



ISSN: 2456-2912

NAAS Rating (2026): 4.61

VET 2026; 11(7): 112-117

© 2026 VET

www.veterinarypaper.com

Received: 24-05-2026

Accepted: 28-06-2026

Shalu Kumari Pathak

Ph.D. Scholar, Animal Genetics
Division, Indian Veterinary
Research Institute, Izatnagar,
Bareilly, Uttar Pradesh, India

Bioinformatic identification and annotation of naturally occurring hyperactivated beta recombinase recognition sites in the Goat genome

Shalu Kumari Pathak

DOI: <https://www.doi.org/10.22271/veterinary.2026.v11.i7b.3585>

Abstract

Hyperactivated serine recombinases are promising genome engineering tools because they mediate precise site-specific DNA recombination without generating DNA double-strand breaks. Identification of naturally occurring genomic recognition sites is essential for assessing their applicability in livestock genome engineering. In the present study, a genome-wide survey of naturally occurring recognition sites for the hyperactivated Beta recombinase was performed in the *Capra hircus* (ARS1.2) reference genome. The genome was computationally scanned using a custom Python-based motif-search algorithm based on the Beta recombinase recognition sequence with a five-base spacer, and candidate sites were identified on both DNA strands. Identified sites were annotated using RefSeq gene annotations, followed by chromosome-wise distribution, genomic feature characterization, and functional enrichment analysis using Gene Ontology (GO) and KEGG pathways. A total of 914 Beta recombinase recognition sites corresponding to 914 unique genomic loci were identified across the goat genome. Of these, 414 sites were located within annotated genes, representing 201 unique genes, while the remaining sites were intergenic. Functional enrichment analysis revealed significant enrichment of genes involved in biological regulation, platelet morphogenesis, synapse organization, protein binding, enzyme activator activity and enzyme regulator activity. These findings provide the first comprehensive genome-wide catalogue of naturally occurring hyperactivated Beta recombinase recognition sites in the goat genome and establish a valuable resource for future recombinase-mediated genome engineering and genetic improvement in goats.

Keywords: *Capra hircus*, Hyperactivated Beta recombinase, Gene Ontology, Functional enrichment

1. Introduction

Precise modification of genomic DNA has transformed animal biotechnology, enabling the development of gene-edited livestock with improved production traits, enhanced disease resistance, and applications in agricultural, pharmaceutical, and biomedical research. Goats, valued for their production of meat, milk, fibre, and other by-products, have also emerged as important animal models owing to their moderate body size, short reproductive cycle, and high milk yield. Consequently, genome engineering technologies have been widely applied in goats to investigate gene function, improve economically important traits, and develop models for biomedical research and gene therapy.

The development of precise and efficient technologies for targeted genome modification remains a central goal of modern biotechnology. Among the available genome-editing approaches, programmable nucleases have emerged as the predominant platforms owing to their ability to introduce site-specific genetic alterations ^[1, 2].

Among the available genome-editing tools, site-specific recombinases (SSRs) enable precise DNA rearrangements by catalysing the cleavage, strand exchange, and re-ligation of double-stranded DNA without relying on endogenous DNA repair pathways or generating insertion/deletion (indel) mutations ^[3, 4, 5]. Site-specific recombinases are DNA-modifying enzymes that catalyse precise genetic recombination at defined recognition sequences. Naturally occurring recombinases mediate targeted DNA rearrangements through specific

Corresponding Author:

Shalu Kumari Pathak

Ph.D. Scholar, Animal Genetics
Division, Indian Veterinary
Research Institute, Izatnagar,
Bareilly, Uttar Pradesh, India

interactions with their cognate recognition sites, enabling the deletion, insertion, inversion, or exchange of DNA segments. These enzymes have become valuable tools for genome engineering because they facilitate predictable and highly specific modifications without relying on endogenous DNA repair pathways^[6]. Site-specific recombinases have become important components of genome engineering strategies owing to their ability to mediate precise and predictable DNA rearrangements. In sheep and goats, recombinase-based genome manipulation has been successfully achieved using systems such as Cre recombinase, FLP recombinase, and PhiC31 integrase^[7]. Consequently, extensive efforts have focused on engineering recombinases with enhanced catalytic activity and expanded target specificity.

Advances in directed evolution, high-throughput screening, and protein engineering have accelerated the development of recombinase variants with modified DNA-binding specificities^[8, 9]. In parallel, detailed characterization of recombinase substrate specificity has provided valuable insights for the rational design and evolution of enzymes capable of recognizing user-defined genomic targets^[10, 11].

Although only the Gin (β)-based RGR platform has been established to date, several additional hyperactivated recombinases have been evaluated in zinc-finger recombinase (ZFR) platforms^[12, 12]. Four additional Gin variants (α , γ , δ , ϵ , and ζ) have also been characterized² and ZFR platforms based on β and Sin recombinases have demonstrated functional activity¹². Target sites for RGR platforms based on recombinases other than Gin (β) can be identified by replacing the central 20-bp pseudo-recognition sequence with the corresponding recombinase-specific sequence¹³. The continued development of hyperactivated recombinases is expected to expand the targeting range and versatility of recombinase-based genome-editing systems.

Accordingly, the present study aimed to generate a genome-wide map of target sites for the hyperactivated Beta recombinase with a spacer length of five base pairs and to identify genes associated with these target sites across the goat genome, thereby facilitating future applications of RNA-guided recombinase-mediated genome editing.

2. Material and Method

2.1 Identification of hyperactivated Beta recombinase recognition sites with a 5-bp spacer in the goat genome

The reference genome assembly of goat (*Capra hircus*) (ARS1.2; RefSeq assembly accession GCF_001704415.2) and the corresponding genome annotation (GFF3) file were obtained from the NCBI RefSeq database. Genome-wide identification of hyperactivated Beta recombinase recognition sites with five base spacer was performed using a custom Python pipeline developed with the Biopython and `regex` libraries. Previously, it has been reported^[12] that the core sequence for hyperactivated recombinase Beta as: NNNV NNR GT WW AC YNN BNNN where N=A, T, C, or G; V=A, C, or G; B=T, C, or G; R=A or G; Y=T or C; W=A or T. Therefore, in present study search pattern was designed according to the experimentally determined DNA recognition specificity of the hyperactivated Beta recombinase described by Sirk *et al.*, (2014). The recognition sequence consisted of fixed terminal nucleotides (5'-CC and GG-3'), variable left and right binding regions, and a conserved 5-bp spacer. Degenerate nucleotide positions were represented using regular-expression character classes to permit all nucleotide combinations experimentally tolerated at each position while

preserving the conserved bases of the recognition motif.

Genome-wide scanning was performed on both the forward and reverse DNA strands. For reverse-strand analysis, the reverse-complement sequence of each chromosome was generated, and the genomic coordinates of reverse-strand matches were converted back to the corresponding coordinates of the reference genome. Each identified recognition site was recorded with its chromosome accession, genomic start and end positions, strand orientation, and complete recognition sequence.

2.2 Chromosomal distribution of Beta recombinase recognition sites

The chromosomal distribution of Beta recombinase recognition sites was determined by counting the number of identified sites on each chromosome. In addition, the distribution of recognition sites across genomic features, including CDS, exons, gene-associated regions, and intergenic regions, was summarized to evaluate their genomic context. All computational analyses were performed using custom Python scripts executed under Python 3.13 on a Windows operating system.

2.3 Genomic features annotation of identified Beta recombinase recognition sites

The identified Beta recombinase recognition sites were subsequently annotated using the corresponding goat genome GFF3 annotation. Each recognition site was compared with annotated genomic features to determine whether it overlapped an annotated gene or was located within an intergenic region. For gene-associated sites, the corresponding gene symbol and strand orientation were recorded. Feature-level annotation was further performed by comparing each recognition site with annotated coding sequences (CDS), exons, and gene regions. Recognition sites overlapping CDS or exon features were classified accordingly, whereas sites overlapping annotated gene regions without intersecting CDS or exon features were classified as gene-associated (intronic/gene region). Sites showing no overlap with annotated genes were classified as intergenic.

2.4 Functional enrichment analysis

Genes containing Beta recombinase recognition sites were subjected to functional enrichment analysis using Gene Ontology (GO) and KEGG pathway analyses through g. Enrichment analyses included GO Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) categories. Significantly enriched GO terms and KEGG pathways were identified using the default multiple-testing correction implemented in g, and terms with adjusted *P* values < 0.05 were considered statistically significant.

3. Results

3.1 Genome-wide identification of hyperactivated Beta recombinase recognition sites

A genome-wide computational survey identified 914 naturally occurring recognition sites for the hyperactivated Beta recombinase in the *Capra hircus* (ARS1.2) reference genome. Screening of both forward and reverse DNA strands revealed that all identified recognition sites represented 914 unique genomic loci, indicating the absence of complementary duplicate matches. The identified sites were distributed throughout the goat genome, including autosomes, sex chromosomes, and unlocalized scaffolds, demonstrating the

widespread A5SA occurrence of sequences compatible with the hyperactivated Beta recombinase recognition motif.

3.2 Chromosome-wise distribution of Beta recombinase recognition sites

The identified Beta recombinase recognition sites were unevenly distributed across the goat chromosomes (Figure 1). Recognition sites were detected on nearly all assembled chromosomes, although their abundance varied considerably among chromosomes. In general, larger chromosomes contained a greater number of recognition sites than smaller chromosomes, reflecting the genome-wide distribution of the Beta recombinase recognition motif.

3.3 Distribution of recombinase recognition sites across genomic features

Annotation against the RefSeq genome annotation identified 414 recognition sites located within annotated genes, corresponding to 201 unique genes, while the remaining 500 sites were located in intergenic regions (Figure 2). Gene-associated recognition sites were distributed among protein-coding genes involved in diverse biological functions, whereas intergenic sites represented the largest genomic category, indicating that naturally occurring Beta recombinase recognition motifs are widely dispersed throughout both coding and non-coding regions of the goat genome.

3.4 Functional enrichment analysis

Functional enrichment analysis of the 201 unique genes containing Beta recombinase recognition sites revealed significant enrichment across the three GO categories: Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) (Table 1 and Figure 3). In the Biological Process (BP) category, the most significantly enriched terms were biological regulation (adjusted $P = 5.93 \times 10^{-4}$; 112 genes) and regulation of biological process (adjusted $P = 2.52 \times 10^{-3}$; 108 genes), indicating that a large proportion of the identified genes participate in diverse regulatory mechanisms. Additional significantly enriched processes included platelet morphogenesis, platelet formation, and synapse organization, suggesting the involvement of these genes in developmental and cellular signaling pathways.

Within the Molecular Function (MF) category, protein binding was the most significantly enriched term (adjusted $P = 7.31 \times 10^{-5}$; 93 genes), demonstrating that many candidate genes encode proteins involved in molecular interactions. Other enriched molecular functions included enzyme activator activity, enzyme regulator activity, GTPase regulator activity, nucleoside-triphosphatase regulator activity, and GTPase activator activity, indicating that the identified genes are associated with the regulation of enzyme activity and intracellular signaling processes.

For the Cellular Component (CC) category, cell junction showed the strongest enrichment (adjusted $P = 2.97 \times 10^{-6}$; 33 genes). Other significantly enriched cellular components included glutamatergic synapse, cell periphery, and synapse, suggesting that many genes are localized to membrane-associated and cell communication structures. The TEAD–YAP complex was also significantly enriched, although represented by only two genes.

Overall, the enrichment analysis indicates that genes containing or located near the identified recombinase recognition sites are predominantly involved in biological regulation, protein interaction, enzyme modulation,

intracellular signaling, and cell junction- and synapse-associated functions. These findings suggest that the identified genomic target sites are associated with genes involved in fundamental cellular and regulatory processes

4. Discussion

The present study provides the first comprehensive genome-wide survey of naturally occurring recognition sites for the hyperactivated Beta recombinase in the *Capra hircus* reference genome. Using a computational motif-search strategy based on the experimentally characterized five-base spacer recognition sequence, a total of 914 unique Beta recombinase recognition sites were identified across the goat genome. The widespread distribution of these sites demonstrates that sequences compatible with the engineered Beta recombinase are naturally present throughout the genome, thereby expanding the repertoire of potential target loci for recombinase-mediated genome engineering in goats.

Unlike programmable nuclease-based genome-editing systems, which rely on the induction of DNA double-strand breaks and endogenous DNA repair pathways, site-specific recombinases catalyze precise DNA rearrangements through conservative recombination between defined recognition sites, enabling predictable genomic modifications with minimal dependence on host repair mechanisms. Consequently, engineered serine recombinases have emerged as attractive alternatives for precise genome engineering applications^[2, 12, 13]. The large number of naturally occurring Beta recognition sites identified in the present study indicates that the goat genome contains numerous candidate loci that could potentially be exploited for targeted recombinase-mediated DNA integration, excision, or inversion.

Genomic annotation revealed that 414 recognition sites were located within annotated genes representing 201 unique genes, whereas the remaining recognition sites were located in intergenic regions. The predominance of intergenic sites is advantageous for future genome engineering applications because intergenic regions are generally less likely to disrupt protein-coding sequences or regulatory elements, making them attractive candidates for targeted insertion of transgenes or other genetic modifications. Nevertheless, the presence of numerous gene-associated recognition sites also provides opportunities for precise manipulation of endogenous genes involved in economically important traits or disease resistance.

Functional enrichment analysis demonstrated that genes containing Beta recombinase recognition sites were significantly enriched for Gene Ontology terms associated with biological regulation, enzyme activator activity, enzyme regulator activity, GTPase regulator activity, nucleoside-triphosphatase regulator activity, and GTPase activator activity, platelet morphogenesis, platelet formation, and synapse organization. These enriched biological processes suggest that Beta recognition sites are widely distributed among genes involved in fundamental cellular and developmental functions rather than being restricted to a particular functional category. Similar observations have been reported for genome-wide surveys of engineered recombinase recognition sequences in livestock genomes, where candidate target sites were broadly distributed across functionally diverse genomic regions^[12, 13].

The enrichment of signaling-related pathways supports the GO enrichment results and indicates that genes harboring Beta recombinase recognition sites participate in complex

biological networks. Importantly, these findings do not imply preferential recombinase activity within these pathways but rather reflect the functional characteristics of genes naturally containing sequences compatible with the Beta recombinase recognition motif. The present study was based on computational prediction of naturally occurring recognition sites using the goat reference genome and experimentally defined Beta recombinase sequence specificity. Consequently, the identified recognition sites represent candidate genomic targets that require experimental validation to determine recombination efficiency, target accessibility, chromatin context, and editing specificity *in vivo*. Future studies combining computational prediction with cell-based recombination assays and genome engineering experiments will be essential to validate these candidate loci and assess their suitability for practical applications. Nevertheless, several challenges remain, including improving recombination efficiency, minimizing off-target activity, understanding the influence of post-translational

modifications on enzyme stability and function [14, 15] and developing safe and efficient delivery strategies for therapeutic applications [16, 17, 18, 19]. Addressing these challenges will be essential for realizing the full potential of engineered site-specific recombinases as versatile tools for genome engineering and gene therapy.

5. Conclusion

This study establishes the first genome-wide catalogue of naturally occurring hyperactivated Beta recombinase recognition sites in the goat genome. The widespread genomic distribution of these sites and their association with genes involved in diverse biological processes provide a valuable resource for future recombinase-mediated genome engineering. These findings expand the available genomic information for engineered serine recombinases and may facilitate the development of precise genome-editing strategies for functional genomics, livestock genetic improvement, and translational biotechnology.

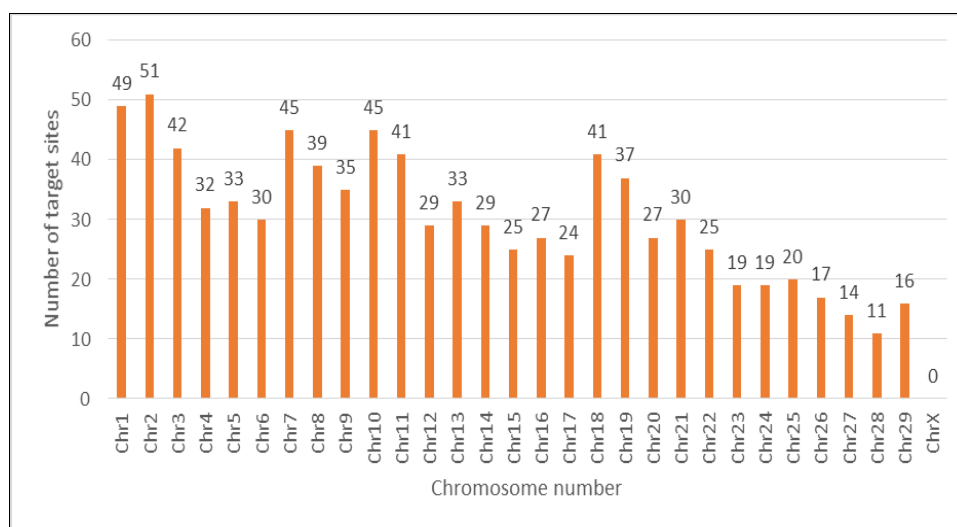


Fig 1: Chromosome-wise distribution of hyperactivated Beta recombinase recognition sites identified in the *Capra hircus* (ARS1.2) reference genome.

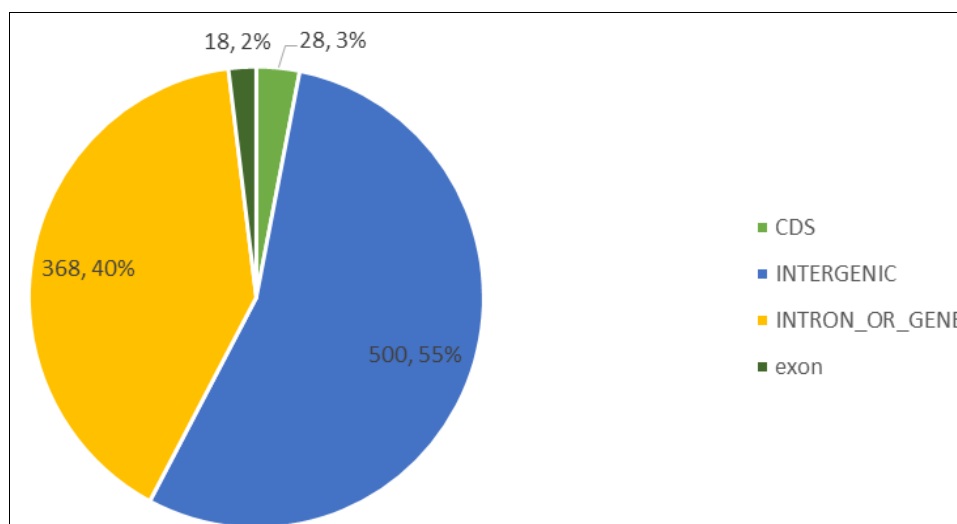
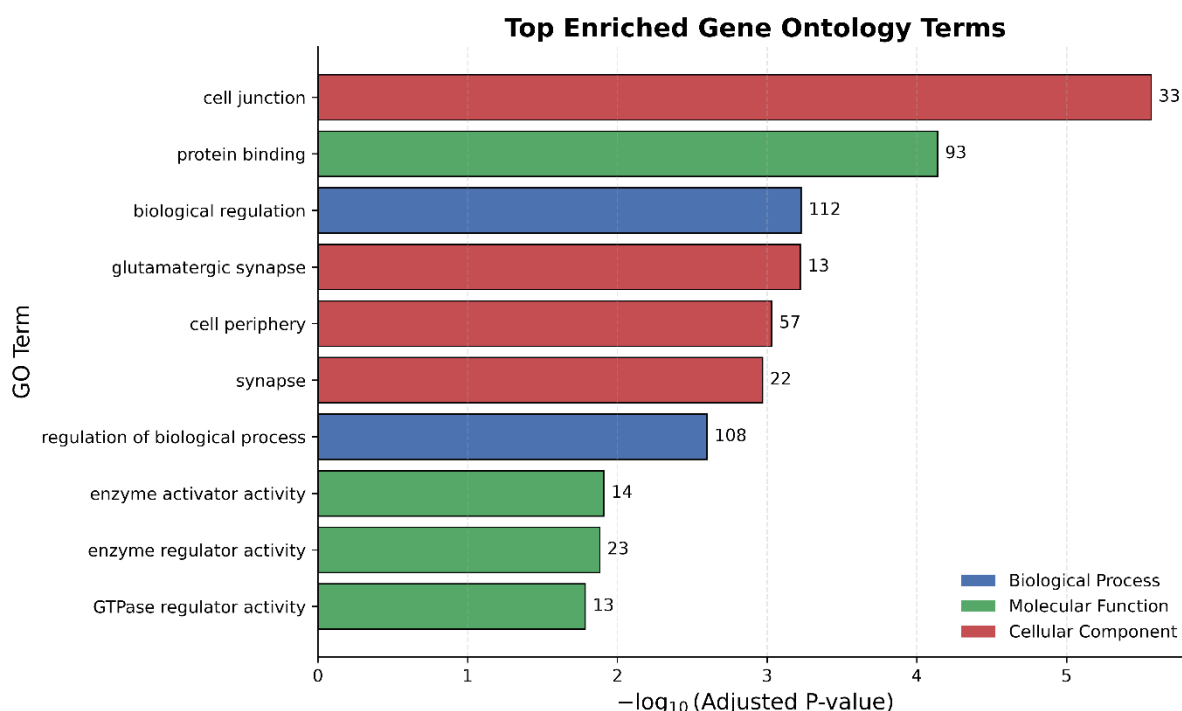


Fig 2: Genomic feature distribution of hyperactivated Beta recombinase recognition sites.

Table 1: Top enriched GO terms associated with genes containing hyperactivated Beta recombinase recognition sites.

GO Category	GO Term	GO ID	Adjusted P-value	$-\log_{10}(\text{Adjusted } P)$	Gene Count
Molecular Function (MF)	protein binding	GO:0005515	7.27×10^{-5}	4.138	93
MF	enzyme activator activity	GO:0008047	0.01233	1.909	14
MF	enzyme regulator activity	GO:0030234	0.01311	1.882	23
MF	GTPase regulator activity	GO:0030695	0.01642	1.785	13
MF	nucleoside-triphosphatase regulator activity	GO:0060589	0.01687	1.773	13
MF	GTPase activator activity	GO:0005096	0.01900	1.721	9
Biological Process (BP)	biological regulation	GO:0065007	5.93×10^{-4}	3.227	112
BP	regulation of biological process	GO:0050789	0.00252	2.598	108
BP	platelet morphogenesis	GO:0036344	0.01703	1.769	4
BP	platelet formation	GO:0030220	0.01703	1.769	4
BP	synapse organization	GO:0050808	0.03399	1.469	12
Cellular Component (CC)	cell junction	GO:0030054	2.72×10^{-6}	5.565	33
CC	glutamatergic synapse	GO:0098978	6.00×10^{-4}	3.222	13
CC	cell periphery	GO:0071944	9.32×10^{-4}	3.031	57
CC	synapse	GO:0045202	0.00108	2.969	22
CC	TEAD-YAP complex	GO:0140552	0.04972	1.303	2

**Fig 3:** Top 10 enriched GO terms represented as a horizontal bar plot of $-\log_{10}(\text{adjusted } P \text{ value})$.

References

- Carroll D. Genome engineering with zinc-finger nucleases. *Genetics*. 2011;188:773-778.
- Gaj T, Mercer AC, Sirk SJ, Smith HL, Barbas CF III. A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res*. 2013;41:3937-3946.
- Nern A, Pfeiffer BD, Rubin GM. Optimized tools for site-specific recombination and genetic manipulation. *Proc Natl Acad Sci U S A*. 2011;108(34):14198-14203.
- Turan S, Bode J. Site-specific recombinases: From tag-and-target- to tag-and-exchange-based genomic modifications. *FASEB J*. 2011;25(12):4088-4107.
- Grindley NDF, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem*. 2006;75:567-605.
- Olorunniji FJ, Rosser SJ, Stark WM. Site-specific recombinases: Molecular machines for the genetic revolution. *Biochem J*. 2016;473:673-684.
- Kalds P, Zhou S, Cai B, Liu J, Wang Y, Petersen B, *et al*. Sheep and goat genome engineering: From random transgenesis to the CRISPR era. *Front Genet*. 2019;10:750.
- Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature*. 2011;472(7344):499-503.
- Leconte AM, Dickinson BC, Yang DD, Chen IA, Allen B, Liu DR. A population-based experimental model for protein evolution: Effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry*. 2013;52(8):1490-1499.
- Hartung M, Kisters-Woike B. [Article title missing.] *J Biol Chem*. 1998;273:22884-22891.
- Whiteson KL, Rice PA. Binding and catalytic contributions to site recognition by FLP recombinase. *J Biol Chem*. 2008;283(17):11414-11423.
- Sirk SJ, Gaj T, Jonsson A, Mercer AC, Barbas CF III. Expanding the zinc-finger recombinase repertoire: Directed evolution and mutational analysis of serine recombinase specificity determinants. *Nucleic Acids Res*.

- 2014;42:4755-4766.
13. Chaikind B, Bessen JL, Thompson DB, Hu JH, Liu DR. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Res.* 2016;44:9758-9770.
 14. Chen XL, Reindle A, Johnson ES. Misregulation of 2 microm circle copy number in a SUMO pathway mutant. *Mol Cell Biol.* 2005;25(10):4311-4320.
 15. Xiong L, Chen XL, Silver HR, Ahmed NT, Johnson ES. Deficient SUMO attachment to Flp recombinase leads to homologous recombination-dependent hyperamplification of the yeast 2 microm circle plasmid. *Mol Biol Cell.* 2009;20(4):1241-1251.
 16. Jo D, Nashabi A, Doxsee C, Lin Q, Unutmaz D, Chen J, *et al.* Epigenetic regulation of gene structure and function with a cell-permeable Cre recombinase. *Nat Biotechnol.* 2001;19(10):929-933.
 17. Peitz M, Pfannkuche K, Rajewsky K, Edenhofer F. Ability of the hydrophobic FGF and basic TAT peptides to promote cellular uptake of recombinant Cre recombinase: A tool for efficient genetic engineering of mammalian genomes. *Proc Natl Acad Sci U S A.* 2002;99(7):4489-4494.
 18. Pfeifer A, Brandon EP, Kootstra N, Gage FH, Verma IM. Delivery of the Cre recombinase by a self-deleting lentiviral vector: Efficient gene targeting in vivo. *Proc Natl Acad Sci U S A.* 2001;98(20):11450-11455.
 19. Wang Y, Krushel LA, Edelman GM. Targeted DNA recombination in vivo using an adenovirus carrying the Cre recombinase gene. *Proc Natl Acad Sci U S A.* 1996;93(9):3932-3936.

How to Cite This Article

Pathak SK. Bioinformatic identification and annotation of naturally occurring hyperactivated beta recombinase recognition sites in the goat genome. *International Journal of Veterinary Sciences and Animal Husbandry.* 2026;11(7):112-117.

Creative Commons (CC) License

This is an open-access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-Share Alike 4.0 International (CC BY-NC-SA 4.0) License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.