



Evolution of immunogenetic components encoding ultralong CDR H3

Jeannine A. Ott¹ · Christian Mitchell¹ · Morgan Sheppard¹ · Thad C. Deiss¹ · J. M. Cody Horton² · Jeremy K. Haakenson^{3,4} · Ruiqi Huang^{3,4} · Abigail R. Kelley³ · Brian W. Davis² · James N. Derr¹ · Vaughn V. Smider^{3,4} · Michael F. Criscitiello^{1,5}

Received: 29 January 2023 / Accepted: 3 April 2023 / Published online: 21 April 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The genomes of most vertebrates contain many V, D, and J gene segments within their Ig loci to construct highly variable CDR3 sequences through combinatorial diversity. This nucleotide variability translates into an antibody population containing extensive paratope diversity. Cattle have relatively few functional VDJ gene segments, requiring innovative approaches for generating diversity like the use of ultralong-encoding IGHV and IGHD gene segments that yield dramatically elongated CDR H3. Unique knob and stalk microdomains create protracted paratopes, where the antigen-binding knob sits atop a long stalk, allowing the antibody to bind both surface and recessed antigen epitopes. We examined genomes of twelve species of Bovidae to determine when ultralong-encoding IGHV and IGHD gene segments evolved. We located the 8-bp duplication encoding the unique TTVHQ motif in ultralong IGHV segments in six Bovid species (cattle, zebu, wild yak, domestic yak, American bison, and domestic gayal), but we did not find evidence of the duplication in species beyond the *Bos* and *Bison* genera. Additionally, we analyzed mRNA from bison spleen and identified a rich repertoire of expressed ultralong CDR H3 antibody mRNA, suggesting that bison use ultralong IGHV transcripts in their host defense. We found ultralong-encoding IGHD gene segments in all the same species except domestic yak, but again not beyond the *Bos* and *Bison* clade. Thus, the duplication event leading to this ultralong-encoding IGHV gene segment and the emergence of the ultralong-encoding IGHD gene segment appears to have evolved in a common ancestor of the *Bos* and *Bison* genera 5–10 million years ago.

Keywords Cattle · Evolution · Antibody · Ultralong CDR H3 · IgH locus · Diversity segment

Introduction

The humoral branch of the adaptive immune system is mediated by B cells and their secreted immunoglobulins (Ig), also referred to as antibodies. There are five antibody isotypes (IgM, IgD, IgG, IgE, and IgA), each composed of two light chains (IgL) and two heavy chains (IgH) linked together by disulfide bonds, with IgH often making the most important contribution to antigen binding. Together, these four protein chains form the characteristic Y-shaped structure of the antibody. IgH and IgL contain both a variable region (F_v , which together comprises the antigen binding paratope) and a constant region (F_c , which confers the isotype, or effector function) (Li et al. 2004). The IgH variable region itself is encoded by variable (V), diversity (D), and joining (J) gene segments (Jones and Gellert 2004) that form functional genes through the process of recombination-activating gene (RAG) mediated somatic recombination (Schatz 2004; Tonegawa 1983). The assembled variable region encodes a

Jeannine Ott and Christian Mitchell share first authorship.

✉ Michael F. Criscitiello
mcriscitiello@cvm.tamu.edu

- ¹ Comparative Immunogenetics Laboratory, Department of Veterinary Pathobiology, School of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA
- ² Department of Veterinary Integrative Sciences, School of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX, USA
- ³ Applied Biomedical Science Institute, San Diego, CA 92127, USA
- ⁴ Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA
- ⁵ Department of Microbial Pathogenesis and Immunology, College of Medicine, Texas A&M Health Science Center, Texas A&M University, Bryan, TX 77807, USA

primary amino acid sequence that is divided into four framework (FR) and three complementarity-determining regions (CDR) that alternate along its length (Nezlin 2019). The paratope is formed by a combination of the three hypervariable CDR loops from both IgH and IgL (six in total) and further diversifies through activation-induced cytidine deaminase (AID)-catalyzed somatic hypermutation (SHM), which mutates the nucleotide sequence within targeted motifs of CDR loops to affinity mature an antibody for its antigen (Conticello et al. 2005; Li et al. 2004).

A diverse antibody repertoire is fundamental to creating a robust defense against pathogens. The genomes of most species, therefore, contain a considerable number of V, D, and J gene segments within their Ig loci to construct the highly variable third CDR (CDR3) sequences through combinatorial and junctional diversity. This nucleotide variability translates into an antibody population containing extensive paratope diversity. Human IgH loci contain 57 functional V (IGHV), 23 functional D (IGHD), and six functional J (IGHJ) gene segments (Mikocziova et al. 2021). Mouse genomes vary by strain, but the IGH locus of five novel mouse strains contains 97–121 IGHV, 9–17 IGHD, and four IGHJ gene segments (Collins et al. 2015; Johnston et al. 2006; Lefranc 2014; Lefranc et al. 2015). These IGHV then are assigned to families (or subgroups) within three distinct clans (I, II, III) based on sequence homology of the first and third FR (FR1 and FR3) that reflects conservation of both the protein sequence and structure across mammalian species (Kirkham et al. 1992; Schroeder et al. 1990). Thus, the utilization of gene segments from across clans would create enormous heterogeneity in the resulting antigen receptors.

In contrast to mice and humans, the genome of cattle (*Bos taurus*) and some other species (i.e., rabbits, *Oryctolagus cuniculus*; pigs, *Sus scrofa*; and sheep, *Ovis aries*) contain relatively few functional IGHV gene segments, thereby constraining the combinatorial diversity typically generated through VDJ recombination (Dufour et al. 1996; Knight and Becker 1990; Ma et al. 2016; Saini et al. 1997). Furthermore, the twelve functional IGHV genes (of 47 total) in the cattle genome all belong to the same IGHV1 subgroup within clan II (Berens et al. 1997; Ma et al. 2016; Niku et al. 2012), and there is evidence that VDJ recombination biases expression towards a single IGHV sequence (IGHV1-10) in conventional antibodies (Deiss et al. 2019). Additionally, while some IGHD gene segments in mice and humans can be used in all three reading frames within functional IgH rearrangements (Jackson 2019; Larimore et al. 2012; Schroeder et al. 2010), all functional cattle antibodies discovered to date incorporate IGHD in only one reading frame. In the utilized reading frame, IGHD segments contain a repeating GYG motif (even for shorter IGHD), while alternate reading frames are severely hydrophobic

or contain stop codons (see Wang et al. (2013)). Thus, cattle and other species with limited recombination potential must employ innovative approaches for augmenting receptor diversity (Berens et al. 1997; Deiss et al. 2019; Lopez et al. 1998; Ma et al. 2016; Wang et al. 2013). One mechanism used by cattle (and other ruminants) is the use of AID-catalyzed SHM in the periphery, extensively mutating CDR H3 of the naïve repertoire within the ileal Peyer's patch and spleen (Liljavirta et al. 2013, 2014; Zhao et al. 2006). Additionally, cattle genomes contain ultralong-encoding IGHV (IGHV1-7) and IGHD (IGHD8-2) gene segments that fold into elongated CDR H3 over 50 amino acids in length (Stanfield et al. 2016; Wang et al. 2013; Warner Jenkins et al. 2022). Thus, cattle generate both a conventional antibody repertoire (comprising about 90% of the total repertoire), in which CDR H3 sequences typically encode up to 40 amino acids, and an ultralong repertoire (about 10% of the repertoire), in which CDR H3 sequences encode 50–70 amino acids (Deiss et al. 2019; Ma et al. 2016; Svilenov et al. 2021; Wang et al. 2013).

A cattle ultralong CDR H3 antibody contains two fairly conserved microdomains, a disulfide-bonded “knob” that sits atop a β -ribbon “stalk,” both of which are formed by recombining specific germline VDJ gene segments into an elongated CDR H3 (Fig. 1). Ultralong IgH are encoded exclusively by IGHV1-7 and IGHD8-2 (typically with IGHJ2-4) gene segments and can be distinguished from other gene segments by several conserved motifs (Berens et al. 1997; Deiss et al. 2019; Warner Jenkins et al. 2022). First, an 8-bp duplication at the 3' end of the V segment extends the nucleotide sequence. This duplication encodes a TTVHQ (threonine-threonine-valine-histidine-glutamine) motif C-terminal to the usual YYC (tyrosine-tyrosine-cysteine) found at the end of a conventional antibody V region and forms a critical β -ribbon ascending strand of the stalk domain. Second, the beginning of IGHD8-2 contains a conserved CPDG (cysteine-proline-aspartic acid-glycine) motif that forms a β -turn at the base of the knob. The remainder of the knob and the beginning of the descending β -ribbon strand of the stalk is encoded solely by the IGHD. Within the knob, repetitive codons for glycine (GGT), tyrosine (TAT), and serine (AGT) residues can be somatically mutated to cysteine (TGT) with a single base change, and rearranged loci undergo extensive SHM within the sequences encoding knob domains, thereby altering the number of noncanonical cysteine residues within the knob (Wang et al. 2013). The paired cysteines form disulfide bonds, designing unique loop patterns that ultimately shape the final paratope diversity (Deiss et al. 2019; Di et al. 2022; Haakenson et al. 2019, 2018; Smider and Smider 2020; Stanfield et al. 2018). Finally, the end of IGHD8-2 encodes a pattern of alternating aromatic tyrosine residues (YxYxY) that (with

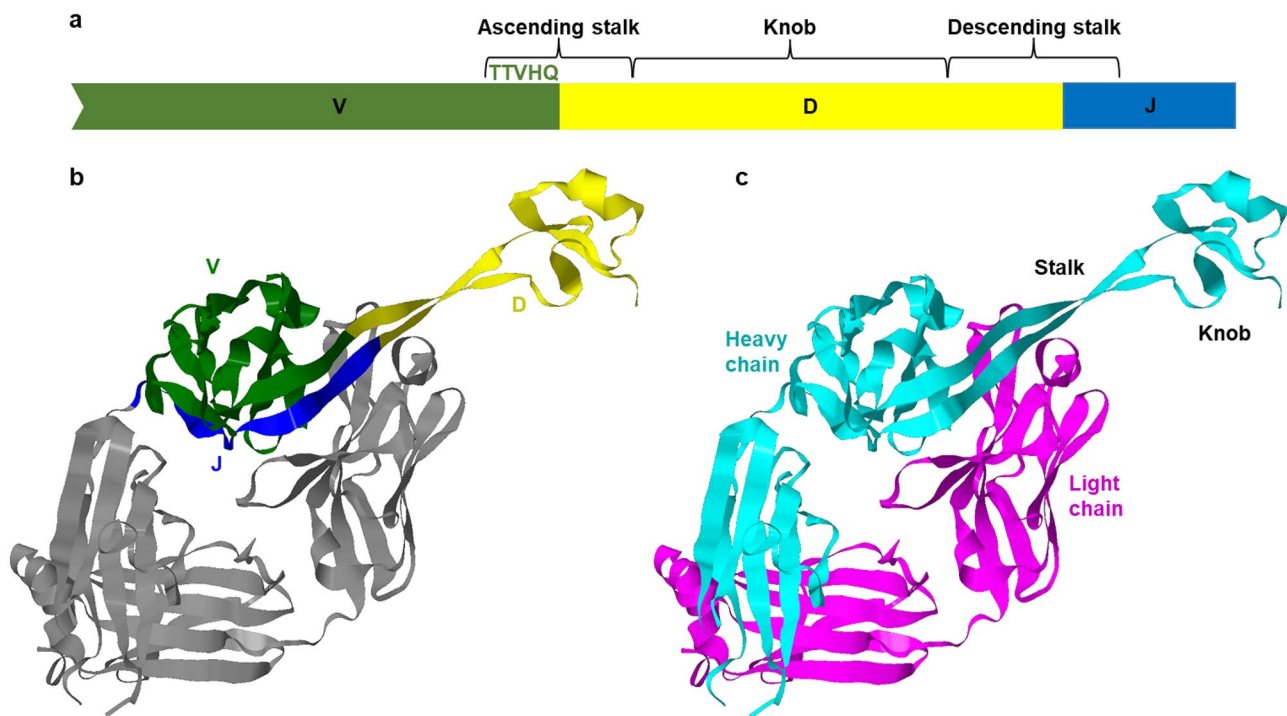


Fig. 1 Structural model illustrating the stalk and knob structure characteristic of cattle ultralong antibodies. **a** The linear form of an ultralong antibody heavy chain variable region with the gene segments color coded (variable, V: green, diversity, D: yellow; joining, J: blue). The portions of the stalk and knob that correspond to these segments

are denoted above the segments. **b** A crystal structure of the bovine ultralong antibody (PDB:6E9H; Dong et al. 2019) with the V, D, and J segments shaded as in **a**. **c** The same crystal structure illustrating the heavy and light chains (heavy chain: cyan; light chain: magenta), with the knob and stalk domains annotated

the beginning of IGHJ2-4) forms the descending strand of the stalk domain and is critical for the integrity of the stalk structure. The ascending and descending strands of the stalk are held together by hydrogen bonds between the eight or twelve amino acid pairs of the strands (Stanfield et al. 2018). While the stalk is important for the overall stability of the ultralong antibody, it is itself supported by the other five CDR, which do not appear to function in epitope binding. Of the six CDR that form the antigen binding site in a conventional antibody, only the knob of an ultralong antibody appears to bind antigen (Stanfield et al. 2020; Svilenov et al. 2021; Wang et al. 2013). Thus, stalk length evolved for optimal folding, stability, and binding to antigen, while disulfide bonds within the knob rigidify the knob structure for stability during antigen binding (Stanfield et al. 2020, 2016; Svilenov et al. 2021).

The knob and stalk microdomains create protracted paratopes, allowing the knob to reach for recessed antigen epitopes (Criscitiello 2021; Stanfield et al. 2020; Svilenov et al. 2021; Wang et al. 2013). Unique antigen receptors with “reach” can have important therapeutic applications. For example, cattle immunized with stabilized HIV spike trimer protein generated broadly neutralizing ultralong antibodies

against HIV (Sok et al. 2017). While clinical applications progress, however, we are only just beginning to understand the genomic mechanisms that yielded ultralong-encoding IGHV and IGHD gene segments in the species that make them (Smider and Smider 2020; Warner Jenkins et al. 2022). The evolutionary origin of these two gene segments—as well as any specific functional purpose of ultralong antibodies in cattle—remains unknown. Until recently, transcripts of expressed ultralong CDR H3 antibody mRNA had only been reported from cattle. However, Wu and others reported expression of ultralong IgH in the domestic yak (*Bos grunniens*) (Wu 2023), suggesting that ultralong-encoding gene segments originated in an ancestor of cattle and yak. We were curious if genomes of other bovid species also contain ultralong-encoding IGHV and IGHD gene segments. Thus, we examined the genomes of cattle, yak, and ten other animals within the Bovidae family to determine if these gene segments were present and if transcripts of ultralong antibodies are expressed in the repertoire of the earliest bovid species whose genome contains these genes. With these data, we can suggest a natural history for the immunogenetic components required to rearrange a mature ultralong CDR H3-encoding immunoglobulin heavy chain.

Methods

Species selection and construction of a phylogenetic tree

We selected species in this study based on two criteria: (1) taxonomic relatedness to the taurine cattle (*Bos taurus*) and (2) availability of an assembled genome for gene segment analysis. Using published phylogenies to inform species placement, we designed a phylogenetic tree of the major clades within the Bovidae family (Chen et al. 2019; Hassanin and Ropiquet 2004; Hernández Fernández and Vrba 2005; Proskuryakova et al. 2019). Additionally, we incorporated the most updated taxonomy from the Integrated Taxonomic Information System (ITIS; www.itis.gov; <https://doi.org/10.5066/F7KH0KBK>). We preferentially chose species closely related to *Bos taurus* over those more distantly related. Thus, we selected four species from the genus *Bos*: (1) *B. taurus indicus* (zebu), (2) *B. frontalis* (domestic gaur), (3) *B. grunniens grunniens* (domestic yak), and (4) *B. grunniens mutus* (wild yak).

For most genera within Bovidae, assembled genomes existed for only one individual species, and many assemblies still had unresolved regions and unplaced contigs, making species selection challenging. However, we were able to select one representative species from each of seven closely related genera within two subfamilies of Bovidae: (1) *Bison bison* (American bison), (2) *Bubalus bubalis* (domestic river buffalo), (3) *Syncerus caffer* (African buffalo), and (4) *Tragelaphus eurycerus* (bongo) from the subfamily Bovinae and (5) *Ovis aries* (domestic sheep), (6) *Ammotragus lervia* (Barbary sheep), and (7) *Capra hircus* (domestic goat) from the subfamily Caprinae (Online Resource 1). These twelve species span eight genera within the family Bovidae. We then refined our phylogenetic tree to include branches for only those species selected for analysis. To simplify nomenclature, we will refer to species by common name or use only their subspecies name (e.g., *Bos mutus*) in this paper (see Online Resource 1).

Identifying IGHV1-7 orthologs in other Bovidae species

We extracted all annotated IGHV gene segments from chromosome 21 of an updated bison genome assembly (Bis-Bis3) (Stroupe et al. 2023) that resolves much of the Ig heavy chain locus. We then aligned all functional IGHV from bison to the ultralong-encoding IGHV gene segment (IGHV1-7) and the most widely utilized conventional IGHV gene segment (IGHV1-10) from cattle. From this alignment, we ascertained which bison sequences were most similar to

both the ultralong and conventional sequences in cattle. We compared these two sets of sequences for similarity in both nucleotide identity and placement of AID-preferred hotspot motifs (WRCH/DGYW).

Using the genomic sequence for exons 1 and 2 (without the intron) of the cattle ultralong IGHV1-7 (accession KT723008.1, chromosome 21:254411–254850) as query, we performed a standard nucleotide BLAST (Altschul et al. 1990; Cunningham et al. 2021; Sayers et al. 2022) against the genome assembly (using either RefSeq Genome database or whole-genome shotgun contigs) of the other ten Bovidae species in this study, optimizing for dissimilar sequences (discontiguous megablast). We downloaded both the complete sequence (chromosome, scaffold, or contig) and the aligned sequences (IGHV exons) for all matches from the top genomic hit. We then mapped the aligned sequences back to the complete sequence using the Geneious mapper with default parameters in Geneious Prime 2022.1.1 (<https://www.geneious.com>). Using the matched regions, we annotated the two exons of each IGHV gene segment and the heptamer of the recombination signal sequence (RSS) at the end of exon 2 and extracted all regions containing both exons and the heptamer. We manually identified intron/exon boundaries between exons 1 and 2 using the cattle sequence as a reference and removed introns. We then aligned all IGHV sequences from each species with the cattle ultralong IGHV1-7 using Clustal Omega, ordering them by the number of differences to the cattle sequence. We selected the nucleotide sequence from each species that was most identical to the cattle ultralong-encoding IGHV, translated each sequence, and aligned both nucleotide and protein sequences for all twelve Bovidae species. We utilized only exon 2 (containing the second part of the leader and the V segment itself) for our analyses, examining all sequences for the 8-bp duplication encoding the signature TTVHQ motif of the cattle IGHV1-7. Additionally, we located an ultralong-encoding IGHV from assembled SRA reads (SRX1265662: SRR2463292) of auroch (*Bos primigenius*), the extinct ancestor of *Bos taurus* and *Bos indicus*.

Identifying IGHD8-2 orthologs in other Bovidae species

From the updated bison genome, we also extracted all annotated IGHD and IGHJ gene segments. We aligned all functional IGHD gene segments to the cattle ultralong-encoding IGHD8-2*01 gene segment and ordered sequences by decreasing length. We then compared the longest IGHD sequence from bison to the cattle IGHD8-2 sequence for both sequence identity and to determine if the bison sequence could encode a stalk and knob structure similar to that seen in cattle. Additionally, we aligned all IGHJ from bison and evaluated sequence conservation.

Simple BLAST searches using the cattle ultralong IGHD8-2 as query did not result in any matches within the evaluated genomic databases. Instead, we downloaded the entire chromosome, scaffold, or contig on which we located an IGHV gene segment and mapped all bison and cattle IGHD gene segments to the genomic sequence for each species using the Geneious mapper tool in Geneious Prime. We included the option to map IGHD segments to all of the best locations to improve identification. Additionally, we searched the genomic region for conserved heptamer and nonamer RSS to identify IGHD segments in genomic regions between IGHV and IGHJ gene segments. We then arranged all IGHD by amino acid length and chose the longest IGHD segment that contained alternating tyrosine (Y) and serine (S) or glycine (G) residues (e.g., YSYGY) for analysis. We were unable to locate any IGHD sequences from *Bos frontalis* (gayal) or *Ammotragus lervia* (Barbary sheep) genomes, likely due to the short scaffold lengths in these two species. Using Clustal Omega in Geneious, we aligned IGHD sequences from all remaining species to the cattle ultralong IGHD8-2 and ordered them by decreasing length.

Amplification of bison IGH transcripts and Amplicon-EZ sequencing

Joe Graham with Wild Horse Graham Ranch (Brazos County, TX, USA) generously donated the spleen from an American bison. We purified spleen total RNA using a RNeasy Mini Kit (Qiagen, Hilden, Germany) and generated cDNA using first strand SuperScript III (Invitrogen, Waltham, MA, USA) reverse transcriptase, following both manufacturer's instructions. We amplified target sequences using DreamTaq PCR master mix (Invitrogen) and primers designed to amplify ultralong IgM, IgG, and IgA isotypes (see Online Resource 2 for primer sequences) through PCR amplification. These primer sets bound more specifically to the ultralong V gene segment (IGHV1-7) than to the conventional V gene segment (IGHV1-10) used most often in cattle (and presumably in bison as well). We validated amplicon lengths using 0.08% agarose gel electrophoresis and extracted all bands of appropriate size (469–516 bp). We purified amplicon DNA from bands using a Pure-link Gel Purification kit (Invitrogen). We measured DNA quantity (ng/μl) with Qubit 4 Fluorometer and DNA quality (260/280) with a Nanodrop ND-1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). We combined 593 ng amplicon cDNA from all three isotypes (109 ng IgM, 234 ng IgG, 250 ng IgA) in 60 μl water for amplicon sequencing (Amplicon-EZ; Azenta Life Sciences, South Plainfield, NJ, USA).

Analysis of bison Amplicon-EZ sequences

We obtained 581,612 sequencing reads via Amplicon-EZ sequencing. Using default options in Geneious Prime (Biomatters, Ltd, Auckland, New Zealand, v2022.1.1), we merged sequences into 275,862 paired reads using the BBMerge paired read merger, removed duplicate reads using the Dedupe duplicate read remover, and extracted all reads greater than 300 bp in length (to ensure reads were long enough to identify both the variable and constant regions). We created a bison VDJ-C germline reference sequence for each isotype by concatenating germline ultralong V, D, and J gene segments (IGHV1-7, IGHD8-2, and IGHJ2-4) with constant regions for IgM, IgG, or IgA. We then simultaneously mapped the remaining 159,776 paired reads (read length 300–491 bp; mean = 388.6 bp) to all three germline isotype reference sequences using default mapper settings in Geneious Prime. Geneious mapper assembled 150,123 reads to three germline references to create three contigs.

We confirmed that amplicons within a contig were ultralong sequences using NCBI BLASTn (Zhang et al. 2000). Using the constant region to validate amplicon isotype, we removed amplicons that did not contain the 8-bp duplication encoding the “TTVHQ” ultralong IGHV sequence motif, realigned remaining nucleotide sequences to the isotype germline reference using the Clustal Omega alignment tool in Geneious, and removed duplicate sequences. We removed 72 (7%) putatively non-functional transcripts (12 IgM; 20 IgG; 40 IgA) from alignments (those containing frame shifts or stop codons within the coding region). We then manually adjusted alignments as needed and annotated the nucleotide boundaries of each V, D, and J gene segment as well as flanking non-template and palindromic (N and P) nucleotides between segments. Ultimately, we analyzed 970 unique bison ultralong sequences (240 IgM; 338 IgG; 392 IgA). We deposited all 970 amplicon sequences into GenBank under accession numbers OQ406489–OQ406728 (IGHM), OQ406729–OQ407066 (IGHG), and OQ407067–OQ407458 (IGHA).

We modeled the stalk and knob structures from four bison ultralong IgG amplicons containing different numbers of paired cysteine residues in the IGHD using SWISS-MODEL (<https://swissmodel.expasy.org/>) (Waterhouse et al. 2018). We based models on cattle templates of ultralong antibody structures (PDB, Protein Data Base; rcsb.org) (Berman et al. 2000) containing the same number of cysteine residues (though not necessarily cysteine position) in the IGHD as the bison amplicon sequences. We imported both bison models and cattle PDB templates into Geneious Prime for visualization of structural components.

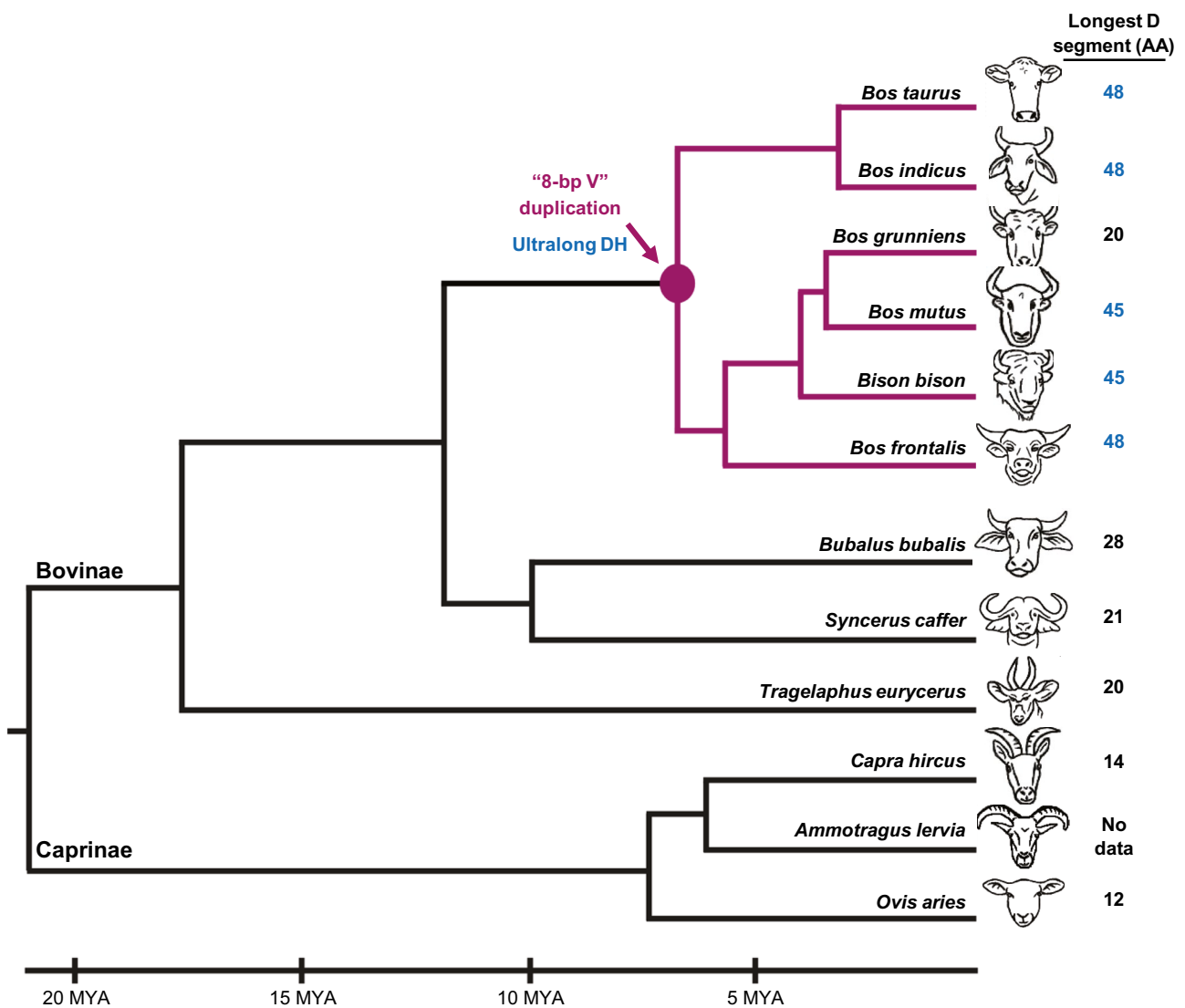


Fig. 5 The evolutionary event leading to the ultralong-encoding IGHV and IGHD gene segments occurred in a common ancestor of the *Bos* and *Bison* group. A phylogenetic tree of the major clades within the Bovidae family represents eight genera. The 8-bp duplication event that led to the evolution of the ultralong IGHV gene segment arose at the base of the *Bos* and *Bison* group, indicated by the circle and clade colored purple. The amino acid length of the longest known IGHD (which forms the knob structure in cattle ultralong antibodies) is shown to the right of each branch next to a line drawing

of each species (see Supplementary Table 1). The origin of ultralong-encoding IGHD gene segments (colored red) likely coincides with the origin of the ultralong-encoding IGHV. We did not find elongated IGHD segments in domestic yak (*Bos grunniens*), and we could not locate any IGHD segments in Barbary sheep (*Ammotragus lervia*), though incomplete genome assemblies limited our exploration. An approximate evolutionary timeline is provided below the tree (MYA, million years ago)

ultralong IGHD sequence for wild yak was 139-bp in length and encoded 45 amino acids (Smider and Smider 2020) and shared 75% amino acid (81% nucleotide) identity with cattle, while the gayal genome contained a 148-bp (48 amino acid) IGHD but was less conserved (64% amino acid/78% nucleotide identity). Finally, the bison genome contained one IGHD gene segment (IGHD8-2) that was similar in both sequence length and identity to the ultralong-encoding IGHV of cattle, with a sequence length of 139-bp encoding 45 amino acids. The bison ultralong IGHD shared 75%

amino acid (81% nucleotide) identity to cattle IGHV8-2 but 91% amino acid (93% nucleotide) identity with wild yak. The longest IGHD gene segment found in the domestic yak encoded an IGHD segment of only 20 amino acids, despite the presence of an ultralong-encoding IGHV. However, we identified only three IGHD gene segments on chromosome 17 in domestic yak and at least five IGHD along the 13.6-Kb long scaffold containing the ultralong-encoding IGHD in wild yak, suggesting that the heavy chain locus assembly may be incomplete in domestic yak.

sequences encoding either a YYCAR/K (conventional) or YYCTTVHQ (ultralong) motif at the 3' end of the V segment, indicating that bison express both conventional and ultralong IGHV in spleen. Amplicon CDR H3 shared only 35–67% nucleotide (25–59% amino acid) sequence identity to our concatenated IgG germline sequence (IGHV1-7-IGHD8-2-IGHJ2-4), suggesting that bison employ AID-mediated SHM to alter the structure of the knob domain in a manner similar to cattle (Online Resource 5). In fact, our amplicon database contains sequences likely in various stages of affinity maturation, with sequences containing similar N and P nucleotides but different somatically mutated IGHD regions (Online Resource 5).

Within our amplicon database, IGHD regions contained between one and nine cysteine residues. However, the majority of sequences contained an even number of cysteines, and more than half of all amplicons (51.5%) contained six cysteine residues (Fig. 7a). To visualize sequences containing different numbers of paired cysteines, we modeled bison amplicon sequences with 2, 4, 6, or 8 IGHD-encoded cysteine residues using cattle ultralong antibody structures as templates (Fig. 7b). These models suggest that paratopes of bison IgH antibodies are likely as diverse as those seen in cattle (Fig. 7c). The number of paired cysteines shaped a wide assortment of folding patterns within the knob region, likely resulting in a varied paratope repertoire.

Discussion

An IGHV gene segment utilized in ultralong CDR H3 antibodies (IGHV1-7 in cattle) is defined by its 8-bp duplication at the terminal end of the IGHV region. This duplication encodes a fairly conserved TTVHQ motif that initiates the ascending loop of the stalk domain in cattle ultralong CDR H3 antibodies (see Fig. 1). This motif interacts with another highly conserved motif within the CDR H1 of ultralong antibodies and helps to stabilize the ultralong CDR H3 structure. Unlike ultralong CDR H3 antibodies, CDR H1 of conventional antibodies with shorter CDR H3 are quite divergent, suggestive of its role in binding diverse epitopes (Wang et al. 2013). To determine if ruminants other than cattle are capable of generating ultralong CDR H3 antibodies, we searched the genomes of twelve species within the family Bovidae for ultralong-encoding IGHV gene segments. We identified the 8-bp duplication encoding the TTVHQ motif in genomic IGHV segments of cattle and five additional species within Bovidae (zebu, domestic yak, wild yak, bison, and gayal). We also identified the 8-bp duplication in auroch, the extinct ancestor to taurine and zebu cattle (Fig. 4). All seven species with ultralong-encoding IGHV fell within the monophyletic group containing the genera *Bos* and *Bison* (Figs. 4 and 5;

Online Resource 1). We found no evidence of the duplication in any more distantly related species.

We defined a “genuine” ultralong IGHD sequence as one that begins with a CPDG motif and ends with a repeating YxYxY motif. Again, these motifs are relatively conserved features of expressed ultralong CDR H3 in cattle. While the CPDG motif forms a β -turn at the base of the knob, the YxYxY motif shapes the descending strand and interacts with the ascending strand of the stalk through hydrogen bonding. These motifs also interact with CDR H1 and CDR H2 and CDR L1 and CDR L3, supporting the base of the stalk (Wang et al. 2013). Additionally, we searched for the presence of repetitive codons for glycine (GGT), tyrosine (TAT), and serine (AGT) residues within the knob domain that could be somatically mutated to cysteine (TGT) with a single base change, altering the number of noncanonical cysteine residues and enabling loop formation within the knob region of the antibody (Haakenson et al. 2019).

IGHD gene segments are naturally short, and highly scaffolded genome (or IgH locus) assemblies for some species, like in African buffalo, bongo, and Barbary sheep, compounded the difficulty in locating them. Regardless, we identified IGHD gene segments encoding 20 amino acids or larger in all nine species (including cattle) within the subfamily Bovinae (Fig. 6), while domestic sheep and goat (subfamily Caprinae species) had IGHD that were much shorter (12 and 14 amino acids, respectively). We were unable to find any IGHD gene segments in the current Barbary sheep genome assembly. Within the *Bos* and *Bison* group, we identified IGHD gene segments encoding 45 to 48 amino acids in all but the domestic yak genome. The longest IGHD we found in the domestic yak genome encoded only 20 amino acids, sharing subspecies status with wild yak (with an IGHD encoding 45 amino acids). Furthermore, Wu et al. (2023) reported ultralong CDR H3 antibody expression in domestic yaks, suggesting the lack of an ultralong IGHD in the domestic yak genome may indicate incomplete IgH locus assembly rather than a true absence of the gene segment itself. The longest IGHD gene segments we located in water buffalo, African buffalo, and bongo genomes only encoded proteins up to 28 amino acids. Furthermore, sequence identity to cattle IGHV1-7 was low, with amino acid conservation ranging from 30% in bongo to 38% in water buffalo (Fig. 6). The shorter IGHD of caprine species also lacked sequence conservation to cattle IGHV1-7 (21 and 25%, respectively).

Genomes of only five Bovidae species (cattle, zebu, wild yak, bison, and gayal), all within the *Bos* and *Bison* group, contained IGHD that matched our criteria for being genuine ultralong-encoding IGHD. Outside this group, there were no IGHD that met all criteria. While it is possible that ultralong-encoding IGHD gene segments could be found in other

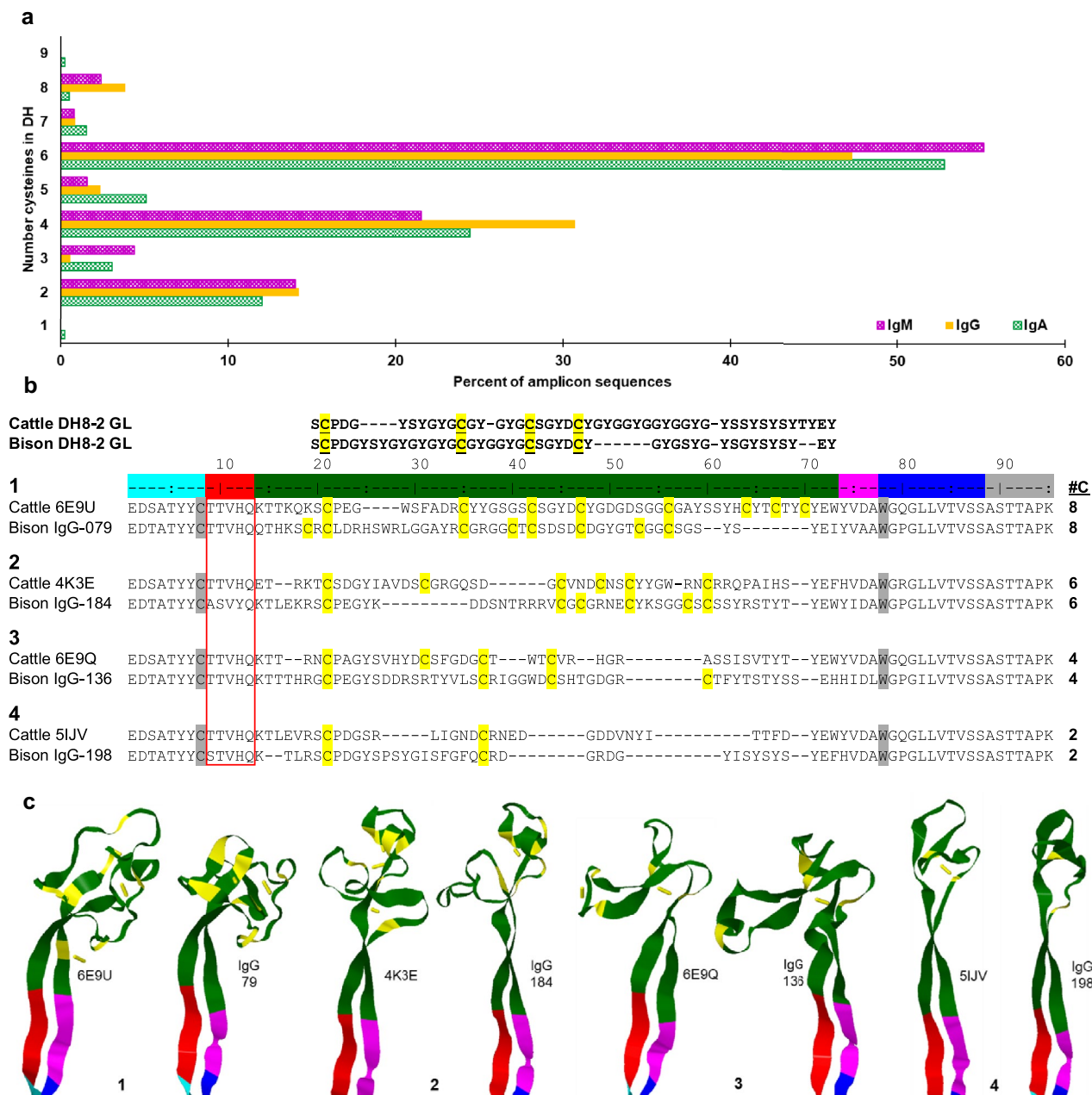


Fig. 7 Modeled structures of bison ultralong IgH antibodies reflect the same knob diversity as seen in cattle. **a** Mutation within bison (*Bison bison*) IGHD usually results in an even number of cysteine (C) residues within the IGHD-encoded knob structure, permitting disulfide (C=C) bonds to form. **b** Modeled bison amplicon sequences aligned with cattle template sequences for each model [1–4]. Conserved cysteine (C) and tryptophan (W) residues at IMGT positions 104 and 118, respectively (highlighted in gray), delineate the boundaries of the CDR H3. Canonical cysteine residues within the CDR H3 are highlighted yellow. The scale is colored as follows: V segment to the cysteine at position 104 of the YYC motif (cyan), TTVHQ motif (red), D segment (green), J segment up to the tryptophan at position 118 of the WGxG motif (magenta), remaining J

segment (blue), constant region (gray). **c** Bison ultralong IgG amplicons modeled against cattle templates illustrate the diversity of knob regions with decreasing numbers of cysteines (1: eight Cs; 2: six Cs; 3: four Cs; 4: two Cs). For each pair of structures, both the cattle template (left) and the bison amplicon sequence (right) contain the same number of cysteines within the IGHD-encoded knob. Disulfide bonds and the paired cysteine residues are colored yellow. Either the PDB entry ID# (cattle templates; protein data base, rcsb.org) or amplicon number (bison IgG sequence) are shown next to each model. Amplicon sequences are colored according to the scale (see **b**) to indicate gene segment and residue position. Values to the right of the sequences indicate the total number of cysteines within CDR H3. For a more detailed summary of templates, see results

bovid species as genome assemblies improve, the concurrence of ultralong-encoding IGHD in only those genomes containing ultralong-encoding IGHV suggests that both of these gene segments most likely evolved together. However, Philp (2018) reported expression of an ultralong CDR H3 IgG phenotype in African buffalo (*Syncerus caffer*) after challenge with foot-and-mouth disease virus, with CDR H3 up to 71 amino acids in length occurring in 5% of the sequenced population. Furthermore, CDR H3 were rich in cysteine, glycine, and tyrosine residues, and IGHD of sequenced transcripts aligned with the ultralong-encoding IGHD gene segment of cattle (Philp 2018). Thus, while we were unable to find either an ultralong-encoding IGHV or IGHD gene segment in the current genome assembly, African buffalo must assemble ultralong CDR H3 using orthologous gene segments or through some alternate mechanism, like IGHD-IGHD fusions. To understand the origin of the ultralong-encoding gene segments more definitely, it is essential that current genome assemblies are improved in species like African buffalo.

To date, studies have reported expressed ultralong-encoding IGHV or IGHD gene segments only in cattle and yak (Saini et al. 1999; Smider and Smider 2020; Wang et al. 2013; Wu et al. 2023). Nucleotide BLASTs of cattle ultralong transcripts against all other Bovidae species produced hits only to transcripts containing conventional IGHV gene segments, indicating that either no other bovid species expresses ultralong V regions (including those species with an ultralong-encoding IGHV) or (more likely) there simply have not been enough studies generating these data in other species. Our amplicon database confirms that bison also express ultralong V regions of IgM, IgG, and IgA isotypes in spleen. All bison amplicons in our database expressed the ultralong variable region phenotype incorporating IGHV1-7, IGHD8-2, and JH2-04 (a or b) gene segments, similar to gene segment usage in cattle ultralong variable regions (Online Resources 4 and 5). Bison IGHV gene segments were highly conserved, with functional protein sequences sharing over 92% identity with each other. Bison IGHV gene segment sequences also were similar across Bovidae, with protein conservation of 91–99% between all twelve species (Fig. 4) and over 97% between bison and all *Bos* species, including to the extinct auroch (*B. primigenius*). As expected, sequence conservation was higher between bison IGHD8-2 and other ultralong-encoding IGHD gene sequences (> 45 amino acids; > 64% identity) than to shorter conventional IGHD (12–28 amino acids; < 38% identity) (Fig. 6).

Bison amplicon sequences shared many similar features of published cattle ultralong amplicons. First, we observed clear deletions within the knob domains of the amplified bison repertoire (Online Resource 5) (Deiss et al. 2019; Dong et al. 2019). These deletions remove interior

nucleotides, leaving the regions encoding the CPDG and YxYxY motifs intact. In cattle, these internal deletions changed the three-dimensional positions of cysteine residues, altering loop lengths and disulfide bonding patterns within the knob structure (Deiss et al. 2019). Alternatively, deletions could be associated with undiscovered IGHD8-2 polymorphisms (Warner Jenkins et al. 2022). We observed similar positional changes in cysteine residues within bison amplicon sequences (Online Resource 5). Second, between the TTVHQ motif of the IGHV and the CPDG motif of the IGHD, bison ultralong amplicons contained loosely conserved IGHV-IGHD junctional sequences composed primarily of adenine bases (60%). These additions are similar to the “conserved short nucleotide sequences” (CSNS) inserted into the IGHV-IGHD junction in adult cattle (Koti et al. 2010). The presence of CSNS-like insertions in bison IGHV-IGHD junctions suggests that bison ultralong antibodies may also engage similar mechanisms as cattle to generate flexibility and diversity in their ultralong antibody repertoire. Lastly, as observed in cattle amplicons, bison IGHD8-2 typically encodes an acidic residue (D/E) subsequent to the YxYxY motif at the terminal end of the CDR H3. This negatively charged residue occurred in 90% of bison amplicons. Located in the descending strand, this acidic residue appears to interact through hydrogen bonding with the first amino acid encoded by CSNS (or N and P) insertions, which in bison is most often a basic lysine residue (K, 35–45%). Thus, it is possible that these CSNS sequences are inserted into IGHV-IGHD junctions to assist structural stability within the stalk domain.

Bison ultralong-encoding IGHV and IGHD are remarkably similar to those found in cattle. Bison IGHV1-7 shared 98% nucleotide and amino acid identity with cattle IGHV1-7, with only seven nucleotide (three amino acids) differences between sequences (Figs. 2 and 4). Cattle and bison IGHD8-2 shared 80% nucleotide identity, with only eleven base pair differences (plus two short deletions) (Figs. 3 and 6). Thus, it is feasible that the presence of these gene segments in bison is a consequence of cattle introgression following hybridization between bison and cattle in the late 1800s (Stroupe et al. 2022). However, we located ultralong-encoding IGHV in all members of the *Bos* and *Bison* group, including *Bos frontalis*, which is evolutionary older than bison, and *Bos primigenius*, the ancestor to extant taurine (*Bos taurus*) and zebu (*Bos indicus*) cattle that went extinct in the early 1600 s (Zeyland et al. 2013). Furthermore, we found ultralong-encoding IGHD in five of these six species. Ultralong-encoding IGHV and IGHD gene segment sequences are highly conserved (Figs. 4 and 6), reflecting a common origin. It is possible that these gene segments emerged in a Bovinae ancestor (at the divergence with Caprinae) and then subsequently were lost in

all non-*Bos* and *Bison* lineages, though this explanation is not the most parsimonious. Accordingly, we propose that both the duplication event leading to this ultralong-encoding IGHV gene segment and the likely emergence of the ultralong-encoding IGHD gene segment evolved in a common ancestor of the *Bos* and *Bison* genera. These results may support a taxonomic change where *Bison bison* is moved into the genus *Bos*.

While no crystallographic or electron microscopy structural data yet exists for bison ultralong CDR H3 antibodies, modeling bison amplicon sequences using cattle ultralong templates indicated that ultralong CDR H3 proteins in bison are probably structurally similar to those of cattle. Side-by-side graphics of cattle templates and bison models illustrated similar folding patterns, where IGHD with the same numbers of cysteines formed generally similar shapes within the knob (Fig. 7). For example, sequences with eight cysteines showed four disulfide bonds and four loops in both template and model. We observed similar patterns between templates and models with six or four cysteines, where three or two disulfide bond(s) and three or two loop(s) formed, respectively. In bison models with two cysteines, the positions of the cysteine residues restricted the formation of the disulfide bond between them. However, except for the first cysteine (within the CPDG motif), cysteine positions within germline-encoded IGHD gene segment sequences between bison models and individual cattle antibody crystal structures do not overlap (see Fig. 3). Because there are no bison ultralong structures to use as templates, it is unclear whether the cysteines could bond in a bison structure. For most models, cysteine position did affect overall paratope shape, corroborating the idea that cysteine position is an important driver of diversity (Haakenson et al. 2019; Warner Jenkins et al. 2022).

Though the conserved CPDG motif occurred only in ultralong-encoding IGHD gene segments, IGHD from both bovine and caprine species incorporated a repeating YxYxY motif at the C-terminal end in at least one IGHD sequence, suggesting it is functionally important in bovine antibodies. Structural models of bison amplicons lacking this CPDG motif nevertheless resemble those models containing the motif (see Fig. 7; Online Resource 6). Furthermore, conventional cattle antibodies typically contain much longer than average CDR H3 compared to other species (Deiss et al. 2019; Wang et al. 2013). Thus, bovine species whose IGHD segments encode 20 to 28 amino acids still construct antibodies with long CDR H3, and these longer loops may have an analogous function as ultralong antibodies in species that do not construct them. It also should be noted that while IGHD gene segments are considerably shorter in caprine species, goats can elongate their CDR H3 by incorporating IGHD-IGHD fusions during recombination (V-D-D-J)

(Du et al. 2018). Thus, the necessity for elongated CDR H3 appears to be a shared condition within Bovidae.

While a specific function of ultralong CDR H3 antibodies is unknown, long CDR H3 (and particularly ultralong CDR H3) in *Bos* and *Bison* may have evolved in response to a rumen microbe or environmental (likely viral) pathogen specific to a bovine subsistence (Deiss et al. 2019; Stanfield et al. 2018; Wang et al. 2013; Wu et al. 2023). Cattle immunized against HIV antigens elicited a rapid and broadly neutralizing ultralong antibody response and are so far the only species capable of this response (Edwards et al. 2021; Sok et al. 2017). As a retrovirus, HIV has a complex structure, with the viral envelope coated in trimeric spikes composed of two glycoproteins—a surface unit (gp120) tethered to the envelope by a transmembrane unit (gp41) and a glycan shield that conceals the inner protein surface from immune recognition. Conventional antibodies have paratopes that are planar or undulating in shape, limiting epitope binding to these surface-exposed sites, constraining their neutralizing ability (Pantophlet and Burton 2006; Wyatt and Sodroski 1998). However, ultralong CDR H3 antibodies, with their knobs extending away from the remaining antibody structure, can penetrate grooves and bind these inner epitopes that are inaccessible to conventional antibodies (Stanfield et al. 2020). Ultralong CDR H3 antibodies have shown similar broadly neutralizing reactivity against *Sarbecovirus* and SARS-CoV-2 (Burke et al. 2022, 2020; Svilenov et al. 2021). It is possible that the positively charged knob region binds to membrane proteins, while the negatively charged conserved acidic residue of the stalk binds to surface proteins, eliciting greater neutralizing capability (Peng et al. 2022; Wang et al. 2013). Considering that cattle (and other ruminants) are exposed to numerous viral pathogens, including retroviruses, ultralong CDR H3 antibodies may have evolved to combat the specific structure of these viruses. To best understand the evolutionary importance of these elongated CDR H3, it is essential that future studies examine how ultralong (and long) CDR H3 antibodies function in the animals that produce them.

The use of AID-mediated SHM to create primary receptor repertoires is not unique to cattle ultralong antibodies. Cattle also diversify their restricted antibody repertoires through AID-mediated SHM within the fetal ileal Peyer's patch and spleen in the absence of exogenous antigens (Liljavirta et al. 2013, 2014; Zhao et al. 2006). Chickens, which undergo a single VDJ recombination event, utilize AID-mediated somatic gene conversion to generate diversity, whereby portions of gene sequences from upstream pseudogenes replace homologous sequences in the recombined variable region (de los Rios et al. 2015; Ratcliffe 2006). Sharks and camelids diversify T cell receptors (TCR) by incorporating AID-mediated somatic mutations during thymic development,

likely to salvage receptors that otherwise would fail selection (Antonacci et al. 2011; Chen et al. 2012; Ciccarese et al. 2014; Ott et al. 2018, 2020). Finally, lamprey and hagfish utilize AID-like enzymes (CD1/CD2) to somatically rearrange variable lymphocyte receptor (VLR) loci to create diverse lymphocyte receptors (Mitchell and Criscitiello 2020; Pancer et al. 2004, 2005).

In addition to conventional antibodies, sharks (IgNAR) and camelids (IGHV_H) construct antibodies with binding regions from a single heavy chain variable domain, limiting the antigen binding surface to only three CDR loops that, like ultralong CDR H3 antibodies, can interact with epitopes in recessed or concave surfaces (reviewed in de los Rios et al. (2015) and Muyldermans and Smider (2016)). Camel IGHV_H, or nanobodies, have longer CDR H3 binding loops that increase flexibility and surface area for antigen binding (Desmyter et al. 1996; Muyldermans et al. 1994). Additionally, sharks (NAR-TCR) and opossum (TCR μ) make TCR with two heavy chain variable domains, with the membrane-proximal V domain serving a supportive role for the membrane-distal antigen-binding V domain (Criscitiello et al. 2006; Morrissey et al. 2021; Parra et al. 2007). These exceptional antibodies and TCRs with small, protruding paratopes have the potential to “reach” for antigen epitopes in places where the more planar shape of conventional antibodies cannot bind (Criscitiello 2021). For this reason, many have been appropriated for experimental human therapeutic or diagnostic applications (de los Rios et al. 2015; Muyldermans and Smider 2016; Sok et al. 2017; Yang and Shah 2020). The cattle ultralong CDR H3 knob (“picobody”) is the smallest and most extended of these examples. However, one distinctive feature of all these unusual receptors and diversifying mechanisms is that we do not fully understand how the receptors or mechanisms function in the immune repertoire of the organisms that employ them. This knowledge not only could improve applications to humans but could provide valuable insight into the evolution of adaptive immune antigen receptor diversification and pathogen recognition.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00251-023-01305-9>.

Declarations

Competing interests The authors declare no competing interests.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)

- Antonacci R, Mineccia M, Lefranc MP, Ashmaoui HM, Lanave C, Piccinni B, Pesole G, Hassanane MS, Massari S, Ciccarese S (2011) Expression and genomic analyses of Camelus dromedarius T cell receptor delta (TRD) genes reveal a variable domain repertoire enlargement due to CDR3 diversification and somatic mutation. *Mol Immunol* 48(12–13):1384–1396. <https://doi.org/10.1016/j.molimm.2011.03.011>
- Berens SJ, Wylie DE, Lopez OJ (1997) Use of a single VH family and long CDR3s in the variable region of cattle Ig heavy chains. *Int Immunol* 9(1):189–199. <https://doi.org/10.1093/intimm/9.1.189>
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28(1):235–242. <https://doi.org/10.1093/nar/28.1.235>
- Burke MJ, Scott JNF, Minshull TC, Gao Z, Manfield I, Savic S, Stockley PG, Calabrese AN, Boyes J (2022) A bovine antibody possessing an ultralong complementarity-determining region CDRH3 targets a highly conserved epitope in sarbecovirus spike proteins. *J Biol Chem* 298(12):102624. <https://doi.org/10.1016/j.jbc.2022.102624>
- Burke MJ, Stockley PG, Boyes J (2020) Broadly neutralizing bovine antibodies: highly effective new tools against evasive pathogens? *Viruses* 12(4):473. <https://doi.org/10.3390/v12040473>
- Chen H, Bernstein H, Ranganathan P, Schluter S (2012) Somatic hypermutation of TCR γ V genes in the sandbar shark. *Dev Comp Immunol* 37(1):176–183. <https://doi.org/10.1016/j.dci.2011.08.018>
- Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, Bibi F, Yang Y, Wang J, Nie W et al (2019) Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. *Science* 364(6446). <https://doi.org/10.1126/science.aav6202>
- Ciccarese S, Vaccarelli G, Lefranc MP, Tasco G, Consiglio A, Casadio R, Linguiti G, Antonacci R (2014) Characteristics of the somatic hypermutation in the Camelus dromedarius T cell receptor gamma (TRG) and delta (TRD) variable domains. *Dev Comp Immunol* 46(2):300–313. <https://doi.org/10.1016/j.dci.2014.05.001>
- Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ (2015) The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. *Philos Trans R Soc Lond B Biol Sci* 370(1676). <https://doi.org/10.1098/rstb.2014.0236>
- Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the AID/APOBEC family of polynucleotide (deoxy) cytidine deaminases. *Mol Biol Evol* 22(2):367–377. <https://doi.org/10.1093/molbev/msi026>
- Criscitiello M, Saltis M, Flajnik M (2006) An evolutionarily mobile antigen receptor variable region gene: doubly rearranging NAR-TcR genes in sharks. *Proc Natl Acad Sci* 103(13):5036–5041. <https://doi.org/10.1073/pnas.0507074103>
- Criscitiello MF (2021) Unusual T cell receptor in opossum. *Science* 371(6536):1308–1309. <https://doi.org/10.1126/science.abg7639>
- Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean Irina M, Austine-Orimoloye O, Azov Andrew G, Barnes I, Bennett R et al (2021) Ensembl 2022. *Nucleic Acids Res* 50(D1):D988–D995. <https://doi.org/10.1093/nar/gkab1049>
- de los Rios M, Criscitiello MF, and Smider VV (2015) Structural and genetic diversity in antibody repertoires from diverse species. *Curr Opin Struct Biol* 33:27–41. <https://doi.org/10.1016/j.sbi.2015.06.002>
- Deiss TC, Vadrnais M, Wang F, Chen PL, Torkamani A, Mwangi W, Lefranc MP, Criscitiello MF, Smider VV (2019) Immunogenetic factors driving formation of ultralong VH CDR3 in Bos taurus antibodies. *Cell Mol Immunol* 16:53–64. <https://doi.org/10.1038/cmi.2017.117>
- Desmyter A, Transue TR, Ghahroudi MA, Dao Thi M-H, Poortmans F, Hamers R, Muyldermans S, Wyns L (1996) Crystal structure of a camel single-domain VH antibody fragment in complex

- with lysozyme. *Nat Struct Mol Biol* 3. <https://doi.org/10.1038/nsb0996-803>
- Di Y, Cai S, Zheng S, Huang J, Du L, Song Y, Zhang M, Wang Z, Yu G, Ren L et al (2022) Reshaping the murine immunoglobulin heavy chain repertoire with bovine DH genes. *Immunology* 165(1):74–87. <https://doi.org/10.1111/imm.13407>
- Dong J, Finn JA, Larsen PA, Smith TPL, Crowe JE (2019) Structural diversity of ultralong CDRH3s in seven bovine antibody heavy chains. *Front Immunol* 10. <https://doi.org/10.3389/fimmu.2019.00558>
- Du L, Wang S, Zhu Y, Zhao H, Basit A, Yu X, Li Q, Sun X (2018) Immunoglobulin heavy chain variable region analysis in dairy goats. *Immunobiology* 223(11):599–607. <https://doi.org/10.1016/j.imbio.2018.07.005>
- Dufour V, Malinge S, Nau F (1996) The sheep Ig variable region repertoire consists of a single VH family. *J Immunol* 156(6):2163–2170 (PMID: 8690905)
- Edwards JM, Heydarchi B, Khoury G, Salazar-Quiroz NA, Gonelli CA, Wines B, Hogarth PM, Kristensen AB, Parsons MS, Purcell DFJ (2021) Enhancement of antibody-dependent cellular cytotoxicity and phagocytosis in anti-HIV-1 human-bovine chimeric broadly neutralizing antibodies. *J Virol* 95(13):e0021921. <https://doi.org/10.1128/jvi.00219-21>
- Haakenson JK, Deiss TC, Warner GF, Mwangi W, Criscitiello MF, Smider VV (2019) A broad role for cysteines in bovine antibody diversity. *Immunohorizons* 3(10):478–487. <https://doi.org/10.4049/immunohorizons.1900058>
- Haakenson JK, Huang R, Smider VV (2018) Diversity in the cow ultralong CDR H3 antibody repertoire. *Front Immunol* 9. <https://doi.org/10.3389/fimmu.2018.01262>
- Hassanin A, Ropiquet A (2004) Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Kouprey, *Bos sauveli* Urbain 1937. *Mol Phylogenet Evol* 33(3):896–907. <https://doi.org/10.1016/j.ympev.2004.08.009>
- Hernández Fernández M, Vrba ES (2005) A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level supertree of the extant ruminants. *Biol Rev Camb Philos Soc* 80(2):269–302. <https://doi.org/10.1017/s1464793104006670>
- Jackson KJL (2019) Study of human antibody responses from analysis of immunoglobulin gene sequences. In Ranganathan S, Gribskov M, Nakai K, Schönbach C (eds.) *Encyclopedia of Bioinformatics and Computational Biology*. Academic Press, Oxford
- Johnston CM, Wood AL, Bolland DJ, Corcoran AE (2006) Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region1. *J Immunol* 176(7):4221–4234. <https://doi.org/10.4049/jimmunol.176.7.4221>
- Jones JM, Gellert M (2004) The taming of a transposon: V(D)J recombination and the immune system. *Immunol Rev* 200:233–248. <https://doi.org/10.1111/j.0105-2896.2004.00168.x>
- Kirkham PM, Mortari F, Newton JA, Schroeder HW Jr (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *Embo J* 11(2):603–609. <https://doi.org/10.1002/j.1460-2075.1992.tb05092.x>
- Knight KL, Becker RS (1990) Molecular basis of the allelic inheritance of rabbit immunoglobulin VH allotypes: implications for the generation of antibody diversity. *Cell* 60(6):963–970. [https://doi.org/10.1016/0092-8674\(90\)90344-e](https://doi.org/10.1016/0092-8674(90)90344-e)
- Koti M, Kataeva G, Kaushik AK (2010) Novel atypical nucleotide insertions specifically at VH-DH junction generate exceptionally long CDR3H in cattle antibodies. *Mol Immunol* 47. <https://doi.org/10.1016/j.molimm.2010.02.014>
- Larimore K, McCormick MW, Robins HS, Greenberg PD (2012) Shaping of human germline IgH repertoires revealed by deep sequencing. *J Immunol* 189(6):3221–3230. <https://doi.org/10.4049/jimmunol.1201303>
- Lefranc M-P (2014) Immunoglobulin and T cell receptor genes: IMGT and the birth and rise of immunoinformatics. *Front Immunol* 5:22–22. <https://doi.org/10.3389/fimmu.2014.00022>
- Lefranc MP, Giudicelli V, Duroux P, Jabado-Michaloud J, Folch G, Aouinti S, Carillon E, Duvergey H, Houles A, Paysan-Lafosse T et al (2015) IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res* 43(Database issue):D413–22. <https://doi.org/10.1093/nar/gku1056>
- Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* 18(1):1–11. <https://doi.org/10.1101/gad.1161904>
- Liljavirta J, Ekman A, Knight JS, Pernthaner A, Iivanainen A, Niku M (2013) Activation-induced cytidine deaminase (AID) is strongly expressed in the fetal bovine ileal Peyer's patch and spleen and is associated with expansion of the primary antibody repertoire in the absence of exogenous antigens. *Mucosal Immunol* 6(5):942–949. <https://doi.org/10.1038/mi.2012.132>
- Liljavirta J, Niku M, Pessa-Morikawa T, Ekman A, Iivanainen A (2014) Expansion of the preimmune antibody repertoire by junctional diversity in *Bos taurus*. *PLoS ONE* 9(6):e99808. <https://doi.org/10.1371/journal.pone.0099808>
- Lopez O, Perez C, Wylle D (1998) A single VH family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunol Rev* 162. <https://doi.org/10.1111/j.1600-065X.1998.tb01429.x>
- Ma L, Qin T, Chu D, Cheng X, Wang J, Wang X, Wang P, Han H, Ren L, Aitken R et al (2016) Internal duplications of DH, JH, and C region genes create an unusual IgH gene locus in cattle. *J Immunol* 196(10):4358–4366. <https://doi.org/10.4049/jimmunol.1600158>
- Mikocziova I, Greiff V, Sollid LM (2021) Immunoglobulin germline gene variation and its impact on human disease. *Genes Immun* 22(4):205–217. <https://doi.org/10.1038/s41435-021-00145-5>
- Mitchell CD, Criscitiello MF (2020) Comparative study of cartilaginous fish divulges insights into the early evolution of primary, secondary and mucosal lymphoid tissue architecture. *Fish Shellfish Immunol* 107:435–443. <https://doi.org/10.1016/j.fsi.2020.11.006>
- Morrissey KA, Wegrecki M, Praveena T, Hansen VL, Bu L, Sivaraman KK, Darko S, Douek DC, Rossjohn J, Miller RD et al (2021) The molecular assembly of the marsupial γ T cell receptor defines a third T cell lineage. *Science* 371(6536):1383–1388. <https://doi.org/10.1126/science.abe7070>
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R (1994) Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng* 7(9):1129–1135. <https://doi.org/10.1093/protein/7.9.1129>
- Muyldermans S, Smider VV (2016) Distinct antibody species: structural differences creating therapeutic opportunities. *Curr Opin Immunol* 40:7–13. <https://doi.org/10.1016/j.coi.2016.02.003>
- Nezlin R (2019) Dynamic aspects of the immunoglobulin structure. *Immunol Invest* 48(8):771–780. <https://doi.org/10.1080/08820139.2019.1597110>
- Niku M, Liljavirta J, Durkin K, Schroeder E, Iivanainen A (2012) The bovine genomic DNA sequence data reveal three IGHV subgroups, only one of which is functionally expressed. *Dev Comp Immunol* 37(3–4):457–461. <https://doi.org/10.1016/j.dci.2012.02.006>
- Ott JA, Castro CD, Deiss TC, Ohta Y, Flajnik MF, Criscitiello MF (2018) Somatic hypermutation of T cell receptor α chain contributes to selection in nurse shark thymus. *eLife* 7:e28477. <https://doi.org/10.7554/eLife.28477>
- Ott JA, Harrison J, Flajnik MF, Criscitiello MF (2020) Nurse shark T cell receptors employ somatic hypermutation preferentially to alter alpha/delta variable segments associated with alpha constant region. *Eur J Immunol* 50(5):1307–1320. <https://doi.org/10.1002/eji.201948495>
- Pancer Z, Amemiya CT, Ehrhardt GRA, Ceitlin J, Larry Gartland G, Cooper MD (2004) Somatic diversification of variable

- lymphocyte receptors in the agnathan sea lamprey. *Nature* 430(6996):174–180. <https://doi.org/10.1038/nature02740>
- Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD (2005) Variable lymphocyte receptors in hagfish. *Proc Natl Acad Sci USA* 102(26):9224–9229. <https://doi.org/10.1073/pnas.0503792102>
- Pantophlet R, Burton D (2006) GP120: Target for neutralizing HIV-1 antibodies. *Annu Rev Immunol* 24:739–769. <https://doi.org/10.1146/annurev.immunol.24.021605.090557>
- Parra ZE, Baker ML, Schwarz R, Deakin J, Lindblad-Toh K, Miller RD (2007) A unique T cell receptor discovered in marsupials. *Proc Natl Acad Sci USA* 104. <https://doi.org/10.1073/pnas.0609106104>
- Peng H-P, Hsu H-J, Yu C-M, Hung F-H, Tung C-P, Huang Y-C, Chen C-Y, Tsai P-H, Yang A-S (2022) Antibody CDR amino acids underlying the functionality of antibody repertoires in recognizing diverse protein antigens. *Sci Rep* 12(1):12555. <https://doi.org/10.1038/s41598-022-16841-9>
- Philp RL (2018) The polyclonal antibody response to FMDV in cattle and African buffalo. Dissertation, University of Glasgow
- Proskuryakova AA, Kulemzina AI, Perelman PL, Yudkin DV, Lemskaya NA, Okhlopkov IM, Kirillin EV, Farré M, Larkin DM, Roelke-Parker ME et al (2019) Comparative chromosome mapping of musk Ox and the X chromosome among some bovidae species. *Genes* 10(11):857. <https://doi.org/10.3390/genes10110857>
- Ratcliffe MJ (2006) Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev Comp Immunol* 30(1–2):101–118. <https://doi.org/10.1016/j.dci.2005.06.018>
- Saini SS, Allore B, Jacobs RM, Kaushik A (1999) Exceptionally long CDR3H region with multiple cysteine residues in functional bovine IgM antibodies. *Eur J Immunol* 29(8):2420–2426. [https://doi.org/10.1002/\(sici\)1521-4141\(199908\)29:08%3c2420::Aid-immu2420%3e3.0.Co;2-a](https://doi.org/10.1002/(sici)1521-4141(199908)29:08%3c2420::Aid-immu2420%3e3.0.Co;2-a)
- Saini SS, Hein WR, Kaushik A (1997) A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. *Mol Immunol* 34(8–9):641–651. [https://doi.org/10.1016/s0161-5890\(97\)00055-2](https://doi.org/10.1016/s0161-5890(97)00055-2)
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S et al (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res* 50(D1):D20–d26. <https://doi.org/10.1093/nar/gkab1112>
- Schatz DG (2004) V(D)J recombination. *Immunol Rev* 200(1):5–11. <https://doi.org/10.1111/j.0105-2896.2004.00173.x>
- Schroeder HW Jr, Hillson JL, Perlmuter RM (1990) Structure and evolution of mammalian VH families. *Int Immunol* 2(1):41–50. <https://doi.org/10.1093/intimm/2.1.41>
- Schroeder HW, Zemlin M, Khass M, Nguyen HH, Schelonka RL (2010) Genetic control of DH reading frame and its effect on B-cell development and antigen-specific antibody production. *Crit Rev Immunol* 30(4):327–344. <https://doi.org/10.1615/critrevimmunol.v30.i4.20>
- Smider BA, Smider VV (2020) Formation of ultralong DH regions through genomic rearrangement. *BMC Immunol* 21(1):30. <https://doi.org/10.1186/s12865-020-00359-8>
- Sok D, Le KM, Vadnais M, Saye-Francisco KL, Jardine JG, Torres JL, Berndsen ZT, Kong L, Stanfield R, Ruiz J et al (2017) Rapid elicitation of broadly neutralizing antibodies to HIV by immunization in cows. *Nature* 548(7665):108–111. <https://doi.org/10.1038/nature23301>
- Stanfield RL, Berndsen ZT, Huang R, Sok D, Warner G, Torres JL, Burton DR, Ward AB, Wilson IA, Smider VV (2020) Structural basis of broad HIV neutralization by a vaccine-induced cow antibody. *Sci Adv* 6(22):eaba0468. <https://doi.org/10.1126/sciadv.aba0468>
- Stanfield RL, Haakenson J, Deiss TC, Criscitiello MF, Wilson IA, Smider VV (2018) Chapter four - the unusual genetics and biochemistry of bovine immunoglobulins. In Alt F (ed.) *Adv Immunol*. Academic Press
- Stanfield RL, Wilson IA, Smider VV (2016) Conservation and diversity in the ultralong third heavy-chain complementarity-determining region of bovine antibodies. *Sci Immunol* 1(1):aaf7962. <https://doi.org/10.1126/sciimmunol.aaf7962>
- Stroupe S, Forgacs D, Harris A, Derr JN, Davis BW (2022) Genomic evaluation of hybridization in historic and modern North American Bison (Bison bison). *Sci Rep* 12(1):6397. <https://doi.org/10.1038/s41598-022-09828-z>
- Stroupe S, Martone C, McCann B, Juras R, Kjällerström HJ, Beard D, Davis BW, Derr JN (2023) Chromosome-level reference genome for North American bison (Bison bison) and variant database aids in identifying albino mutation in bison. *G3-GENES GENOM GENET* (in press)
- Svilenov HL, Sacherl J, Protzer U, Zacharias M, Buchner J (2021) Mechanistic principles of an ultra-long bovine CDR reveal strategies for antibody design. *Nat Commun* 12(1):6737. <https://doi.org/10.1038/s41467-021-27103-z>
- Tonegawa S (1983) Somatic generation of antibody diversity. *Nature* 302(5909):575–581. <https://doi.org/10.1038/302575a0>
- Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, Torkamani A, Raudsepp T, Mwangi W, Criscitiello MF et al (2013) Reshaping antibody diversity. *Cell* 153(6):1379–1393. <https://doi.org/10.1016/j.cell.2013.04.049>
- Warner Jenkins G, Safonova Y, Smider VV (2022) Germline-encoded positional cysteine polymorphisms enhance diversity in antibody ultralong CDR H3 regions. *J Immunol*. <https://doi.org/10.4049/jimmunol.2200455>
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46(W1):W296–W303. <https://doi.org/10.1093/nar/gky427>
- Wu M, Zhao H, Tang X, Zhao W, Yi X, Li Q, Sun X (2023) Organization and complexity of the yak (*Bos Grunniens*) immunoglobulin loci. *Front Immunol* 13. <https://doi.org/10.3389/fimmu.2022.876509>
- Wyatt R, Sodroski J (1998) The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280(5371):1884–1888. <https://doi.org/10.1126/science.280.5371.1884>
- Yang EY, Shah K (2020) Nanobodies: next generation of cancer diagnostics and therapeutics. *Front Oncol* 10. <https://doi.org/10.3389/fonc.2020.01182>
- Zeyland J, Wolko L, Bocianowski J, Szalata M, Słomski R, Dzieduszycki AM, Ryba M, Przystałowska H, Lipiński D (2013) Complete mitochondrial genome of wild aurochs (*Bos primigenius*) reconstructed from ancient DNA. *Pol J Vet Sci* 16(2):265–273. <https://doi.org/10.2478/pjvs-2013-0037>
- Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* 7(1–2):203–214. <https://doi.org/10.1089/10665270050081478>
- Zhao Y, Jackson SM, Aitken R (2006) The bovine antibody repertoire. *Dev Comp Immunol* 30(1–2):175–186. <https://doi.org/10.1016/j.dci.2005.06.012>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.