



Organization and Complexity of the Yak (*Bos Grunniens*) Immunoglobulin Loci

Mingli Wu¹, Haidong Zhao¹, Xiaoqin Tang¹, Wanxia Zhao¹, Xiaohua Yi¹, Qi Li¹ and Xiuzhu Sun^{1,2*}

¹ College of Animal Science and Technology, Northwest A&F University, Yangling, China, ² College of Grassland Agriculture, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:

Robert David Miller,
University of New Mexico,
United States

Reviewed by:

Brad Magor,
University of Alberta, Canada
Michael F Criscitiello,
Texas A&M University, United States

*Correspondence:

Xiuzhu Sun
sunxiuzhu@nwfau.edu.cn

Specialty section:

This article was submitted to
Comparative Immunology,
a section of the journal
Frontiers in Immunology

Received: 23 February 2022

Accepted: 11 April 2022

Published: 09 May 2022

Citation:

Wu M, Zhao H, Tang X, Zhao W, Yi X,
Li Q and Sun X (2022) Organization
and Complexity of the Yak (*Bos
Grunniens*) Immunoglobulin Loci.
Front. Immunol. 13:876509.
doi: 10.3389/fimmu.2022.876509

As important livestock in Qinghai-Tibet Plateau, yak provides meat and other necessities for Tibetans living. Plateau yak has resistance to diseases and stress, yet is nearly unknown in the structure and expression mechanism of yak immunoglobulin loci. Based on the published immunoglobulin genes of bovids (cattle, sheep and goat), the genomic organization of the yak immunoglobulin heavy chain (IgH) and immunoglobulin light chain (IgL) were described. The assemblage diversity of IgH, Ig λ and Ig κ in yak was similar to that in bovids, and contributes little to the antibody lineage compared with that in humans and mice. Somatic hypermutation (SHM) had a greater effect on immunoglobulin diversity in yak than in goat and sheep, and in addition to the complementarity-determining region (CDR), some loci in the framework region (FR) also showed high frequency mutations. CDR3 diversity showed that immunological lineages in yak were overwhelmingly generated through linkage diversity in IgH rearrangements. The emergence of new high-throughput sequencing technologies and the yak whole genome (2019) publication have greatly improved our understanding of the immune response in yaks. We had a more comprehensive analysis of yak immunoglobulin expression diversity by PE300, which avoided the disadvantage of missing low-frequency recombination in traditional Sanger sequencing. In summary, we described the schematic structure of the genomic organization of yak IgH loci and IgL loci. The analysis of immunoglobulin expression diversity showed that yak made up for the deficiency of V(D)J recombinant diversity by junctional diversity and CDR3 diversity. In addition, yak, like cattle, also had the same ultra-long IgH CDR3 (CDR3H), which provided more contribution to the diverse expression of yak immunoglobulin. These findings might provide a theoretical basis for disease resistance breeding and vaccine development in yak.

Keywords: yak, IgH, Ig λ , Ig κ , structure of the genomic, diversity of expression

1 INTRODUCTION

Humoral immunity is an immune mechanism by which B lymphocytes produce antibodies to achieve the purpose of protection. The key process of humoral immunity is the production of highly effective and short-lived effector B lymphocytes, which secrete antibodies to remove antigens. B lymphocytes, also called “B cells” could generate nearly an infinite number of antibodies with diverse binding sites through combinatorial rearranging of large numbers of duplicated genes that encode antibody specificity, by modifying the joining sites using exonucleases and terminal deoxynucleotidyl transferase (TdT) and then targeting these rearrangements for SHM using nucleases like activation-induced cytidine deaminase (AID) (1). There was a specific B cell receptor (BCR) on the surface of each B cell, which directly binds the antigen to humoral immunity. When an antigen invades the body, only B cells with complementary BCR can combine with the antigen, then receive selective stimulation, and then clone on a large scale (clonal selection theory) (2). Classical immunoglobulin has a Y-shaped structure which consists of two identical IgL and two identical IgH (3). The variable region of IgH and Ig κ /Ig λ constitute the epitopes of immunoglobulin, which directly determines the specificity and affinity of the antibody to antigen by binding with antigen. The constant region mainly regulates immune cells and effector molecules or cells. The physicochemical and structural properties of CDRs between Ig κ and Ig λ were quite different, but there was no significant difference in their pairing with IgH (4).

In response to a variety of pathogenic microorganisms in the environment, the body has evolved a variety of mechanisms, such as V(D)J recombination, gene conversion (GCV), SHM and class switch recombination (CSR), resulting in immunoglobulin diversity (5). In addition to CSR targeting constant regions, the other four mechanisms enrich the types of antibodies by increasing the diversity of variable regions. Recombination-activating genes 1/2 (RAG 1/2) and high-mobility group (HMG) protein initiate V(D)J recombination, specifically recognize recombination signal sequence (RSS) on the flank of the recombinant gene and form RAG complex, and then cut the RSS, repair the DNA double strand break (DSB) and complete rearrangement (6–9).

The schematic structure of the genomic organization and diversity of immunoglobulin in some ruminants had been reported. In 2016, Dr. Ma completed the genome organization structure of bovine IgH. The IgH consists of 43 VH genes, 23 DH genes and 12 JH genes. The bovine IgH had an ultra-long CDR3H not found in other species, encoding up to 65 amino acids (AA) (10, 11). Ultra-long CDR3H had previously been found in IgNAR of sharks and was thought to exist only in cartilaginous fish. But then longer ultra-long CDR3H was found in cattle, which presents a unique ‘stalk and knob’ structure in the crystal of ultra-long CDR3H antibodies. Different paired combinations of cysteine (Cys) in CDR3H form disulfide bonds, thus forming a variety of different spatial configurations of antigen binding sites under the same CDR3H sequence (12, 13). Goat IgH contains eight VH genes, three DH genes

and six JH genes, and only three VH genes and three DH genes and two JH genes were involved in the VDJ recombination. The composition of IgH was similar between sheep and goats, but the number of germ line genes is much less than that of cattle. In bovids, due to the low level of κ gene expression in cattle, there were few studies on the κ chain in cattle, Dr. Qin and Dr. Yu completed the study on the structure and expression diversity of immunoglobulin κ gene in sheep and goat in 2015 and 2020, respectively (14, 15). The sheep λ chain consists of 32 V λ genes, three J λ genes and two λ genes, and only seven V λ genes have potential functions. The goat λ chain consists of 35 V λ genes, three J λ genes and three λ genes, just like sheep, there are only seven V λ genes with potential function. The number of V, D and J genes germline genes in cattle was slightly higher than that in sheep and goats, but far lower than that in humans and mice.

Known as the boat of the plateau, yak live on the Qinghai-Tibet plateau at an altitude of 3 km above sea level and showed satisfactory cold and disease resistance. Therefore, the study of the structure and diversity mechanism of the yak immunoglobulin gene could be of academic and practical significance. Although there was a breakthrough in the structure of immunoglobulin in cattle and sheep, the insufficient basic knowledge of yak immune function hindered our ability to breed and maintain optimal yak health. Therefore, the study of the immunoglobulin gene map and diversity mechanism of yak is of great significance for disease resistance breeding, disease prevention and vaccine production of yak.

2 MATERIALS AND METHODS

All experimental procedures were carried out accordance according to the Regulations on the Administration of Affairs Concerning Experimental Animals approved by the State Council of the People’s Republic of China. The research was approved by the Institutional Animal Care and Use Committee of Northwest A&F University.

2.1 Analysis of the yak VH, V λ and V κ Segments

The yak genome (*Bos grunniens*, assembly BosGru3.0) was provided by NCBI (<https://www.ncbi.nlm.nih.gov/genome/?term=yak>). Immunoglobulin V gene was searched in yak genome with NCBI BLAST (Basic Local Alignment Search Tool) using the immunoglobulin V genes of cattle, goat, sheep, human and mouse as templates (16). The same method was used to obtain the immunoglobulin constant region genes in yak. FUZZNUC (<http://embossgui.sourceforge.net/demo/fuzznuc.html>) was used to find RSS to locate the D genes and J genes. V gene domains (FRs or CDRs) and classifications (potential functional gene, open reading frame (ORF), and pseudo gene) were divided according to IMGT (<https://www.imgt.org/>) (17).

2.2 Sequence Computations

DNA sequence editing, alignments and comparisons were performed using the SnapGene (18).

2.3 Phylogenetic Analysis

The phylogenetic tree was constructed using Mega version 7.0 (www.mega.com/) (19). ClustalW was used to compare multiple DNA sequences of the constructed tree. Each VH, V λ and V κ subgroup was randomly selected as a sequence among functional genes. The VH, V λ and V κ sequences used in this study (excluding yak) were as follows: human (*Homo sapiens*): VH1, L22582; VH2, X62111; VH3, X92206; VH4, Z12367; VH5, M99686; VH6, X92224; VH7, L10057; V κ 1, V01577; V κ 2, X63403; V κ 3, X12686; V κ 4, Z00023; V κ 5, X02485; V κ 6, X63399; V κ 7, X12682; V λ 1, Z73654; V λ 2, X97462; V λ 3, X97464; V λ 4, X57828; V λ 5, Z73672; V λ 6, Z73673; V λ 7, X14614; V λ 8, Z73650; V λ 9, Z73675; V λ 10, Z73676; V λ 11, D86996 (4, 20, 21). Mouse (*Mus musculus*): VH1, M34982; VH2, J00502; VH3, K01569; VH4, AC079273; VH5, X00163; VH6, AC073590; VH7, X03253; VH8, AC079181; VH9, L14368; VH10, AF064444; VH11, AC073563; VH12, AC073590; VH13, X55935; VH14, X55934; VH15, AC090843; VH16, AC073563; V κ 1, AJ231198; V κ 2, J00562; V κ 3, Y15968; V κ 4, AJ231209; V κ 5, AJ235974; V κ 6, AJ235962; V κ 7, AF044198; V κ 8, AJ235948; V κ 9, V00804; V κ 10, AF029261; V κ 11, AJ231256; V κ 12, AJ235951; V κ 13, AJ231273; V κ 14, AJ231241; V κ 15, AJ231269; V κ 16, AJ235936; V κ 17, AJ231258; V κ 18, AJ235966; V κ 19, AJ235935; V λ 1, J00590; V λ 3, M34597; V λ 4, M94349 (22, 23). Sheep (*Ovis aries*): VH1, H6; VH2, H7; VH3, H203; VH4, H16; VH5, H37; VH6, H168; VH7, H242; VH8, H17; VH9, H23 is from Charlton's article; V κ 1, AF038133; V κ 2, AF038134; V κ 3, AF038137; V κ 4, AF038138; V λ 1, AF040908; V λ 2, AF040921; V λ 3, AF040914; V λ 4, AF040907; V λ 5, AF040906; V λ 6, AF038145 (14, 15, 24). Horse (*Equus caballus*): VH1, DQ125413; VH2, HM175886; VH3, HM176050; V κ 1, HM176113; V κ 9, HM176157; V κ 10, V κ 12, V κ 13, V κ 15, V κ 19, CM000391; V λ 1, KR190470; V λ 2, KR190471; V λ 3, KR190472; V λ 4, KR190473; V λ 8, KR190477; V λ 10, KR190479; V λ 14, KR190483; V λ 25, KR190494; V λ 81, KR190550; V λ 77, KR190546; V λ 20, KR190489 (25, 26). Cattle (*Bos taurus*): immunoglobulin rearranged VH, KT723008; V λ 1, HM596891; V λ 2, U32258; V λ 3, HM596897 (11, 27–29). Chinese alligator (*Alligator sinensis*): VH1–8, JQ479335 (30). Green anole (*Anolis carolinensis*): V κ 1, EU419646; V κ 2, EU419647; V κ 3, EU419651; V κ 24, EU419668; V κ 7, EU419655 (31). Chicken (*Gallus domesticus*): VH1, AB233003; V λ 1, AB061561 (32, 33). African clawed frog (*Xenopus laevis*): VH1, Y00380; VH2, M30025; VH3, M24675; VH4, M24676; VH5, M24677; VH6, M24678; VH7, M24679; VH8, M24680; VH9, M24681; V κ 1, L15570; type III V1, L76575; type III V2, L76577; type III V3, L76586; type III V4, L76579; type III V5, L76576 (34–36). zebrafish (*Danio rerio*): VH1, BX649502 (37). Common carp (*Cyprinus carpio*): Cpi1, AB091112; Cpi2, AB091113; Cpe1, AB073328; Cpf1, AB073332; Cpg1, AB073335; Cpc1, AB035728; Cpb4, AB035730 (38, 39). Nurse shark (*Ginglymostoma cirratum*): VH1, M92851; NS3, EF114759; NS4, GU109471; NS5, AY720853 (7, 40).

2.4 Animals, RNA Isolations

In this study, three healthy and unrelated yaks (3 - 4 years old) were selected. RNA from spleens was extracted using the traditional Trizol method (TaKaRa, Dalian) (41).

2.5 Cloning of the Expressed Yak IgH, Ig λ and Ig κ Chain Genes by 5' Race PCR and Sequencing

The SMARTer RACE 5'/3' Kit (Takara, Dalian) was used to amplify the IgH, Ig λ and Ig κ of yak. The specific primers for IgH chain were as follows: yak-IgH-F, 5'-AAGCAGTGGTATCAACGCAGAGT-3'; yak-IgH-R, 5'-AGCTCACGCAGGACACCAG-3'. The specific primers for Ig κ chain were as follows: yak-Ig κ -F, 5'-AAGCAGTGGTATCAACGCAGAGT-3'; yak-Ig κ -R, 5'-AGATGGATGGCTGAGCATCA-3'. The specific primers for Ig λ chain were as follows: yak-Ig λ -F, 5'-AAGCAGTGGTATCAACGCAGAGT-3'; yak-Ig λ -R, 5'-GTGACCGAGGGTGC GGACTT-3'. Since the length of DH genes were random, and PE300 sequencing requires that the fragment length should not exceed 550bp, IgH was sequenced by sanger sequencing, and Ig λ and Ig κ were sequenced by PE300. The IgH PCR product was cloned into pMD-19T vector (TaKaRa, Dalian) and the sequence was determined by sanger sequencing. 100 clones of each sample were sequenced. The PCR products of Ig λ and Ig κ were directly sequenced by PE300.

2.6 V(D)J Combinatorial Diversity Analysis, SHM Statistics and CDR3 Calculations

All items were clustered with germline V genes, and each germline V gene was a cluster. The clone (from FR1 to FR3) that most matched the germ line fragment was analyzed by VH expression and base substitution, and the variation sites and frequencies were calculated by Mega version 7.0 and Excel software.

Editing and alignment of nucleotide sequences were described in Section 2.4. Mega version 7.0 was used to determine the types of V(D)J gene in all cloned fragments and calculated the combinatorial diversity of all clones. According to the grouping of V gene by Mega version 7.0, the base substitution of clone sequence (FR1–FR3) was analyzed by germline V gene as a template, the nucleotide variation frequencies and types of all sites were calculated. CDR3 region refers to the sequence between 'YYC' and 'WGXX' ('FGXX' in IgL) (42). 'YYC', 'WGXX' and 'FGXX' are relatively conservative, but with a little variation. It is necessary to truncate the CDR3 region of the cloned fragment according to the actual conditions. The random deletion sequence and random insertion sequence (non-templated (N) nucleotides and palindromic (P) nucleotides) were calculated according to the germline V, D and J genes.

High-throughput sequencing and Sanger sequencing were completed by Shanghai Biotech Company. The general process of high-throughput sequencing is as follows: (1) Perform 5' RACE according to experimental requirements to recover PCR products and send samples to the sequencing department; (2) PCR products of the samples were tested for quality, and the library was prepared after qualified: Illumina bridge PCR compatible primers were introduced, and the DNA fragments were purified; (3) In order to obtain uniform long cluster effect and high-quality sequencing data, Qubit 3.0 fluorescence quantifier was used for library concentration determination and quality inspection. After passing the library, the library was sequenced on the computer, and PE300 double-terminal sequencing program was run to obtain 300 bp double-terminal

sequencing reads and splice them. (4) Conduct bioinformatics analysis after obtaining sequencing data spliced from the company: Reads were screened, and those without amplified primers (F: AAGCAGTGGTATCAACGCAGAGT, yak-IgH-R: AGCTCACGCAGGACACCAG, yak-Ig κ -R: AGATGGATGGC TGAGCATCA, yak-Ig λ -R: GTGACCGAGGGTGCGGACTT), and with sequencing length less than 400 bp were removed. After obtaining coarse screening data, the sequencing sequences were analyzed as follows; IMGT was used to compare the sequencing data with the reference V gene, D gene and J gene, and the sequences of the V gene and J gene could be compared at the same time for subsequent analysis; MEGA (version7.0) was used to divide all sequences into V gene, (D gene), J gene and junction fragment; The neighbor-joining tree was constructed by MEGA (version7.0) to identify the template V gene, D gene and J gene corresponding to all sequences. V gene is downloaded into Excel format after ClustaIW alignment for SHM calculation; Blast (version2.2.30) was used to find the alignment information of V, D and J sequences, and output N and P nucleotides to generate specific information; (5) Statistical classification of immunoglobulin expression spectrum data: The number of sequences using different V, (D) and J clans in each sample was counted without removing repeated reads, and the results were converted to the percentage of the number of sequences using each V clan in the total number of sequencing; CDR3 sequence of each read was counted, the length of each CDR3 fragment was calculated, and the CDR3 length distribution data of each sample was analyzed; The number of bases in each reads nucleotide and the distribution of N and P nucleotide in each individual were calculated; The SHM frequency was counted. The formula was used in this study: SHM frequency = the number of observed mutant bases/the total number of sequenced bases; SHM analysis, classify observable mutant bases, count the number of mutations from A, T, C and G to the other three bases, and convert the results into the percentage of the number of each mutation in observable mutant bases; (6) GraphPad Prism 6 was used for mapping and statistical analysis.

3 RESULTS

3.1 Structure of the Genomic Organization and Phylogenetic Analysis of Yak IgH

IgH spans about 475 kb on chromosome 17 in yak, including 42 VH genes, 55 DH genes, three JH genes, one μ gene, one δ gene, three γ genes, one ϵ gene and one α gene (Figure 1A). Only eight VH genes had potential functions, including leader exons, complete translation initiation codons, splice fragments and downstream RSSs, but did not include stop codon or frameshift mutations. VH2 gene was ORF because the 104th AA is not conserved Cys. Five VH genes structures were incomplete, and the remaining 27 VH genes were terminated prematurely (Figure 1B). VH genes were divided into three clans, and the potentially functional VH genes were mainly concentrated in clan II (Figure 1C). All three JH genes had no features of 'WG χ G' or 'TVSS' (Figure 1D). In addition, all 55 DH genes exist in at least one optional reading frame, indicating their potential capabilities. The homology of yak VH genes with human, mouse, cattle, sheep, horse, green anole lizard, chicken,

xenopus laevis, zebrafish and nurse shark were analyzed by MrBayes 3.1 software. The tree was constructed from FR1-FR3 nucleotide sequence. Figure 1E shows that the evolution of the three VH clans of yak is most closely related to the three VH clans of cattle, followed by the clan I, clan III and clan VII of sheep. The evolutionary relationship of yak constant region gene was similar to VH gene, and yak constant region genes had the highest similarity to cattle, followed by yak (Figure 1F).

3.2 Diversity of IgH

3.2.1 VDJ junctions in the expressed IgH chains

311 (n1 = 103, n2 = 107, n3 = 101) IgH sequences were cloned from three yak spleen samples. The VH, DH and JH genes of the rearranged sequence were grouped by germline genes with potential function through Mega version 7.0. Because there is no potential functional JH gene found in the yak genome, bovine JH genes were used as germline JH genes. Yak had the same preferences for VH, DH and JH, and all three individuals only use the VH clan II (Figure 2A). Due to the high similarity of the VH genes in VH clan II and the existence of SHM in VH gene, it is impossible to accurately determine which germline VH all clones belong to. The DH gene mainly expresses DH28 and DH34 (Figure 2B). JH gene was not expressed in three yak germline JH genes, among which the highest expression frequency of the bovine JH10 gene was 100%, 97.2%, 91%. Three clones expressed the bovine JH6 gene in individual 2, six clones expressed the bovine JH6 gene and one in individual 3. The bovine JH3 gene was expressed in six clones (Figure 2C).

3.2.2 CDR3H diversity

As shown in Figure 3A that the cloned sequences were aligned by germline VHs, DHs and JHs. There were few combinations of VDJ expressed, and most clones express the combination of 'VH clan II-DH28/34- bovine JH10'. It showed that the diversity of yak IgH combinations is single. All clones were grouped and aligned manually. Some sequences were randomly deleted at the 3' end of V gene and the 5' end of J gene. As shown in Figures 3B, C, the length of the deleted fragment at the 3' end of V gene was mainly 0 bp, 2 bp and 3 bp. The length of the 5' end deletion fragment of the JH gene was mainly concentrated on 0 bp, 3 bp and 6 bp. The sequence between the conserved AA 'YYC' of the VH gene and the conserved AA 'WG χ G' of the JH gene was called the CDR3H region, including the 3' end of VH gene, complete DH gene and the 5' end of JH gene. The CDR3H length of yak ranges from 5 to 43 AA (with an average of 19.28 AA) (Figure 3D).

3.2.3 SHM analysis of VH (FR1-FR3)

VH15 was used as a template for SHM analysis with all cloned sequences. The variation frequency of each point was shown in Figure 4A. The variation frequency of FR1 was the lowest among the five regions, and the high frequency variation region was concentrated in the CDR1-CDR2 region. Several high-frequency variation sites were found at the 3' end of FR3. According to the statistics of single base variation types of all loci, the base substitution between A and G was significantly higher than other base substitution forms (Figure 4B-E), accounting for

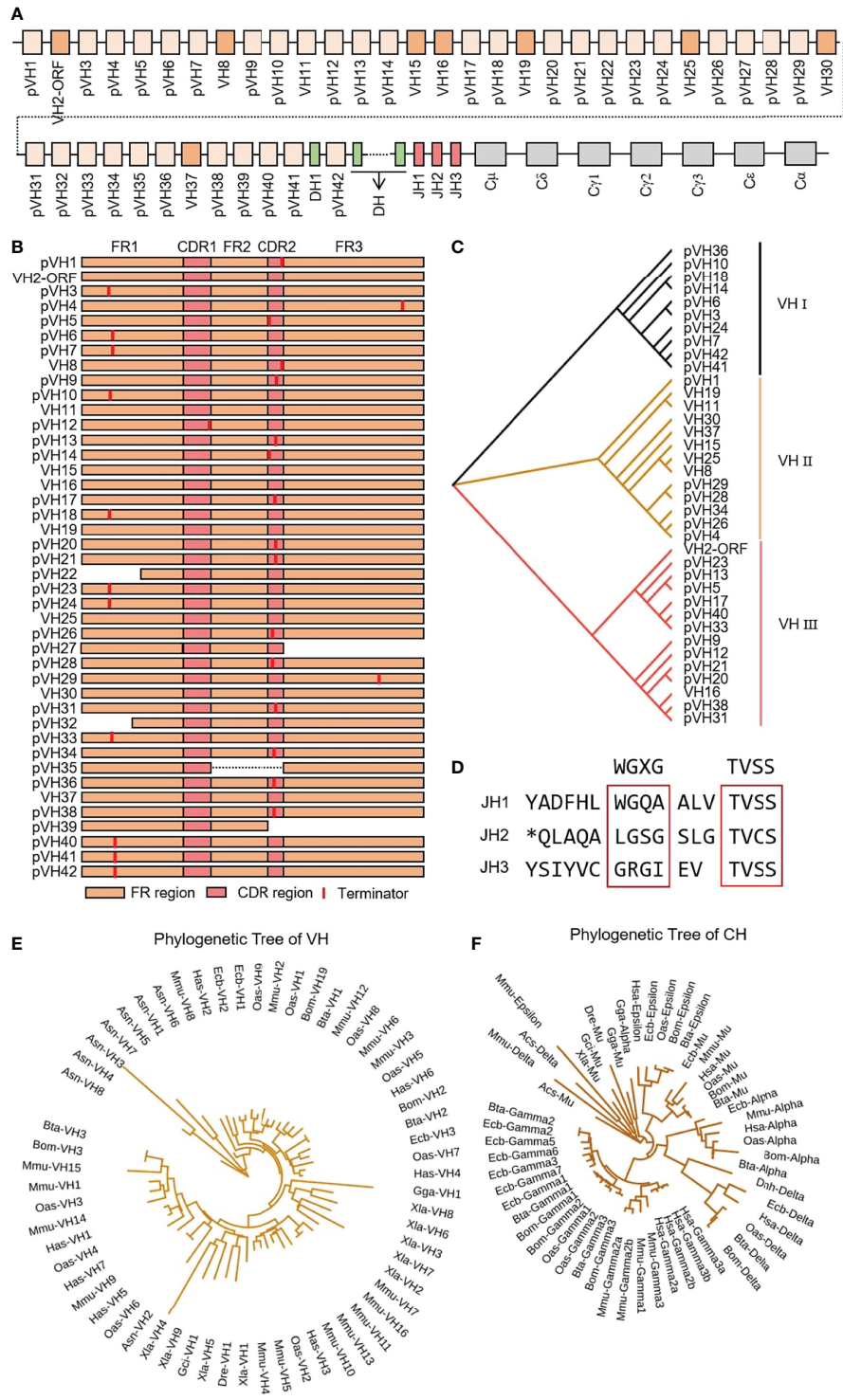


FIGURE 1 | Schematic structure of the genomic organization of Yak IgH and its relatedness with representative vertebrate species. **(A)** Schematic structure of the genomic organization of Yak IgH; **(B)** The structural variation of Yak VH; **(C)** The phylogenetic tree of yak VH; **(D)** The structural variation of Yak JH. *: termination codon **(E)** The phylogenetic tree of VH in yak and representative vertebrate species; **(F)** The phylogenetic tree of different CH gene family between yak and representative vertebrate species.

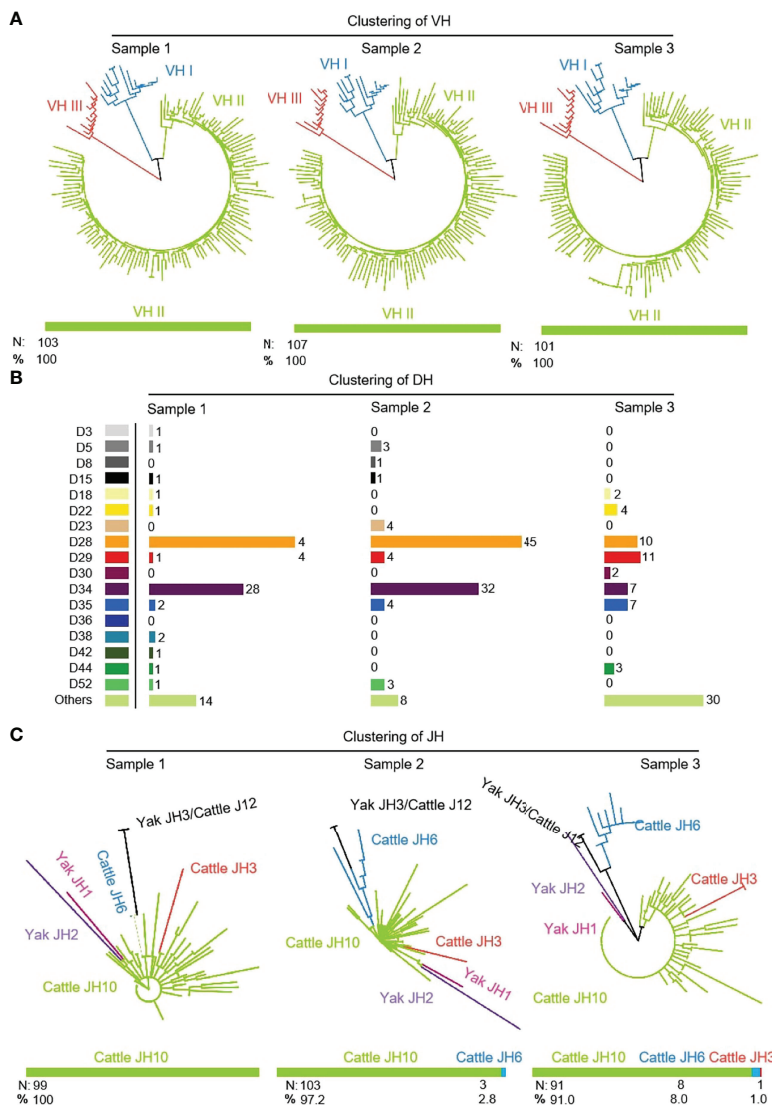


FIGURE 2 | The VH, DH and JH genes usage in Yak IgH. **(A)** Expression of the yak VH; **(B)** Expression of the yak DH; **(C)** Expression of the yak JH.

7.1%, 10.2% and 7.3% of the variation frequency of the three individuals respectively, which was consistent with the SHM model of cattle and sheep.

3.3 Structure of the Genomic Organization and Phylogenetic Analysis of Yak Ig κ

The yak Ig κ was located on chromosome 9 (GenBank: CM016698) with full-length 175 kb (Figure 5A). Yak Ig κ contained 23 V κ genes, of which nine had potential functions, the 23rd AA of both ORFs was not conserved Cys, five had large fragment deletions and the rest of the fragments were terminated prematurely (Figure 5B). According to the sequence similarity, all V κ s could be divided into three clans, which were similar to sheep V κ s (Figure 5C). Five J κ genes followed V κ s, which J κ 3 didn't contain a conserved 'FGXG' motif (Figure 5D). κ was located downstream of J κ , with a

conserved 23 bp nucleotide interval on the 5' end. MrBayes 3.1 software was used to analyze the clan of V κ gene. The tree was constructed from the nucleotide sequence FR1-FR3. k gene of yak was most closely related to cattle, followed by sheep and horse, which had the same relationship with IgH (Figure 5E). As shown in Figure 5F, yak had the highest homology with cattle. The V κ I of yak was clustered with V κ 3 of sheep, V κ 2 gene clan of yak was clustered with V κ 4 of sheep, and V κ 3 gene clan of yak was clustered with V κ 2 of sheep. The κ gene of yak had the highest homology with cattle, followed by sheep and horses (Figure 5F).

3.4 Diversity of Ig κ

3.4.1 V-J junctions in the expressed Ig κ chain

3939 (n1 = 1676, n2 = 1149 and n3 = 1114) recombinant Ig κ sequences were cloned from three yak spleen samples. Six kinds

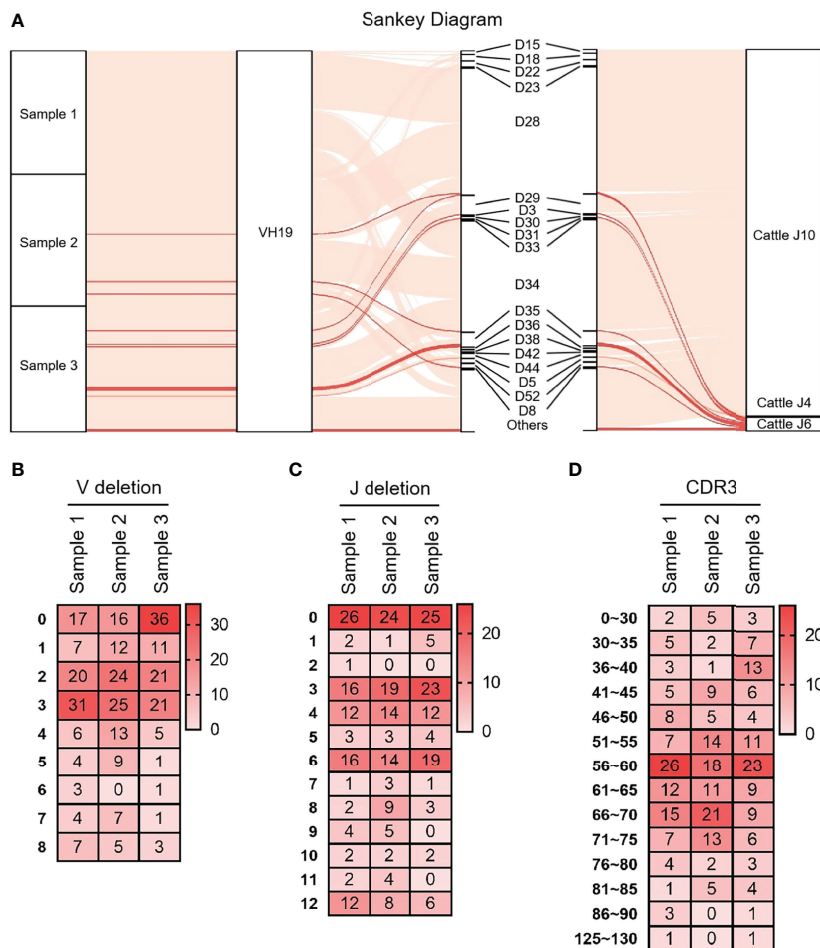


FIGURE 3 | The diversity of VDJ recombination and CDR3 length in yak IgH. **(A)** The sankey diagram of VDJ recombination in yak IgH; **(B)** the random deletion nucleotides of VH gene in yak IgH; **(C)** the random deletion nucleotides of JH gene in yak IgH; **(D)** the nucleotides of CDR3 in yak IgH.

of $V\kappa$ were expressed in all clones, namely $V\kappa 9$, $V\kappa 11$, $V\kappa 13$, $V\kappa 16$, $V\kappa 18$ and $V\kappa 19$, respectively. The expression rate of $V\kappa 11$ was 1.22% and 0.4% in individuals 2 and 3, respectively. The highest expression of $V\kappa 18$ was 80.6%, 45.4% and 66.2%, respectively (**Figure 6A**). The results of the $J\kappa$ cluster showed that $J\kappa 2$ was mainly used in three individuals, and the utilization rates were above 68%. The frequency of $J\kappa 4$ expression was second only to $J\kappa 2$. It was found that $J\kappa 1$ and $J\kappa 3$ were used in a few clones. The $J\kappa$ distribution of the three yaks was similar (**Figure 6B**).

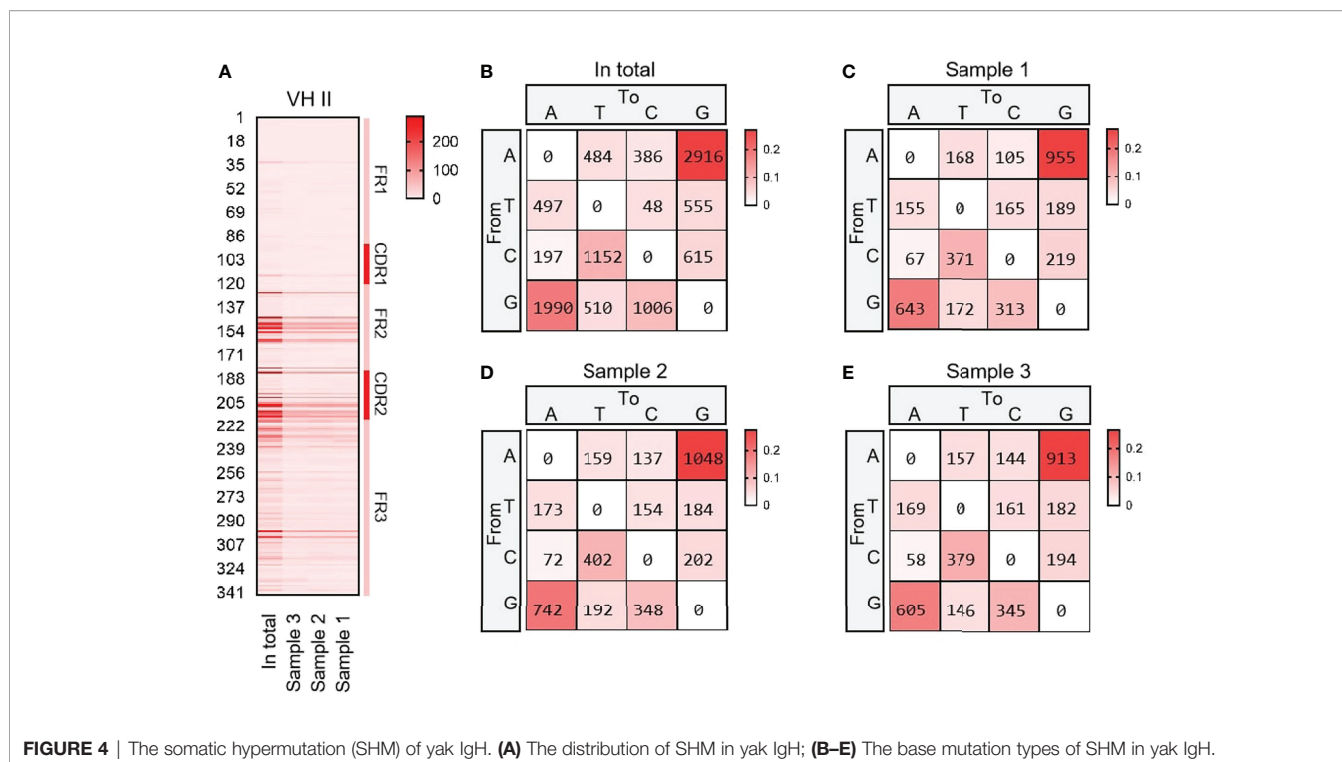
3.4.2 CDR3 κ diversity

As shown in **Figure 7A** that the cloned sequences were aligned by germline $V\kappa$ s and $J\kappa$ s. There were few recombinant sequences of expressed $V\kappa$ - $J\kappa$, and the combination ' $V\kappa 18$ - $J\kappa 2$ ' was expressed in most clones. These results showed that $Ig\kappa$ was biased in the process of recombination, but its diversity was higher than that of IgH . All clones were manually grouped and compared, and random deletions of the 3' end of $V\kappa$ gene and the 5' end of $J\kappa$ gene were counted. As shown in **Figure 7B**, the random deletion

length of 3' end of $V\kappa$ and the 5' end of $J\kappa$ was mainly 1-3 bp. Random addition of non-template sequences between the $V\kappa$ gene and $J\kappa$ gene, including N nucleotide and P nucleotide. Statistics showed that the insertion of N and P nucleotides mainly focus on 0 - 1 bp. The sequence between the conserved AA 'YYC' of $V\kappa$ and the conserved AA 'FGXG' of $J\kappa$ was called CDR3 κ region, which included the 3' end of $V\kappa$, the 5' end of $J\kappa$ and the random insertion sequence between $V\kappa$ and $J\kappa$. The length of CDR3 κ in yak was mainly 8 - 9 AA.

3.4.3 SHM analysis in the expressed $V\kappa$ region

As shown in **Figures 8A-F**, six $V\kappa$ genes were expressed, namely $V\kappa 9$, $V\kappa 11$, $V\kappa 13$, $V\kappa 16$, $V\kappa 18$ and $V\kappa 19$. All clones were grouped into six $V\kappa$ genes, and SHM analyzed were performed. The change frequency of each spot is shown in **Figures 8A-F**. The high-frequency changes were not completely concentrated in the CDRs, and high-frequency mutations also appeared at some sites in the FR2 region. According to the statistics of the types of base substitution at all sites, the frequency of base exchange between 'A' and 'G' was



the highest, followed by that between ‘C’ and ‘T’ (**Figures 8G–J**). This study found that the SHM pattern of yak was the same as that of humans, mice and sheep, which indicated that SHM was evolutionarily conservative.

Then, the frequency of C/G mutation in AID hotspot WRCY/DGYW was analyzed, that is, the frequency of C/G mutation in the hotspot and the frequency of the total mutant base. The mutation frequencies of IgH, Ig κ and Ig λ were highest when C mutated to T and G mutated to A (**Figure 9**).

3.5 Structure of the Genomic Organization and Phylogenetic Analysis of Yak Ig λ

Yak Ig λ was located on chromosome 16 (GenBank: CM016705) with full-length 1287 kb, including 45 V λ genes, nine J λ genes and eight λ genes. The genes from pV λ 1 to λ 7 were closely distributed, only 346 kb, and the downstream genes are far away and the transcription direction is opposite to the chromosome direction (**Figure 10A**). Based on the classification rules of the IMGT database, 18 of the 45 V λ genes had potential functions, four were ORFs, eight were structurally incomplete, and 15 were pseudo genes (**Figure 10B**). According to the homology of nucleic acid sequence, all V λ genes could be divided into five clans. The V λ genes with potential functions were mainly concentrated in clan I and clan II (**Figure 10C**). Nine J λ genes appeared upstream of eight λ genes. Unlike human, mouse and cattle, the J λ genes in yak did not exist in the form of (J λ - λ) $_n$ pairing, but randomly distributed in the upstream of λ . J λ 1, J λ 2, J λ 7, J λ 8 and J λ 9 did not contain conservative ‘FGXG’ (**Figure 10D**). Phylogenetic analysis of V λ genes and λ genes

in yak was carried out by Bayesian method, and the serial numbers of related genes were shown in 2.3. Yak λ was grouped with cattle, sheep and horse λ (**Figure 10E**). The tree was constructed by the nucleotide sequences FR1–FR3. A gene with potential functions was randomly selected from each V λ clan, and an evolutionary tree system was constructed. As shown in **Figure 10F**, a close relationship between yak and cattle, horse and sheep, was observed, and yak V λ clan I clustered with horse V λ 3, yak V λ clan II clustered with horse V λ 20, yak V λ clan III clustered with sheep V λ 2, yak V λ clan IV clustered with horse V λ 10, yak V λ clan V clustered with cattle V λ 2.

3.6 Diversity of Ig λ

3.6.1 V–J junctions in the expressed Ig λ chain

16464 ($n_1 = 5551$, $n_2 = 4916$, $n_3 = 5997$) Ig λ recombinant sequences were cloned from 3 yak spleen samples. All clones expressed 10 kinds of V λ , namely V λ 8, V λ 11, V λ 20, V λ 25, V λ 28, V λ 31, V λ 36, V λ 37, V λ 40 and V λ 41 respectively (**Figures 11A, B**). Among the three individuals, the expression of V λ 11 (yellow) was the highest, followed by V λ 20 (black). V λ 11 and V λ 20 belong to the clan V, and among the eight kinds of V λ genes with low expression, except for V λ 33 (blue, clan III), the expression level was less than 10%. All V λ genes were expressed in similar proportion in the three individuals. J λ 3 was expressed by all clones (**Figure 11C**).

3.6.2 CDR3 λ diversity

Ten V λ genes were expressed and one J λ was expressed, so there were 10 types of V λ -J λ recombination types in the clones, with

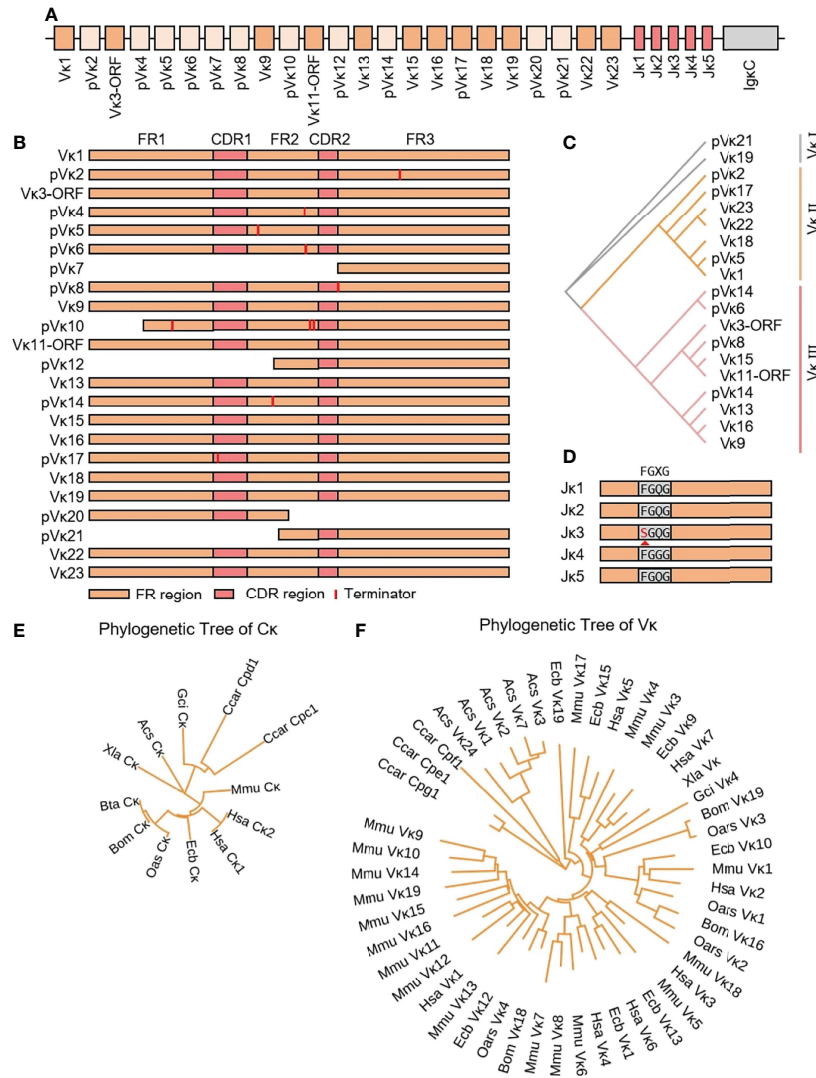


FIGURE 5 | Schematic structure of the genomic organization of Yak IgL(κ) and its relatedness with representative vertebrate species. **(A)**: Schematic structure of the genomic organization of Yak IgL(κ); **(B)** The structural variation of Yak V κ ; **(C)** The phylogenetic tree of yak V κ ; **(D)** The structural variation of Yak J κ ; **(E)** The phylogenetic tree of κ in yak and representative vertebrate species; **(F)** The phylogenetic tree of different V κ gene family between yak and representative vertebrate species.

the highest recombination probability of V λ 14-J λ 1 (**Figure 12A**). The recombination probability of V λ clan V was the highest, followed by V λ clan III and V λ clan I. All clones were manually grouped and compared, and the random deletions of V λ 3' end and J λ 5' end were counted. As shown in **Figure 12B**, the random deletion lengths of the 3' end of V λ and the 5' end of J λ were mainly 0 - 3 bp. The non-template sequences between V λ and J λ were randomly added, including N nucleotides and P nucleotides. The statistical results showed that the insertions of N and P nucleotide were mainly concentrated in 0 - 2 bp. The sequence between the conserved AA 'YYC' of V λ and the conserved AA 'FGXG' of J λ was called CDR3 λ region, which includes the 3' end of V λ , the 5' end of J λ and the random insertion sequence between V λ and J λ . The length of yak CDR3 λ

was mainly 10 - 11 AA, which was longer than that of CDR3 κ (**Figure 12B**).

3.6.3 SHM analysis in the expressed V λ region

10 V λ genes were expressed, and all clones were grouped according to 10 V λ genes, which were analyzed by SHM. The change frequency of each point is shown in **Figures 13A-J**. The high-frequency changes were not completely concentrated in the CDRs, and high-frequency mutations also appeared in FRs. According to the statistics of the base substitution types of all sites, the base exchange frequency between A and G was the highest (**Figures 13K-N**). In this study, it was found that the SHM pattern of yak Ig λ was the same as that of IgH and Ig κ .

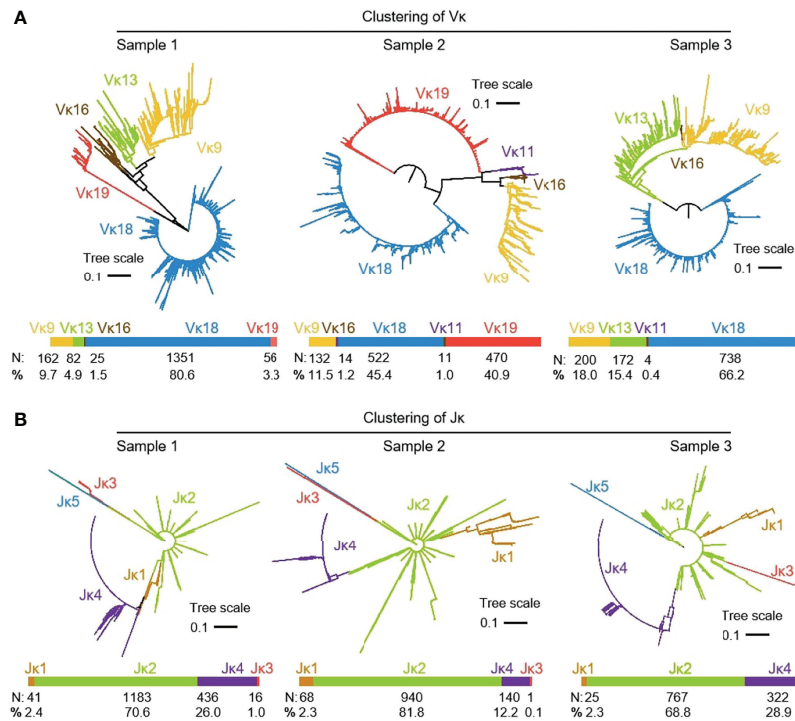


FIGURE 6 | The Vκ and Jκ genes usage in yak IgL (κ). **(A)** Expression of the yak Vκ; **(B)** Expression of the yak Jκ.

4 DISCUSSION

It is critical to understand the natural immune mechanisms that fight infection and how vaccination, biosafety, nutrition, livestock and management practices could be used to maintain and enhance immune protection (43). The sequence and structure of IgH constant genes in yak were similar to those in bovids (11). Interestingly, two μ genes with high similarity were found in the genome of cattle, but only one μ gene had been found in the genome of other bovids (yak, sheep and goat) (11, 24, 44, 45). Three JH genes were found in the yak genome, but none of them were transcribed. The three JH genes did not have typical 'WGXXG' and 'TVSS' AA, which might be pseudogenes (45). The results showed that more than 95% of recombinant sequences in yak expressed bovine JH10 gene and 4% of recombinant sequences expressed bovine JH6 gene. By comparing JH gene sequences of yak and cattle, it was found that yak JH1 gene was highly similar to bovine JH1 gene, yak JH2 gene was highly similar to bovine JH5 gene, and yak JH3 gene was similar to bovine JH12 gene. Therefore, we speculated that the sequence between bovine JH5-JH6- μ 1-...-JH7- JH8- JH9- JH10- JH11 might be missing in yak genome. The IgH span of yak was 475 kb, that of cattle was 678 kb, and that of JH5-JH6- μ 1-...-JH7-JH8-JH9-JH10-JH11 was about 200 kb, which was consistent with the difference between the length of cattle and yak IgH. In the study of VDJ recombination of cattle, JH10 was recombined with the μ 2 gene, and JH6 was recombined with μ 1 gene. For these reasons, some JH genes and one μ gene might be

missing in the yak genome. There was also a strong preference for yak DH genes in VDJ recombination. DH 28 and DH 34 were mainly used in VDJ recombination, which had obvious (YG)n AA characteristics, which was consistent with the characteristics of bovine DH8 gene (11). There were 42 germline VH genes in the yak genome, only 9 VH genes with potential function, all of which belong to clan II. The genes in clan II were about 90% homologous. Yak V gene contributed less to VDJ recombinant diversity, which was similar to horse. There were 54 species VH genes in horse, only VH4 and VH5 (clan II) were mainly used in VDJ recombination. **Figure 1E** showed that horse VH clan II and Yak VH clan II belong to the same branch (26). Since the VDJ recombination of bovine IgH was not reported. The three VH genes involved in VDJ recombination in goat belong to the clan I, but the VH genes used in goat were more abundant and less conservative than in other species, so the VH genes in goats contribute more to the diversity of VDJ recombination in bovids (24, 45). In summary, recombinant diversity of VDJ recombination in bovids contributes limited immunoglobulin diversity compared to human.

The previous researchers did not study the diversity of Igκ expression because the expression level of Igκ was extremely low. In this study, the Igκ of yak was amplified successfully, and the structure and diversity of Igκ were analyzed. The Igκ of yak had a typical structure of the genomic organization, containing only one κ gene, 5 Jκ genes and 23 Vκ genes. The κ genes were conserved in most mammals and were highly similar in sequence and number, but rabbits had two κ genes in particular. Yak,

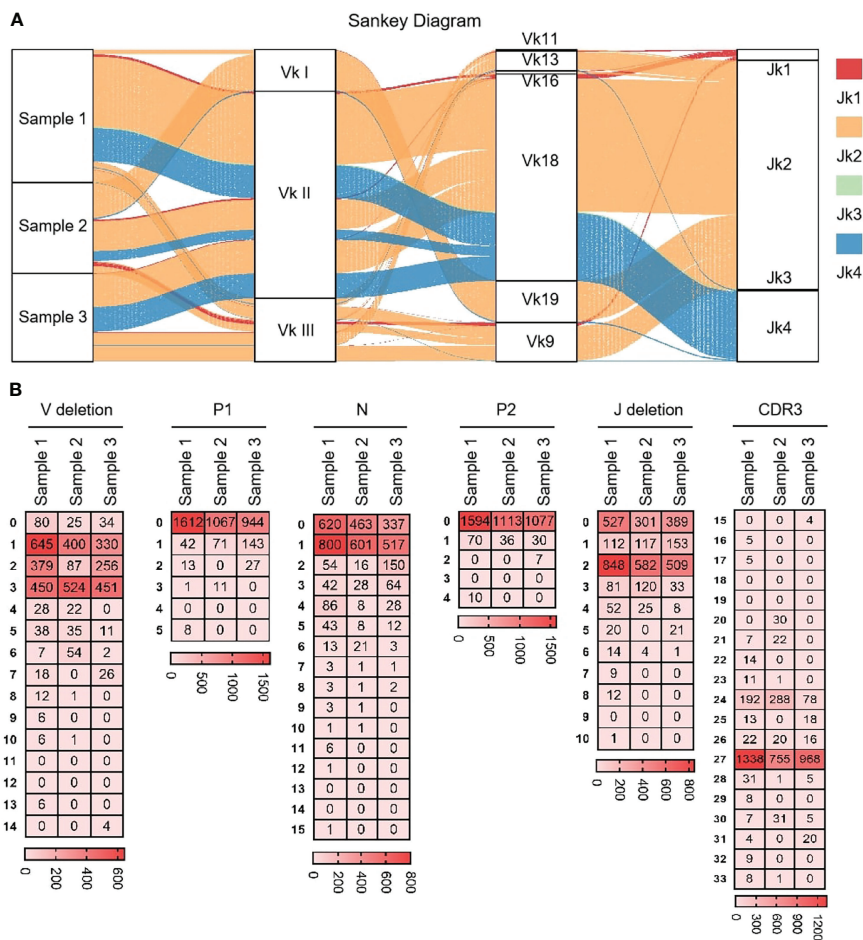


FIGURE 7 | The diversity of VJ recombination and CDR3 length in yak IgL(κ). **(A)** The sankey diagram of VJ recombination in yak IgL(κ); **(B)** the nucleotides distribution of V deletion, P1 nucleotide, N nucleotide, P2 nucleotide, J deletion and CDR3 in yak IgL(κ).

mouse, rabbit, horse and pig had five Jκ genes, while goat and sheep had only three Jκ genes and one was 19 bp irregular RSS (20, 25, 46, 47). Jκ1-4 was used in yak and horse VJ recombination and was similar in type and frequency of Jκ used (26). In this study, high-throughput sequencing found that six Vκ genes and four Jκ genes were used in VJ recombination, and Vκ16 and Jκ3 with extremely low frequency (less than 1%) were also found. The abundance of VJ recombination of Igκ in yak was higher than that in sheep and goats. Previous studies have found that Igλ of all species followed the rule of (Jλ-Cλ)_n. The structure of the genomic organization of Igλ was Vλn-(Jλ-Cλ)₇-Vλn in horse, Vλn-(Jλ-Cλ)₄ in cattle, Vλn-(Cλ-Jλ)₃-λ4 in pigs, Vn-(Jλ-Cλ)₂, in sheep and Vλn-(Jλ-Cλ)₃ in goat (15, 26). Interestingly, we found that yak Igλ did not follow such a rule, and the structure of the genomic organization of yak Igλ was Jλ1-Jλ2-λ1-λ2-Jλ3-Jλ4-Jλ5-λ3-Jλ6-λ4-Jλ7-λ5-Jλ8-λ6-Jλ9-λ7-λ8. 11 Vλ genes and four Jλ genes (Jλ1, Jλ4, Jλ6, Jλ7) were used in VJ recombination of Equine Igλ (25, 26). Cattle mainly expresses Vλ1a, Vλ1b, Jλ2 and Jλ3. Pig only Jλ2-λ2 and Jλ3-λ3 were used in recombination (47, 48). Only the combination of

Jλ1-λ1 was found in the recombination sequence for sheep (14). 11 Vλ genes and four Jλ genes were used for VJ recombination of Equine Igλ (25, 26). The bovine Vλ1a, Vλ1b, Jλ2 and Jλ3 gene were mainly expressed. Jλ2-Cλ2 and Jλ3-Cλ3 have been used for recombination in pigs. Only Jλ1-Cλ1 combinations were found in the recombination sequence of sheep. Six Vλ genes and four Jλ genes were involved in the VJ recombination of yak (27, 29). The abundance of yak VJ recombination was close to that of horse but higher than that of goat, sheep and cattle.

Yak IgH, Igλ and Igκ showed a significant preference for J gene selection when V(D)J was recombined, and the IgH also had obvious preference for DH genes. Studies have shown that a single DH gene fragment could form its own CDR3H pedigree, which was sufficient to meet the needs of B cell development and immune function. However, multiple DH gene fragments are still necessary for full immune function. This study also found that the retention and development of CDR3H in specific species has its own characteristics in terms of length and AA utilization. CDR3H includes 3' end of VH gene, DH gene, 5' end of JH gene, V-D and D-J junction sequences (N nucleotides and P

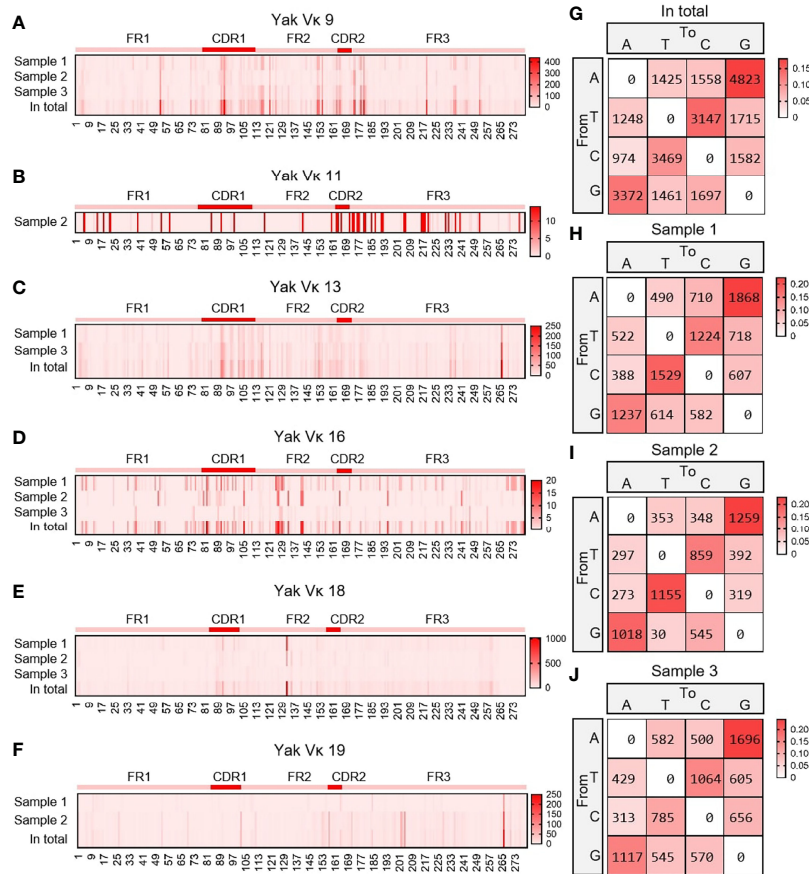


FIGURE 8 | The somatic hypermutation (SHM) of yak IgL(κ). (A–F) The distribution of SHM in yak IgL(κ); (G–J) The base mutation types of SHM in yak IgL(κ).

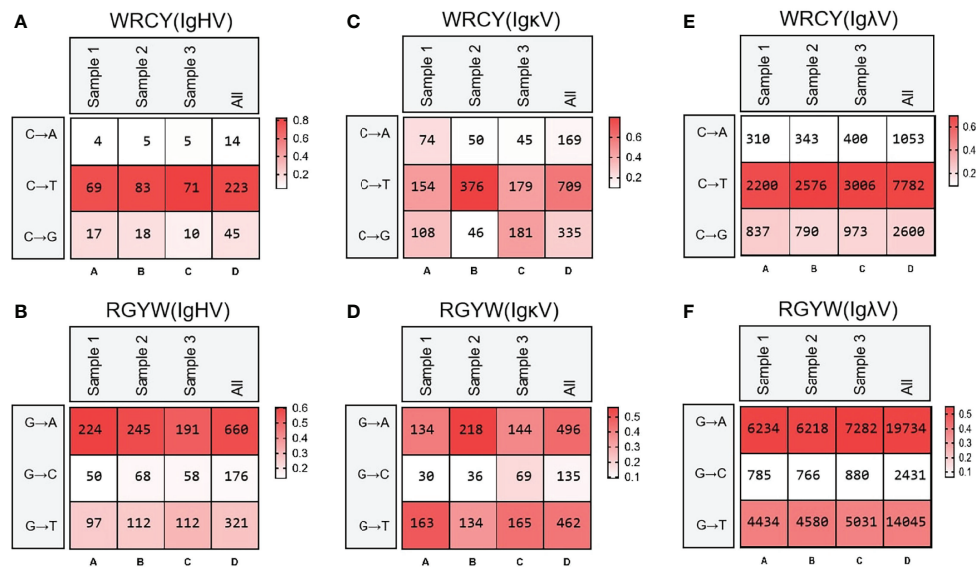


FIGURE 9 | Hot spot mutation frequency of SHM in IgH, Igκ, Igλ. (A) IgH SHM in “WRCY”; (B) IgH SHM in “RGYW”; (C) Igκ SHM in “WRCY”; (D) Igκ SHM in “RGYW”; (E) Igλ SHM in “WRCY”; (F) Igλ SHM in “RGYW”.

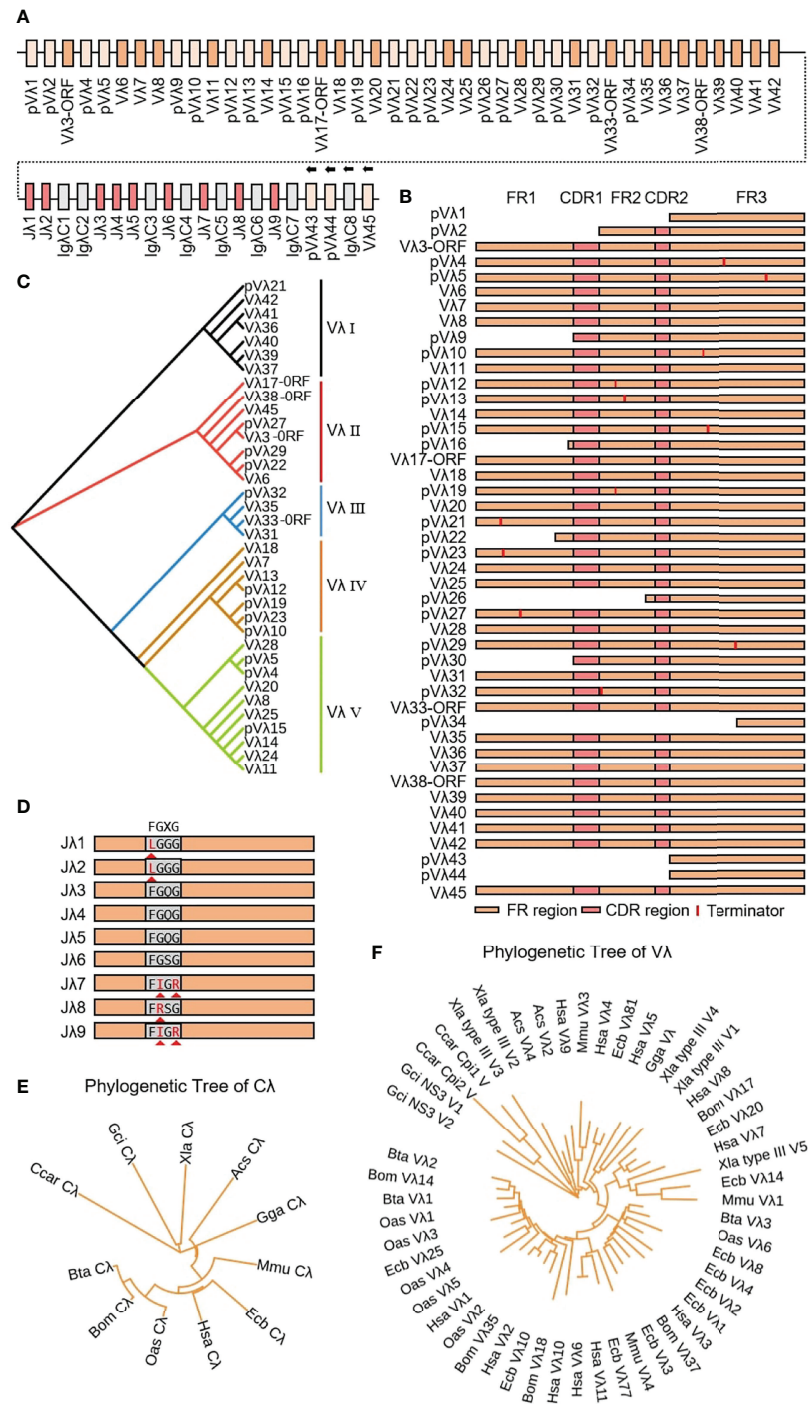


FIGURE 10 | Schematic structure of the genomic organization of Yak IgL(λ) and its relatedness with representative vertebrate species. **(A)** Schematic structure of the genomic organization of Yak IgL(λ); **(B)** The structural variation of Yak Vλ; **(C)** The phylogenetic tree of yak Vλ; **(D)** The structural variation of Yak Jλ; **(E)** The phylogenetic tree of Cλ in yak and representative vertebrate species; **(F)** The phylogenetic tree of different Vλ gene family between yak and representative vertebrate species.

nucleotides). The random deletion of the 3' end of V gene and the 5' end of J gene, and the random insertion of N and P nucleotides were counted in this study (49). The random deletion of V gene mainly concentrated in 0 - 3 bp, with the largest

deletion of 8 bp. The random deletion of J gene was mainly 0 bp, 3 bp, 4 bp and 6 bp, and the longest deletion was 12 bp. Compared with human (6.5 ± 1.7 bp), mouse (5.8 ± 1.7 bp) and cattle (8.8 ± 3.7 bp), the contribution of VH gene fragment to

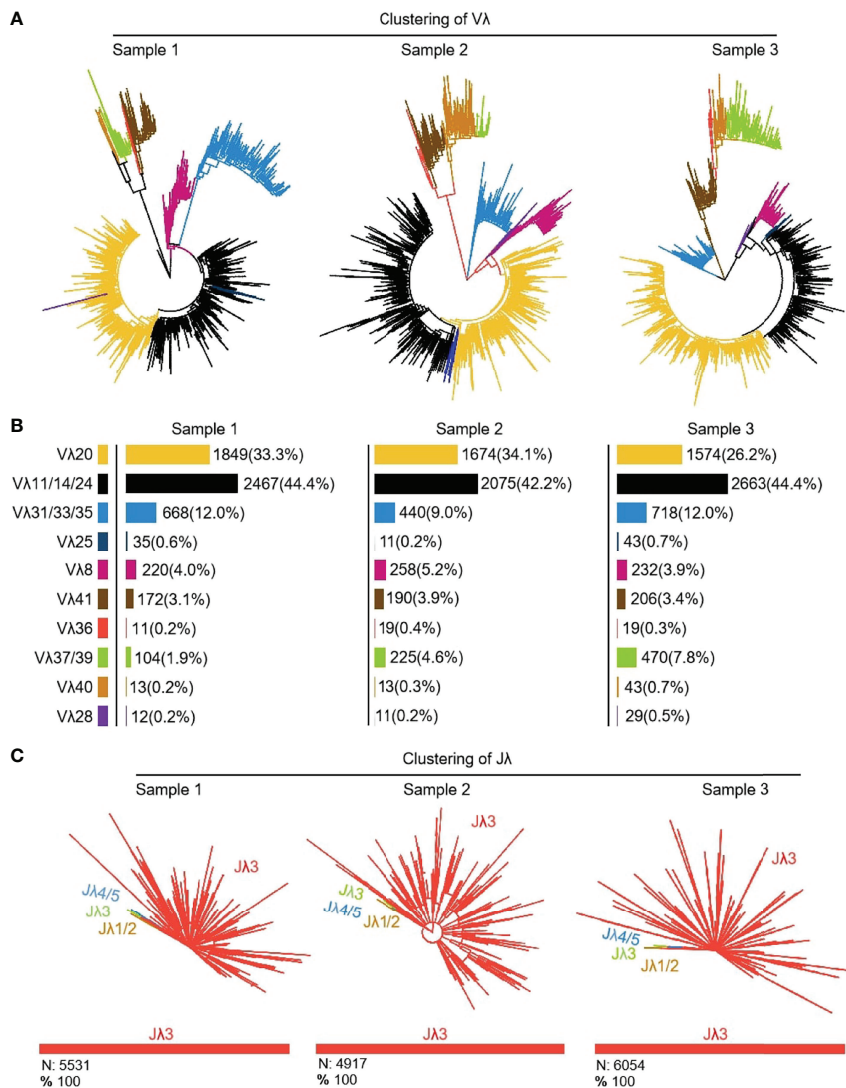


FIGURE 11 | The V λ and J λ genes usage in yak IgL (λ). **(A, B)** Expression of the yak V λ ; **(C)** Expression of the yak J λ .

CDR3H length was smaller in yak (5.5 ± 2.1 bp) (11, 20, 50, 51). The contribution of Yak JH gene to CDR3H length was 7.6 ± 3.7 bp, and that of cattle was 15.1 ± 4.0 bp. The most notable and significant difference is the contribution length of DH to CDR3H. The DH contribution length of human and mouse was 14.3 ± 5.5 bp and 10.8 ± 4.7 bp, respectively. The contribution of bovine $\mu 1$ and $\mu 2$ to CDR3H length was 27.5 ± 8.5 bp and 43.1 ± 25.3 bp, respectively (52). The length of DH contribution of yak was 39.9 ± 13.7 bp, which was between bovine $\mu 1$ and $\mu 2$. Previous studies have found that cattle have an ultra-long CDR3H, and the longest CDR3H reaching 65 AA, which is not available in other species. By cloning yak IgH, we found the average length of CDR3H in yak was 58.0 ± 15.3 bp. Unfortunately, the longest CDR3 fragment did not match the DH expressed in the genome, which might be due to the missing of some fragment in genome splicing or the large variation of

SHM in the CDR3H region. It is also possible that D-D fusion exists (53, 54). It was important since D-D fusions often result in long (at least 24 AA) and ultralong CDR3Hs (55). However, there was no evidence to prove whether D-D fusion really exists and the mechanism of D-D fusion, so D-D fusion was not further analyzed in this study (56). Previous studies had shown that ultra-long CDR3H was considered to be a unique and diverse form of bovine immunoglobulin production, but in this study, we confirmed that ultra-long CDR3H also exists in yak, and CDR3H includes V gene and was also realized by ultra-long DH. Dr. Ma found that the bovine DH gene of ultra-long CDR3H had a 'YxYx' feature, which also appeared in the ultra-long CDR3H sequence of yak (11). CDR3H was located in the center of the antigen binding site and plays a key role in determining the specificity and affinity of antibodies (43). DH gene was the main component of CDR3H and largely

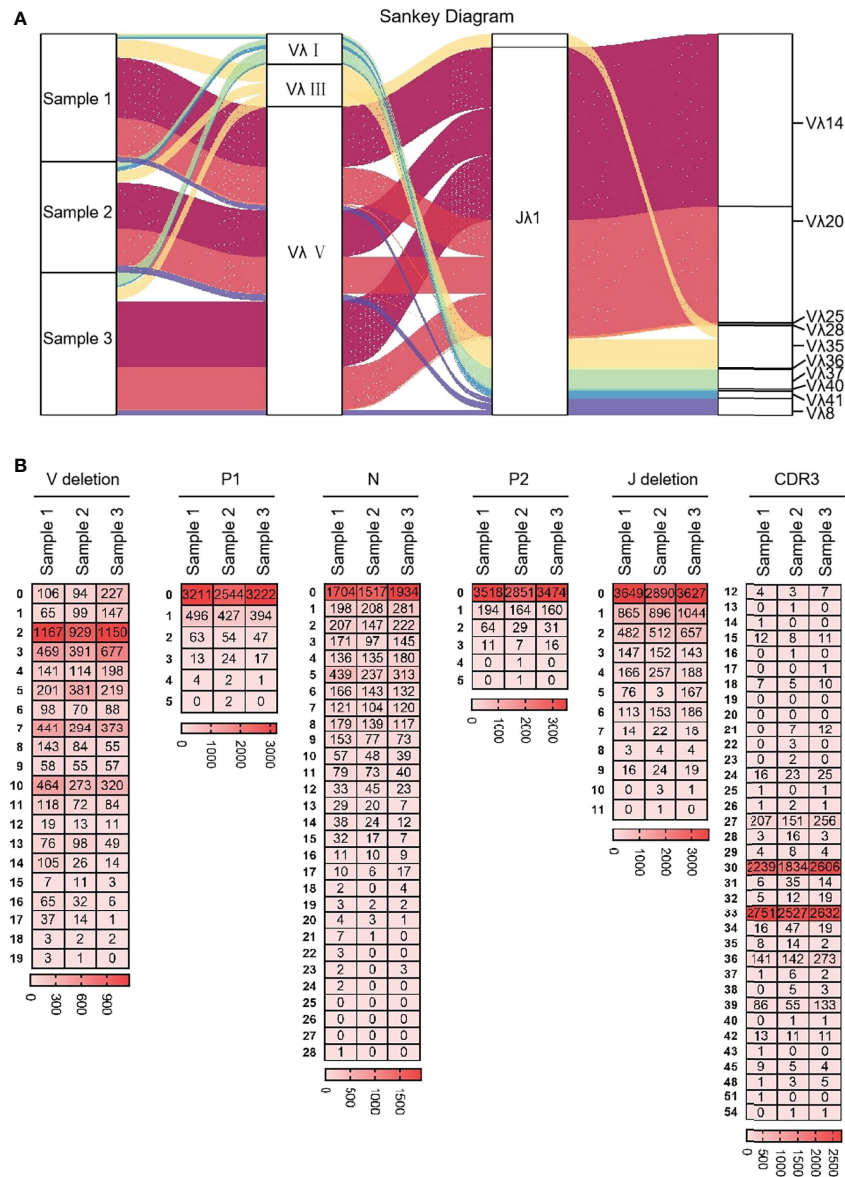


FIGURE 12 | The diversity of VJ recombination and CDR3 length in yak Ig(L). **(A)** The sankey diagram of VJ recombination in yak Ig(L); **(B)** the nucleotides distribution of V deletion, P1 nucleotide, N nucleotide, P2 nucleotide, J deletion and CDR3 in yak Ig(L).

determines the length and AA composition of CDR3H. As the core element of VDJ recombination, DH gene seems to have different adaptations during the evolution of different animals. As the core element and contributor of CDR3H, DH genes showed large sequence variation among species and were less conserved than VH and JH genes. Thus, the DH gene in a particular species may be the key to producing more protective antibodies and fewer autoreactive antibodies in that species. The antibody of ultra-long CDR3H had a special structure and the ability to bind to special antigens. Additionally, ultra-long CDR3H resulted in “microfolds” allowing bovine IgH to bind antigens that would otherwise be inaccessible (12). The number

of Yak VH, DH and JH genes was limited, and it was thought that IgH diversity was achieved through frequent recombination and endogenous mutations in the CDR3H region (57). It was still interesting to know exactly what advantages ultra-long CDRH3 confers on immunoglobulin function. For cattle and yaks, ultra-long CDR3H might have developed as another mechanism for maximizing diversification in the face of limited combinations of V(D)J. Alternatively, the ultra-long CDR3H might have evolved to optimize binding to antigens from rumen microbes or certain bovine specific pathogens which were not typically encountered by other vertebrates (12, 58, 59).

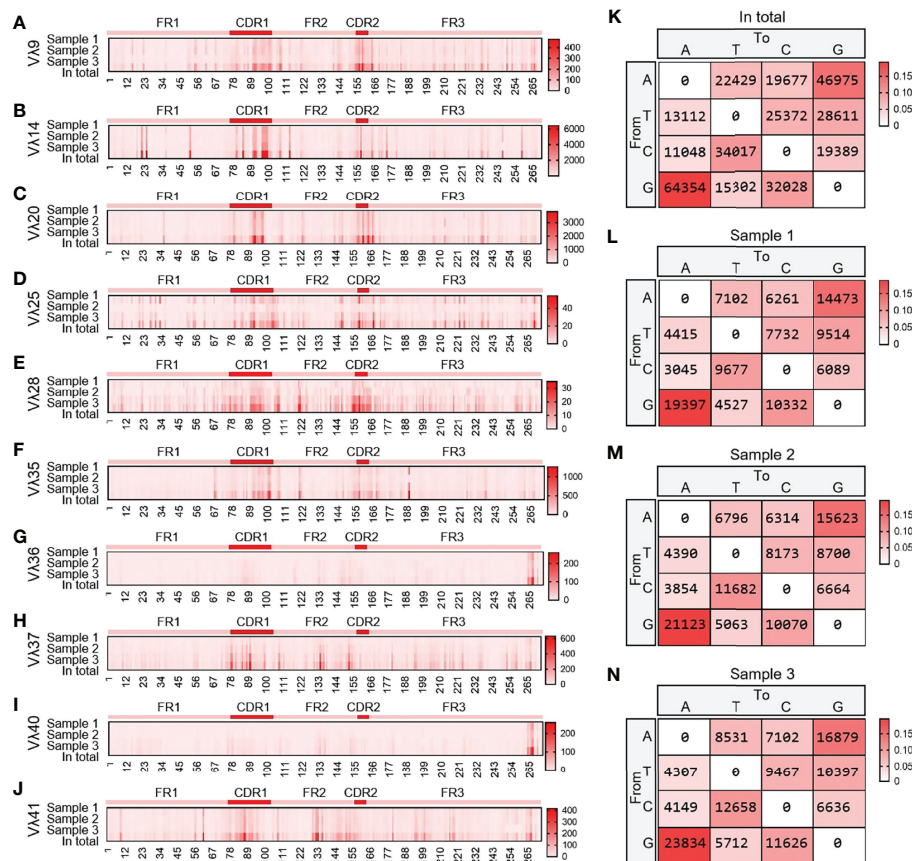


FIGURE 13 | The somatic hypermutation (SHM) of yak IgLλ. **(A–J)** The distribution of SHM in yak IgLλ; **(K–N)** The base mutation types of SHM in yak IgLλ.

SHM was an important mechanism of antibody diversity after V(D)J recombination. The AA changes caused by the mutation could improve the affinity of antibodies (5, 60). The SHM of yak IgH had a preference, and the mutation frequency from A to G and G to A was the highest, followed by the mutation between A and T, which was the same as that of cattle, mice, zebrafish, etc., which indicates that the mutation pattern of SHM is evolutionarily conservative (11, 37). However, it is worth noting that in addition to CDR1 and CDR2, high-frequency SHM loci also exist on FR2, possibly because some germ line VH genes exist in unsplicing genome fragments, resulting in an incomplete template library. In addition, studies have shown that SHM is produced in order to increase the diversity of fetal bovine antibodies, and it has also been reported that the point mutations produced by SHM were mainly concentrated at 100–200 bp away from the gene transcription start site, which might be the reason why SHM was not concentrated in CDRs regions in yak (61, 62). The SHM pattern of the yak was different from sheep and goats. The average SHM of the VH gene, Vλ gene and Vκ gene was 8.26%, 6.61% and 16.75%. The SHM frequencies of two horses used in previous study were 8.35% and 7.79% for the VH, 10.24% and 10.27% for the Vλ, and 4.17% and 5.51% for the Vκ (26). The SHM frequencies of the yak VH gene were 7.1%,

10.2% and 7.3%, Vλ gene SHM frequencies were 6.92%, 8.26% and 7.58%, and Vκ gene SHM frequencies were 2.24%, 2.13% and 2.85%, respectively. The SHM frequency of yak H chain and λ chain was similar to that of other large livestock, but the SHM frequency of κ clan was much lower than that of other large livestock. SHM was an irreversible process, and the accumulated SHM were all favorable mutations, and SHM frequency was higher in species with more perfect germinal centers.

In this study, the diversity of yak IgH was analyzed by high-throughput sequencing - PE300, which greatly increased the number of reads analyzed (nλ=16464, nκ=3939), and some low-frequency recombinant sequences of VJ were found, such as Vλ25-Jλ3, Vλ28-Jλ3, Vλ36-Jλ3 and Vλ40-Jλ3. PE300 could avoid the omission of low-frequency expression genes due to the small number of samples. The differences between the findings of existing studies on Igκ in pigs might be influenced by the fact that too few items were sequenced. (47, 63). PE300 was not used for the IgH diversity study in this study, because yak had extremely long CDR3 (59, 64, 65).

By comparing the production mechanism of immunoglobulin diversity in different species, it was found that humans and mice had rich V(D)J recombination diversity, and they could produce rich immunoglobulin types through recombination to resist the

invasion of different antigens (66). For species with less V(D)J recombination types, such as rabbit, it compensated for this limitation through GCV of different somatic cells and high levels of SHM, resulting in a more diverse antibody lineage than human or mouse. Structurally, rabbit antibodies rely more on IgL for antigen specificity, using longer CDR3L rings and interdomain disulfide bonds (67). In camels, B cells produce heavy chain only antibodies (HCAs) in addition to regular antibodies with a IgH/IgL pairing configuration. These HCAs contained some unusual structural features that further diversify the camel's antibody library (68). In this study, by analyzing the structure and expression diversity of immunoglobulin gene in the yak, it was found that yak compensated for the restriction of little V(D)J recombination by ultra-long CDR3H and abundant SHM, which was consistent with the previous production mode of immunoglobulin diversity in cattle. However, the mechanism that relies on ultra-long CDR3H to increase immunoglobulin diversity had not been found in other species. Therefore, we hypothesized that the ultra-long CDR3H was the unique mechanism of the rich immunoglobulin diversity acquired during the evolution of the bovid. It was critical to understand the natural immune mechanisms that fight infection and how vaccination, biosafety, nutrition, livestock and management practices could be used to maintain and enhance immune protection. It was hoped that the study on yak immunoglobulin could provide basic theoretical knowledge for yak breeding and immunity research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA805364.

REFERENCES

- Honjo T, Kinoshita K, Muramatsu M. Molecular Mechanism of Class Switch Recombination: Linkage With Somatic Hypermutation. *Annu Rev Immunol* (2002) 20:165–96. doi: 10.1146/annurev.immunol.20.090501.112049
- Rawlings DJ, Metzler G, Wray-Dutra M, Jackson SW. Altered B Cell Signalling in Autoimmunity. *Nat Rev Immunol* (2017) 17(7):421–36. doi: 10.1038/nri.2017.24
- Schroeder HW Jr., Cavacini L. Structure and Function of Immunoglobulins. *J Allergy Clin Immunol* (2010) 125(2 Suppl 2):S41–52. doi: 10.1016/j.jaci.2009.09.046
- Townsend CL, Laffy JM, Wu YB, Silva O'Hare J, Martin V, Kipling D, et al. Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions. *Front Immunol* (2016) 7:388. doi: 10.3389/fimmu.2016.00388
- Chi X, Li Y, Qiu X. V(D)J Recombination, Somatic Hypermutation and Class Switch Recombination of Immunoglobulins: Mechanism and Regulation. *Immunology* (2020) 160(3):233–47. doi: 10.1111/imm.13176
- Agrawal A, Schatz DG. RAG1 and RAG2 Form a Stable Postcleavage Synaptic Complex With DNA Containing Signal Ends in V(D)J Recombination. *Cell* (1997) 89:43–53. doi: 10.1016/S0092-8674(00)80181-6
- Hsu E. Assembly and Expression of Shark Ig Genes. *J Immunol* (2016) 196(9):3517–23. doi: 10.4049/jimmunol.1600164

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee of Northwest A&F University.

AUTHOR CONTRIBUTIONS

HZ, MW, and XT analyzed the data; MW, XT, WZ, and QL wrote the manuscript; HZ, MW, WZ, QL, and XY carried out the experiment; XY and XS reviewed and edited the manuscript; HZ designed the experiment. All authors contributed to the interpretation of the results and writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

Project supported by National Natural Science Foundation of China (31972557).

ACKNOWLEDGMENTS

We thank professor Ping Yan (Lanzhou Institute of Husbandry and Pharmaceutical Sciences) for sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.876509/full#supplementary-material>

Supplement 1 | DEGs list.

- Carmona LM, Schatz DG. New Insights Into the Evolutionary Origins of the Recombination-Activating Gene Proteins and V(D)J Recombination. *FEBS J* (2017) 284(11):1590–605. doi: 10.1111/febs.13990
- Yu K, Lieber MR. Current Insights Into the Mechanism of Mammalian Immunoglobulin Class Switch Recombination. *Crit Rev Biochem Mol Biol* (2019) 54(4):333–51. doi: 10.1080/10409238.2019.1659227
- Liljavirta J, Niku M, Pessa-Morikawa T, Ekman A, Iivanainen A. Expansion of the Preimmune Antibody Repertoire by Junctional Diversity in Bos Taurus. *PLoS One* (2014) 9(6):e99808. doi: 10.1371/journal.pone.0099808
- Ma L, Qin T, Chu D, Cheng X, Wang J, Wang X, et al. Internal Duplications of DH, JH, and C Region Genes Create an Unusual IgH Gene Locus in Cattle. *J Immunol* (2016) 196(10):4358–66. doi: 10.4049/jimmunol.1600158
- Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, et al. Reshaping Antibody Diversity. *Cell* (2013) 153(6):1379–93. doi: 10.1016/j.cell.2013.04.049
- Haakenson JK, Huang R, Smider VV. Diversity in the Cow Ultralong CDR H3 Antibody Repertoire. *Front Immunol* (2018) 9:1262. doi: 10.3389/fimmu.2018.01262
- Qin T, Liu Z, Zhao H. Organization and Genomic Complexity of Sheep Immunoglobulin Light Chain Gene Loci. *Immunol Letters* (2015) 168(2):313–8. doi: 10.1016/j.imlet.2015.10.010
- Yu X, Du L, Wu M, Wu J, He S, Yuan T, et al. The Analysis of Organization and Diversity Mechanism in Goat Immunoglobulin Light Chain Gene Loci. *Immunobiology* (2020) 225(2):151889. doi: 10.1016/j.imbio.2019.11.024

16. Saeed U, Usman Z. *Biological Sequence Analysis*. Brisbane (AU: Codon Publications) (2019).
17. Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT Unique Numbering for Immunoglobulin and T Cell Receptor Variable Domains and Ig Superfamily V-Like Domains. *Dev Comp Immunol* (2003) 27 (1):55–77. doi: 10.1016/S0145-305X(02)00039-3
18. Jiang Z, Wu J, Jiang M, Xie Y, Bu W, Liu C, et al. A Novel Technique for Constructing Infectious Cloning of Type 3 Porcine Circovirus. *Front Microbiol* (2020) 11:1067. doi: 10.3389/fmicb.2020.01067
19. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evolution* (2016) 33 (7):1870–4. doi: 10.1093/molbev/msw054
20. Schroeder HWJr. Similarity and Divergence in the Development and Expression of the Mouse and Human Antibody Repertoires. *Dev Comp Immunol* (2006) 30(1-2):119–35. doi: 10.1016/j.dci.2005.06.006
21. Heilig R, Eckenberg R, Petit JL, Fonknechten N, Da Silva C, Cattolico L, et al. The DNA Sequence and Analysis of Human Chromosome 14. *Nature* (2003) 421(6923):601–7. doi: 10.1038/nature01348
22. Elemento O, Lefranc MP. IMGT/PhyloGene: An on-Line Tool for Comparative Analysis of Immunoglobulin and T Cell Receptor Genes. *Dev Comp Immunol* (2003) 27(9):763–79. doi: 10.1016/S0145-305X(03)00078-8
23. Johnston CM, Wood AL, Bolland DJ, Corcoran AE. Complete Sequence Assembly and Characterization of the C57BL/6 Mouse Ig Heavy Chain V Region. *J Immunol* (2006) 176(7):4221–34. doi: 10.4049/jimmunol.176.7.4221
24. Charlton KA, Moyle S, Porter AJ, Harris WJ. Analysis of the Diversity of a Sheep Antibody Repertoire as Revealed From a Bacteriophage Display Library. *J Immunol* (2000) 164(12):6221–9. doi: 10.4049/jimmunol.164.12.6221
25. Walther S, Rusitzka TV, Diesterbeck US, Czerny CP. Equine Immunoglobulins and Organization of Immunoglobulin Genes. *Dev Comp Immunol* (2015) 53(2):303–19. doi: 10.1016/j.dci.2015.07.017
26. Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, et al. A Comprehensive Analysis of Germline and Expressed Immunoglobulin Repertoire in the Horse. *Dev Comp Immunol* (2010) 34(9):1009–20. doi: 10.1016/j.dci.2010.05.003
27. Chen L, Li M, Li Q, Yang X, An X, Chen Y. Characterization of the Bovine Immunoglobulin Lambda Light Chain Constant IGLC Genes. *Veterinary Immunol Immunopathol* (2008) 124(3-4):284–94. doi: 10.1016/j.vetimm.2008.04.012
28. Zhao Y, Kacsokovics I, Rabbani H, Hammarstrom L. Physical Mapping of the Bovine Immunoglobulin Heavy Chain Constant Region Gene Locus. *J Biol Chem* (2003) 278(37):35024–32. doi: 10.1074/jbc.M301337200
29. Pasman Y, Saini SS, Smith E, Kaushik AK. Organization and Genomic Complexity of Bovine Lambda-Light Chain Gene Locus. *Veterinary Immunol Immunopathol* (2010) 135(3-4):306–13. doi: 10.1016/j.vetimm.2009.12.012
30. Cheng G, Gao Y, Wang T, Sun Y, Wei Z, Li L, et al. Extensive Diversification of IgH Subclass-Encoding Genes and IgM Subclass Switching in Crocodylians. *Nat Commun* (2013) 4:1337. doi: 10.1038/ncomms2317
31. Wu Q, Wei Z, Yang Z, Wang T, Ren L, Hu X, et al. Phylogeny, Genomic Organization and Expression of Lambda and Kappa Immunoglobulin Light Chain Genes in a Reptile, Anolis Carolinensis. *Dev Comp Immunol* (2010) 34 (5):579–89. doi: 10.1016/j.dci.2009.12.019
32. Arakawa H, Kuma K, Yasuda M, Ekino S, Shimizu A, Yamagishi H. Effect of Environmental Antigens on the Ig Diversification and the Selection of Productive V-J Joints in the Bursa. *J Immunol* (2002) 169(2):818–28. doi: 10.4049/jimmunol.169.2.818
33. Guan X, Wang J, Ma L, Wang X, Cheng X, Han H, et al. Multiple Germline Functional VL Genes Contribute to the IgL Repertoire in Ducks. *Dev Comp Immunol* (2016) 60:167–79. doi: 10.1016/j.dci.2016.02.029
34. Yamawaki-Kataoka Y, Honjo T. Nucleotide Sequences of Variable Region Segments of the Immunoglobulin Heavy Chain of *Xenopus laevis*. *Nucleic Acids Res* (1987) 15(14):5888. doi: 10.1093/nar/15.14.5888
35. Haire RN, Ota T, Rast JP, Litman RT, Chan FY, Zon LI, et al. A Third Ig Light Chain Gene Isotype in *Xenopus laevis* Consists of Six Distinct VL Families and is Related to Mammalian Lambda Genes. *J Immunol* (1996) 157(4):1544–50.
36. Qin T, Ren L, Hu X, Guo Y, Fei J, Zhu Q, et al. Genomic Organization of the Immunoglobulin Light Chain Gene Loci in *Xenopus tropicalis*: Evolutionary Implications. *Dev Comp Immunol* (2008) 32(2):156–65. doi: 10.1016/j.dci.2007.05.007
37. Danilova N, Bussmann J, Jekosch K, Steiner LA. The Immunoglobulin Heavy-Chain Locus in Zebrafish: Identification and Expression of a Previously Unknown Isotype, Immunoglobulin Z. *Nat Immunol* (2005) 6(3):295–302. doi: 10.1038/nri1166
38. Wang Z, Wu Y, Hu Q, Li Y. Differences on the Biological Function of Three Ig Isotypes in Zebrafish: A Gene Expression Profile. *Fish Shellfish Immunol* (2015) 44:283–6. doi: 10.1016/j.fsi.2015.02.030
39. Tomana M, Ishikawa J, Imai E, Moritomo T, Nakao M, Yano T. Characterization of Immunoglobulin Light Chain Isotypes in the Common Carp. *Immunogenetics* (2002) 54(2):120–9. doi: 10.1007/s00251-002-0447-7
40. Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E. Hypermutation in Shark Immunoglobulin Light Chain Genes Results in Contiguous Substitutions. *Immunol* (2002) 16(4):571–82. doi: 10.1016/S1074-7613(02)00300-X
41. Rio DC, Ares MJr., Hannon GJ, Nilsen TW. Purification of RNA Using TRIzol (TRI Reagent). *Cold Spring Harbor Protoc* (2010) 2010(6):pdb.prot5439. doi: 10.1101/pdb.prot5439
42. Hu X, Zhang J, Wang J, Fu J, Li T, Zheng X, et al. Landscape of B Cell Immunity and Related Immune Evasion in Human Cancers. *Nat Genet* (2019) 51(3):560–7. doi: 10.1038/s41588-018-0339-x
43. Vlasova AN, Saif LJ. Bovine Immunology: Implications for Dairy Cattle. *Front Immunol* (2021) 12:643206. doi: 10.3389/fimmu.2021.643206
44. Jenne CN, Kennedy LJ, Reynolds JD. Antibody Repertoire Development in the Sheep. *Dev Comp Immunol* (2006) 30(1-2):165–74. doi: 10.1016/j.dci.2005.06.009
45. Du L, Wang S, Zhu Y, Zhao H, Basit A, Yu X, et al. Immunoglobulin Heavy Chain Variable Region Analysis in Dairy Goats. *Immunobiology* (2018) 223 (11):599–607. doi: 10.1016/j.imbio.2018.07.005
46. Mage RG, Lanning D, Knight KL. B Cell and Antibody Repertoire Development in Rabbits: The Requirement of Gut-Associated Lymphoid Tissues. *Dev Comp Immunol* (2006) 30(1-2):137–53. doi: 10.1016/j.dci.2005.06.017
47. Zhang M, Li Z, Li J, Huang T, Peng G, Tang W, et al. Revisiting the Pig IGHC Gene Locus in Different Breeds Uncovers Nine Distinct IGHC Genes. *J Immunol* (2020) 205(8):2137–45. doi: 10.4049/jimmunol.1901483
48. Butler JE, Wertz N, Sinkora M. Antibody Repertoire Development in Swine. *Annu Rev Anim Biosci* (2017) 5:255–79. doi: 10.1146/annurev-animal-022516-022818
49. Fugmann SD, Lee AI, Shockett PE, Villey IJ, Schatz DG. The RAG Proteins and V(D)J Recombination: Complexes, Ends, and Transposition. *Annu Rev Immunol* (2000) 18:495–527. doi: 10.1146/annurev.immunol.18.1.495
50. Koti M, Kataeva G, Kaushik AK. Organization of D(H)-Gene Locus is Distinct in Cattle. *Developments Biologicals* (2008) 132:307–13. doi: 10.1159/000317176
51. Koti M, Kataeva G, Kaushik AK. Novel Atypical Nucleotide Insertions Specifically at VH-DH Junction Generate Exceptionally Long CDR3H in Cattle Antibodies. *Mol Immunol* (2010) 47(11-12):2119–28. doi: 10.1016/j.molimm.2010.02.014
52. Ma L. *Genomic Organization of the IgH Genes and Their Repertoire Generation in Holstein Cows*. [Master's Thesis] (2016) [Beijing]: China Agricultural University.
53. Meek KD, Hasemann CA, Capra JD. Novel Rearrangements at the Immunoglobulin D Locus. Inversions and Fusions Add to IgH Somatic Diversity. *J Exp Med* (1989) 170(1):39–57. doi: 10.1084/jem.170.1.39
54. Walther S, Czerny CP, Diesterbeck US. Exceptionally Long CDR3H are Not Isotype Restricted in Bovine Immunoglobulins. *PLoS One* (2013) 8(5):e64234. doi: 10.1371/journal.pone.0064234
55. Briney BS, Willis JR, Crowe J. Human Peripheral Blood Antibodies With Long HCDR3s are Established Primarily at Original Recombination Using a Limited Subset of Germline Genes. *PLoS One* (2012) 7(5):e36750. doi: 10.1371/journal.pone.0036750
56. Safonova Y, Pevzner PA. V(DD)J recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDR3s. *Genome research* (2020) 30(11):1547–58. doi: 10.1101/gr.259598.119

57. Zhuang Y, Futse JE, Brown WC, Brayton KA, Palmer GH. Maintenance of Antibody to Pathogen Epitopes Generated by Segmental Gene Conversion is Highly Dynamic During Long-Term Persistent Infection. *Infect Immunology* (2007) 75(11):5185–90. doi: 10.1128/IAI.00913-07
58. Stanfield RL, Haakenson J, Deiss TC, Criscitiello MF, Wilson IA, Smider VV. The Unusual Genetics and Biochemistry of Bovine Immunoglobulins. *Adv Immunol* (2018) 137:135–64. doi: 10.1016/bs.ai.2017.12.004
59. Deiss TC, Vadnais M, Wang F, Chen PL, Torkamani A, Mwangi W, et al. Immunogenetic Factors Driving Formation of Ultralong VH CDR3 in Bos Taurus Antibodies. *Cell Mol Immunol* (2019) 16(1):53–64. doi: 10.1038/cmi.2017.117
60. Azhar A, Begum NA, Husain A. Nucleotide Pool Imbalance and Antibody Gene Diversification. *Vaccines* (2021) 9(10):1050. doi: 10.3390/vaccines9101050
61. Chaudhuri J, Tian M, Khuong C, Chua K, Pinaud E, Alt FW. Transcription-Targeted DNA Deamination by the AID Antibody Diversification Enzyme. *Nature* (2003) 422(6933):726–30. doi: 10.1038/nature01574
62. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc Natl Acad Sci USA* (2003) 100(7):4102–7. doi: 10.1073/pnas.0730835100
63. Butler JE, Wertz N, Sun X. Antibody Repertoire Development in Fetal and Neonatal Piglets. XIV. Highly Restricted IGKV Gene Usage Parallels the Pattern Seen With IGLV and IGHV. *Mol Immunol* (2013) 55(3-4):329–36.
64. Saini SS, Allore B, Jacobs RM, Kaushik A. Exceptionally Long CDR3H Region With Multiple Cysteine Residues in Functional Bovine IgM Antibodies. *Eur J Immunol* (1999) 29(8):2420–6. doi: 10.1002/(SICI)1521-4141(199908)29:08<2420::AID-IMMU2420>3.0.CO;2-A
65. Stanfield RL, Wilson IA, Smider VV. Conservation and Diversity in the Ultralong Third Heavy-Chain Complementarity-Determining Region of Bovine Antibodies. *Sci Immunol* (2016) 1(1):aaf7962. doi: 10.1126/sciimmunol.aaf7962
66. Dong J, Finn JA, Larsen PA, Smith T, Crowe JE Jr. Structural Diversity of Ultralong CDRH3s in Seven Bovine Antibody Heavy Chains. *Front Immunol* (2019) 10:558. doi: 10.3389/fimmu.2019.00558
67. Weber J, Peng H, Rader C. From Rabbit Antibody Repertoires to Rabbit Monoclonal Antibodies. *Exp Mol Med* (2017) 49(3):e305. doi: 10.1038/emmm.2017.23
68. Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *The EMBO journal* (2020) 19(5):921–30. doi: 10.1093/emboj/19.5.921

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wu, Zhao, Tang, Zhao, Yi, Li and Sun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.