The Journal of Immunology

Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle

Li Ma, Tong Qin, Dan Chu, Xueqian Cheng, Jing Wang, Xifeng Wang, Peng Wang, Haitang Han, Liming Ren, Robert Aitken, Lennart Hammarström, Ning Li and Yaofeng Zhao

J Immunol 2016; 196:4358-4366; Prepublished online 6 April 2016;
doi: 10.4049/jimmunol.1600158
http://www.jimmunol.org/content/196/10/4358

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle

Li Ma,*[,1] Tong Qin,[†,1] Dan Chu,* Xueqian Cheng,* Jing Wang,* Xifeng Wang,* Peng Wang,* Haitang Han,* Liming Ren,* Robert Aitken,[‡] Lennart Hammarström,[§] Ning Li,* and Yaofeng Zhao*

It has been suspected for many years that cattle possess two functional IgH gene loci, located on *Bos taurus* autosome (BTA) 21 and BTA11, respectively. In this study, based on fluorescence in situ hybridization and additional experiments, we showed that all functional bovine IgH genes were located on BTA21, and only a truncated μCH2 exon was present on BTA11. By sequencing of seven bacterial artificial chromosome clones screened from a Hostein cow bacterial artificial chromosome library, we generated a 678-kb continuous genomic sequence covering the bovine IGHV, IGHD, IGHJ, and IGHC genes, which are organized as IGHVn-IGHDn-IGHJn-IGHM1-(IGHDP-IGHV3-IGHDn)$_3$-IGHJn-IGHM2-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA. Although both of two functional IGHM genes, IGHM1 and IGHM2, can be expressed via independent VDJ recombinations, the IGHM2 can also be expressed through class switch recombination. Likely because more IGHD segments can be involved in the expression of IGHM2, the IGHM2 gene was shown to be dominantly expressed in most tissues throughout different developmental stages. Based on the length and identity of the coding sequence, the 23 IGHD segments identified in the locus could be divided into nine subgroups (termed IGHD1 to IGHD9). Except two members of IGHD9 (14 nt in size), all other functional IGHD segments are longer than 30 nt, with the IGHD8 gene (149 bp) to be the longest. These remarkably long germline IGHD segments play a pivotal role in generating the exceptionally great H chain CDR 3 length variability in cattle. *The Journal of Immunology*, **2016, 196: 4358–4366.**

Like humans and mice, cattle express all of the five classes of IgH genes: IGHM (1), IGHG (2–4), IGHE (5), IGHA (6), and IGHD (7). Uniquely, cattle have been thought to possess two IgH loci, located at *Bos taurus* autosome (BTA) 21 and BTA11, respectively (8–11). The IgH locus in BTA21 contains all of the functional IgH isotype-encoding genes, which are arranged in the order 5′-IGHM-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA-3′ (11), whereas the IgH locus in BTA11 was reported to harbor an IGHJ locus and an IGHM-like sequence, IGHML1 (8–10). The IGHML1 was not thought to be functional until 2009, when it was found that there are two functional IGHM loci in the bovine genome, both of which are needed to be inactivated for complete bovine B cell deficiency (12). This raises a fundamental and interesting question: how do the two IGHM loci interact with each other to guarantee B cell development and a normal immune response in cattle.

IgM plays a crucial role in B cell development and the immune response (13). The membrane-bound IgM functions as a BCR and is the first expressed IgH isotype during B cell ontogeny, and its expression is essential for B cell survival, development (14), and subsequent expression of other IgH classes. In early stages of B cell maturation, a successful expression of IgM H chain gene in one chromosome would turn off recombination of IGHV with rearranged D-J in the second homologous chromosome, thus ensuring the monospecificity of the BCR, a phenomenon known as "allelic exclusion" (15–17). This process is not confined solely to allelic genes located on homologous chromosomes; it also regulates the activity of IgH genes located on different chromosomes. A number of studies using transgenic mice have revealed that the expression of exogenous transgenic IgH genes could cause a strong inhibition of endogenous IgH genes and disordered B cell development (17–20). Thus far, cattle are the only mammalian species reported to have two functional IgH loci on different chromosomes, and it remains unclear how these two gene loci interact when expressed.

Generation of a sufficient Ab repertoire is of great importance for a species to respond specifically to countless Ags. Several key molecular mechanisms including V(D)J recombination and somatic hypermutation are fundamental to the Ab repertoire diversification in mammals (21–23). As compared with humans and mice, cattle are characterized by having a single functional IGHV family with a very limited number of IGHV genes (24–28), which is similar to rabbits (29–31) and pigs (32, 33). The present understanding of the mechanisms involved in bovine Ab diversification is still based on the partial characterization of germline

*State Key Laboratory of Agrobiotechnology, China Agricultural University, Beijing 100193, People's Republic of China; [†]Embryo Biotechnology and Reproduction Laboratory, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, People's Republic of China; [‡]Faculty of Health and Life Sciences, York St John University, York YO31 7EX, United Kingdom; and [§]Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska University Hospital Huddinge, SE-141 86 Stockholm, Sweden

[1]L.M. and T.Q. contributed equally to this work.

ORCIDs: 0000-0002-6904-631X (L.M.); 0000-0003-3874-8068 (R.A.); 0000-0003-4499-8610 (Y.Z.).

Received for publication January 27, 2016. Accepted for publication March 9, 2016.

This work was supported by the National Natural Science Foundation of China (Grants 31530070 and 31301011).

The sequences presented in this article have been submitted to the National Center for Biotechnology Information's GenBank (http://www.ncbi.nlm.nih.gov/) under accession number KT723008.

Address correspondence and reprint requests to Prof. Yaofeng Zhao, State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, People's Republic of China. E-mail address: yaofengzhao@cau.edu.cn

The online version of this article contains supplemental material.

Abbreviations used in this article: BAC, bacterial artificial chromosome; BTA, *Bos taurus* autosome; CDRH3, H chain CDR 3; CH, H chain C region; CSR, class switch recombination; FISH, fluorescence in situ hybridization; IPP, ileal Peyer's patch.

IGHV, IGHD, and IGHJ gene segments. To date, ~10 functional IGHV genes, 14 IGHD, and 4 functional IGHJ segments have been defined (9, 24, 25, 28, 34–36). All the functional IGHV genes and some other potentially functional IGHV genes (either additional IGHV genes or allelic variants) belong to a single family (IGHV1) (25, 27). All members of the two additional IGHV families (IGHV2, IGHV3) identified in cattle are pseudogenes (27). Among the IGHD segments, three IGHD segments with unusual lengths (149, 154, and 121 bp) have been reported and might contribute significantly to the length of H chain CDR 3 (CDRH3) (25). The identification of exceptionally long germline IGHD segments is consistent with the finding that bovine IgH chains possess ultralong CDRH3s (35–38), which was seen as compensation for the limited diversity generated by VDJ recombination. A recent study suggested that the Abs with ultralong CDRH3s can generate structural diversity via combinations of somatically generated disulfides (39). A number of studies have concluded that the bovine Ab repertoire is primarily diversified by somatic hypermutation and junctional flexibility caused by nucleotide deletions and additions during the recombination process (25, 26, 36, 40, 41). However, lack of the sufficient information (such as number and physical positions) of the bovine IGHV, IGHD segments has been an obstacle to a complete elucidation of Ab repertoire diversification in cattle. Our aim was thus to assemble definitive data on the genomic organization of the bovine IgH locus (or loci) and to dissect the processes underlying the Ab repertoire diversification.

Materials and Methods

Southern blotting–based screening of the bovine bacterial artificial chromosome library

Filters for the Holstein bull bacterial artificial chromosome (BAC) library RPCI-42 were purchased from the BACPAC Resources Center (http://bacpac.chori.org/home.htm), Children's Hospital Oakland Research Institute. The IGHV1 gene family, IGHM, IGHG, and IGHA gene–specific probes were labeled with digoxigenin-11-dUTP using a PCR digoxigenin probe synthesis kit (Roche, Basel, Switzerland). PCR primers were: VHI-F: 5′-GTG GAC CCT CCT CTT TGT GCT G-3′; VHI-R: 5′-ACA GTA GTA TGT GGC TGT GTC CTC-3′; Cu-F: 5′-TTG TCC CGC CTC GCA ACA GC-3′; Cu-R: 5′-GGG GGA TGG TGA AGA CCC CGA-3′; Cg-F: 5′-TCC AGT CCC GAC GAC GCC AA-3′; Cg-R: 5′-CAC GTG ACC TCG GGC GTT CC-3′; Ca-F: 5′-GCT TCT TCC CGT CGG CAC CC-3′; Ca-R: 5′-TCG CAG TTC GGC ACG CAA CA-3′. Hybridization and detection were performed using the DIG-High Prime DNA Labeling and Detection Starter Kit II following the manufacturer's instructions.

Quantitative RT-PCR

Total RNA was isolated from various tissues using an RNeasy Mini Kit (Qiagen, Valencia, CA), and RNA quantity and quality were assessed with a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA). First-strand cDNA was synthesized using M-MLV Reverse Transcriptase (Promega, Madison, WI) and oligodeoxythymidine [oligo$(dT)_{20}$] primers. The two IGHM genes expression assays were performed using LightCycler 480 SYBR Green I Master mix (Roche, Basel, Switzerland) with primers to GAPDH (For: 5′-AGA TGG TGA AGG TCG GAG TG-3′; Rev: 5′-GAA GGT CAA TGA AGG GGT CA-3′), IGHM1 (For: 5′-CAG AAG TGC TGT CCC CCA-3′; Rev: 5′-TGT TTG GGG CTG AAG TCC GT-3′), or IGHM2 (For: 5′-CGG AAG TGC TGT CCC CAG-3′; Rev: 5′-TGT TTG GGG CTG AAG TCC GT-3′) under the following cycling conditions: 95°C 5 min, then 40 cycles of 95°C 10 s, 62°C 10 s, 72°C 10 s.

Fluorescence in situ hybridization analysis

BAC DNA was isolated using a QIAGEN Large-Construct Kit (Qiagen, Valencia, CA) according to the manufacturer's protocols. The BACs selected as probes were labeled with ChromaTide Alexa Fluor 568-5-dUTP or ChromaTide Alexa Fluor 488-5-dUTP (Thermo Fisher Scientific, Waltham, MA) using the BioPrime DNA Labeling System (Invitrogen, Waltham, MA). The probes were hybridized to cattle metaphase chromosomes prepared from the bovine fibroblasts, and the cattle chromosomes were counterstained with DAPI. The hybridization results for each experiment were captured and analyzed with an Olympus BX63 fluorescence microscope equipped with CellSens Dimension-V1.4 software. Chromosome identification followed international nomenclature (42). The detailed fluorescence in situ hybridization (FISH) protocols were described previously (43).

Annotation of the Bos taurus *IgH gene locus*

The *Bos taurus* Ig C region genes were retrieved by comparing with GenBank files. The National Center for Biotechnology Information's GenBank accession numbers of the sequences used were as follows: IGHM, AY230207, U63637; IGHD, AF515672, AFJ19352.1; IGHG1, X16701; IGHG2, M36946; IGHG3, U63638, U63639; IGHE, U63640; IGHA, AF109167. The IGHV, IGHD, and IGHJ gene segments were identified using online software (http://embossgui.sourceforge.net/demo/fuzznuc) and by searching for adjacent recombination signal sequences allowing for six mismatches.

Analysis of the recombined switching fragments

The nested PCR amplifications and analysis of the recombined S regions were conducted as described previously (44) The primers used were as follows: class switch recombination (CSR)-Su1-1F: 5′-AGG GGC GAG CGG GCT CCC AGG-3′; CSR-Su1-2F: 5′-CAT GTC TGT GTT CGC GGT TCA-3′; CSR-Su2-1R: 5′-TCT CTC TGG GAC AGT GGA GCG-3′; CSR-Su2-2R: 5′-CCC TCA CTG CCC GCG ATT-3′; CSR-Sg-1R: 5′-ATT CCA GCG CCT GCT CAG CTG-3′; CSR-Sg-2R: 5′-CTA ATA CCC TGA ACT CCC AG-3′; CSR-Su2-1F: 5′-GGA GGC ATC GGG CAC CCA-3′; CSR-Su2-2F: 5′-TCA CGT CTG TGA GGG CAG CC-3′.

Preparation and sequencing analysis of fetus and adult IGHV cDNA libraries

Spleen and bone marrow samples were collected from two Holstein fetuses at 110 d gestation. Spleen and lymph node samples were collected from three adult Holstein cows that were 12–18 mo old. The synthesized cDNA samples originating from the different tissues were used as templates, and the IGHV regions of the two IgM subclasses were amplified using KOD plus (TOYOBO, Osaka, Japan) with primers to IGHM1 (For: 5′-ATG AAC CCA CTG TGG ACC CT-3′; Rev: 5′-AGC ACT TCT GCC TCT GGA GT-3′) or IGHM2 (For: 5′-ATG AAC CCA CTG TGG ACC CT-3′; Rev: 5′-AGC ACT TCC GCC TCT GTA GC-3′) under the following conditions: 94°C for 5 min; 35 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 40 s and a final extension at 72°C for 7 min. The PCR products were purified using a gel extraction kit (Omega, Norcross, GA) and subsequently cloned into the pMD-19T vector (Takara, Dalian, China) and sequenced.

The partial, frameshifted, and stop codon–containing sequences were discarded, and unintentional duplicates were further removed using the CD-HIT Web server (45). DNA and protein sequence editing, alignments, and comparisons were performed using DNASTAR's Lasergene software suite, MEGA 6. The sequence data from the cDNA libraries were analyzed with IBM SPSS Statistics 20.0 and Microsoft Excel 2013.

Human and mouse sequence retrieval and statistical analysis

All of the available human and mouse V-region sequences were downloaded from the IMGT/LIGM-DB database (http://www.imgt.org/ligmdb/). The removal of nonfunctional, duplicated sequences and subsequent detailed statistical analysis for functional and unique sequences were conducted by searching IMGT/HighV-QUEST (http://www.imgt.org/HighV-QUEST/search.action) (46).

Results

Screening of the bovine BAC library

The Holstein bull BAC library RPCI-42 (47) was primarily screened by Southern blotting using the IGHV, μCH3, γCH2, and αCH2 as probes; then all of the available positive clones were confirmed by PCR, and the insert size of these clones was determined by pulsed field gel electrophoresis. In total, 23 IGHV, 11 IGHM, 5 IGHG, and 1 IGHA gene positive clones were confirmed. Of these clones, six were shown to be positive for both the IGHV and the IGHM genes, three contained the IGHM and IGHG genes, and one clone contained the IGHE and IGHA genes.

The IGHM gene positive clones could be further identified as IGHM (AY23027)-containing or IGHML1 (U63637)-containing

clones based on the sequence differences between the two genes. Seven clones were found to contain IGHML1, whereas four were confirmed as IGHM gene–containing BACs.

Chromosomal localization of the bovine IgH gene locus

Previous studies have suggested that the bovine genome harbors two IgH loci: the main IgH locus, with a IGHJ locus, a IGHM gene, and all other IgH isotype-encoding genes, has been mapped to BTA21, whereas a second IGHJ locus, together with the bovine IGHM-like sequence IGHML1, was thought to be anchored to BTA11 (8–11). To validate this conclusion, three IGHV-positive BAC clones (RP42-49A20, 152O19, 195P14), one IGHM clone (RP42-90B11), one IGHML1 (RP42-498B11) and INRA-944D11, a clone reported to carry the second IGHJ locus and IGHML1 gene (9), were selected as probes for FISH analysis. Surprisingly, all of the BAC clones were mapped to BTA21 and no obvious signal was observed in BTA11 or other chromosomes (Fig. 1). This provides strong evidence that only a single IgH locus exists in the bovine genome.

Genomic organization of the bovine IgH gene locus

The bovine IGHC gene locus in BTA21 was previously characterized on an ~150-kb contiguous fragment of DNA in the following order: 5′-IGHJ-IGHM-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA-3′ (11). Interestingly, our FISH results showed that the IGHML1 gene and the second IGHJ locus were also assigned to BTA21, which indicated that previous investigations were incomplete. To map the entire IGHC locus, we selected seven IgH gene–positive BAC clones (RP42-195P14, 49A20, 498B11, 567N23, 90B11, 4E14, and INRA-944D11) for sequencing, which generated a 678-kb contiguous genomic sequence covering part of the IGHV region and IGHD, IGHJ, and IGHC genes in their entirety (Fig. 2).

Two IGHM genes (termed IGHM1 and IGHM2 hereafter) were found to be located in tandem in the IgH locus; their coding sequences are identical to IGHML1 (U63637) and the IGHM (AY2302071), respectively. The IGHM1 gene (U63637) was located in the upstream region of the IGHC locus, whereas the IGHM2 gene (AY2302071) was situated in the 3′-portion of the locus; the two previously characterized IGHJ loci were also found (9). As might be anticipated from the location of IGHM1 and IGHM2 genes, the IGHJ locus expressed infrequently (J_H^{low}; AY149283) was located 5 kb upstream of the IGHM1 gene, and the J_H^{high} locus (AY158087) was found 5 kb upstream of the IGHM2 gene. Now that the structure and the relative position of the two IGHJ loci are defined by our experiments, we rename the IGHJ segments IGHJ1 to IGHJ12 according to the 5′-3′ order, where only IGHJ4, IGHJ6 (components of the J_H^{low} locus), IGHJ10, and IGHJ12 (components of the J_H^{high} locus) are functional.

Five IGHD segments were found upstream of the J_H^{low} locus, whereas three "IGHDP-IGHV3-IGHDn" clusters were identified between the IGHM1 and J_H^{high} locus. Among the three clusters, IGHDP1 (located immediately downstream of the IGHM1 gene) shared >90% sequence identity at the genomic level with the functional IGHD gene, and has a 1-bp insertion in the H chain C region (CH) 2 exon that causes a frameshift. The other two pseudo-IGHD genes (termed IGHDP2 and IGHDP3) are fragmentary, in which the 5′-225 bp sequence of the CH1 exon and the second hinge exon are lost.

Four IGHD loci were found in the bovine IgH locus, from which 23 IGHD segments were identified (Supplemental Fig. 1). Based on the length and identity of the coding sequence, the IGHD segments could be divided into nine subgroups (termed IGHD1–IGHD9). Among these 23 IGHD segments, 16 appear to be functional, as all members of IGHD2 and IGHD5 subgroups are pseudogenized because of highly diverged recombination signal sequences. Except two members of IGHD9 (14 nt in size), all other 14 functional IGHD segments are longer than 30 nt, with the IGHD8 segment (149 bp) to be the longest (Supplemental Fig. 1). The previously reported 154- and 121-bp-long IGHD segments were not found (25), and they may be allelic variants of IGHD8.

Within the ~284-kb IGHV region, we identified 15 IGHV1, 17 IGHV2, and 14 IGHV3 (including 3 IGHV3 in the downstream "IGHDP-IGHV3-IGHDn" clusters). Consistent with previous studies (24, 26–28), only the IGHV1 family is found to be functional, and among the 15 IGHV1 genes, 3 were also designated as pseudogenes because of either in-frame stop codons or sequence deletions.

The downstream IGHC locus is identical to the previous description (11) except for some divergence in the distance between

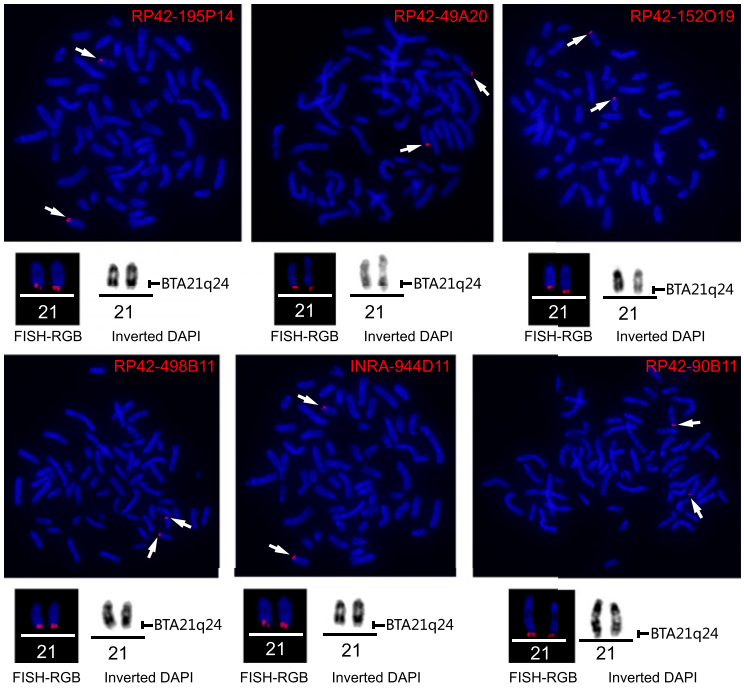


FIGURE 1. FISH localization of BAC clones on BTA21. The arrows in the cattle metaphase spread and enlarged chromosome 21 show the location of the bovine IgH locus in BTA21q24 by FISH analyses with the BAC clones RP42-195P14, 49A20, 152O19, 498B11, 90B11, and 4E14.

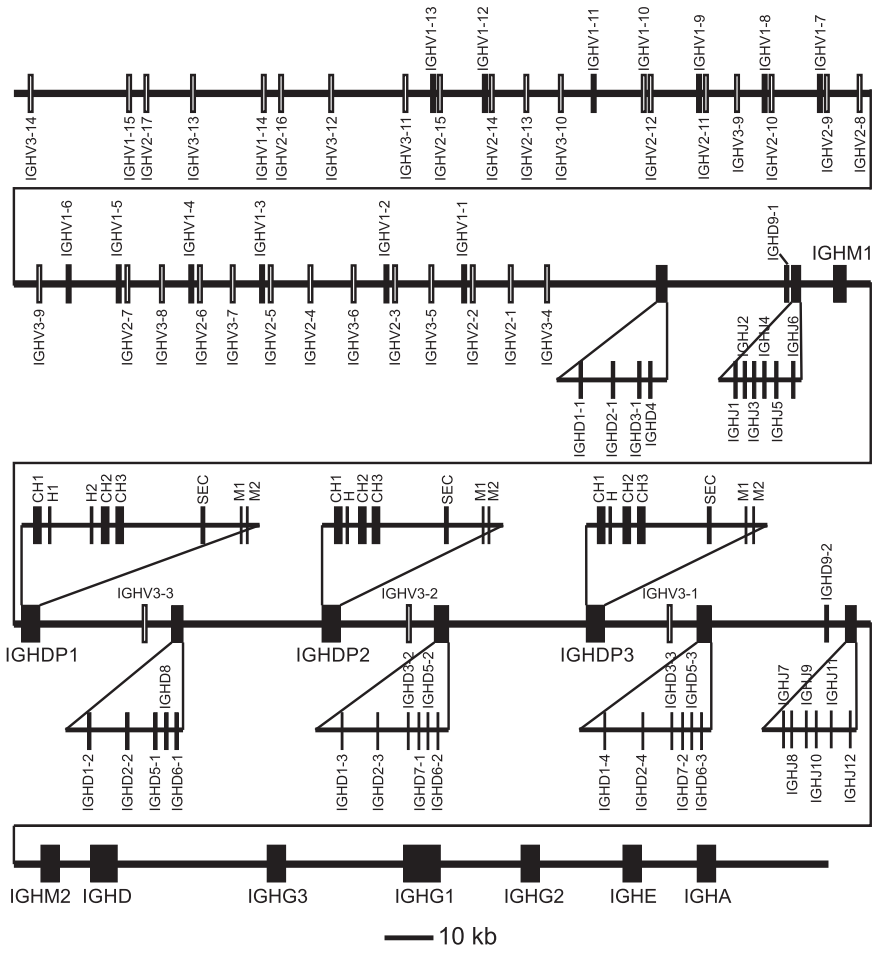


FIGURE 2. Physical map of the bovine IgH gene locus. Open boxes indicate the IGHV pseudogenes. H, hinge region-encoding exon; IGHDP1, IGHDP2, IGHDP3, the pseudo-Ig C region H chain δ genes; M, membrane region–encoding exon.

the adjacent constant genes. The above analysis allowed us to deduce the genomic organization of the bovine IgH gene locus as IGHVn-IGHDn-IGHJn-IGHM1-(IGHDP-IGHV3-IGHDn)$_3$-IGHJn-IGHM2-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA.

No functional IgH locus exists on BTA11

All of the bovine IGHC genes are located in the single IgH locus of BTA21, including the IGHM1 gene, which was previously thought to be located on BTA11. To determine whether an extra IGHC locus exists in BTA11, we first conducted a BLAST search against the bovine genome (genome assembly: UMD3.1, http://asia.ensembl.org/Bos_taurus/Tools/Blast?db=core) using the germline IGHM2 gene as the query. As a result, a 262-bp-long sequence with a high homology (98.9%) to the μCH2 exon was found in BTA11.

A sense primer derived from the 5′-end of the μCH2 exon and antisense primers for the amplification of the partial and the intact μCH2 exons were designed for the BAC screening to identify the BAC clones containing the μCH2-like sequence, and four BAC clones, RP42-6O15, 175K20, 253E1, and 452M15, were shown to contain the μCH2-like sequence (Fig. 3A). Using RP42-452M15 as a probe, the FISH results indicated that the μCH2-like sequence was indeed located at BTA11 (Fig. 3B). Then, we used genome walking and obtained a 587-bp 5′- and a 2246-bp 3′-flanking sequence. The sequence analysis indicated that no other IGHM exons or IgH genes existed in BTA11 (Fig. 3C).

Expression analysis of IGHM1 and IGHM2 genes

Because both the IGHM1 and the IGHM2 genes have IGHD and IGHJ loci in their respective upstream sequences, independent D-J recombination events may occur when the two genes are expressed. To analyze the IGHD and IGHJ usage of IGHM1 and IGHM2, 249 IGHM1 and 287 IGHM2 unique cDNA sequences were cloned from fetal spleen and ileal Peyer's patches (IPP). The fetal spleen and IPP were used because the IgH cDNA sequences derived from these tissues are less mutated by somatic hypermutation than IgH cDNA sequences derived from adult cattle tissues, and it is thus easier to determine the germline origin of IGHV, IGHD, and IGHJ segments in each IgH transcript. It was shown that IGHD and IGHJ assignments exhibited a strong preference in both IgM subclasses: IGHD4 (97.0%) and IGHJ6 (100.0%) were dominantly used by IGHM1, whereas in IGHM2, IGHJ10 accounted for 99.0% of all sequences, and IGHD6 (47.2%), IGHD7 (26.4%), and IGHD4 (12.1%) dominated the VDJ recombination. These data suggest that the IGHM1 and IGHM2 expression can be accomplished through independent VDJ recombination.

Because IGHM2 is situated downstream of the IGHM1, and the two IGHM genes have their own switch regions, it is possible that the IGHM2 can also be expressed through CSR, in the case that the IGHM1 is expressed first. To determine whether CSR is involved in IGHM2 expression, we generated specific nested PCR primers for the amplification of the recombined Sμ1-Sμ2, Sμ1-Sγ1, and Sμ2-Sγ1 fragments (Fig. 4A). Using spleen genomic DNA as a template, we amplified the desired recombination products with these primers. A total of 13 unique Sμ1-Sμ2, 12 Sμ1-Sγ1, and 20 Sμ2-Sγ1 recombined fragments were sequenced (Fig. 4B, Supplemental Fig. 2), which clearly demonstrated that the bovine IGHM2 gene could also be expressed through a CSR process.

We further examined the relative expression levels of two IGHM genes in different tissues and in distinct developmental stages using quantitative real-time PCR (Fig. 5). IGHM2 was shown to be

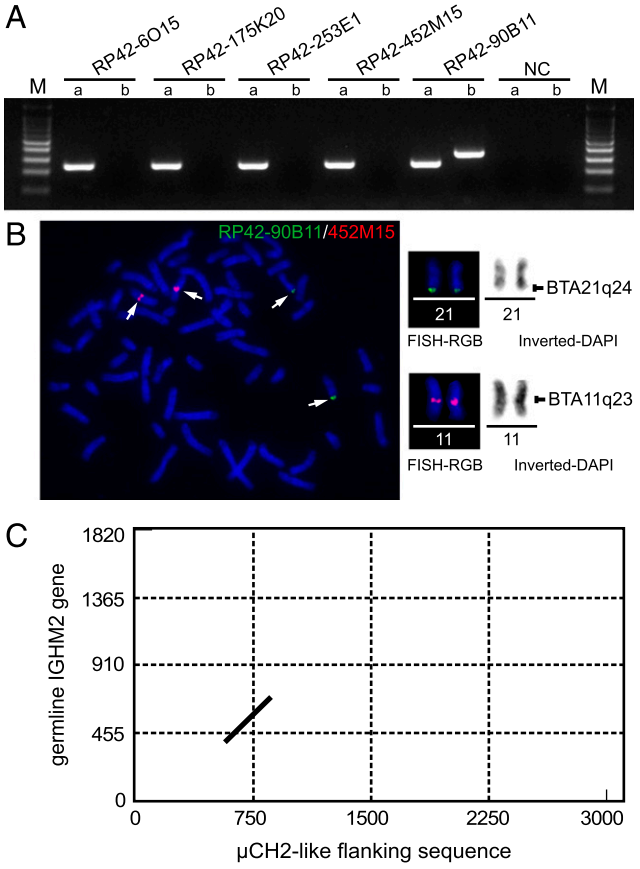


FIGURE 3. FISH localization of BAC clones on BTA11. (**A**) Four BAC clones containing a μCH2-like sequence were identified. a, the μCH2-like sequence (262 bp); b, the intact μCH2 exon (349 bp). RP42-90B11 is used as a positive control for both the partial and the intact μCH2 exon, and water is loaded as the negative control (NC). (**B**) The μCH2-like sequence is physically located in BTA11. RP42-90B11 (green) is used as a probe to denote chromosome 21, and arrows in the metaphase spread and enlarged chromosome 11 indicate the location of the μCH2-like sequence in BTA11 with RP42-452M15 (red) as the probe. (**C**) A dot plot comparison of the partial μCH2 exon flanking sequence with the germline IGHM2 gene, window: 30; percentage: 80.

dominantly expressed in almost all the tissues throughout the developmental stages.

Junctional diversity generated by VDJ recombination

To analyze the contribution of VDJ recombination to Ab diversity, the earlier stated 249 IGHM1 and 287 IGHM2 unique sequences cloned from fetal spleen and IPP cDNA were used for analysis. To determine the extent of bovine VDJ junctional flexibility compared with other species, 9062 human and 6830 mouse unique sequences retrieved from the IMGT/LIGM-DB database (48) were added for comparison. The V, D, and J gene segments, as well as the N and P nucleotide additions to the entire length of the CDRH3 (H chain CDR3) nucleotide sequence, were analyzed (Fig. 6). Sequences with P nucleotides accounted for 39.4, 28.9, 32.0, and 40.72% in bovine IGHM1, IGHM2, human, and mouse, respectively. The average numbers of P nucleotide additions were close in each group (bovine IGHM1, 0.6 ± 0.9; IGHM2, 0.5 ± 1.0; human, 0.6 ± 1.2; mouse, 0.7 ± 1.0). N nucleotide addition was observed in almost all of the human CDRH3 sequences (98.1%), which is significantly higher than that of the other three groups (bovine IGHM1, 79.5%; IGHM2, 84.7%; mouse, 82.8%). In addition, the average number of N nucleotides was markedly larger in human sequences (12.8 ± 8.1) as compared with any other group (bovine IGHM1 4.8 ± 4.9; IGHM2, 5.4 ± 4.7; mouse, 4.4 ± 3.9).

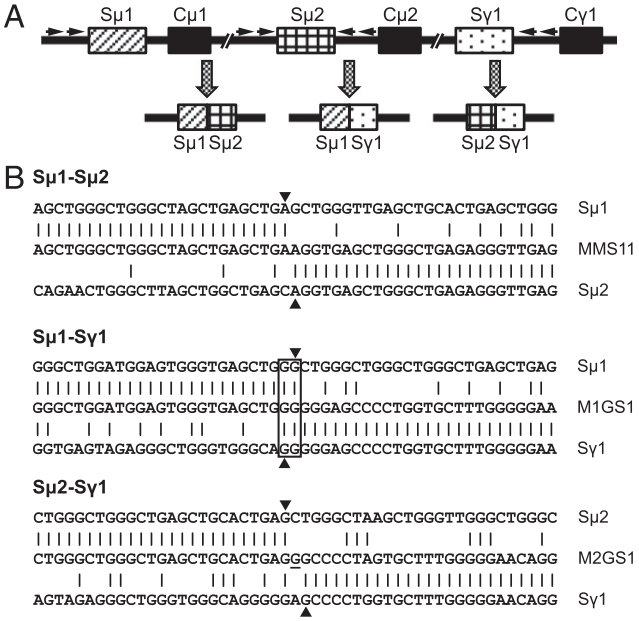


FIGURE 4. Amplification and analysis of the recombined switch fragments Sμ1-Sμ2, Sμ1-Sγ1, and Sμ2-Sγ1. (**A**) Schematic map of the PCR strategy for the amplification of the recombined switch fragments. Arrows indicate the positions of the primers. (**B**) Recombined switch fragments amplified by nested PCR. In each alignment, the middle sequence is the cloned PCR product, and the upper and lower sequences are the switch regions involved in the recombination. Triangles indicate the breakpoints of the switch regions, and the boxes show the overlaps.

The nucleotide contribution of IGHV to the CDRH3s was relatively larger in cattle (IGHM1, 8.8 ± 3.8; IGHM2, 9.5 ± 4.2) compared with that in human and mouse (6.5 ± 1.7 and 5.8 ± 1.7, respectively). The IGHJ gene segment donated sequences were observed to be longest in bovine IGHM1 (15.1 ± 4.0) and shortest in IGHM2 (9.7 ± 3.4). Notably, the most obvious differences lie in the IGHD donation: human and mouse had relatively limited lengths of sequences contributed by the IGHD segments (14.3 ± 5.5 and 10.8 ± 4.7, respectively), whereas obviously longer and more flexible sequence donations of IGHD were observed in bovine IGHM1 (27.5 ± 8.5) and especially in IGHM2 (43.1 ± 25.3).

Length distribution of CDRH3s

For analysis of the length distribution of the CDRH3 regions and subsequent fundamental characteristics, 364 IGHM1 and 404 IGHM2 V-region sequences derived from spleens of three adult cows were examined. Unlike the nearly normal distribution in human and mouse, bovine IGHM1 and IGHM2 demonstrated bimodal distributions, as the extremely short CDRH3s (6–7 aa) were also relatively preferred (IGHM1, 10.4%; IGHM2, 5.4%) in addition to their most frequent CDRH3 length (IGHM1, 15 aa, 19.8%; IGHM2, 23 aa, 9.2%) (Fig. 7). The mean CDRH3 length (amino acid numbers) of bovine IGHM1 was close to the human (15.1 ± 4.5 versus 15.5 ± 4.0) and significantly longer than the mouse (11.5 ± 2.5). As for the bovine IGHM2, the mean length of CDRH3 was found to be markedly longer than the other three groups (19.4 ± 6.8).

In all of the fetal and adult V-region sequences examined earlier, the exceptionally long CDRH3s were exclusively found in IGHM2. The proportion of ultralong CDRH3s (fetal, 11/287, 3.8%; adult, 3/404, 0.7%) was much lower than previous observation (~9%) in bovine B cell hybridomas derived from mitogen-activated peripheral blood B cells (37). This variation could be expected as different materials and methods were used in two experiments.

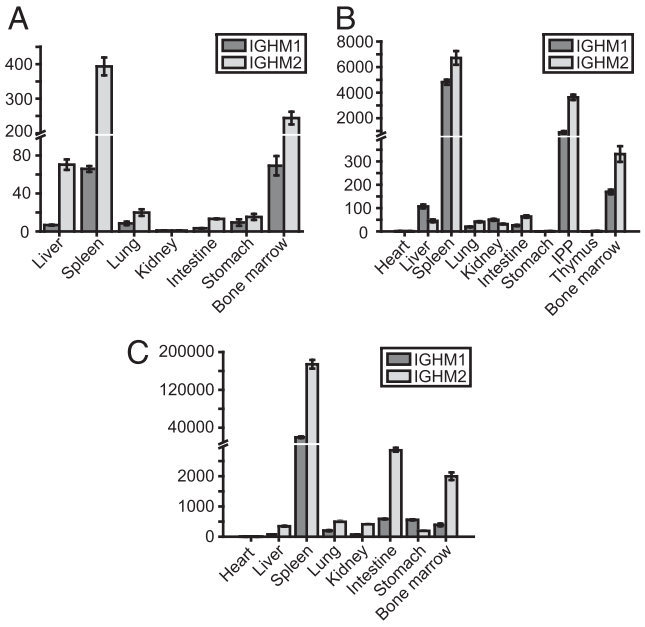


FIGURE 5. Quantitative RT-PCR analysis of IGHM1 and IGHM2 gene expression in the Holstein cow. Data are representative of three independent experiments. Vertical lines are the SD of the mean. (**A**) Expression of IGHM1 and IGHM2 in a 110–gestation d fetus. (**B**) Expression of IGHM1 and IGHM2 in a 1-d neonate. (**C**) Expression of IGHM1 and IGHM2 in a 12-mo adult.

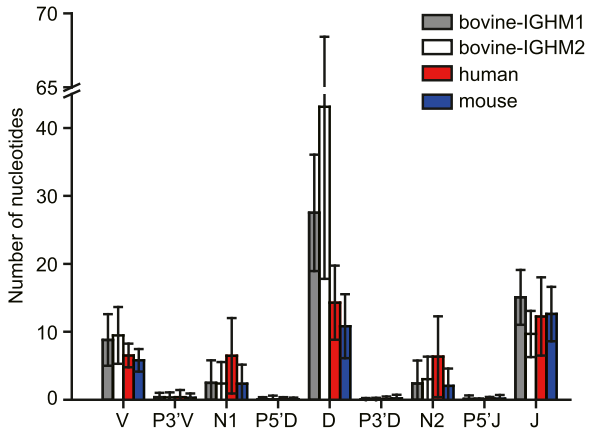


FIGURE 6. Nucleotide contributions of V, D, J, and N, P nucleotides to the length of CDRH3. N1 and N2 are the nucleotide insertions at V-D and D-J junctions, respectively; P3′V are the insertions of palindromic sequences encoded by the 3′ ends of V gene segments; P5′D, P3′D are the insertions of palindromic sequences encoded by the 5′ and 3′ ends of D gene segments, respectively; P5′J are the insertions of palindromic sequences encoded by the 5′ ends of J gene segments.

Analysis of exceptionally long CDRH3s

To characterize the generation of exceptionally long CDRH3s, we analyzed 178 unique functional CDRH3 (46–65aa) sequences cloned from the bone marrow and spleen cDNA of two Holstein cows. Sequence analysis showed that all clones were the products of IGHV1-1+IGHD8+IGHJ10 rearrangements exclusively, as conserved "TTVHQ" of IGHV1-1, "CPDG," and/or "YxYxY" motifs of IGHD8 were identified.

At the genomic level, the VDJ recombinants were cloned using adult bone marrow genomic DNA as template, and 25 unique clones encoding ultralong CDRH3s (>40 aa) were obtained. The germline IGHV genes involved in VDJ rearrangement were found not to be restricted to IGHV1-1 (data not shown). Although most of the ultralong CDRH3s used IGHD8, with the conserved "CDPG" and/or "YxYxY" motifs identified at the 5′ and 3′ portions, respectively, IGHD4, IGHD6, and IGHD7 could also contribute ultralong CDRH3s, as the number of N and P nucleotide insertions could even exceed >100 bp (data not shown). Furthermore, no signs of D-D fusions were observed.

Amino acid usage in the CDRH3 regions

The amino acid distribution and Shannon entropy (amino acid variability in a given position in aligned sequences) (49) for the bovine IGHM1 and IGHM2 CDRH3 sequences of representative lengths were analyzed separately (Fig. 8A, Supplemental Fig. 3). An examination of the Shannon entropy plots showed limited variability at the IMGT positions 105–106 and 115–117, especially at the positions 106 (K/R) and 116 (D). These highly conserved amino acid residues are important for stabilizing the CDRH3 base by forming a salt bridge (50). Although both bovine IGHM1 and IGHM2 showed comparable high entropy values across positions 107–114 (Fig. 8A), the cysteine residues could be found at most of the positions between 107 and 114 in bovine IGHM2 CDRH3, whereas the use of cysteine residues in the counterparts of the IGHM1 CDRH3 loop was extremely restricted.

To further characterize the CDRH3 repertoire composition, the overall amino acid usage in bovine, human, and mouse CDRH3s was subsequently analyzed between positions 107–114 (Fig. 8B). Positions 105–116 and 115–117 were excluded from the statistics because of the high maintenance of germline-encoded motifs. Several fundamental differences were observed between the CDRH3 amino acid contents of the bovine IGHM2 and bovine IGHM1/human/mouse. First, the usage of cysteine residues in the bovine IGHM2 CDRH3 repertoire was much higher than in IGHM1 and human/mouse (bovine IGHM2, 7.9%; bovine IGHM1, 2.5%; human, 1.4%; mouse, 0.25%). Second, the bovine IGHM2 repertoire mostly showed a preference for glycine (21.0%) relative to bovine IGHM1 (16.9%), human (16.0%), and mouse (16.4%), which may increase the structural flexibility of the CDRH3 loop (51, 52). Third, the proline content in the human CDRH3 (4.7%) was significantly higher than any of the other groups (bovine IGHM1, 1.5%; bovine IGHM2, 1.5%; mouse, 2.2%), and a previous study indicated that the frequency of proline also increases as the human CDRH3 length gets longer (52), whereas it may not be necessary for the structural stabilization of the bovine CDRH3 loops.

We subsequently analyzed cysteine usage in the predicted bovine V-region CDR amino acid sequences. In the CDRH3 of the bovine IGHM2 repertoire, 93 of 404 sequences (23.0%) contained no Cys, 116 (28.7%) contained two Cys, and sequences with four or six Cys accounted for 0.5%. There is still a large proportion of the CDRH3s (47.3%) that contains an odd number of Cys residues (one, 169/404, 41.8%; three, 22/404, 5.4%), which was compensated for to a small degree by the extra noncanonical Cys in the CDRH1 (2.2%) or CDRH2 (1.0%). Instead, it is of marked interest that the CDRH3s containing one or three Cys residues were almost always accompanied by a single Cys in the IMGT position 52 of framework region 2 (114/191) or position 66 of framework region 3 (53/191) (Supplemental Fig. 4). As with the ultralong CDRH3s ranging in length from 46 to 65 aa, a strong preference for an even number of Cys (176/178, 98.87%) was observed (data not shown). In the CDRH3 of the bovine naive IGHM1 repertoire, most of the sequences (293/362, 80.9%) contained no Cys and only 27 (7.5%) sequences contained 2 Cys residues.

Discussion

In this study, we confirm that a single bovine IgH locus containing two functional IGHM genes maps to BTA21 and only a μCH2-like

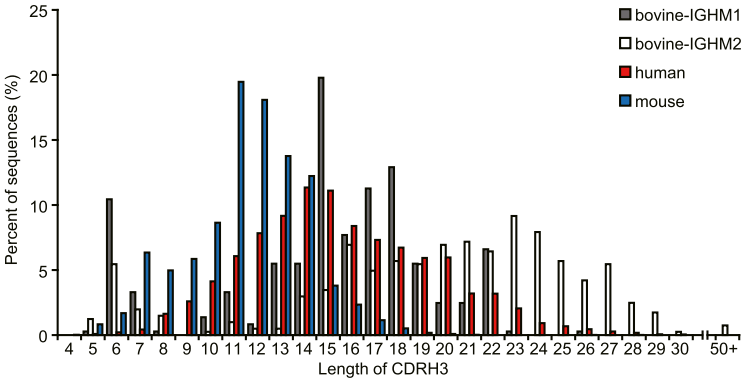


FIGURE 7. CDRH3 length distribution in different species. The CDR3 border is determined based on the IMGT database (IMGT positions 105–117).

sequence exists in BTA11. According to this study, the bovine IgH locus is organized as IGHVn-IGHDn-IGHJn-IGHM1-(IGHDP-IGHV3-IGHDn)3-IGHJn-IGHM2-IGHD-IGHG3-IGHG1-IGHG2-IGHE-IGHA, in which a genetic duplication event appears to have shaped the bovine IgH gene locus. As a result, the V, D, J, and C region genes are not rigidly arranged in order like in other mammalian IgH gene loci, but are rather organized as the zebrafish IgH gene locus (53) or mammalian TCRα/δ gene loci (54), where C region genes are separated by D and/or J segments. This genomic structure allows distinct constant genes to be expressed through independent V(D)J recombination.

This study helps to clarify several key and perplexing questions associated with the bovine IgH genes. The existence of more than one IGHM gene in a single IgH locus has never been seen in other mammalian species and is also rare in tetrapods, because only in crocodilians are three functional IGHM genes found to be located in a single IgH gene locus (44). The finding that two bovine IGHM genes in the same locus answers clearly why both of the genes must be inactivated to inhibit the expression of bovine IgH genes (12). We show that these two IGHM genes can be expressed through independent V(D)J recombination, and the IGHM2 can also be expressed through IGHM1-IGHM2 switching. Because both the IGHM1 and the IGHM2 genes are located upstream of other bovine C region genes, inactivation of one of them would still maintain the functionality of the bovine IgH locus, for example, expression of IgM to drive B cell development and subsequent expression of other IGHC genes through CSR.

Although both the bovine IGHM genes can be functionally expressed, quantitative real-time PCR analyses showed that IGHM2 was dominantly expressed in most tissues throughout different developmental stages. This could be because varying numbers of IGHD segments are respectively involved in the expression of two IGHM genes. Although all 16 functional IGHD segments can be associated with expression of the IGHM2, only 4 of them, all located upstream of the IGHM1, can be involved in the expression of the IGHM1. The theoretical IGHD usage ratio is more or less close to the observed expressional ratios between the two genes.

Compared with humans and mice, cattle are known to have a very limited number of functional IGHV segments. However, it remains unclear what compensating mechanisms this species could

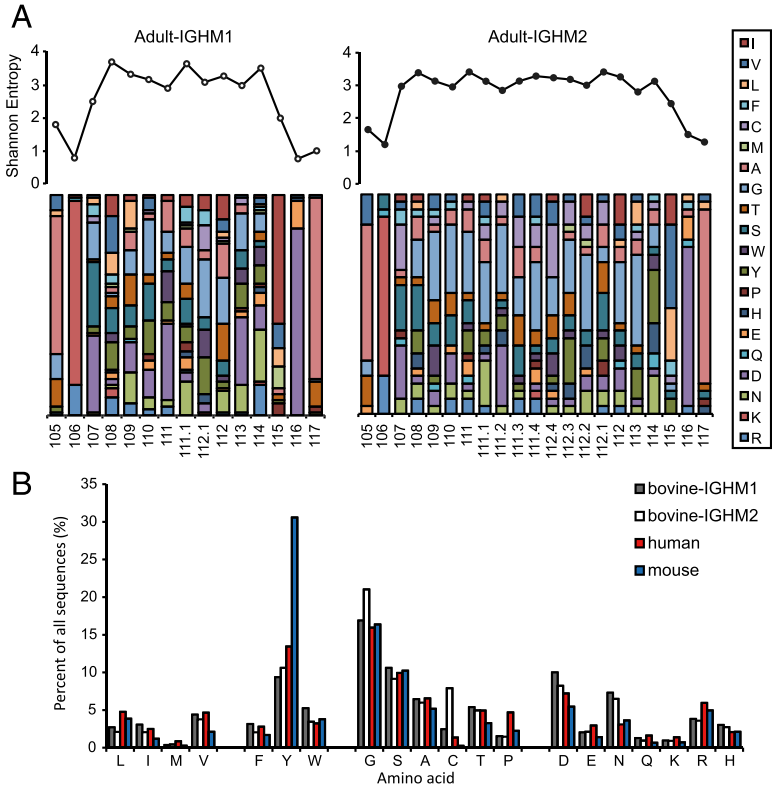


FIGURE 8. Amino acid composition analyses of bovine CDRH3. (**A**) CDRH3 Shannon entropy analysis (*upper panel*) and amino acid frequencies (*lower panel*) at positions 105–117 with median lengths (adult IGHM1, 15 aa, $n = 72$; adult IGHM2, 21 aa, $n = 29$). (**B**) A comparison of the amino acid usage in CDRH3 (IMGT positions 107–114) of the bovine adult IGHM1 (gray), IGHM2 (white), humans (red), and mice (blue). Amino acids are grouped according to their chemical characteristics as follows: hydrophobic, LIMV; aromatic, FYW; small molecular, GSACTP; charged, DENQKRH.

use to generate sufficient IgH sequence diversity. This study reveals that the bovine IGHD segments play a significant role in this process. Although the number of functional IGHD segments in the bovine IgH gene locus is only 16, comparable with 14 in mice and much less than the 27 in humans, a remarkable characteristic of the bovine germline IGHD segments is their length. Whereas the average IGHD lengths in humans and mice are 22.56 ± 6.82 and 16.93 ± 2.37 bp, respectively, the average length of the 16 functional bovine IGHD segments is 47.88 ± 32.05 bp. Except the IGHD9 segments, all other bovine IGHD segments are longer than 31 bp. Although all the 16 bovine IGHD segments are involved in the expression of bovine IGHM2, only 4 of them, located upstream of the IGHM1 gene and with an average length of 31.00 ± 12.36 bp, can be used in the expression of IGHM1 (data not shown). A comprehensive analysis of CDRH3 length in cattle, humans, and mice suggests that there is strong positive correlation between the average lengths of CDRH3 and IGHD segments. Obviously, the bovine IGHM2 H chains show the longest average CDRH3 length. Compared with IgH chains of humans, mice, and even bovine IGHM1, the bovine IGHM2 H chains not only show longer CDRH3 on average, but also much greater length variability in the CDRH3. As shown in Fig. 7, the bovine IGHM2 H chain CDRH3 length can range from 5 to >60 aa. Considering the important role of CDRH3 in Ag binding, this length variability would markedly increase the sequence/structural diversity of bovine Ig H chains. Restricted by a limited number of functional IGHV segments, cattle may mainly rely on the vast CDRH3 length variability to generate a sufficient recombinatorial Ab repertoire. Except for the longer germline IGHD segments, N or P nucleotide additions and exonuclease activity also contribute to this process. Although most exceptionally long CDRH3s are generated by the longest IGHD (IGHD8), some other IGHD segments can also be used to generate exceptionally long CDRH3s together with additions of N and/or P nucleotides. Moreover, the novel Ab diversification mechanism via "CSNS" insertions, specifically at the V(H)–D(H) junction (36), may also be involved in generating exceptionally long CDRH3 over 46 codons in cattle Abs. Short CDRH3s are obviously attributed to trimming of V, D, J ends during the recombination process, as suggested by a previous study (25).

Compared with bovine IGHM1 and human/mouse, bovine IGHM2 showed a higher usage of Gly and Cys. The higher incorporation of Gly may be necessary for bovine IGHM2 to increase flexibility and complexity in the secondary structure of the long CDRH3 loops, and Cys residues are required for long CDRH3 loop structural stability by forming inter- and/or intra-CDR disulfide bridges. In the ultralong CDRH3s analyzed, we found that an even number of Cys residues was strongly preferred (98.9%), which suggests that ultralong CDRH3s (>40 aa) are under strong selection for structural stability maintained by a disulfide bridge. Similarly, an interloop disulfide bond is also frequently observed between the CDRH1 and CDRH3 in dromedary VHHs, which could enhance the structural stability and somewhat reduce the conformational flexibility of the long CDR3 loop (55).

In summary, we show in this study that two functional bovine IGHM genes are located in tandem in a single IgH gene locus on BTA21. Based on the genomic IgH sequence data, we confirmed that albeit having a very limited number of functional germline IGHV gene segments, cattle rely on long germline IGHD segments to generate remarkable CDRH3 length variability forming the Ab recombinatorial repertoire.

Disclosures

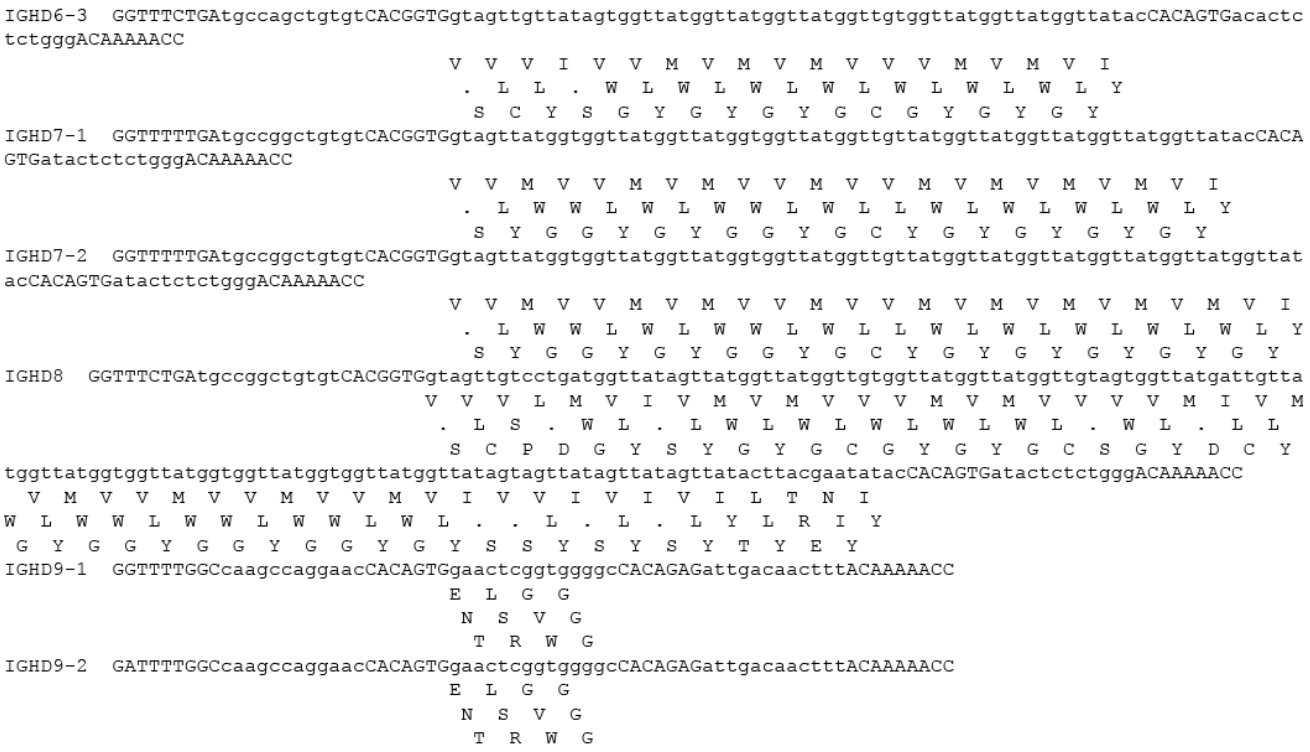
The authors have no financial conflicts of interest.

References

1. Mousavi, M., H. Rabbani, L. Pilström, and L. Hammarström. 1998. Characterization of the gene for the membrane and secretory form of the IgM heavy-chain constant region gene (C mu) of the cow (*Bos taurus*). *Immunology* 93: 581–588.
2. Kacskovics, I., and J. E. Butler. 1996. The heterogeneity of bovine IgG2–VIII. The complete cDNA sequence of bovine IgG2a (A2) and an IgG1. *Mol. Immunol.* 33: 189–195.
3. Rabbani, H., W. R. Brown, J. E. Butler, and L. Hammarström. 1997. Polymorphism of the IGHG3 gene in cattle. *Immunogenetics* 46: 326–331.
4. Symons, D. B., C. A. Clarkson, and D. Beale. 1989. Structure of bovine immunoglobulin constant region heavy chain gamma 1 and gamma 2 genes. *Mol. Immunol.* 26: 841–850.
5. Mousavi, M., H. Rabbani, and L. Hammarström. 1997. Characterization of the bovine epsilon gene. *Immunology* 92: 369–373.
6. Brown, W. R., H. Rabbani, J. E. Butler, and L. Hammarström. 1997. Characterization of the bovine C alpha gene. *Immunology* 91: 1–6.
7. Zhao, Y., I. Kacskovics, Q. Pan, D. A. Liberles, J. Geli, S. K. Davis, H. Rabbani, and L. Hammarstrom. 2002. Artiodactyl IgD: the missing link. *J. Immunol.* 169: 4408–4416.
8. Tobin-Janzen, T. C., and J. E. Womack. 1992. Comparative mapping of IGHG1, IGHM, FES, and FOS in domestic cattle. *Immunogenetics* 36: 157–165.
9. Hosseini, A., G. Campbell, M. Prorocic, and R. Aitken. 2004. Duplicated copies of the bovine JH locus contribute to the Ig repertoire. *Int. Immunol.* 16: 843–852.
10. Hayes, H. C., and E. J. Petit. 1993. Mapping of the beta-lactoglobulin gene and of an immunoglobulin M heavy chain-like sequence to homoeologous cattle, sheep, and goat chromosomes. *Mamm. Genome* 4: 207–210.
11. Zhao, Y., I. Kacskovics, H. Rabbani, and L. Hammarstrom. 2003. Physical mapping of the bovine immunoglobulin heavy chain constant region gene locus. *J. Biol. Chem.* 278: 35024–35032.
12. Kuroiwa, Y., P. Kasinathan, T. Sathiyaseelan, J. A. Jiao, H. Matsushita, J. Sathiyaseelan, H. Wu, J. Mellquist, M. Hammitt, J. Koster, et al. 2009. Antigen-specific human polyclonal antibodies from hyperimmunized cattle. *Nat. Biotechnol.* 27: 173–181.
13. Ehrenstein, M. R., and C. A. Notley. 2010. The importance of natural IgM: scavenger, protector and regulator. *Nat. Rev. Immunol.* 10: 778–786.
14. Lam, K. P., R. Kühn, and K. Rajewsky. 1997. In vivo ablation of surface immunoglobulin on mature B cells by inducible gene targeting results in rapid cell death. *Cell* 90: 1073–1083.
15. Melchers, F., E. ten Boekel, T. Yamagami, J. Andersson, and A. Rolink. 1999. The roles of preB and B cell receptors in the stepwise allelic exclusion of mouse IgH and L chain gene loci. *Semin. Immunol.* 11: 307–317.
16. Geier, J. K., and M. S. Schlissel. 2006. Pre-BCR signals and the control of Ig gene rearrangements. *Semin. Immunol.* 18: 31–39.
17. Zhang, M., X. Cheng, D. Chu, J. Liang, Y. Sun, L. Ma, B. Xu, M. Zheng, M. Wang, L. Ren, et al. 2014. Depletion of conventional mature B cells and compromised specific antibody response in bovine immunoglobulin μ heavy-chain transgenic mice. *Front. Agr. Sci. Eng.* 1: 158–173.
18. Nussenzweig, M. C., A. C. Shaw, E. Sinn, D. B. Danner, K. L. Holmes, H. C. Morse, III, and P. Leder. 1987. Allelic exclusion in transgenic mice that express the membrane form of immunoglobulin mu. *Science* 236: 816–819.
19. Iglesias, A., M. Lamers, and G. Köhler. 1987. Expression of immunoglobulin delta chain causes allelic exclusion in transgenic mice. *Nature* 330: 482–484.
20. Roth, P. E., L. Doglio, J. T. Manz, J. Y. Kim, D. Lo, and U. Storb. 1993. Immunoglobulin gamma 2b transgenes inhibit heavy chain gene rearrangement, but cannot promote B cell development. *J. Exp. Med.* 178: 2007–2021.
21. Jung, D., C. Giallourakis, R. Mostoslavsky, and F. W. Alt. 2006. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* 24: 541–570.
22. Komori, T., A. Okada, V. Stewart, and F. W. Alt. 1993. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science* 261: 1171–1175.
23. Xu, Z., H. Zan, E. J. Pone, T. Mai, and P. Casali. 2012. Immunoglobulin class-switch DNA recombination: induction, targeting and beyond. *Nat. Rev. Immunol.* 12: 517–531.
24. Berens, S. J., D. E. Wylie, and O. J. Lopez. 1997. Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *Int. Immunol.* 9: 189–199.
25. Liljavirta, J., M. Niku, T. Pessa-Morikawa, A. Ekman, and A. Iivanainen. 2014. Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS One* 9: e99808 doi:10.1371/journal.pone.0099808.
26. Lopez, O., C. Perez, and D. Wylie. 1998. A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunol. Rev.* 162: 55–66.
27. Niku, M., J. Liljavirta, K. Durkin, E. Schroderus, and A. Iivanainen. 2012. The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. *Dev. Comp. Immunol.* 37: 457–461.
28. Saini, S. S., W. R. Hein, and A. Kaushik. 1997. A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. *Mol. Immunol.* 34: 641–651.
29. Becker, R. S., M. Suter, and K. L. Knight. 1990. Restricted utilization of VH and DH genes in leukemic rabbit B cells. *Eur. J. Immunol.* 20: 397–402.
30. DiPietro, L. A., P. Sethupathi, M. Kingzette, S. K. Zhai, M. Suter, and K. L. Knight. 1992. Limited repertoire of utilized VH gene segments in a VHa3-allotype-suppressed rabbit. *Int. Immunol.* 4: 555–561.
31. DiPietro, L. A., J. A. Short, S. K. Zhai, A. S. Kelus, D. Meier, and K. L. Knight. 1990. Limited number of immunoglobulin VH regions expressed in the mutant rabbit "Alicia". *Eur. J. Immunol.* 20: 1401–1404.

32. Sun, J., C. Hayward, R. Shinde, R. Christenson, S. P. Ford, and J. E. Butler. 1998. Antibody repertoire development in fetal and neonatal piglets. I. Four VH genes account for 80 percent of VH usage during 84 days of fetal life. *J. Immunol.* 161: 5070–5078.
33. Eguchi-Ogawa, T., N. Wertz, X. Z. Sun, F. Piumi, H. Uenishi, K. Wells, P. Chardon, G. J. Tobin, and J. E. Butler. 2010. Antibody repertoire development in fetal and neonatal piglets. XI. The relationship of variable heavy chain gene usage and the genomic organization of the variable heavy chain locus. *J. Immunol.* 184: 3734–3742.
34. Sinclair, M. C., J. Gilchrist, and R. Aitken. 1997. Bovine IgG repertoire is dominated by a single diversified VH gene family. *J. Immunol.* 159: 3883–3889.
35. Shojaei, F., S. S. Saini, and A. K. Kaushik. 2003. Unusually long germline DH genes contribute to large sized CDR3H in bovine antibodies. *Mol. Immunol.* 40: 61–67.
36. Koti, M., G. Kataeva, and A. K. Kaushik. 2010. Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Mol. Immunol.* 47: 2119–2128.
37. Saini, S. S., B. Allore, R. M. Jacobs, and A. Kaushik. 1999. Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur. J. Immunol.* 29: 2420–2426.
38. Saini, S. S., and A. Kaushik. 2002. Extensive CDR3H length heterogeneity exists in bovine foetal VDJ rearrangements. *Scand. J. Immunol.* 55: 140–148.
39. Wang, F., D. C. Ekiert, I. Ahmad, W. Yu, Y. Zhang, O. Bazirgan, A. Torkamani, T. Raudsepp, W. Mwangi, M. F. Criscitiello, et al. 2013. Reshaping antibody diversity. *Cell* 153: 1379–1393.
40. Kaushik, A. K., M. E. Kehrli, Jr., A. Kurtz, S. Ng, M. Koti, F. Shojaei, and S. S. Saini. 2009. Somatic hypermutations and isotype restricted exceptionally long CDR3H contribute to antibody diversification in cattle. *Vet. Immunol. Immunopathol.* 127: 106–113.
41. Larsen, P. A., and T. P. Smith. 2012. Application of circular consensus sequencing and network analysis to characterize the bovine IgG repertoire. *BMC Immunol.* 13: 52.
42. Cribiu, E. P., D. Di Berardino, G. P. Di Meo, A. Eggen, D. S. Gallagher, I. Gustavsson, H. Hayes, L. Iannuzzi, C. P. Popescu, J. Rubes, et al. 2001. International System for Chromosome Nomenclature of Domestic Bovids (ISCNDB 2000). *Cytogenet. Cell Genet.* 92: 283–299.
43. Raudsepp, T., and B. P. Chowdhary. 2008. FISH for mapping single copy genes. *Methods Mol. Biol.* 422: 31–49.
44. Cheng, G., Y. Gao, T. Wang, Y. Sun, Z. Wei, L. Li, L. Ren, Y. Guo, X. Hu, Y. Lu, et al. 2013. Extensive diversification of IgH subclass-encoding genes and IgM subclass switching in crocodilians. *Nat. Commun.* 4: 1337 doi:10.1038/ncomms2317.
45. Huang, Y., B. Niu, Y. Gao, L. Fu, and W. Li. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26: 680–682.
46. Alamyar, E., P. Duroux, M. P. Lefranc, and V. Giudicelli. 2012. 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.
47. Warren, W., T. P. Smith, C. E. Rexroad, III, S. C. Fahrenkrug, T. Allison, C. L. Shu, J. Catanese, and P. J. de Jong. 2000. Construction and characterization of a new bovine bacterial artificial chromosome library with 10 genome-equivalent coverage. *Mamm. Genome* 11: 662–663.
48. Giudicelli, V., P. Duroux, C. Ginestoux, G. Folch, J. Jabado-Michaloud, D. Chaume, and M. P. Lefranc. 2006. IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences. *Nucleic Acids Res.* 34: D781–D784.
49. Shannon, C. E. 1997. The mathematical theory of communication. 1963. *MD Comput.* 14: 306–317.
50. Shirai, H., A. Kidera, and H. Nakamura. 1996. Structural classification of CDR-H3 in antibodies. *FEBS Lett.* 399: 1–8 doi:10.1016/S0014-5793(96)01252-5.
51. Novotný, J., R. Bruccoleri, J. Newell, D. Murphy, E. Haber, and M. Karplus. 1983. Molecular anatomy of the antibody binding site. *J. Biol. Chem.* 258: 14433–14437.
52. Zemlin, M., M. Klinger, J. Link, C. Zemlin, K. Bauer, J. A. Engler, H. W. Schroeder, Jr., and P. M. Kirkham. 2003. Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J. Mol. Biol.* 334: 733–749.
53. Danilova, N., J. Bussmann, K. Jekosch, and L. A. Steiner. 2005. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* 6: 295–302.
54. Rowen, L., B. F. Koop, and L. Hood. 1996. The complete 685-kilobase DNA sequence of the human beta T cell receptor locus. *Science* 272: 1755–1762.
55. Govaert, J., M. Pellis, N. Deschacht, C. Vincke, K. Conrath, S. Muyldermans, and D. Saerens. 2012. Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. *J. Biol. Chem.* 287: 1970–1979.

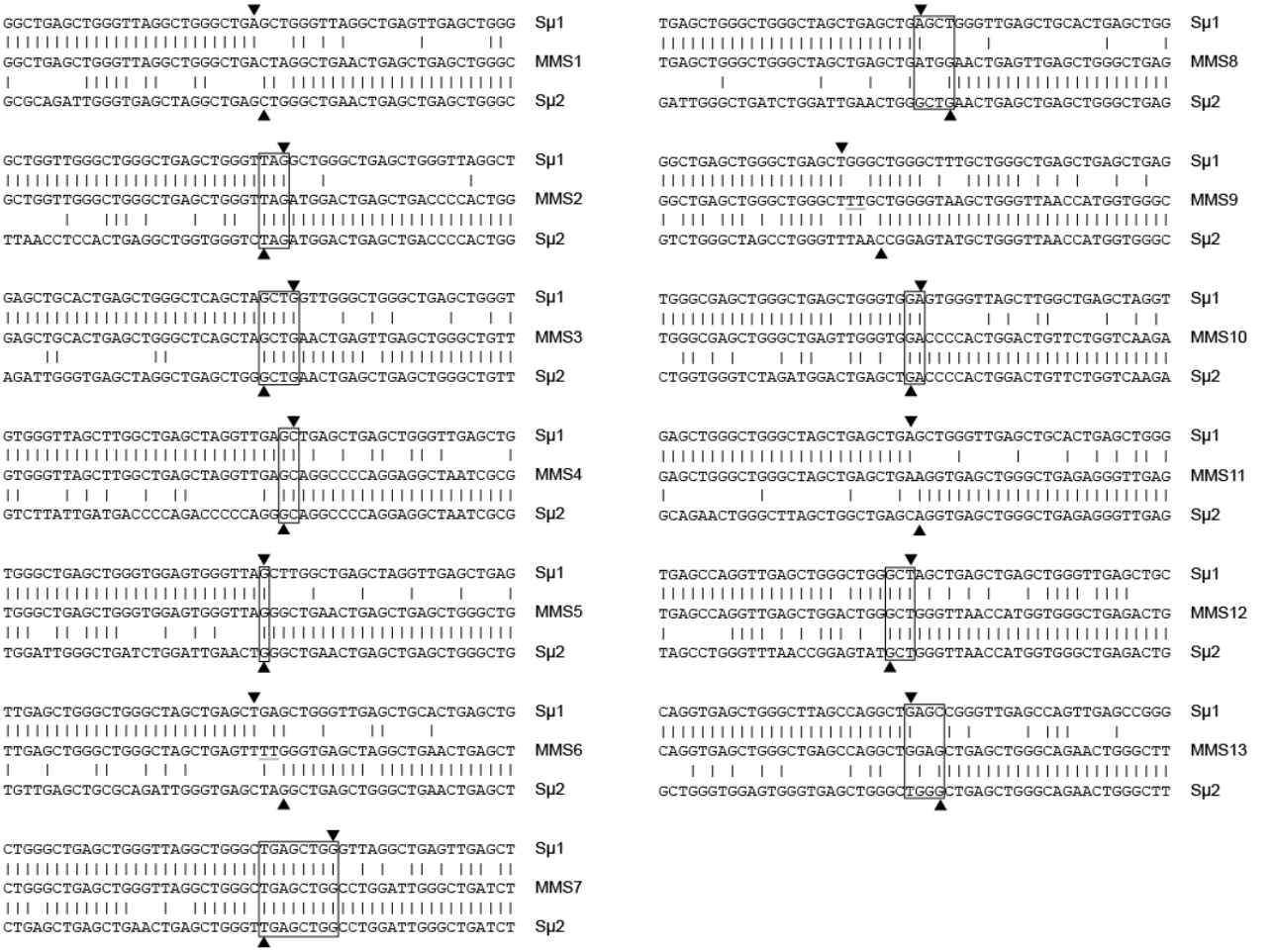
```
IGHD1-1  GGATTTTGAgggtgtgtgtgtCACTGTGagaataccgtgatgatggttactgctacaccCACAGTGactcaggccctgACATAAAGC
                                  R  I  P  .  .  W  L  L  L  H
                                   E  Y  R  D  D  G  Y  C  Y  T
                                    N  T  V  M  M  V  T  A  T
IGHD1-2  GGATTTTGAgggtgtgcgtgtCACCCTGagaatatcgtgatgatggttactgctacaccCACAGTGactcaggccctgACATAAAGT
                                  R  I  S  .  .  W  L  L  L  H
                                   E  Y  R  D  D  G  Y  C  Y  T
                                    N  I  V  M  M  V  T  A  T
IGHD1-3  GGATTTTGAgggtgtgcgtgtCACCGTGagactatcgtgatgatggttactgctacaccCACAGTGactcaggccctgACATAAAGT
                                  R  L  S  .  .  W  L  L  L  H
                                   D  Y  R  D  D  G  Y  C  Y  T
                                    T  I  V  M  M  V  T  A  T
IGHD1-4  GGATTTTGAgggtgtgcgtgtCACCGTGagaatatcgtgatgatggttactgctacaccCACAGTGactcaggccctgACATAAAGT
                                  R  I  S  .  .  W  L  L  L  H
                                   E  Y  R  D  D  G  Y  C  Y  T
                                    N  I  V  M  M  V  T  A  T
IGHD2-1  GCTTTTTGCcaagggctctacTGCGGTGttactatagtgaccacCACAGTGagacatggggcaGCAAACCCT
                                  L  L  .  .  P
                                   Y  Y  S  D  H
                                    T  I  V  T
IGHD2-2  GCTTTTTGCcaagggctctacTGCGGTGttactatagtgaccacCACAGTGagacatggggcaGCAAACCCT
                                  L  L  .  .  P
                                   Y  Y  S  D  H
                                    T  I  V  T
IGHD2-3  GCTTTTTGCcaagggctctacTGCGGTGttactatagtgaccacCACAGTGagacatggggcaGCAAACCCT
                                  L  L  .  .  P
                                   Y  Y  S  D  H
                                    T  I  V  T
IGHD2-4  GCTTTTTGCcaagggctctacTGCGGTGttactatagtgaccacCACAGTGagacatggggcaGCAAACCCT
                                  L  L  .  .  P
                                   Y  Y  S  D  H
                                    T  I  V  T
IGHD3-1  GGTTTCTGAtgccggctgtgtCACGGTGgtattgtggtagctattgtggtagttattatggtacCACAGTGacactgtccaggACAGAAACC
                                  V  L  W  .  L  L  W  .  L  L  W  Y
                                   Y  C  G  S  Y  C  G  S  Y  Y  G
                                    I  V  V  A  I  V  V  V  I  M  V
IGHD3-2  GGTTTCTGAtgccggctgtgtCACGGTGgtattgtggtagctattgtggtagttattatggtacCACAGTGacactgtccaggACAGAAACC
                                  V  L  W  .  L  L  W  .  L  L  W  Y
                                   Y  C  G  S  Y  C  G  S  Y  Y  G
                                    I  V  V  A  I  V  V  V  I  M  V
IGHD3-3  GGTTTCTGAtgccggctgtgtCACGGTGgtattgtggtagctattgtggtagttattatggtacCACAGTGacactgtccaggACAGAAACC
                                  V  L  W  .  L  L  W  .  L  L  W  Y
                                   Y  C  G  S  Y  C  G  S  Y  Y  G
                                    I  V  V  A  I  V  V  V  I  M  V
IGHD4    GGTTTCTGAtgccggctgtgtCATGGTGgtagttatagtggttatggttatggttatagttatggttatacCACAGTGacactctctgggACAAAAACC
                                  V  V  I  V  V  M  V  M  V  I  V  M  V  I
                                   .  L  .  W  L  W  L  W  L  .  L  W  L  Y
                                    S  Y  S  G  Y  G  Y  G  Y  S  Y  G  Y
IGHD5-1  GGTTTCTGAtgccggctgtgtTGTGGTGatgatacgataggtgtggttgtagttattgtagtgttgctacCACAGTGatgctctcagtgTCAGAAACC
                                  M  I  R  .  V  W  L  .  L  L  .  C  C  Y
                                   .  Y  D  R  C  G  C  S  Y  C  S  V  A
                                    D  T  I  G  V  V  V  V  I  V  V  L  L
IGHD5-2  GGTTTCTGAtgccggctgtgtTGTGGTGatgatacgataggtgtggttttagttattgtagtgttgctacCACAGTGacgctctcagtgTCAGAAACC
                                  M  I  R  .  V  W  F  .  L  L  .  C  C  Y
                                   .  Y  D  R  C  G  F  S  Y  C  S  V  A
                                    D  T  I  G  V  V  L  V  I  V  V  L  L
IGHD5-3  GGTTTCTGAtgccggctgtgtTGTGGTGatgatacgataggtgtggttttagttattgtagtgttgctacCACAGTGacgctctcagtgTCAGAAACC
                                  M  I  R  .  V  W  F  .  L  L  .  C  C  Y
                                   .  Y  D  R  C  G  F  S  Y  C  S  V  A
                                    D  T  I  G  V  V  L  V  I  V  V  L  L
IGHD6-1  GGTTTCTGAtgccggctgtgtCACGGTGgtagttgttatagtggttatggttatggttgtggttatggttatggttatgattatacCACAGTGacactc
tctgggACAAAAACC
                                  V  V  V  I  V  V  M  V  M  V  V  V  M  V  M  V  M  I  I
                                   .  L  L  .  W  L  W  L  W  L  W  L  W  L  W  L  .  L  Y
                                    S  C  Y  S  G  Y  G  Y  G  C  G  Y  G  Y  G  Y  D  Y
IGHD6-2  GGTTTCTGAtgccagctgtgtCACGGTGgtagttgttatagtggttatggttatggttatggttgtggttatggttatggttatacCACAGTGacactc
tctgggACAAAAACC
                                  V  V  V  I  V  V  M  V  M  V  M  V  V  V  M  V  M  V  I
                                   .  L  L  .  W  L  W  L  W  L  W  L  W  L  W  L  W  L  Y
                                    S  C  Y  S  G  Y  G  Y  G  Y  G  C  G  Y  G  Y  G  Y
```



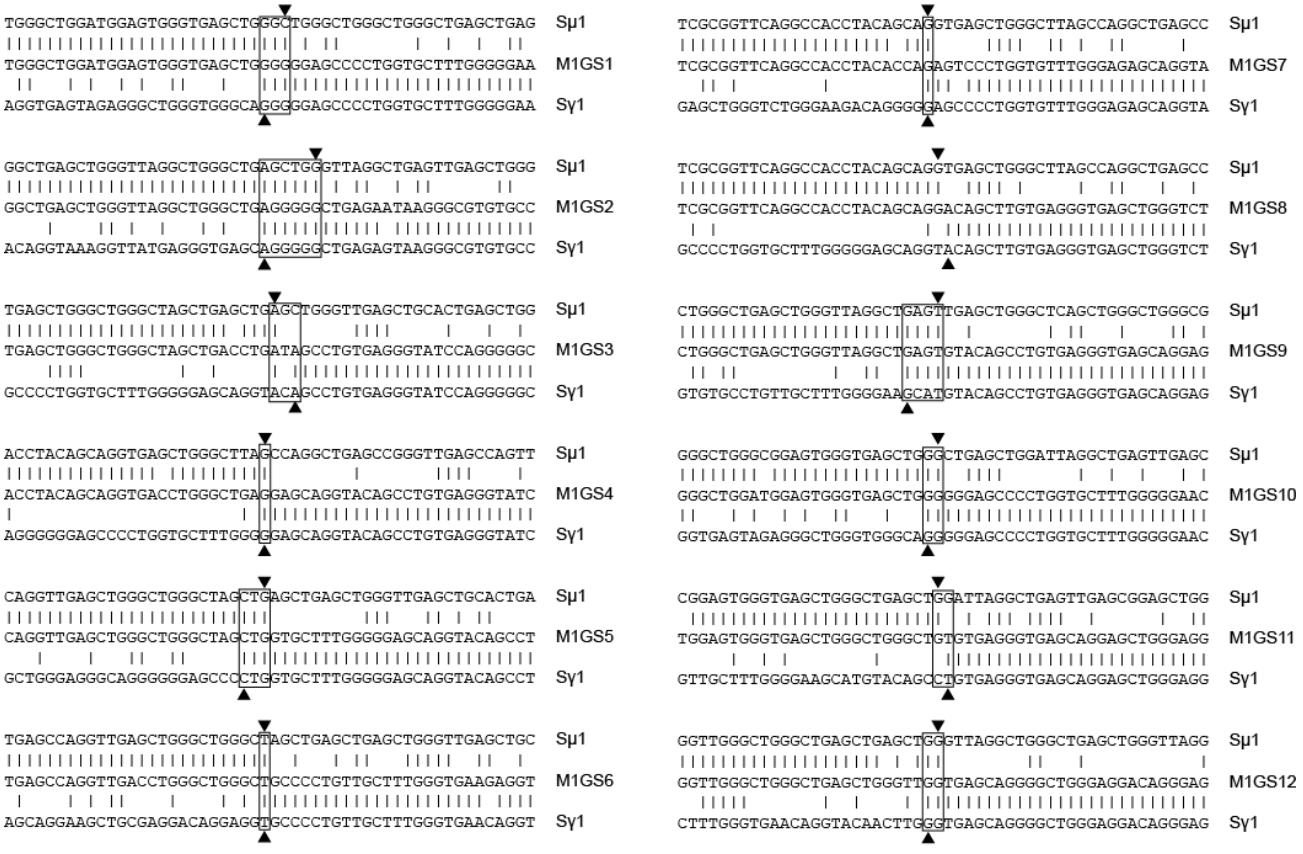
Supplemental Figure 1

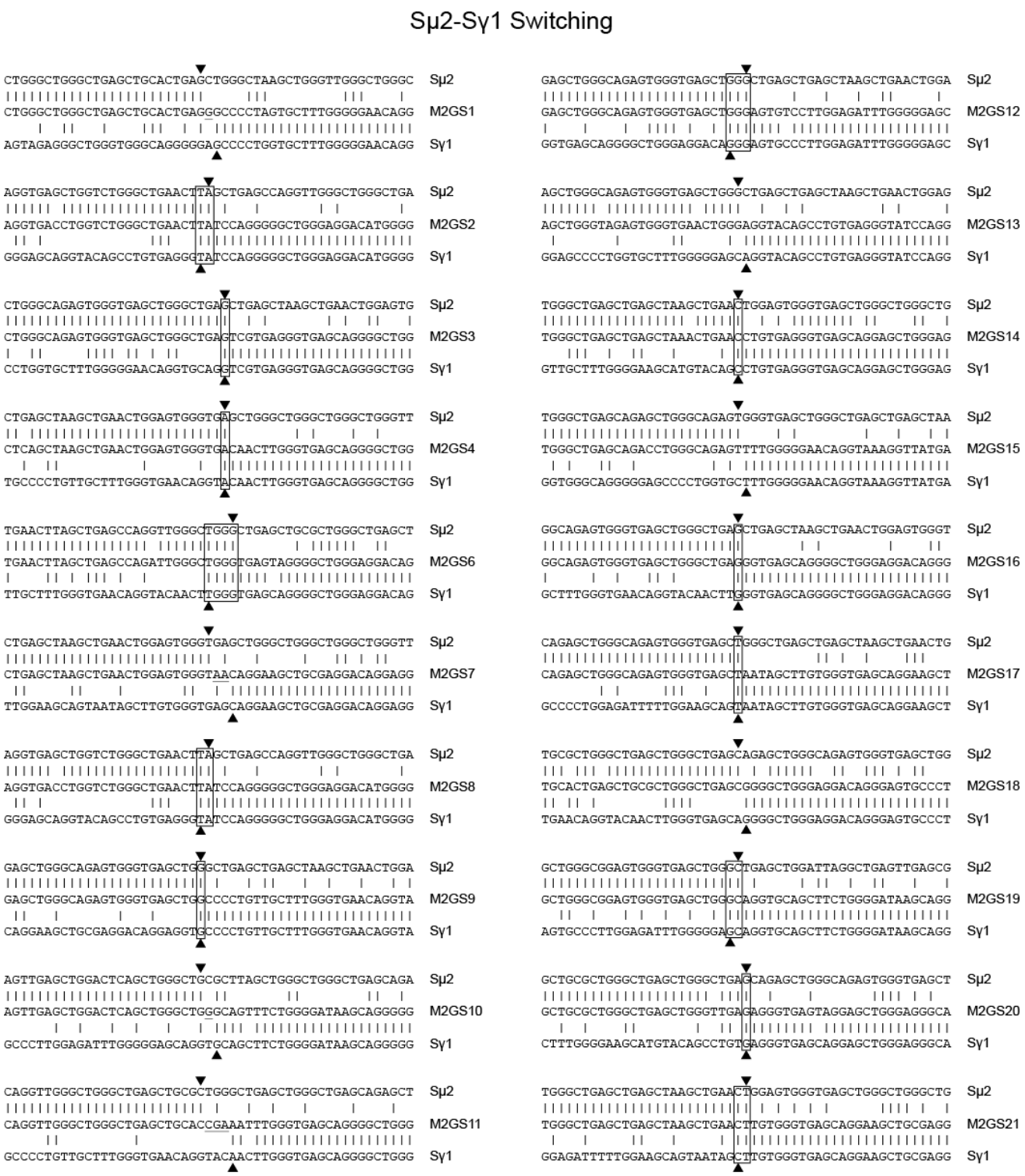
IGHD gene segments of the Holstein cow. Recombination signal sequences are indicated as capital letters, and the amino acid sequences translated in three reading frames are shown below each IGHD segment.

Sμ1-Sμ2 Switching



Sμ1-Sγ1 Switching

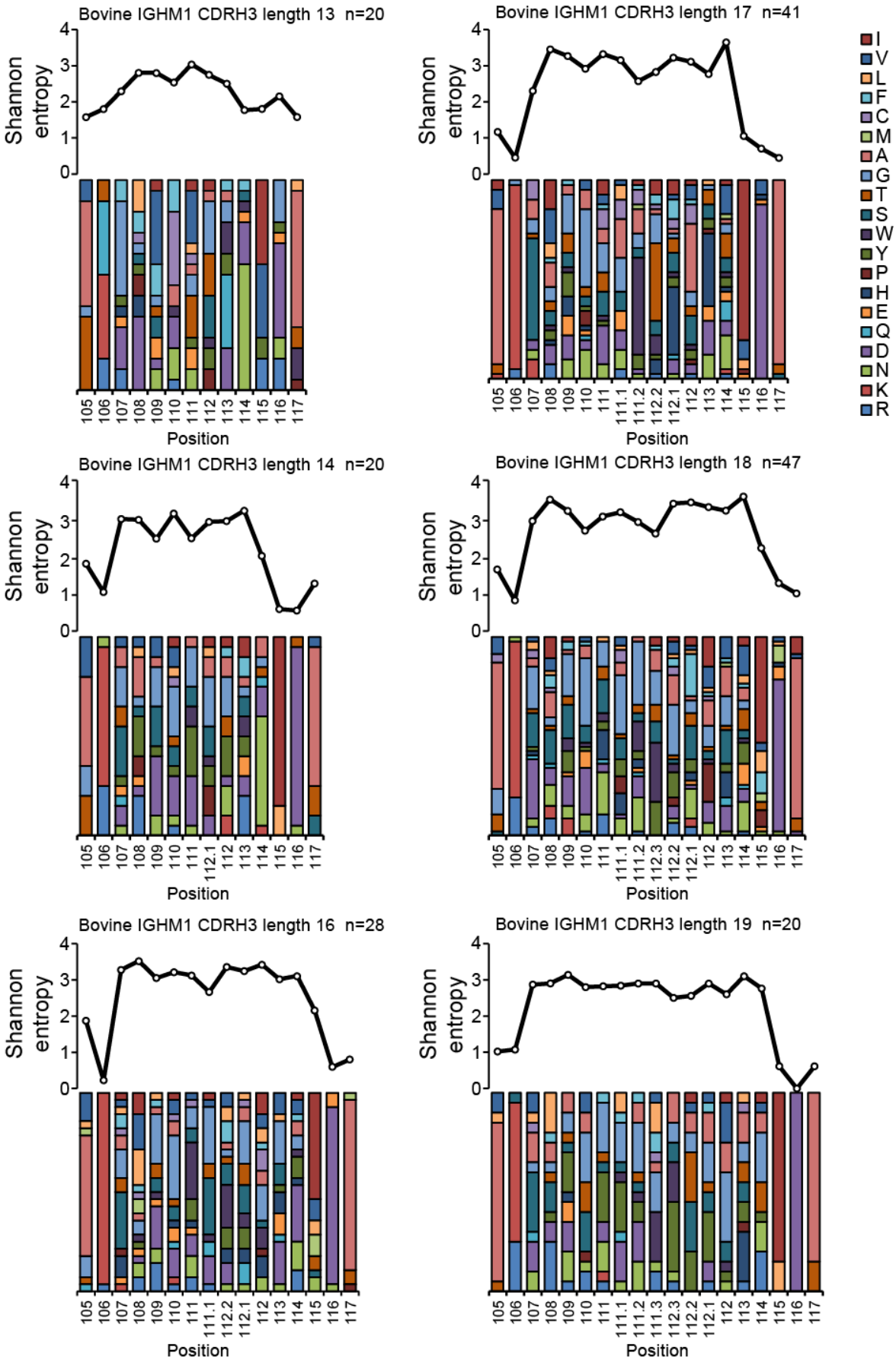




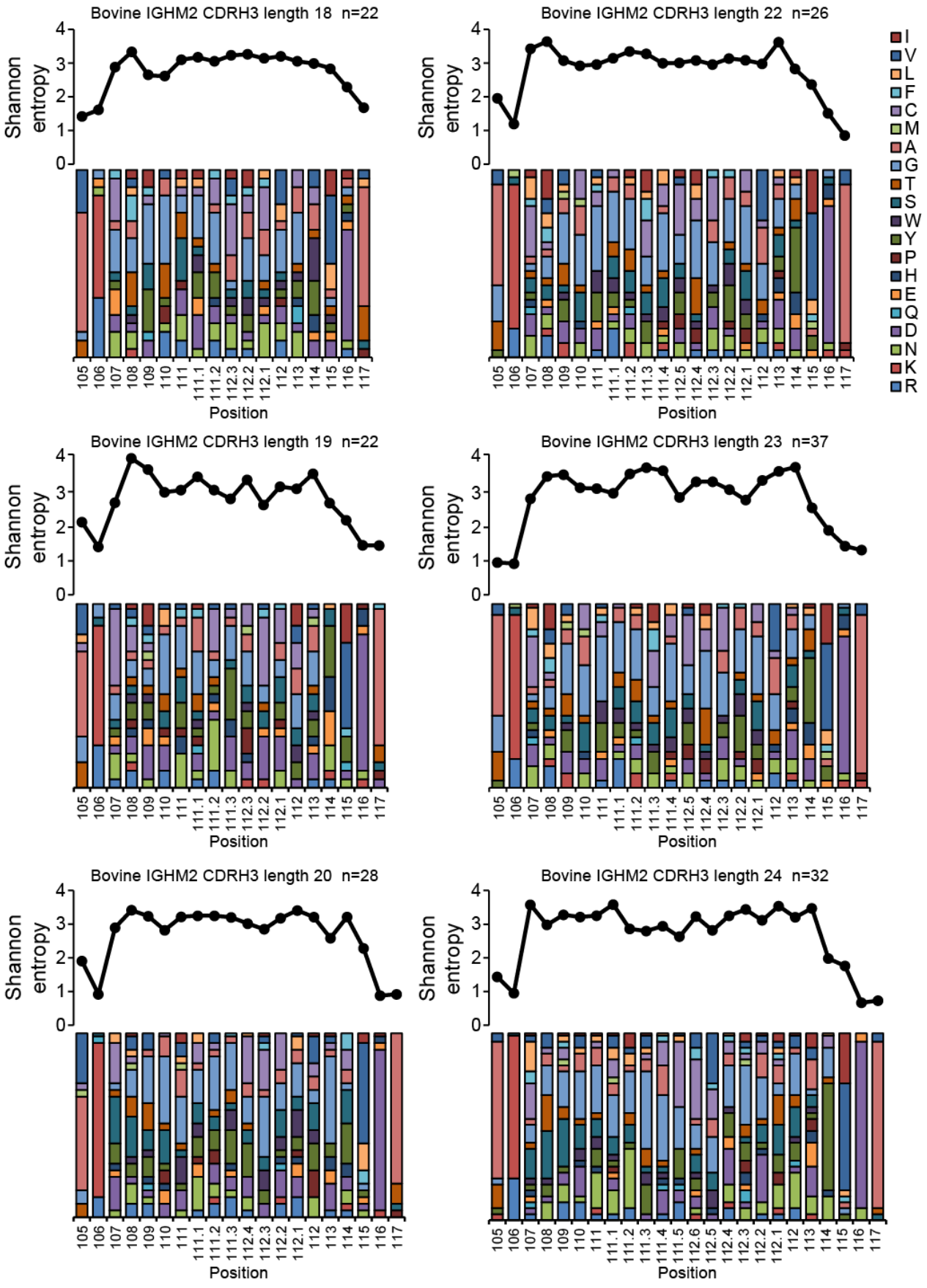
Supplemental Figure 2 (related to Figure 4B)

Analysis of the junction sites of the cloned recombined switch fragments. In each alignment, the middle sequence is the cloned PCR product, and the upper and lower sequences are the switch regions involved in the recombination. The triangles indicate the breakpoints of the switch regions, and the boxes show the overlaps.

A


Bovine IGHM1 CDRH3 length 13 n=20
Bovine IGHM1 CDRH3 length 17 n=41
Bovine IGHM1 CDRH3 length 14 n=20
Bovine IGHM1 CDRH3 length 18 n=47
Bovine IGHM1 CDRH3 length 16 n=28
Bovine IGHM1 CDRH3 length 19 n=20
Shannon entropy
Position
I V L F C M A G T S W Y P H E Q D N K R

B


Bovine IGHM2 CDRH3 length 18 n=22
Bovine IGHM2 CDRH3 length 22 n=26
Bovine IGHM2 CDRH3 length 19 n=22
Bovine IGHM2 CDRH3 length 23 n=37
Bovine IGHM2 CDRH3 length 20 n=28
Bovine IGHM2 CDRH3 length 24 n=32
Shannon entropy
Position
I V L F C M A G T S W Y P H E Q D N K R

Supplemental Figure 3 (related to Figure 8A)

CDRH3 Shannon entropy analysis (upper panel) and amino acid frequencies (lower panel) at positions 105-117 for (A) bovine IGHM1 and (B) IGHM2 of representative lengths**.**

A

	FR2	CDR2	FR3	CDR3	length
90A_66	VSWVRQAPGKALECLGG	INSAGST	YYNPALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	GKSSTRGSLCYYDDAW	15
90A_1	VTWVRQAPGKALECLGG	ISGEGGT	IYNPALKSRLTITKDNSRSQVSLSVSNVTPEDTATYYC	ARTYASGNWHCSDVDTW	16
A2B_30	VNWVRQAPGKALECLGG	INGDGYT	GYNSALKSRLSITKDSSKSQVSLSVSSVTPEDTATYYC	GKSAWTYGATGCNHVEAW	17
90A_21	VGWVRQAPGKALECLGG	ISAGGST	TYNPALKSRLSITKDNSKSQVSLSVNSVTTEDTATYYC	TKGVGTDYNNACDGWIDSW	18
A2C_67	VSWLRQAPGKALECLGG	IGSGGST	VYNRALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	VKSVIDHPCYGCCSDHIDAW	19
A2B_49	VGWVRQAPGKALECLGG	VSGGGST	TYNPVLTSRLSITKDNSKSEVSLSLSSVTSDDTATYYC	VRCWNHACNERCWRVDYVDTW	20
90A_27	VAWVRQSPGKAPECLGG	ISSDGST	GYNPTLKSRLGITKDNSKSQVSLSVSSIAAEDTATYYC	AKDAWTGNDGGTYHCGGNLDAW	21
90A_36	VGWVRQAPGKALECLGG	VRNSGTI	GYNPGLKSRLSVTKDDSKNQVSLSINGVTTEDTATYYC	VKSIDNRGWSCSCGFCAGHVDAW	22
A2B_7	VSWIRQAPGKALECLGG	INGGGST	NYNPALKSRLSITKDNSKSQVSLSVSSVTPEDMATYYC	GRDLDRTGCMCCDFDYSNEYVVTW	23
A2B_79	VSWVRQAPGKALECLGG	ANSGGST	VYNPALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	AKQNCGNVNTGGCVCPGIGPNVDAW	24
90A_47	VSWVRQTPGKALECLSS	ISKDGTT	YYNPALKSRLSITKDNSKSQVSLSVNSVTPEDTATYYC	AKDSGAGSASRGCRDWGRTYSNFDTW	25
A2B_37	VSWVRQAPGKALECLGG	ITGGGDT	DYNPALKSRLSITKDNSQSQVSLLVSSVTPEDTATYYC	AKATYGGAAGYGYDCYVLGSRNYIDAW	26
90A_44	VSWVRQAPGKALECLGA	ISNGGRI	YYNPALKSRLSITKDNSKSQVSLSVSSVTPDDTATYYC	AKGGVETYGSICYSDTGSGIDFTYVDAW	27
90A_82	VGWVRQAPGKALECLGS	ISRRGRT	GYNPALKSRLSITKDNSENRVSLSVGSVTTEDTATYYC	AKLNDAGDNVGSYGCYDYDYSQYEYEDAW	28
90A_14	VDWVRQAPGKALECLGG	ITSGGKS	GYNPALKSRLSITKDNSKSQVSLSVSSVSPEDTATYYC	AKTTYTEDDVHGDGCHGFGPVTDRSYVDAW	29
A2C_87	VNWVRQAPGKTLECLGG	ISSGGTT	GISPTLKSRLSITKDNSKSQVSLSLSSVTPEDTATYYC	VKMSSASGLDRYGDYGCLGLGGGGSHWVDAW	30

B

	FR2	CDR2	FR3	CDR3	length
SP90VA_33	VGWVRQASGKTLEWVGG	IRSSGTT	CLNPALKSRLSITKDNSKSQVSLSISSVTPEDTATYYC	VRIRTADAGCGYDVDAW	16
A2B_42	VGWIRQAPGKALEWIGG	HHSGGGT	CLNSALKSRLSITKDNSKSQVSLSVSSVTNEDTATYYC	AKYRFGGTCGTDFGIDAW	17
A2B_24	VHWVRQAPGKALEWLGG	INGGGVP	CLGPTLKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	AKNAGGRDYNCAGNHVDTW	18
90A_23	VAWVRQAPGKALEWVGG	ISGGGKT	CLNAALKSRLSITKENSKSQVSLLVNSVTTEDTATYYC	VKDFTGSDNSWVCGYKYDAW	19
90A_51	VGWVRQAPGKALEWVGG	IKRSGST	CLQPALKSRLSITKDNSKSQVSLSVSMVTPEDTATYYC	VKSYRGDYGDCSWSSINVDTW	20
90A_62	VGWVRQAPGKALEWVGS	IDNYGIT	CLNAALKSRLTITKDNSKSQVSLSVRSVTPEDTATYYC	ARPGTRYYTGGVYCTGADYEAW	21
A2C_59	VGWFRQAPGKALEWLST	MKFDGNT	CHNPALKARLSITKDNSKSQVSLSLSSVTPEDTATYYC	AKAVATGTDCWGPRGGVGNVDAW	22
90A_71	VGWVRQAPGKALEWIGT	ITRTGST	CLNPALKSRLGITKDNSKSQVSLSVSSVTTEDTATYYC	AKLYNSNTYEYGCNRGRSFYVGTW	23
A2B_33	VGWVRQAPGKALEWVGG	IGSGGGI	CLNPALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	ARRVGGGSRDCYDSRIVYGPNIDAW	24
VA_50	VGWVRQAPGKALEWIGF	IGTDGRT	CLNSALKSRLSITKDNSKSQVSLSVSSVTPEDTATYYC	AKGVCGGCWYGSYCAEGHGYGHIDAW	25
A2B_51	VGWVRQAPGKALEWIGS	VHRSGTT	CLNPALKSRLSITKDNSKSQVSLSVKGVTFDDTATYFC	AKNSGSGYSGGGCYGYSGGGGEYVDTW	26
A2C_58	IGWVRQAPGKALEWVGD	IDHNGET	CYNPALKSRLSITTDIHQSQVSLSLSSVTIEDAATYYC	AKFYGTIRFGSAACWTDIRPYVTYVDAW	27
SP90VA_11	VGWVRQAPGKALEWVGG	VGSDGST	CLNPALKSRLSITKDNSKSQVYLSVSSVTNDDTATYYC	AKGSHGGVRDGNDGCLDYLYGWNQYVDAW	28

Supplemental Figure 4

Representative sequences with an odd number of Cys residues in various lengths of CDRH3s accompanied by a germline-encoded Cys in (A) position 52 of FR2 or (B) position 66 of FR3. The boundaries of CDRH3 are highlighted in cyan, and the Cys residues in position 52 of FR2 or position 66 of FR3 are colored red, and the noncanonical Cys residues in CDRH3 are highlighted in yellow.