x.

17. B. Y. Lee, T. Y. Hur, Y. H. Jung, and H. Kim, "Identification of BoLA-DRB3.2 Alleles in Korean Native Cattle (Hanwoo) and Holstein Populations Using a Next Generation Sequencer," *Animal Genetics* 43, no. 4 (2012): 438–441, https://doi.org/10.1111/j.1365-2052.2011.02264.x.

18. T. E. G. Andrade, M. S. Peña, J. Fiorotti, et al., "Graduate Student Literature Review: The DRB3 Gene of the Bovine Major Histocompatibility Complex—Discovery, Diversity, and Distribution of Alleles in Commercial Breeds of Cattle and Applications for Development of Vaccines," *Journal of Dairy Science* 107, no. 12 (2024): S0022-0302(24)00989–5, https://doi.org/10.3168/jds.2023-24628.

19. M. J. T. Van Eijk, J. A. Stewart-Haynes, and H. A. Lewin, "Extensive Polymorphism of the BOLA-DRB3 Gene Distinguished by PCR-RFLP," *Animal Genetics* 23 (1992): 483–496, https://doi.org/10.1111/j.1365-2052.1992.tb00168.x.

20. S. Yohe and B. Thyagarajan, "Review of Clinical Next-Generation Sequencing," *Archives of Pathology & Laboratory Medicine* 141, no. 11 (2017): 1544–1557, https://doi.org/10.5858/arpa.2016-0501-RA.

21. Bovine Genome Sequencing and Analysis Consortium, C. G. Elsik, R. L. Tellam, et al., "The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution," *Science* 324, no. 5926 (2009): 522–528, https://doi.org/10.1126/science.1169588.

22. R. Kawahara-Miki, K. Tsuda, Y. Shiwa, et al., "Whole-Genome Resequencing Shows Numerous Genes With Nonsynonymous SNPs in the Japanese Native Cattle Kuchinoshima-Ushi," *BMC Genomics* 12 (2011): 1–8, https://doi.org/10.1186/1471-2164-12-103.

23. J. W. Choi, X. Liao, S. Park, et al., "Massively Parallel Sequencing of Chikso (Korean Brindle Cattle) to Discover Genome-Wide SNPs and InDels," *Molecules and Cells* 36 (2013): 203–211, https://doi.org/10.1007/s10059-013-2347-0.

24. K. T. Lee, W. H. Chung, S. Y. Lee, et al., "Whole-Genome Resequencing of Hanwoo (Korean Cattle) and Insight Into Regions of Homozygosity," *BMC Genomics* 14 (2013): 519, https://doi.org/10.1186/1471-2164-14-519.

25. A. Das, F. Panitz, V. R. Gregersen, et al., "Deep Sequencing of Danish Holstein Dairy Cattle for Variant Detection and Insight Into Potential Loss-Of-Function Variants in Protein Coding Genes," *BMC Genomics* 16 (2015): 1043, https://doi.org/10.1186/s12864-015-2249-y.

26. B. J. Hayes and H. D. Daetwyler, "1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes," *Annual Review of Animal Biosciences* 7 (2019): 89–102, https://doi.org/10.1146/annurev-animal-020518-115024.

27. F. Zhang, K. Qu, N. Chen, et al., "Genome-Wide SNPs and InDels Characteristics of Three Chinese Cattle Breeds," *Animals* 9 (2019): 596, https://doi.org/10.3390/ani9090596.

28. F. Mertes, A. Elsharawy, S. Sauer, et al., "Targeted Enrichment of Genomic DNA Regions for Next-Generation Sequencing," *Briefings in Functional Genomics* 10, no. 6 (2011): 374–386, https://doi.org/10.1093/bfgp/elr033.

29. I. Kozarewa, J. Armisen, A. F. Gardner, B. E. Slatko, and C. L. Hendrickson, "Overview of Target Enrichment Strategies," *Current Protocols in Molecular Biology* 112 (2015): 7–21, https://doi.org/10.1002/0471142727.mb0721s112.

30. E. Samorodnitsky, B. M. Jewell, R. Hagopian, et al., "Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing," *Human Mutation* 36 (2015): 903–914, https://doi.org/10.1002/humu.22825.

31. B. O. Makanjuola, F. Miglior, E. A. Abdalla, C. Maltecca, F. S. Schenkel, and C. F. Baes, "Effect of Genomic Selection on Rate of Inbreeding and Coancestry and Effective Population Size of Holstein and Jersey Cattle Populations," *Journal of Dairy Science* 103, no. 6 (2020): 5183–5199, https://doi.org/10.3168/jds.2019-18013.

32. A. C. Doublet, P. Croiseau, S. Fritz, et al., "The Impact of Genomic Selection on Genetic Diversity and Genetic Gain in Three French Dairy Cattle Breeds," *Genetics, Selection, Evolution* 51, no. 1 (2019): 52.

33. A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data," *Bioinformatics* 30, no. 15 (2014): 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.

34. H. Li and R. Durbin, "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform," *Bioinformatics* 25, no. 14 (2009): 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.

35. H. Li, B. Handsaker, A. Wysoker, et al., "The Sequence Alignment/Map Format and SAMtools," *Bioinformatics* 25, no. 16 (2009): 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.

36. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, et al., "Integrative Genomics Viewer," *Nature Biotechnology* 29 (2011): 24–26, https://doi.org/10.1038/nbt.1754.

37. S. Purcell, B. Neale, K. Todd-Brown, et al., "Plink: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses," *American Journal of Human Genetics* 81 (2007): 559–575, https://doi.org/10.1086/519795.

38. J. Barrett, B. Fry, J. Maller, and M. Daly, "HAPLOVIEW: Analysis and Visualization of LD and Haplotype Maps," *Bioinformatics* 21 (2005): 263–265.

39. B. D. Rosen, D. M. Bickhart, R. D. Schnabel, et al., "De Novo Assembly of the Cattle Reference Genome With Single-Molecule Sequencing," *GigaScience* 9, no. 3 (2020): giaa021, https://doi.org/10.1093/gigascience/giaa021.

40. S. Yohe, A. Hauge, K. Bunjer, et al., "Clinical Validation of Targeted Next-Generation Sequencing for Inherited Disorders," *Archives of Pathology & Laboratory Medicine* 139, no. 2 (2015): 204–210, https://doi.org/10.5858/arpa.2013-0625-OA.

41. H. L. Rehm, S. J. Bale, P. Bayrak-Toydemir, et al., "ACMG Clinical Laboratory Standards for Next-Generation Sequencing," *Genetics in Medicine* 15, no. 9 (2013): 733–747, https://doi.org/10.1038/gim.2013.92.

42. P. Stothard, J. W. Choi, U. Basu, et al., "Whole Genome Resequencing of Black Angus and Holstein Cattle for SNP and CNV Discovery," *BMC Genomics* 12 (2011): 559, https://doi.org/10.1186/1471-2164-12-559.

43. B. Zhan, J. Fadista, B. Thomsen, J. Hedegaard, F. Panitz, and C. Bendixen, "Global Assessment of Genomic Variation in Cattle by Genome Resequencing and High-Throughput Genotyping," *BMC Genomics* 12 (2011): 557, https://doi.org/10.1186/1471-2164-12-557.

44. N. Iqbal, X. Liu, T. Yang, et al., "Genomic Variants Identified From Whole-Genome Resequencing of Indicine Cattl Breeds From Pakistan," *PLoS One* 14, no. 4 (2019): e0215065, https://doi.org/10.1371/journal.pone.0215065.

45. A. A. Zwane, R. D. Schnabel, J. Hoff, et al., "Genome-Wide SNP Discovery in Indigenous Cattle Breeds of South Africa," *Frontiers in Genetics* 10 (2019): 273, https://doi.org/10.3389/fgene.2019.00273.

46. N. Masharing, M. Sodhi, D. Chanda, et al., "ddRAD Sequencing Based Genotyping of Six Indigenous Dairy Cattle Breeds of India to Infer Existing Genetic Diversity and Population Structure," *Scientific Reports* 13, no. 1 (2023): 9379, https://doi.org/10.1038/s41598-023-32418-6.

47. M. C. McClure, D. Bickhart, D. Null, et al., "Bovine Exome Sequence Analysis and Targeted SNP Genotyping of Recessive Fertility Defects BH1, HH2, and HH3 Reveal a Putative Causative Mutation in SMC2 for HH3," *PLoS One* 9, no. 3 (2014): e92769, https://doi.org/10.1371/journal.pone.0092769.

48. L. Jiang, X. Liu, J. Yang, et al., "Targeted Resequencing of GWAS Loci Reveals Novel Genetic Variants for Milk Production Traits," *BMC Genomics* 15 (2014): 1105, https://doi.org/10.1186/1471-2164-15-1105.

49. C. Charlier, W. Li, C. Harland, et al., "NGS-Based Reverse Genetic Screen for Common Embryonic Lethal Mutations Compromising Fertility in Livestock," *Genome Research* 26, no. 10 (2016): 1333–1341, https://doi.org/10.1101/gr.207076.116.

50. A. Arizmendi, G. Rudd Garces, J. A. Crespi, et al., "Analysis of Doberman Pinscher and Toy Poodle Samples With Targeted Next-Generation

Sequencing," *Gene* 853 (2023): 147069, https://doi.org/10.1016/j.gene.2022.147069.

51. B. Czech, M. Frąszczak, M. Mielczarek, and J. Szyda, "Identification and Annotation of Breed-Specific Single Nucleotide Polymorphisms in *Bos Taurus* Genomes," *PLoS One* 13, no. 6 (2018): e0198419.

52. J. Bohmanova, M. Sargolzaei, and F. S. Schenkel, "Characteristics of Linkage Disequilibrium in North American Holsteins," *BMC Genomics* 11 (2010): 421, https://doi.org/10.1186/1471-2164-11-421.

53. R. Suezawa, H. Nikadori, and S. Sasaki, "Genetic Diversity and Genomic Inbreeding in Japanese Black Cows in the Islands of Okinawa Prefecture Evaluated Using Single-Nucleotide Polymorphism Array," *Animal Science Journal* 92 (2021): e13525, https://doi.org/10.1111/asj.13525.

54. C. Xie and M. T. Tammi, "CNV-Seq, a New Method to Detect Copy Number Variation Using High-Throughput Sequencing," *BMC Bioinformatics* 10 (2009): 80, https://doi.org/10.1186/1471-2105-10-80.

55. C. Alkan, B. P. Coe, and E. E. Eichler, "Genome Structural Variation Discovery and Genotyping," *Nature Reviews. Genetics* 12, no. 5 (2011): 363–376, https://doi.org/10.1038/nrg2958.

56. K. T. Ballingall, B. S. Marasa, A. Luyai, and D. J. McKeever, "Identification of Diverse BoLA DQA3 Genes Consistent With Non-allelic Sequences," *Animal Genetics* 29 (1998): 123–129.

57. A. Gelhaus, B. Förster, C. Wippern, and R. D. Horstmann, "Evidence for an Additional Cattle DQA Locus, BoLA-DQA5," *Immunogenetics* 49 (1999): 321–327, https://doi.org/10.1007/s002510050499.

58. D. E. Goszczynski, M. M. Halstead, A. D. Islas-Trejo, H. Zhou, and P. J. Ross, "Transcription Initiation Mapping in 31 Bovine Tissues Reveals Complex Promoter Activity, Pervasive Transcription, and Tissue-Specific Promoter Usage," *Genome Research* 31, no. 4 (2021): 732–744.

59. M. V. Ripoli, P. Peral-García, F. N. Dulout, and G. Giovambattista, "Polymorphism in the Bovine BoLA-DRB$_3$ Upstream Regulatory Regions Detected Through PCR-SSCP and DNA Sequencing," *Gene* 339 (2004): 71–78, https://doi.org/10.1016/j.gene.2004.06.002.

60. D. E. Goszczynski, M. V. Ripoli, S. N. Takeshima, L. Baltian, Y. Aida, and G. Giovambattista, "Haplotype Determination of the Upstream Regulatory Region and the Second Exon of the BoLA-DRB$_3$ Gene in Holstein Cattle," *Tissue Antigens* 83, no. 3 (2014): 180–183, https://doi.org/10.1111/tan.12293.

**Supporting Information**

Additional supporting information can be found online in the Supporting Information section.