

Access to ultra-long IgG CDRH3 bovine antibody sequences using short read sequencing technology

Samuel O. Oyola ^{*,3}, Sonal P. Henson ^{*,3}, Benjamin Nzau ¹, Elizabeth Kibwana ², Vishvanath Nene

International Livestock Research Institute (ILRI), P.O. Box 30709 -00100, Nairobi, Kenya

ARTICLE INFO

Keywords:

Ultra-long bovine antibodies
CDRH3
High throughput sequencing
Illumina technology
Cattle breeds

ABSTRACT

The advances in high-throughput DNA sequencing and recombinant antibody technologies has presented new methods for characterizing antibody repertoires and significantly increased our understanding on the functional role of antibodies in immunity and their use in diagnostics, vaccine antigen design and as biological therapeutics. A subset of *Bos taurus* antibodies possesses unique ultra-long third complementary-determining region of the heavy chain (CDRH3) and are of special interest because they are thought to have unique functional abilities of broadly neutralizing properties – a functional role that has not been fully explored in vaccine development. Next generation sequencing technologies that are widely used to profile immunoglobulin (Ig) repertoires are based on short-read methods such as the Illumina technology. Although this technology has worked well in sequencing Ig V-D-J regions of most jawed vertebrates, it has faced serious technical challenges with sequencing regions in bovine Ig bearing ultra-long CDRH3 sequences, which are longer than 120 bp. To overcome this limitation, we have developed a sequencing strategy based on nested PCR products that allows sequence assembly of full-length bovine Ig heavy-chain (IgH) V-D-J regions. We have used this strategy to sequence IgH V-D-J regions of two *Bos indicus* breeds, Ankole and Boran. We confirm the presence of ultra-long CDRH3 sequences in IgG transcripts in both African cattle breeds, and provide preliminary evidence for differences and preferences in germline V_H, D_H and J_H allele gene usage as well as differences in the length of the VH region in the two bovine breeds. Our method provides tools that should allow more robust analyses of ultra-long CDRH3 sequences aiding antibody and epitope discovery in different cattle breeds and their role in mediating immunity.

1. Introduction

An adaptive immune system that comprises B and T cells is a principle defense mechanism of mammals. It consists of humoral immunity mediated by antibodies produced by B-cells, and cell-mediated immunity by T cells (Rubelt et al., 2017). An effective antibody response is partly dependent on B-cells generating diverse B-cell receptors (BCRs), which collectively form an individual's B-cell receptor repertoire (BCRR). Traditionally, antibody responses have been analyzed using low-resolution approaches such as spectratyping and Sanger sequencing (Six et al., 2013). Advances in high throughput sequencing technologies has made it possible to study BCRRs at a higher resolution enabling interrogation into diversity of the receptor repertoire, isotype

composition, B-cell clonal expansion and mutation levels (Wardemann and Busse, 2017). These tools and analytical methods provide more insights into the state and composition of the adaptive antibody system, which can be exploited for developing effective biopharmaceuticals.

Structurally, a BCR is a membrane bound immunoglobulin (Ig) and consists of paired heavy (IgH) and light (IgL) chains derived from somatic rearrangement of variable (V), diversity (D), and joining (J) gene segments for IgH chains and VJ genes for IgL chains creating V_H and V_L variable domains joined to an Ig constant region. Somatic recombination of the germline V(D)J gene segments during B-cell development together with antigen-driven somatic hypermutations (SHM) are the major contributing factors to the huge diversity of the antibody repertoire (~1 × 10¹² in humans) that principally reside in the Ig V_H and V_L

* Corresponding authors.

E-mail addresses: s.oyola@cgiar.org (S.O. Oyola), s.henson@cgiar.org (S.P. Henson).

¹ Current address: The Pirbright Institute, Surrey GU24 0NF, UK.

² Current address: KEMRI-Wellcome Trust, P.O Box 43640 – 00100, Nairobi, Kenya.

³ These authors contributed equally.

<https://doi.org/10.1016/j.molimm.2021.08.017>

Received 18 January 2021; Received in revised form 23 August 2021; Accepted 24 August 2021

Available online 28 August 2021

0161-5890/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

domains (Georgiou et al., 2014). BCRs are the progenitors of soluble antibodies, capable of recognizing and binding to a vast array of antigens, which undergo isotype switch of the Ig constant region adding further functional properties to an antibody response.

The current reference bovine germline IgH genetic locus comprises 12 functional V_H, 23 D_H and 4 J_H gene segments (Liljavirta, 2014; Ma et al., 2016), in comparison to humans who have 36–49 V_H, 23 D_H and 6 J_H gene segments. A reduced number of Ig gene segments in bovines compared to humans are also seen in the lambda and kappa IgL loci and bovines have been observed to preferentially use lambda IgL. Although not fully proven, bovines most likely rely on activation-induced cytidine deaminase (AID) to generate sequence diversity and compensate for a much lower combinatorial diversity brought about by the limited number of Ig gene segments, (Haakenson et al., 2018; Liljavirta et al., 2013).

The main structural segments from paired V_H and V_L variable domains that bind to an epitope are referred to as complementarity determining regions (CDRs), of which there are three in each variable domain. Each CDR is flanked by a framework region (FWR) that is relatively conserved in sequence. The third CDR of V_H (CDRH3) plays a dominant role in antigen recognition and as such is the most diverse of all the CDRs. Hence, numerous BCRR studies focus on comparative sequence analysis of CDRH3 regions (Shi et al., 2014; Xu and Davis, 2000). Human and mouse CDRH3 sequences are on average 14 and 11 amino acid residues in length, respectively (Shi et al., 2014). Bovine CDRH3 sequences are on average 22 amino acid residues in length (Wang et al., 2013; Zhao et al., 2006) but a subset of antibodies has ultra-long CDRH3 regions that can reach over 70 amino acid residues in length (Saini et al., 1999). These are derived from recombination of the IGHV1-7 and IGH D8-2 gene segments (Deiss et al., 2017; Koti et al., 2010; Wang et al., 2013).

Sequence studies have identified a presence of an elevated number of non-germline encoded cysteine residues within ultra-long CDRH3 domains. These cysteine residues generate mini-domains as folded knobs through disulfide-bonding that protrude out of the antibody as revealed by crystal structure data (Wang et al., 2013), which may structurally facilitate access to otherwise hidden epitopes. In this class of antibodies, the V_L variable domain does not appear to contribute to epitope specificity and the IgL chain appears to play a structural role. The different shapes and structural orientations of these mini-domains has been proposed to contribute to the complexity and diversity of bovine antibody repertoires (Wang et al., 2013).

The use of short-read sequencing technologies such as Illumina has the level of base calling accuracy needed to deconvolute the huge diversity of immune sequence repertoire (Bolotin et al., 2012). Illumina MiSeq sequencing technology can cover up to 600 bp (2 × 300 bp) and has been used successfully for sequencing human and mouse BCRR V_H regions, which are mainly up to ~530 bp in length. However, bovine V_H regions with ultra-long CDRH3 domains exceed 600 bp in length and are difficult to sequence with Illumina technology following standard procedures. Hence, studies on bovine immune repertoires aiming to capture the full diversity of V_H sequences have relied on long-read sequencing platforms such as Pacific Biosciences (Deiss et al., 2017; Larsen and Smith, 2012).

Several research groups have developed methods that utilize short-read platforms to sequence BCR transcripts with lengths exceeding Illumina recommended raw-read length. Most of these protocols have utilized sophisticated techniques such as the use of Tn5 transposase and molecular identifiers to split transcripts into shorter fragments before sequencing using short-read sequencers (Cole et al., 2016). Others have used library preparation methods that includes circularization of fragments and biotin pulldown (Hiatt et al., 2010; Lundin et al., 2013). Although success has been reported, these protocols have proven to be expensive, less robust and often pose technical challenges in assembling bioinformatic analysis pipelines for full length transcript sequence reconstruction and annotation.

We have initiated bovine antibody discovery with the aim of establishing robust methodologies for sequence characterization of bovine V_H regions using Illumina short-read technology. Here, we describe a method that allows capture of IgG sequences bearing CDR3H domains including those with ultra-long sequences by assembly of sequence data from nested PCR products. Data from the two African indigenous cattle used in developing this methodology has shown differences in usage of V_H, D_H, J_H genes however, the trend will need confirmation with more individual animals per breed.

2. Material and methods

2.1. Cattle

Samples of blood for preparation of peripheral blood mononuclear cells (PBMCs) were taken from two adult male African *Bos indicus* breeds – 1 year old Boran and 11 year old Ankole - sourced from the ILRI farm. Animal experiments and routine maintenance was in accordance with procedures approved by ILRI's Institute Animal Care and Use Committee (IACUC experiment reference 2016.24). As a routine treatment, animals were vaccinated against foot and mouth disease (FMD). The Ankole sample was collected one day post FMD vaccination and used for PCR-1, PCR-2 and PCR-3. The Boran sample for PCR-1, PCR-2 and PCR-3 was collected eight weeks post FMD vaccination. Ficoll-based density centrifugation was used for separation of PBMCs from cattle blood mixed with Alsever's solution. Purified PBMC were aliquoted in portions of 1 × 10⁶ cells and frozen in phosphate buffered saline (PBS) at negative 80 °C until further processing.

2.2. Repertoire sequencing

Whole PBMC were defrosted and total RNA extracted using Trizol reagent or the RNeasy Mini Kit (Qiagen). Complementary DNA synthesis was performed on 500 ng RNA using the SuperScript III reverse transcriptase system (Invitrogen, Grand Island, NY, USA) following manufacturer instructions, with oligo dT primers (Applied Biosciences, Warrington, UK). Reverse transcription was performed at 42 °C for 60 min and 95 °C for 10 min. Bovine Ig heavy chain transcripts were amplified from complementary DNA using KAPA HiFi PCR amplification kit (KAPA Biosystems, South Africa). The PCR master mix contained 1 × KAPA HiFi master mix, 0.2 μM of each primer pair and 10 μl of the previous 1st-strand cDNA synthesized product, in a 50-μl total reaction volume. The primer pair used were (Table 1): VH family specific forward (F1) and total IgG specific reverse (R1) to generate PCR-1, F1 forward and heavy-chain constant (CH1) conserved sequence in framework 4 (FRW4) reverse (R2) to generate PCR-2 and CDR1 conserved forward primer specific to IGHV1-7 allele (F2) and R2 to generate PCR-3 products (see Table 1 and Fig. 1). The primers for PCR-2 and PCR-3 were modified to contain Illumina P7 and P5 sequences on the reverse and forward respectively. Amplification reactions were performed using the following thermocycling conditions: 1 min at 98 °C for the initial denaturation, followed by 30 cycles of 10 s at 98 °C and 1 min at 65 °C and a final extension for 5 min at 65 °C. The PCR products were purified and size-selected using Agencourt Ampure XP beads (Beckman Coulter Inc., Beverly, MA) following the manufacturer's instructions. Purified products were further amplified using Illumina indexing primers for no more than 12 cycles using similar PCR reagents and conditions as above

Table 1
PCR primers used to amplify bovine IgG.

| Primer | Sequence (5'-3') | Location |
|--------|-------------------------|--|
| F1 | ACCCACTGTGGACCTCCTC | Upstream of FR1 in V _H region |
| R1 | GGCACCCGAGTTCAGGTCA | IgG specific CH region |
| R2 | DRGADSYRGRGGGTAGRSYTCGG | 5' end of CH region |
| F2 | TTGAGCGACAAGGCTGTAGGCTG | CDR1-FR2 junction |

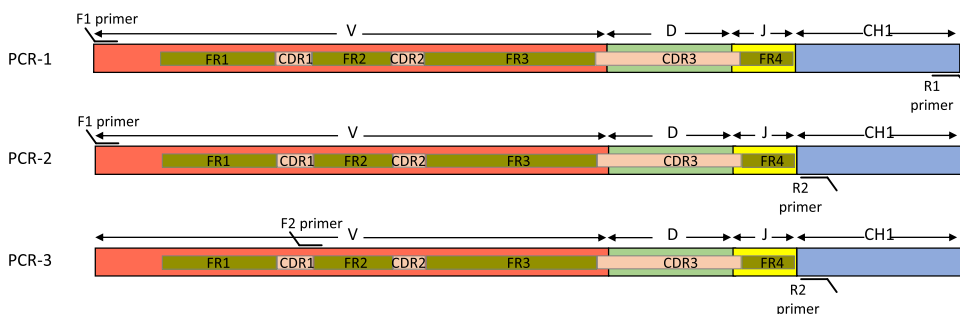


Fig. 1. Amplification strategy used to capture ultra-long antibody sequences. Using bovine 1st strand cDNA as template, PCR is performed using forward (F1) and reverse (R1) primers specific to bovine conserved V_H leader and IgG constant region (CH1) respectively. The product of PCR-1 was used as a template for two subsequent PCRs - i) PCR-2 using F1 primer and a conserved sequence at 5' end of CH1 as reverse primer (R2), ii) PCR-3 using a forward (F2) primer conserved in the CDRH1 region of the IGHV1-7 gene (F2) and R2.

to obtain Illumina-compatible libraries. After indexing PCR, the products were purified using XP beads before quantification using KAPA Illumina library quantification kit (KAPA Biosystems, South Africa). Quantified libraries were normalized and pooled for multiplex sequencing using one 2 × 300 bp MiSeq run (Illumina, San Diego, CA, USA).

2.3. Sequence analysis

Overall quality of de-multiplexed raw reads from Illumina MiSeq was assessed using Fastqc (www.bioinformatics.babraham.ac.uk/projects/fastqc/). Scripts within pRESTO v1.0.0 (Vander Heiden et al., 2014) were used for paired read assembly, trimming, quality filtering and de-duplicating of assembled reads as follows: Paired raw reads were assembled by alignment with a minimum overlap of 8 bp, alignment error (max_error) < 0.3 and alignment significance score (alpha) < 1e-05. Assembled sequences with mean phred quality score less than 20 were removed. Assuming the following lengths based on IMG T numbering scheme (Lefranc et al., 2003): FRW1 (75 nt), CDR1 (24 nt), FRW2 (51 nt), CDR2 (21 nt), FRW3 (114 nt), CDRH3 (12 nt), FRW4 (33 nt), C-region (94 nt, 15 nt, 15 nt for PCR-1, PCR-2, PCR-3, respectively), F1 primer (53 nt), F2 primer (23 nt), R1 primer (25 nt), R2 primer (19 nt), the shortest amplicon we can expect to assemble with an 8 bp overlap for PCR-1 is 511 bp, for PCR-2 is 438 bp and for PCR-3 is 309 bp. As such, transcript length cut-offs of 490 bp, 420 bp and 290 bp for PCR-1, PCR-2 and PCR-3 respectively, were applied to eliminate assemblies of short poor quality reads but at the same time allow assemblies which may have length variation in a structural region to pass. Forward and reverse primers were identified (max_error = 0.2) and removed. The remaining assembled sequences were deduplicated and submitted for annotation to IMG T HighV-Quest server (version 3.4.9) using the bovine IgH germline database. Annotated sequences were filtered further using Change-O (Gupta et al., 2015) to retain only those with germline gene assignments for all genes making up the variable region – V_H, D_H and J_H.

Annotations of sequences bearing the motif 'CXVXQ' at the beginning of their CDRH3 region were corrected to IGHV1-7 if they were assigned to another gene. In-house scripts were used to achieve this. Criteria for identifying full-length sequences were based on (Lefranc et al., 2003) and were as follows – FRW1 length 75–78 bp, FWR2 length 48–51 bp, FWR3 length 108–117 bp, FWR4 length 30–36 bp, CDRH1 length 15–36 bp, CDRH2 length 21–30 bp.

Gene usage and mutation frequency analyses were carried out on full length productive sequences using *countGenes* and *observedMutations* methods of Alakazam and SHazam R packages (Gupta et al., 2015). Entropy was calculated on full-length structural regions for short, medium and ultra-long sequences using the *bio3d* package in R. Sequences of the following lengths were considered for entropy calculation - 25 AA for FWR1, 8 AA for CDRH1, 17 AA for FWR2, 7 AA for CDRH2 and 38 AA for FWR3. Average number of sequences used for calculations were as follows: 105,230 and 177,914 short, 471,134 and 641,060 medium and 770 and 1656 ultra-long Ankole and Boran, respectively.

3. Results

3.1. Sequencing of bovine IgGH V_H domains using Illumina short-read technology

Sequencing the V_H variable domain of bovine BCRs using short-read Illumina technology (2 × 300 bp paired-end) is challenging due to the presence of ultra-long CDRH3 segments in some IgH chains which can reach 200bp in length. To ensure that we captured the entire range of lengths in the bovine IgH repertoire using the Illumina platform, we designed a nested PCR strategy and sequencing approach (Fig. 1). A pan IgG isotype-specific PCR (PCR-1) was performed on first-strand cDNA template using a 5' end primer specific to a conserved V_H leader sequence (F1), and a 3' end primer (R1) specific to the CH1 domain of all bovine IgG transcripts (Walther et al., 2013) (Table 1). Using the PCR-1 product as a template, we generated two nested PCR products, PCR-2 and PCR-3. For PCR-2 we used primer F1 and a new primer R2 that we designed adjacent to a conserved nucleotide motif GCCTCCACC (Fig. S1) at the beginning of the IgG constant region defined in Walther et al. (2016). For PCR-3 we used primer F2 that targets the conserved CDR1 sequence motif specific to bovine IGHV1-7 gene described in Deiss et al. (2017), and primer R2. This strategy was used to amplify IgGH V_H regions of two *Bos indicus* African breeds, Ankole and Boran, from the same biological sample. All three PCR products from each breed were sequenced. By combining sequence reads from PCR1, PCR-2 and PCR-3, we were able to capture bovine BCR sequences covering CDRH3 ranging from 4 to 75 amino acids in length.

3.2. Only ~60 % of paired end reads of PCR-1 assembled into complete transcripts

From PCR-1 product, we obtained ~2.2 M paired raw reads for Ankole and ~2.7 M paired raw reads for Boran in total (Table 2). Forward and reverse reads were assembled if they overlapped by at least 8 bp with alignment error less than 0.3 and alignment significance score less than 1e-05. We obtained 872,471 (40 %) and 1,434,209 (54 %) assembled paired raw reads for Ankole and Boran samples, respectively. 69 % and 58 % of Ankole and Boran assembled sequences, respectively, passed the QC to yield complete IGHV transcripts. After deduplication, 575,110 and 796,995 high quality unique IGHV transcripts were obtained for Ankole and Boran samples, respectively. These were annotated against the *Bos taurus* germline gene database using the IMG T/HighV-QUEST online server (Alamyar et al., 2012b) and filtered to retain only those that had all three germline V_H, D and J_H genes assigned. 82 % and 83 % of the Ankole and Boran transcripts, respectively, encoded complete V_H open reading frames, and were thus considered productive Ig sequences (Alamyar et al., 2012a). Further analysis of the productive IGHV repertoire obtained from the PCR-1 product of both samples showed that only 4% of Ankole and 2.6 % of Boran sequences had CDRH3 longer than 22 amino acids – the length at which the most pronounced drop in sequences was observed - indicating that a large number of bovine IGHV transcripts categorized as long and

Table 2

Sequencing and read processing outputs of PCR-1, PCR-2 and PCR-3 for Ankole and Boran datasets. Raw read pairs were assembled. Assembled reads with > Q20 average read quality and sequence lengths exceeding 490 bp, 420 bp and 290 bp for PCR-1, PCR-2, PCR-3, respectively were retained. Assembled reads that passed quality filters were de-duplicated to retain unique sequences. Unique assemblies were annotated by IMGT/HighV-Quest and productive sequences were analysed further.

| Sample | Library | Raw read pairs | Assembled read pairs | Assembled read pairs that passed filters | Unique assemblies | Productive V _H region |
|--------|---------|----------------|----------------------|--|-------------------|----------------------------------|
| Ankole | PCR-1 | 2,157,754 | 872,471 (40 %) | 600,550 (69 %) | 575,110 | 471,637 (82 %) |
| Ankole | PCR-2 | 638,094 | 519,352 (81 %) | 166,584 (32 %) | 155,441 | 105,983 (68 %) |
| Ankole | PCR-3 | 902,896 | 827,823 (92 %) | 543,669 (66 %) | 383,253 | 331,410 (86 %) |
| Boran | PCR-1 | 2,672,666 | 1,434,209 (54 %) | 833,427 (58 %) | 796,995 | 665,451 (83 %) |
| Boran | PCR-2 | 792,500 | 656,202 (83 %) | 217,911 (33 %) | 206,137 | 156,934 (76 %) |
| Boran | PCR-3 | 620,713 | 557,251 (90 %) | 364,816 (65 %) | 272,087 | 233,660 (86 %) |

ultra-long CDRH3 were missing and could not be assembled from sequencing PCR-1.

3.3. The nested PCR sequencing strategy improves assembly and coverage of the bovine CDRH3 repertoire

Sequences from the nested PCR products, PCR-2 and PCR-3, were analyzed following the same procedure as described above. As shown in Table 2, paired read assembly improved by ~ 2-fold compared to the PCR-1 products. The proportion of raw reads successfully assembled increased from 40 % and 54 % for PCR-1 to 92 % and 90 % for Ankole and Boran, respectively. After filtering to remove short and low-quality transcripts, 69 % and 58 % of PCR-1, 32 % and 33 % of PCR-2 and 66 % and 65 % of PCR-3 transcripts were retained for Ankole and Boran, respectively.

PCR-2 enabled assembly of longer complete V_H regions (Fig. S2). For

PCR-1 the median V_H length captured was 360 bp for Ankole and 349 bp for Boran, whereas for PCR-2 the product was significantly longer at 384 bp for Ankole (ANOVA $P = 2 \times 10^{-3}$) and 387 bp for Boran (ANOVA $P = 1 \times 10^{-7}$). This affected the length distribution of V_H regions by shifting the graph slightly to the right for PCR-2 (Fig. S2b). The most abundant V_H length captured in PCR-1 was 376 bp for Ankole and 379 bp for Boran (excluding the peak at 352 bp – see discussion). In comparison, for PCR-2 the most abundant V_H length was 388 bp and 385 bp for Ankole and Boran, respectively. Primer F2 used for PCR-3 begins at 138 bp downstream of primer F1 used for PCR-1 and PCR-2. Consequently, sequencing PCR-3 permitted capture of ultra-long CDRH3 sequences.

3.4. Assembly of sequences from all PCR products reveal bovine Ig BCRRs

A total of 909,030 and 1,056,045 unique Ankole and Boran productive Ig transcript sequences were identified, respectively. Of those,

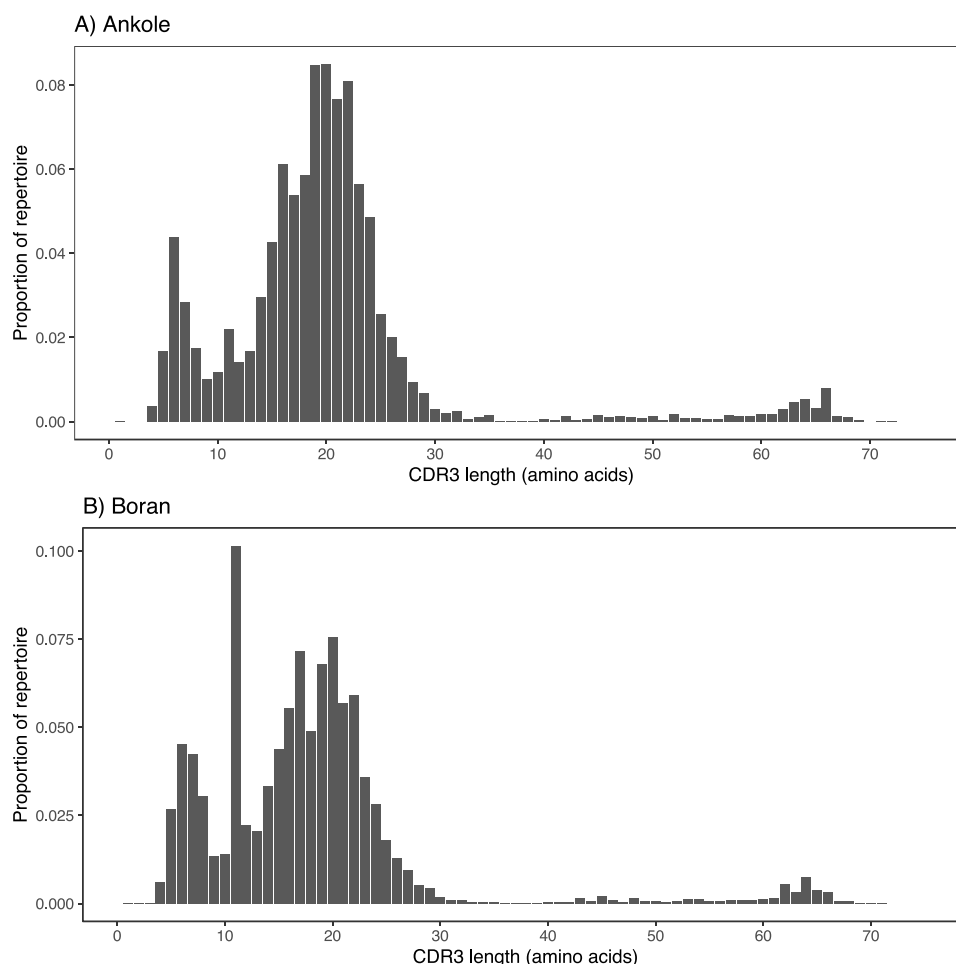


Fig. 2. Distribution of CDRH3 lengths in 909,030 and 1,056,045 unique productive Ankole and Boran immunoglobulin sequences, respectively.

567,051 (62.3 %) Ankole and 807,600 (76.5 %) Boran transcripts had complete FR1–FR4 regions. As such, these transcripts were deemed to be of full length. Fig. 2 shows the distribution of CDRH3 lengths derived from the total unique sequenced repertoire of Ankole and Boran. A bimodal distribution with a long tail was seen for both the repertoires. An abundance of transcripts with CDRH3 length of 11 AA was observed in the Boran repertoire. This peak, which comprised more than 10 % of the repertoire, was a result of an expansion of three clusters of 95 % identical sequences of CDRH3 length 11 AA. This expansion may represent an immunological response unknown to us. Transcripts with CDRH3 length 20 AA were abundant in both Ankole and Boran (excluding the expanded clones), representing 8.5 % and 7.6 % of the repertoires, respectively.

3.5. Germline gene usage differs in the two bovine breeds

3.5.1. Annotation of transcripts derived from IGHV1-7 gene

The IGHV1-7 gene segment has been shown to have an extended 3' end sequence arising from an 8 bp duplication which encodes the motif 'CTTVHQ' that is unique to this gene segment and lies at the start of the CDRH3 region (Deiss et al., 2017). In our analysis we noticed that several CDRH3 sequences encoded a similar motif but were assigned a V_H gene other than IGHV1-7 by IMGT/HighV-Quest. Amino acid diversity represented as sequence logos and a stacked bar chart of the first six amino acids of short, medium and ultra-long CDRH3 sequences

revealed that the first cysteine (C) of the CDRH3 region, the valine (V) at position 4 and glutamine (Q) at position 6 are conserved and occur at a frequency of > 90 % in the IGHV1-7 gene (Fig. S3). A conservative motif pattern of 'CXXVXQ' (where X represents any amino acid) was derived based on these observations and used to re-assign IGHV1-7 to CDRH3 sequences with this motif. The V_H gene of 784 Ankole and 1600 Boran sequences were re-assigned to IGHV1-7 gene.

3.5.2. Gene usage is more evenly distributed in the two African indigenous breeds

Bovines have only a single functional IGHV family with 12 functional V_H genes known currently (Liljavirta, 2014; Ma et al., 2016). To determine the germline gene usage in the two bovine breeds, we analyzed the full-length productive sequences of Ankole and Boran using the IMGT/HighV-QUEST online server. In Ankole and Boran repertoires, a more even distribution of V_H gene usage was observed (Table S1). Notable differences in gene usage between Ankole and Boran were found for IGHV1-20 and IGHV1-30 alleles, both of which were more abundant in Ankole, whereas IGHV1-7 was more abundant in Boran. Data from more animals per breed will be needed to establish the observed trends.

3.5.3. IGHJ2-4 allele is preferentially expressed among the J_H genes with no clear preference in D_H gene allele expression

In bovine, four of the 12 joining (J_H) genes are functional at the IGHJ

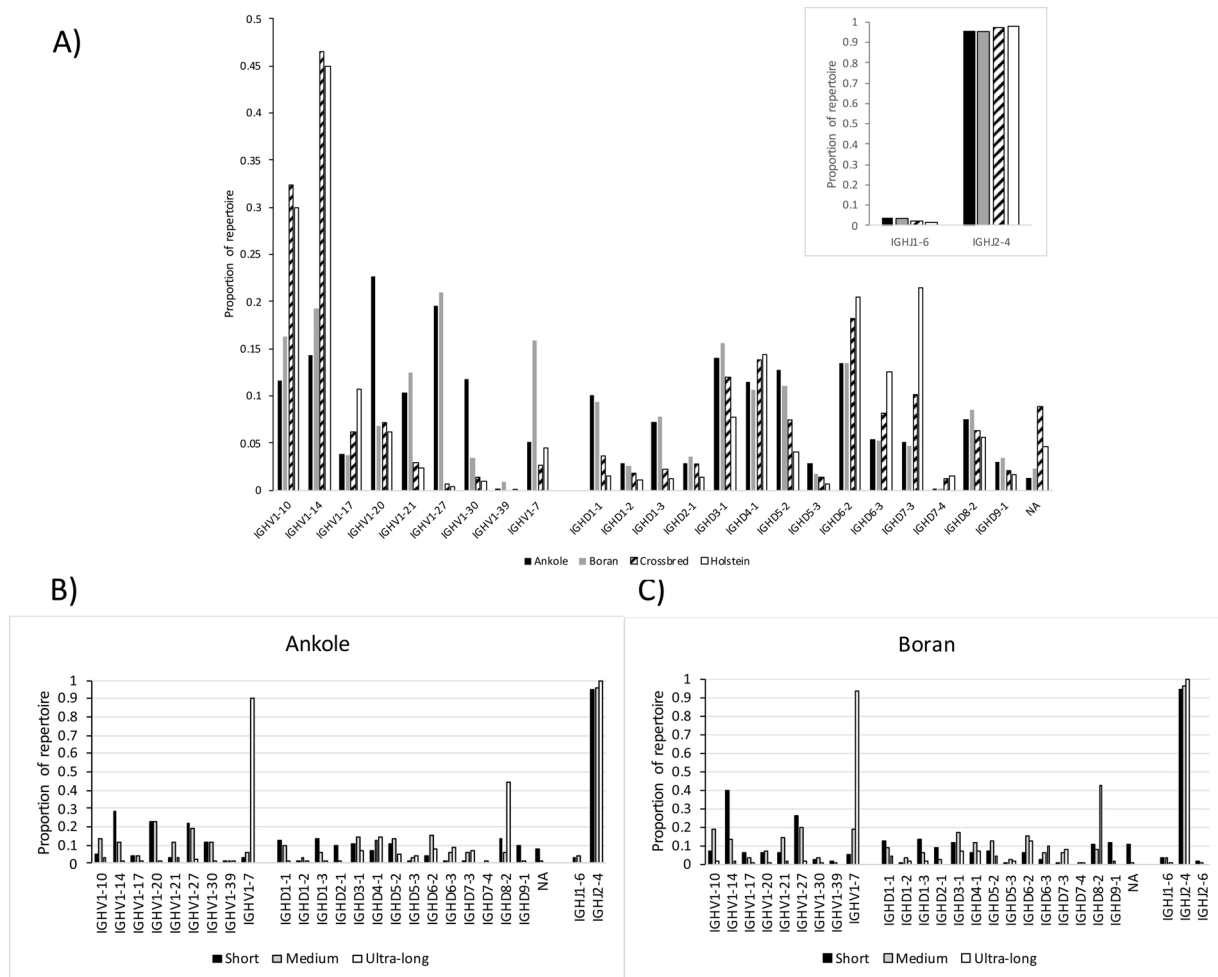


Fig. 3. Germline allele usage shown as a percentage of total unique IgH sequences from PCR-1 and PCR-2. Only genes with frequency of >1 % are shown. V_H genes IGHV1-25, IGHV1-32 and IGHV1-37 were found at lower than 1 % abundance in the datasets and so were not included in the figure. A) compares proportions between Ankole and Boran. Inset graph shows abundance of IGHJ genes found at >1 % frequency. B) and C) compare proportions in antibodies with short (< = 10 AA), medium (11–39 AA) and ultra-long (> = 40 AA) CDRH3 sequence in Ankole and Boran, respectively.

locus (Ma et al., 2016). Reports on the bovine J_H gene usage show preferential expression of the IGHJ2-4 gene (Liljavirta et al., 2014; Ma et al., 2016; Walther et al., 2016). We observed the same pattern in our data, whereby 95 % or greater of the repertoires were assigned to the IGHJ2-4 gene (Fig. 3, inset). All the remaining J_H genes had less than 1 % assignment except IGHJ1-6 which had 2~4 % of sequences assigned to it.

DH gene profiles were similar between Ankole and Boran repertoires. Of the D_H gene sequences, IGHD8-2 has been shown to have the longest nucleotide sequence (bp) and is one of the alleles shown to be responsible for encoding the ultra-long species of the bovine IgH antibodies (Deiss et al., 2017). Similar proportions of IGHD8-2 gene assignment were observed in all four repertoire datasets, which is within the range (less than 10 %) of the proportion of reads reported to have derived from ultra-long antibodies in the literature (Saini et al., 1999; Wang et al., 2013).

3.6. Differences in V_H gene distribution by CDRH3 length

In addition to gene usage differences observed between bovine breeds, variations were also seen between full-length antibodies with short (≤ 10 AA), medium (11–39 AA) and ultra-long (≥ 40 AA) CDRH3 length (Fig. 3b & c). IGHV1-14 was the most abundant V_H gene found in short antibodies. In Ankole, expression of IGHV1-20 and IGHV1-27 was also high for both short and medium length antibodies. This result was in contrast to that reported in Deiss et al. (2017) study where they found that >70 % of antibodies with <40 AA CDRH3 length utilized IGHV1-10 gene. Such dominant use of a gene was not observed in either of the African breeds from this dataset. Lower levels of expression of IGHV1-20 and IGHV1-30 were observed in Boran compared to Ankole. V_H genes IGHV1-25, IGHV1-32 and IGHV1-37 were found at lower than 1 % abundance in the datasets (Table S1) which may indicate that they are less frequently used in these breeds. Due to their low frequency, this data was omitted in Fig. 3. In the ultra-long subset of antibodies (≥ 40 AA), IGHV1-7 gene was observed at 90 % and 93 % for Ankole and Boran respectively. The remaining small fractions of the ultra-long transcripts were assigned other V_H genes as they encode small variations to the 'CXXVXQ' pattern used to identify transcripts derived from IGHV1-7. Furthermore, as has been shown before (Deiss et al., 2017) and is evident from our data, even though IGHV1-7 doesn't exclusively encode ultra-long antibodies, it is also not

preferentially used in antibodies with short CDRH3 as demonstrated by the small proportion of short antibodies assigned to the IGHV1-7 gene (3.3 % in Ankole, 5.2 % in Boran).

There was no significant difference in J_H gene usage between short and medium length antibodies, however, ultra-long antibodies were expressed predominantly with the IGHJ2-4 gene. Amongst the D_H genes, IGHD8-2 was the most abundant in both Ankole and Boran ultra-long antibodies. Differential usage of D_H genes was observed in short and medium antibodies in both the breeds.

Amongst the short antibodies, the most frequent V_H - D_H gene pairing observed for Ankole was between IGHV1-14 and IGHD3-1 at 4 % and for Boran was between IGHV1-14 and IGHD9-1 at 5.8 % (Fig. S4). In medium length antibodies, pairing of IGHV1-20 with IGHD6-2 was highest, at 3.3 %, for Ankole, while that of IGHD3-1 with IGHV1-7 was highest at 4 % for Boran. For ultra-long antibodies, the most abundant pairing was between IGHV1-7 and IGHD8-2 at 37.6 % for Ankole and 39.3 % for Boran. In both the repertoires the DH gene for a large proportion of short antibodies was left unassigned due to lack of sequence similarity.

3.7. The mutation frequency in Ig V_H regions is related to CDRH3 length

In both Ankole and Boran sequences, we observed a significantly higher mutation rate in the CDRs than in FWRs (ANOVA $P < 2.2e-16$) (Figs. 4, S5), in keeping with what has been observed in other studies (Liljavirta et al., 2013). When grouped by CDRH3 length we observed that mutation frequency was significantly lower in sequences with ultra-long CDRH3 length in the five structural regions - FWR1, FWR2, FWR3, CDRH1 and CDRH2 - for both breeds (ANOVA $P < 2.2e-16$) compared to sequences with short and medium CDRH3 lengths. This corroborates previous observations (Deiss et al., 2017; Haakenson et al., 2018) that, for the ultra-long antibodies, antigen binding occurs mainly at the CDRH3 loop which drives diversity in the region, while other regions are conserved for structural purposes. Short and medium antibodies had significantly (ANOVA $P < 2.2e-16$) more total mutations than ultra-long antibodies. FWR of all groups of antibodies had on average fewer mutations (range: 0.013–0.07) than CDR (range: 0.03–0.16) (Table S2). Entropy analyses over the V_H gene complemented the findings from mutation frequency analysis (Fig. S6). We found a statistically significant increase in entropy in CDR compared to FWR for all groups, including the ultra-long class.

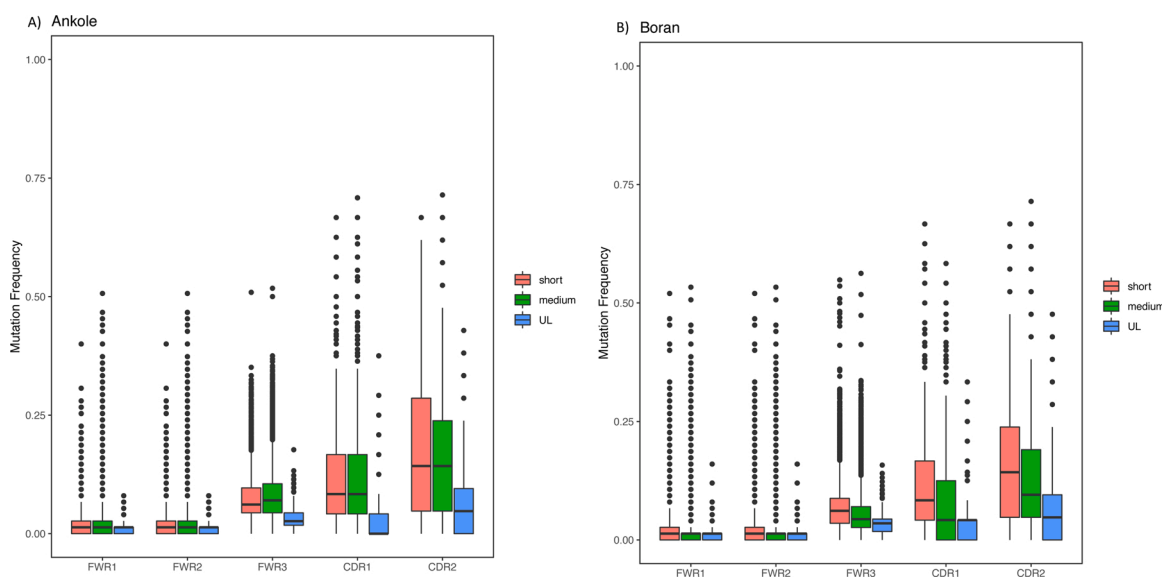


Fig. 4. Box and whisker plots showing mutation frequencies observed in framework (FWR1, FWR2, FWR3) and complementarity determining regions (CDRH1, CDRH2) of antibodies with short (≤ 10 AA), medium (11–39 AA) and ultra-long (≥ 40 AA) CDRH3 sequence in Ankole and Boran. Horizontal line inside a box gives the median mutation frequency.

3.8. Analysis of CDRH3 sequence in ultra-long antibodies

Analyses of CDRH3 sequences of the ultra-long antibodies from this study show characteristics that largely corroborate with observations made in other studies. IGHV1-7 has an 8 bp duplication at the 3' end which causes a frameshift mutation that extends the V_H region by three residues (Deiss et al., 2017; Haakenson et al., 2018). These conserved residues make the CXXVXQ motif which initiate the ascending β -strand of the “stalk” of the CDRH3 of an ultra-long antibody. Also highly conserved in our data is the IGHD8-2-derived CPDG motif which comprises the cysteine that forms a spatially conserved di-sulphide bond in the “knob” region (Fig. 5). Germline IGHD8-2 gene encodes four cysteines that can form di-sulphide bonds. We found four well conserved cysteines in both Ankole and Boran ultra-long CDRH3. Non-germline cysteines that were introduced through somatic hypermutation were also present in CDRH3 of both breeds. The number of cysteines increased with CDRH3 length and there was a higher proportion of even number of cysteines than odd number (Fig. S7). The germline alternating tyrosines at the 3' end of CDRH3 that encode the descending β strand of the “stalk” were also fairly conserved.

4. Discussion

A key utility of high throughput sequencing technologies to immunogenetic studies is to identify and distinguish functional antibody sequences from a total BCR repertoire. This is useful in elucidating B-cell responses to infection or vaccination. In this study we have attempted to identify and characterize bovine antibodies from two animals from different African *Bos indicus* breeds. Bovine antibodies are of special interest, because a small subset of their heavy chain antibodies bear exceptionally long CDRH3 (Vadnais and Smider, 2016) thought to have broadly neutralizing properties (Sok et al., 2017).

This work confirms previous reports that bovine ultra-long CDRH3

loops can reach over 70 residues in length. Such long V_H regions pose a technical challenge to the current high throughput short-read sequencing technologies such as Illumina, which currently is the most accurate sequencing technology with the lowest error rate suitable for immune repertoire studies (Bolotin et al., 2012). In order to overcome this hurdle, we devised a nested PCR amplification and sequencing approach that enabled assembly of a range of different V_H lengths. Analysis on the CDRH3 length distribution showed that ~5 % of the antibodies bear ultra-long CDRH3 lengths ranging from 40 to 75 amino acids. This finding is consistent with previous observations showing that the ultra-long CDRH3 population is less than 10 % of the entire bovine BCR repertoire. The most abundant CDRH3 lengths in this bovine data were within 20 amino acids which in humans and mice are typically 8–16 amino acids.

Methods for V(D)J germline allele assignment is an area that is still under development, and validity of the current assignments is only statistical (Yaari and Kleinstein, 2015). Furthermore, they may exhibit species-specific inaccuracies going by what we have observed from our bovine data, in which ~13 % of the sequences derived from IGHV1-7 alleles were assigned to others by IMGT/HighV-Quest. Immune repertoire studies in humans have demonstrated that IGHV gene usage shifts in response to influenza immunization and varies by ethnicity (Avnir et al., 2016; Pappas et al., 2014; Liu and Lucas, 2003; Atkinson et al., 1996). Our data suggests a difference in VDJ gene usage of heavy chain Ig transcripts between the two bovine breeds. We show that the African *B. indicus* breeds show no significant bias to any specific V_H gene. Although this observation is preliminary due to limitation in the number of animals used, species-specific variation in germline alleles has been observed in macaques and in-bred mice (Corcoran et al., 2016; Collins et al., 2015). Less than 40 % of unique alleles were shared between two in-bred mice, while less than 5 % of them were shared between in-bred mice of different species. To establish any genetic or functional implications of the current observations, more data from naïve and multiple

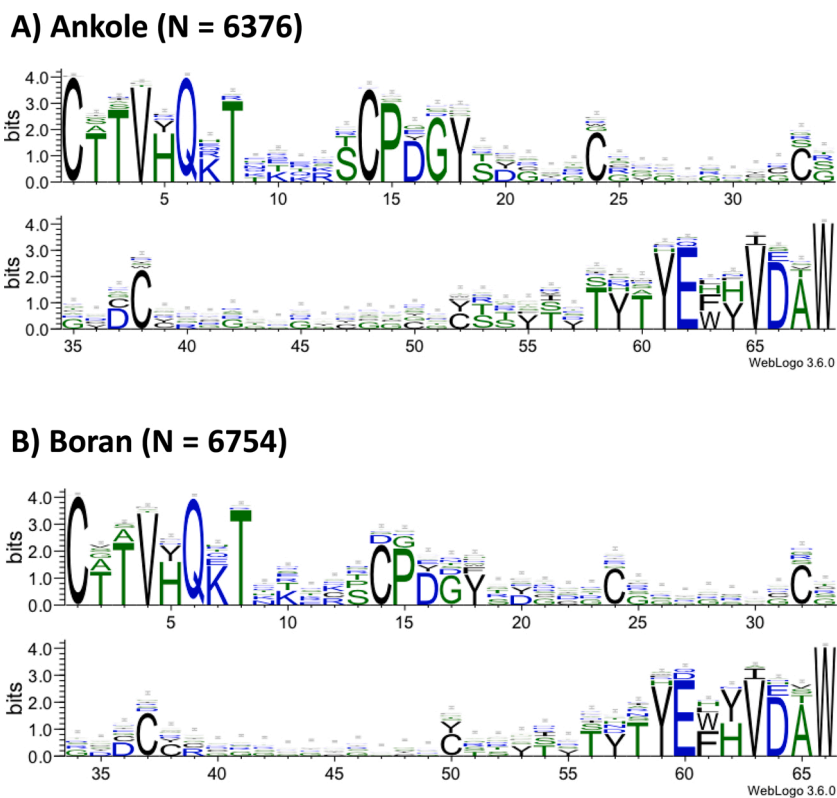


Fig. 5. Conservation of residues in the junction region shown as sequence logos. Junction region for the most abundant ultra-long CDRH3 length shown for all. Number of sequences comprised in the sequence logo is given in brackets in the headings. A) Ankole - 66 AA, B) Boran - 64 AA. Of note are CTTVHQ motif from IGHV1-7, CPDG, three conserved cysteine residues and the alternating 3' tyrosine from IGHD8-2, and DAW motif from IGHJ2-4.

repertoires per breed will be necessary.

D_H gene assignment to mature transcripts based on sequence similarity methods has been deemed to be challenging because D_H genes are typically short and highly mutated through rearrangements and somatic hypermutation. As such, the D_H gene usage proportions reported in this study should be interpreted with caution. The accuracy of gene assignment using sequence similarity methods is dependent on the completeness of the germline gene database used. Germline genes used for the annotation in IMGT/HighV-Quest are determined from Holstein cattle (Ma et al., 2016; Koti et al., 2010) which is a taurine breed whereas the African cattle breeds in this study are of the *Bos indicus* species. Allelic variation has been observed even between individual mice (Corcoran et al., 2016) so it is highly likely that allelic difference exists between cattle breeds from different subspecies. Preliminary analysis to identify novel IGHV alleles in IgM repertoires of these African indicine breeds reveal that they may encode several new alleles (data not shown).

Somatic hypermutations on the BCR have been shown to target particular DNA motifs (hotspots) and the mutation frequency is not uniform across the V(D)J sequence. Generally, mutation frequency in the CDR that binds the antigen is higher than the FWR that maintains the antibody structure (Yaari and Kleinstein, 2015). CDR regions are less structurally constrained and increased mutation rate indicates antigen-driven selection pressure to generate high affinity antibodies. In bovine ultra-long antibodies low variability in FWR and CDR1 and CDR2 is expected as a result of restricted use compared to the CDR3. Our data corroborate what has been reported previously that bovine antibodies bearing short and medium CDRH3 lengths have generally higher mutation frequency than their counterparts with ultra-long CDRH3 lengths.

The current BCR dataset on the two African indicine breeds has also confirmed previous observation with taurine breeds on the distribution pattern of cysteine residues in antibodies bearing ultra-long CDRH3. In both Ankole and Boran heavy chain transcripts, we observed that the number of cysteines increased with CDRH3 length and, more interestingly, that the proportion of even numbers of cysteine residues were significantly higher than those with odd numbers. This distribution pattern of cysteine residues in antibodies bearing ultra-long CDRH3 sequences supports the previous hypothesis on their role in generating mini-domains as folded knobs through disulfide-bonding. The different folding knobs is thought to contribute to the complexity and diversity of cow antibody binding domain repertoire (Wang et al., 2013).

Humans and mice can generate long CDRH3 antibodies with broad HIV and influenza viral strain neutralizing activity (Walker et al., 2009). However, these are rare events with CDRH3 lengths of about 28 amino acid residues. In contrast, immunization of cattle with HIV Env trimer antigens (JR-FL gp120 and BG505 SOSIP), readily resulted in the generation of much more potent ultra-long CDRH3 antibodies with broadly neutralizing activity (Sok et al., 2017). Hence, ultra-long CDRH3 antibodies and knob domains are of interest in the design of novel therapeutic reagents and for the identification of novel epitopes. While ultra-long CDRH3 cattle antibodies have found new biotechnology applications, their general role in mediating immunity in cattle remains to be elucidated.

Taken together, we have used a short-read high throughput sequencing technology to study bovine antibody repertoires from two African indigenous cows. We report a method of successfully sequencing ultra-long bovine Ig VH transcripts using short-read Illumina technology. This method could alleviate the technical challenges encountered in profiling bovine ultra-long antibody repertoire using the low-error rate short-read sequencing technologies. Based on the current attempt to annotate bovine IgH VDJ genes, we report a preliminary observation on VDJ germline usage and distribution by the two African breeds. This preliminary study will pave the way for more robust analysis using more individual animals per breed. Our data corroborate previous observations on unique features of bovine antibodies bearing ultra-long CDRH3. The methodology presented here will be useful in bovine antibody and

epitope discovery and in vaccine studies to help elucidate the functional role of antibodies with different CDRH3 lengths.

Author statement

Vishvanath Nen: Conceptualization, Supervision, Editing and Funding acquisition.

Samuel O. Oyola: Investigation, Methodology, Validation, Software, Writing-Original Draft, supervision and Data analysis.

Sonal P. Henson: Formal analysis, Data Curation, Validation, Visualization, Review & Editing.

Benjamin Nzau: Investigation

Elizabeth Kibwana: Investigation.

Data availability

Raw reads have been deposited in SRA under project ID numbers PRJNA690660.

Funding

This work was funded by the Bill and Melinda Gates Foundation (BMGF) and the Department for International Development (DFID) of the United Kingdom [OPP1078791], the Norman Borlaug Commemorative Research Initiative, an initiative between the Feed the Future program of United States Agency for International Development (USAID), USA and United States Department of Agriculture (USDA)-Agricultural Research Service, USA (58-5348-2-117F) and through funding from the CGIAR Research Program on Livestock.

Declaration of Competing Interest

The authors declare there is no conflict of interest.

Acknowledgments

We thank Dr Lucilla Stienaa (ILRI), Dr Anna Lacasta (ILRI) and Stephen Munyao (ILRI) for useful discussions and advice. We would like to thank Dr John Hammond (The Pirbright Institute, UK) for exchange of data and primer information.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.molimm.2021.08.017>.

References

- Alamyar, E., Duroux, P., Lefranc, M.P., Giudicelli, V., 2012a. IMGT® tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* 882, 569–604. https://doi.org/10.1007/978-1-61779-842-9_32.
- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., Lefranc, M.P., 2012b. IMGT/HighV-quest: the IMGT® web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* 8 (1).
- Atkinson, M.J., Cowan, M.J., Feeney, A.J., 1996. New alleles of IGKV genes A2 and A18 suggest significant human IGKV locus polymorphism. *Immunogenetics* 44 (2), 115–120. <https://doi.org/10.1007/S002510050098>, 1996 44:2.
- Avnir, Y., Watson, C.T., Glanville, J., Peterson, E.C., Tallarico, A.S., Bennett, A.S., et al., 2016. IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci. Rep.* 6 <https://doi.org/10.1038/srep20842>.
- Bolotin, D.A., Mamedov, I.Z., Britanova, O.V., Zvyagin, I.V., Shagin, D., Ustyugova, S.V., Turchaninova, M.A., Lukyanov, S., Lebedev, Y.B., Chudakov, D.M., 2012. Next generation sequencing for TCR repertoire profiling: platform-specific features and correction algorithms. *Eur. J. Immunol.* 42 (11), 3073–3083. <https://doi.org/10.1002/eji.201242517>.
- Cole, C., Volden, R., Dharmadhikari, S., Scelfo-Dalbey, C., Vollmers, C., 2016. Highly accurate sequencing of full-length immune repertoire amplicons using Tn5-enabled

- and molecular identifier-guided amplicon assembly. *J. Immunol.* 196, 2902–2907. <https://doi.org/10.4049/jimmunol.1502563>.
- Collins, A.M., Wang, Y., Roskin, K.M., Marquis, C.P., Jackson, K.J.L., 2015. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos. Trans. Biol. Sci.* 370 (1676) <https://doi.org/10.1098/rstb.2014.0236>.
- Corcoran, M.M., Phad, G.E., Bernat, N.V., Stahl-Hennig, C., Sumida, N., Persson, M.A.A., Martin, M., Hedestam, G.B.K., 2016. Production of individualized v gene databases reveals high levels of immunoglobulin genetic diversity. *Nat. Commun.* 7, 13642. <https://doi.org/10.1038/ncomms13642>.
- Deiss, T.C., Vadnais, M., Wang, F., Chen, P.L., Torkamani, A., Mwangi, W., Lefranc, M.-P., Criscitiello, M.F., Smider, V.V., 2017. Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cell. Mol. Immunol.* 16 (1), 64–75. <https://doi.org/10.1038/cmi.2017.117>.
- Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., Quake, S.R., 2014. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32, 158–168. <https://doi.org/10.1038/nbt.2782>.
- Gupta, N.T., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Yaari, G., Kleinstein, S. H., 2015. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* 31, 3356–3358. <https://doi.org/10.1093/bioinformatics/btv359>.
- Haakenson, J.K., Huang, R., Smider, V.V., 2018. Diversity in the cow ultralong CDR H3 antibody repertoire. *Front. Immunol.* 9, 1–10. <https://doi.org/10.3389/fimmu.2018.01262>.
- Hiatt, J.B., Patwardhan, R.P., Turner, E.H., Lee, C., Shendure, J., 2010. Parallel, tag-directed assembly of locally derived short sequence reads. *Nat. Methods* 7, 119–122. <https://doi.org/10.1038/nmeth.1416>.
- Koti, M., Kataeva, G., Kaushik, A.K., 2010. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Mol. Immunol.* 47, 2119–2128. <https://doi.org/10.1016/j.molimm.2010.02.014>.
- Larsen, P.A., Smith, T.P.L., 2012. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC Immunol.* 13, 1–12. <https://doi.org/10.1186/1471-2172-13-52>.
- Lefranc, M.P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., Lefranc, G., 2003. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.* 27 (1), 55–77. [https://doi.org/10.1016/S0145-305X\(02\)00039-3](https://doi.org/10.1016/S0145-305X(02)00039-3).
- Liljavirta, J., 2014. Bovine B Cells: Antibody Repertoire Diversification in Fetal Cattle. *Liljavirta, J., Ekman, A., Knight, J.S., Pernthaner, A., Iivanainen, A., Niku, M., 2013. Activation-induced cytidine deaminase (AID) is strongly expressed in the fetal bovine ileal Peyer's patch and spleen and is associated with expansion of the primary antibody repertoire in the absence of exogenous antigens. Mucosal Immunol.* 6, 942–949. <https://doi.org/10.1038/mi.2012.132>.
- Liljavirta, J., Niku, M., Pessa-Morikawa, T., Ekman, A., Iivanainen, A., 2014. Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS One* 9 (6). <https://doi.org/10.1371/journal.pone.0099808>.
- Liu, L., Lucas, A.H., 2003. IGH V3-23*01 and its allele V3-23*03 differ in their capacity to form the canonical human antibody combining site specific for the capsular polysaccharide of *Haemophilus influenzae* type b. *Immunogenetics* 55 (5), 336–338. <https://doi.org/10.1007/S00251-003-0583-8>, 2003 55:5.
- Lundin, S., Gruselius, J., Nystedt, B., Lexow, P., Käller, M., Lundeberg, J., 2013. Hierarchical molecular tagging to resolve long continuous sequences by massively parallel sequencing. *Sci. Rep.* 3 <https://doi.org/10.1038/srep01186>.
- Ma, L., Qin, T., Chu, D., Cheng, X., Wang, J., Wang, X., Wang, P., Han, H., Ren, L., Aitken, R., Hammarström, L., Li, N., Zhao, Y., 2016. Internal duplications of DH, JH, and C region genes create an unusual IGH gene locus in cattle. *J. Immunol.* 196, 4358–4366. <https://doi.org/10.4049/jimmunol.1600158>.
- Pappas, L., Foglierini, M., Piccoli, L., Kallewaard, N.L., Turrini, F., Silacci, C., Fernandez-Rodriguez, B., Agatic, G., Giacchetto-Sasselli, I., Pellicciotta, G., Sallusto, F., Zhu, Q., Vicenzi, E., Corti, D., Lanzavecchia, A., 2014. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* 516 (7531), 418–422. <https://doi.org/10.1038/nature13764>, 2014 516:7531.
- Rubelt, F., Busse, C.E., Bukhari, S.A.C., Bürckert, J.P., Mariotti-Ferrandiz, E., Cowell, L. G., Watson, C.T., Marthandan, N., Faison, W.J., Hershberg, U., Laserson, U., Corrie, B.D., Davis, M.M., Peters, B., Lefranc, M.P., Scott, J.K., Breden, F., Luning Prak, E.T., Kleinstein, S.H., 2017. Adaptive Immune Receptor Repertoire Community recommendations for sharing immune-repertoire sequencing data. *Nat. Immunol.* 18, 1274–1278. <https://doi.org/10.1038/ni.3873>.
- Saini, S.S., Allore, B., Jacobs, R.M., Kaushik, A., 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur. J. Immunol.* 29, 2420–2426. [https://doi.org/10.1002/\(SICI\)1521-4141\(199908\)29:08<2420::AID-IMMU2420>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A).
- Shi, B., Ma, L., He, X., Wang, X., Wang, P., Zhou, L., Yao, X., 2014. Comparative analysis of human and mouse immunoglobulin variable heavy regions from IMGT/LIGM-DB with IMGT/HighV-QUEST. *Theor. Biol. Med. Model.* 11 (1) <https://doi.org/10.1186/1742-4682-11-30>.
- Six, A., Mariotti-Ferrandiz, M.E., Chaara, W., Magadan, S., Pham, H.P., Lefranc, M.P., Mora, T., Thomas-Vaslin, V., Walczak, A.M., Boudinot, P., 2013. The past, present, and future of immune repertoire biology - the rise of next-generation repertoire analysis. *Front. Immunol.* 4, 1–16. <https://doi.org/10.3389/fimmu.2013.00413>.
- Sok, D., Le, K.M., Vadnais, M., Saye-Francisco, K.L., Jardine, J.G., Torres, J.L., Berndsen, Z.T., Kong, L., Stanfield, R., Ruiz, J., Ramos, A., Liang, C.H., Chen, P.L., Criscitiello, M.F., Mwangi, W., Wilson, I.A., Ward, A.B., Smider, V.V., Burton, D.R., 2017. Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature* 548 (7665), 108–111. <https://doi.org/10.1038/nature23301>.
- Vadnais, M.L., Smider, V.V., 2016. ScienceDirect *Bos taurus* ultralong CDR H3 antibodies. *Curr. Opin. Struct. Biol.* 38, 62–67. <https://doi.org/10.1016/j.sbi.2016.05.004>.
- Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'connor, K.C., Hafler, D.A., Vigneault, F., Kleinstein, S.H., 2014. PRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* 30 (13), 1930–1932. <https://doi.org/10.1093/bioinformatics/btu138>.
- Walker, L.M., Phogat, S.K., Chan-Hui, P.-Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek, M.D., Fling, S., Mitcham, J.L., Lehrman, J.K., Priddy, F.H., Olsen, O.A., Frey, S.M., Hammond, P.W., Investigators, P.G.P., Kaminsky, S., Zamb, T., Moyle, M., Koff, W.C., Poignard, P., Burton, D.R., 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* (80-) 326, 285–289. <https://doi.org/10.1126/SCIENCE.1178746>.
- Walther, S., Czerny, C.P., Diesterbeck, U.S., 2013. Exceptionally long CDR3H are not isotype restricted in bovine immunoglobulins. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0064234>.
- Walther, S., Tietze, M., Czerny, C.P., König, S., Diesterbeck, U.S., 2016. Development of a bioinformatics framework for the detection of gene conversion and the analysis of combinatorial diversity in immunoglobulin heavy chains in four cattle breeds. *PLoS One* 11 (11). <https://doi.org/10.1371/journal.pone.0164567>.
- Wang, F., Ekiert, D.C., Ahmad, I., Yu, W., Zhang, Y., Bazirgan, O., Torkamani, A., Raudsepp, T., Mwangi, W., Criscitiello, M.F., Wilson, I.A., Schultz, P.G., Smider, V. V., 2013. Reshaping antibody diversity. *Cell* 153, 1379–1393. <https://doi.org/10.1016/j.cell.2013.04.049>.
- Wardemann, H., Busse, C.E., 2017. Novel approaches to analyze immunoglobulin repertoires. *Trends Immunol.* 38, 471–482.
- Xu, J.L., Davis, M.M., 2000. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. *Immunity* 13 (1), 37–45. [https://doi.org/10.1016/S1074-7613\(00\)00006-6](https://doi.org/10.1016/S1074-7613(00)00006-6).
- Yaari, G., Kleinstein, S.H., 2015. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* 7 <https://doi.org/10.1186/s13073-015-0243-2>, 121.
- Zhao, Y., Jackson, S.M., Aitken, R., 2006. The bovine antibody repertoire. *Dev. Comp. Immunol.* 30 (1-2), 175–186. <https://doi.org/10.1016/j.dci.2005.06.012>.