Immunobiology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Immunobiology

journal homepage: www.elsevier.com/locate/imbio

Immunoglobulin heavy chain variable region analysis in dairy goats

Lijuan Du[a,b], Shuhui Wang[a], Yanjiao Zhu[a], Haidong Zhao[a], Abdul Basit[b], Xiaohui Yu[a], Qingwang Li[a], Xiuzhu Sun[a,*]

[a] College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, 712100, China
[b] State Key Laboratory of Agrobiotechnology, College of Biological Sciences, China Agricultural University, Beijing 100193, China

ARTICLE INFO

Keywords:
Goat
Immunoglobulin heavy chain
Genomic organization
Complementarity-Determining region 3
Somatic hypermutation
V(D)J recombination

ABSTRACT

Based on the goat genome database, we have annotated the genomic organization of the goat immunoglobulin heavy chain variable region. The goat IgH locus is present on seven genome scaffolds, and contains ten V_H, three D_H and six J_H segments. After the exclusion of three shorter segments, the V_H genes were divided into two gene families based on sequence similarity. By analyzing the IgH cDNA sequences, we further identified that V_H2 (54.2%), D_H1 (61.7%) and J_H1 (60.5%) segments were most frequently utilized in the expression of the immunoglobulin variable region, and that point mutations introduced by somatic hypermutation were the major mutation present in these expressed variable region. Compared with human and horses, D_H-D_H fusion occurred at a higher frequency in goat V(D)J recombination. These results provided variable insights into goat immunoglobulin heavy chain variable region genome loci and repertoire diversity.

1. Introduction

A typical immunoglobulin molecule (Ig) in most jawed vertebrates is composed of two heavy (H) and two light (L) chains. The H and L chains are both comprised of variable and constant regions. The IgH variable region consists of variable (V_H), diversity (D_H) and joining (J_H) gene segments. Their three-dimensional structure is arranged in a staggered order, with four framework regions (FRs) and three complementarity-determining regions (CDRs). The FR sequences are highly conserved among different species. The CDR regions, however, provide diverse antigen-binding sites through sequence variation (Chothia and Lesk, 1987; Wang et al., 2013).

The variable regions of both the H and L chains can be diversified through V(D)J recombination, somatic hypermutation (SHM) and gene conversion (GC). V(D)J recombination is a site-specific somatic process, primarily diversifies the expressed Ig variable region. The discontinuousV, (D), and J gene segments are flanked by recombination signal sequences (RSS), and are assembled into a variable region through non-homologous end joining during B-cell development (Bassing et al., 2002; Sun et al., 2012). Therefore, the number of germline V_H, D_H and J_H gene segments, and the distinctions between these sequences play important role in this diversity mechanism. In species with a large number of germline V_H segments, such as human (123) and mice (approximately 200), V(D)J recombination is the major Ig diversity mechanism (Heilig et al., 2003; Hendricks et al., 2010; Johnston et al., 2006; Matsuda et al., 1998). However, in some species, particularly birds (chicken (Reynaud et al., 1989), duck (McCormack et al., 1989), goose (Huang et al., 2016)), V(D)J recombinatorial diversity is limited. These species mainly depend on GC-based recombination to generate primary antibody repertoire using rearranged functional V(D)J junctions.

SHM is well characterized as a secondary diversity mechanism in many cases, improving antigen binding affinity. It occurs in the rearranged variable regions in germinal center B cells (MacLennan, 2005). After imprecise V(D)J junction, non-templated mutations introduced by SHM, including substitutions, insertions or deletions are amplified in the rearranged variable region (Briney and Crowe, 2013; Peled et al., 2008). Some species, such as cattle and pig, have limited a number of

Abbreviations: Ig, immunoglobulin; V_H, immunoglobulin heavy chain variable region variable gene; D_H, immunoglobulin heavy chain variable region diversity gene; J_H, immunoglobulin heavy chain variable region joining gene; V(D)J, junction of immunoglobulin heavy chain variable, (diversity) and joining regions; SHM, somatic hypermutation; GC, gene conversion; FR, framework region; CDR, complementarity-determining region; ORF, open reading frame; IMGT, the international ImMunoGeneTics information system; RSS, recombination signal sequence; PCR, Polymerase chain reaction; p-V_H, immunoglobulin heavy chain variable region pseudo-variable gene; ψJ_H, immunoglobulin heavy chain variable region pseudo-joining gene; IGHV, immunoglobulin heavy chain variable region variable gene family

* Corresponding author.
E-mail address: sunxiuzhu208@163.com (X. Sun).

https://doi.org/10.1016/j.imbio.2018.07.005
Received 24 October 2017; Received in revised form 12 January 2018; Accepted 5 July 2018

functional variable region genes for V(D)J recombination, and SHM, compensates for this, driving diversification post-rearrangement (Berens et al., 1997; Butler and Wertz, 2012; Saini and Kaushik, 2002).

Goats (Capra hircus) are one of the most important and economically impactful domestic animals. Goat antibodies are commonly used in biotechnology. They are used to prepare diagnostic reagents of immunochemical techniques. However, limited information is available about goat immunoglobulin genomic organization and diversification mechanisms. Based on the recently completed *C.* hircus genome database (Dong et al., 2013), here we report the Ig H chain variable region genome complexity and the diversity mechanisms of goat IgH gene diversification.

2. Materials and methods

2.1. Sample collection

Three healthy adult *Xinong Saanen* 12-month-old dairy goats were randomly selected at the breeding farm at Northwest A&F University Veterinary (Yangling, Shaanxi Province, China). The spleen of each goat was collected and processed for RNA extraction.

2.2. Sequence analysis of the goat germline Ig variable region

The goat genome sequence and is available on the National Center for Biotechnology Information website (https://www.ncbi.nlm.nih.gov/assembly/GCA_000317765.2) (Dong et al., 2013). Nine families of sheep V_H gene and goat IGHV cDNA clones (GenBank accession numbers EU344746-EU344750) were used as queries to search goat genome scaffolds containing Ig genes. A single sequence was randomly selected from each sheep V_H gene family. The V_H segments were predicted using IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/). The FRs and CDRs of V_Hs were classified using international ImMunoGeneTics information system (IMGT) numbering system (http://imgt.cines.fr/) (Lefranc et al., 2003). Potentially functional, open reading frame (ORF), and pseudo-V_H segments were identified according to the IMGT database (Lefranc et al., 2003). FUZZNUC (http://emboss.bioinformatics.nl/cgi-bin/emboss/fuzznuc) was used for RSS identification in the D_H and J_H segments.

2.3. RNA extraction and recovery of V-D-J transcripts

Total RNA was extracted from each spleen using an Ultrapure RNA kit (Cowin, Beijing, China)and cDNA was generated using the PrimeScriptII 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China), according to the manufacturers' instructions. Polymerase chain reaction (PCR) specific primers for V_HF (5′-ACTGTGGACCCTCCTCTTTGTG-3′) and J_HR (5′-GAGGAGACGGTGACCAGGAGT-3′) were used to amplify the V_H-D_H-J_H genes. V_HF was designed based on a highly conserved region of the V_H leader sequence (GenBank accession number EU344748). J_HR was complementary to the last seven codons of the goat J_H segment (GenBank accession number EU182620). Amplification was performed by using PrimeSTAR HS (Premix) (TaKaRa, Dalian, China). The cycling conditions were: 30 cycles of 98 °C for 10 s, 55 °C for 5 s and 72 °C for 40 s. PCR products were resolved on a 1.0% agarose gel and purified using the TaKaRa MiniBEST DNA Fragment Purification Kit Ver.4.0 (TaKaRa, Dalian, China). The extracted fragments were ligated into the pMD19-T vector (TaKaRa, Dalian, China) and transformed into *Escherichia coli* JM109 competent cells (TaKaRa, Dalian, China). Individual clones containing the correctly sized insert were expanded in lysogeny broth containing ampicillin (0.1%) and confirmed by DNA sequencing (Sangon Biotech, Shanghai, China).

2.4. Sequence computations and identification of V_H gene families

DNA and protein sequence editing and alignment were performed with DNASTAR software and ClustalX2. Germline V_H genes with > 75% nucleotide identity were considered to belong to the same gene family, whereas those with < 70% similarity were classified into different subgroups. Genes with sequence similarity of 70–75% were inspected on a case-by-case basis (Schroeder et al., 1990).

2.5. Phylogenetic analysis

The phylogenetic tree was constructed using randomly selected V_H segments from different species including: Human (Homo sapiens): IGHV1-69, L22582; IGHV2-5, X62111; IGHV3-72, X92206; IGHV4-b, Z12367; IGHV5-51, M99686; IGHV6-1, X92224; IGHV7-4-1, L10057. Mouse (Mus musculus): IGHV1S52, M34982; IGHV2-2, J00502; IGHV3S1, K01569; IGHV4-2, AJ851868; IGHV5-9, X00163; IGHV6-3, AC073590; IGHV7-1, AJ851868; IGHV8-5, AC074329; IGHV9-1, AC073563; IGHV10-3, AF064444; IGHV11-1, AC073563; IGHV12-3, AC073590; IGHV13-1, AC073589; IGHV14-4, AC073589; IGHV15-2, AC090843; IGHV16-1, AC073563. Cattle (Bos taurus): U49765; Pig (Sus scrofa): IGHV1S4, AF064687. Nurse shark (Ginglymostoma cirratum): IGHV1S1, M92851 (Sun et al., 2010). Sheep (Ovis aries): IGHV1S2, Z49180; IGHV2, H37; IGHV3, H204; IGHV4, H297; IGHV5, H281; IGHV6, H168; IGHV7, H225; IGHV8, H17; IGHV9, H23 (Charlton et al., 2000). The tree was constructed with MrBayes 3.1 and visualized in TREEVIEW (Page, 1996; Ronquist and Huelsenbeck, 2003).

2.6. V(D)J recombination analysis and SHM statistics

Nucleotide sequence editing and alignment were performed as in Section 2.4. Clones (FR1-FR3) best matched the germline segments were selected for V_H expression and base substitution analyses, and nucleotide differences were counted manually. The nucleotides between V_H and J_H in all cDNA clones were aligned with three germline D_H coding sequences. One mismatched base was allowed in clusters of over six nucleotides. Duplicate clones were eliminated.

3. Results

3.1. Genomic organization of the goat IgH locus

BLAST analysis using sheep and goat V_H gene segments as queries revealed that the goat IgH locus was located on seven genome scaffolds (Fig. 1, unmapped on chromosomes). A total of ten V_H, three D_H and six J_H segments were identified. Of the ten, three of the V_H genes were potentially functional containing leader exons, intact translation initiation codons, splice junctions, and downstream RSSs, but containing no in-frame stop codons or frameshifts (Brodeur and Riblet, 1984). One of the V_H segment was designated as an ORF as it lacked 15 nucleotides at the beginning of the leader sequences, and the highly conversed cysteine at amino acid position 23 was replaced by phenylalanine (Lefranc et al., 2003) (Supplementary Fig. 1). The remaining six V_H segments contained either in-frame stop codons or were partial segments of approximately 200 bp, and were considered pseudogenes (Lefranc, 1998). The short fragments were excluded, and the remaining sequences were classified into families 1 and 2 based on sequence similarity. Family 1 consisted of the three potentially functional V_Hs and the single ORF, while family 2 contained the three pseudogenes (Table 1).

Three D_H segments were identified in a 1.3 kb region on scaffold1847 (Fig. 1A). These segments were flanked on either side by RSSs with 12-bp spacers (Fig. 2A). Furthermore, all three D_H segments existed within at least one alternative reading frame, indicating their potential functionality. The lengths of the potential coding regions of