



ISSN: 2456-2912
NAAS Rating (2026): 4.61
VET 2026; 11(7): 31-35
© 2026 VET
www.veterinarypaper.com
Received: 08-04-2026
Accepted: 14-05-2026

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

Subodh Kumar
Principal Scientist, Animal
Genetics Division, Indian
Veterinary Research Institute,
Izatnagar, Bareilly, Uttar
Pradesh, India

Arvind A Sonwane
Principal Scientist, Animal
Genetics Division, Indian
Veterinary Research Institute,
Izatnagar, Bareilly, Uttar
Pradesh, India

Corresponding Author:
Shalu Kumari Pathak
Ph.D. Scholar, Animal Genetics
Division, Indian Veterinary
Research Institute, Izatnagar,
Bareilly, Uttar Pradesh, India

Identification and genomic characterization of hyperactivated gin recombinase recognition sites with a six-base-pair spacer in the goat (*Capra hircus*) genome

Shalu Kumari Pathak, Subodh Kumar and Arvind A Sonwane

DOI: <https://www.doi.org/10.22271/veterinary.2026.v11.i7a.3559>

Abstract

Site-specific recombinases enable precise genome engineering through sequence-specific DNA recombination. Hyperactivated Gin recombinase recognizes degenerate DNA target sequences, increasing the likelihood of endogenous recognition sites within mammalian genomes. However, a genome-wide survey of Gin recognition sites in the goat (*Capra hircus*) genome has not been reported. The goat reference genome (ARS1.2; GCF_001704415.2) and corresponding GFF3 annotation were obtained from the NCBI Genome database. Custom Python scripts were used to identify hyperactivated Gin recombinase recognition sites with a six-base spacer on both DNA strands. Identified sites were annotated against genomic features, and chromosome-wise and genomic feature distributions were analyzed. A total of 68 strand-specific recognition sites, corresponding to 34 unique genomic loci, were identified across the goat genome. Recognition sites were distributed across multiple chromosomes and occurred in both coding and non-coding genomic regions, including CDS, exons, genes, and intergenic regions. This study presents the first genome-wide catalogue of hyperactivated Gin recombinase recognition sites in the goat genome. The identified endogenous target sites provide a useful resource for recombinase-mediated genome engineering and functional genomics applications in goats.

Keywords: *Capra hircus*, Gin recombinase, site-specific recombinase, genome-wide analysis, genome annotation, recombinase recognition sites, genome engineering, bioinformatics

1. Introduction

Genome editing expands the toolkit for precision animal breeding, complementing and surpassing aspects of conventional selection while supporting sustainable production across environmental, economic, and social dimensions [1]. Recombinases are valuable genome-editing tools that mediate precise DNA rearrangements; however, their application is limited by the need for complex directed-evolution approaches to alter target specificity and their lack of user programmability. To overcome these limitations, hyperactivated recombinases with relaxed substrate specificity have been developed. In addition, the RNA-guided recombinase (RGR) platform was introduced to provide programmable recombination. The RGR system recognizes a target sequence comprising a minimal pseudo-core recombinase site separated by a 5- to 6-bp spacer and flanked by two guide RNA-directed Cas9-binding sites followed by protospacer adjacent motifs (PAMs).

Currently, only one RGR platform based on the hyperactivated Gin (β) recombinase has been developed, although several hyperactivated recombinases have been evaluated as components of zinc-finger recombinase (ZFR) platforms [2, 3]. In addition to Gin (β), four other hyperactivated Gin variants (α , γ , δ , ϵ , and ζ) have been characterized and demonstrated the functionality of ZFR platforms based on the β and Sin recombinases [2, 3]. The development of additional hyperactivated recombinases is expected to broaden the targeting range of recombinase- and RGR-based genome-editing systems. Furthermore, target sites for RGR platforms using recombinases other than Gin (β) can be identified by replacing the central 20-bp pseudo-recognition sequence with the corresponding recombinase-specific sequence [4].

The RGR platform has considerable potential for precise genome editing in both biomedical

and livestock applications. For example, the African swine fever virus resilience-associated SNP allele (T1591C) in the p65 gene has been introgressed from warthog into conventional swine cells [5], while the mutation responsible for Isoleucyl-tRNA synthetase (IARS) syndrome in Japanese Black cattle has been corrected using CRISPR/Cas9-assisted genome editing [6]. These studies highlight the potential of programmable genome-editing systems for correcting disease-associated mutations.

The RGR platform represents a promising tool for future livestock improvement because of its efficiency, precision, and safety. However, its successful implementation in animal breeding depends on two major prerequisites: the identification of quantitative trait nucleotides (QTNs) through genomics, genome-wide association studies (GWAS), and genomic selection (GS), and the development of comprehensive genome-wide maps of RGR target sites. The present study addresses the latter challenge by generating a genome-wide map of target sites for the hyperactivated Gin recombinase with a spacer length of six base pairs, identifying associated genes across the caprine genome.

2. Materials and Methods

The reference genome assembly of goat (*Capra hircus*) (ARS1.2; RefSeq assembly accession GCF_001704415.2) was downloaded in FASTA format together with the corresponding genome annotation (GFF3) file from the NCBI RefSeq database. Genome-wide identification of hyperactivated Gin recombinase recognition sites was performed using a custom Python script developed with the Biopython and 'regex' libraries.

Hyperactivated Gin recombinase recognizes a target sequence with the general architecture CCN(30-31)-AAASSWSSTTT-N(30-31)-GG, where the central 12-bp degenerate motif is flanked by variable spacer regions [4]. The search pattern was designed according to the DNA recognition sequence of the hyperactivated Gin recombinase and consisted of a 78-bp target site with fixed terminal nucleotides (5'-CC and GG-3'), two 31-bp variable binding regions, and a conserved 10-bp central core sequence represented as AAA [GC] [GC] [AT] [AT] [GC] [GC]TTT. Degenerate nucleotide positions were represented using regular-expression character classes, allowing exhaustive identification of all genomic sequences matching the specified recognition motif.

Genome-wide scanning was performed on both the forward and reverse DNA strands. For reverse-strand analysis, the reverse complement of each chromosome sequence was generated, and genomic coordinates of reverse-strand matches were converted back to the corresponding positions in the reference genome. Each identified target site was recorded with its chromosome accession, genomic start and end coordinates, strand orientation, and matched DNA sequence. Candidate recognition sites were subsequently annotated using the corresponding goat genome GFF3 annotation file. Each target site was compared with annotated genomic features to determine whether it overlapped an annotated gene or occurred within an intergenic region. For gene-associated sites, the corresponding gene symbol and strand orientation were recorded. Genomic feature annotation was further refined by comparing each target site with coding sequence (CDS), exon, and gene annotations. Sites overlapping CDS or exons were classified accordingly, whereas sites overlapping gene regions without intersecting annotated CDS or exons were classified as gene-associated (intronic/gene region).

Sites with no overlap to annotated genes were classified as intergenic.

The genomic distribution of identified Gin recognition sites was summarized according to chromosome and genomic feature class. Candidate sites were ranked according to their genomic context, with intergenic sites considered the highest-priority candidates for downstream analyses because they are less likely to disrupt annotated protein-coding genes. All analyses were performed using custom Python scripts executed under Python 3.13 on a Windows operating system.

3. Results and Discussion

In the present study, a comprehensive genome-wide survey of naturally occurring recognition sites for the hyperactivated Gin recombinase with six base spacer was performed in the *Capra hircus* (ARS1.2) reference genome. Using a computational motif-search strategy, numerous candidate Gin recognition sites were identified across the goat genome, demonstrating that sequences conforming to the recombinase recognition motif are widely distributed. In present genome-wide search of the *Capra hircus* reference, 34 unique genomic recognition sites matching the predefined hyperactivated Gin recombinase recognition sequence were identified. Scanning of both DNA strands yielded 68 strand-specific matches, 34 on forward and 34 on reverse strand which corresponded to the same 34 genomic loci found in opposite orientations due to the reverse-complement symmetry of the recognition motif. Each recognition site was recorded with its chromosome accession, genomic coordinates, strand orientation and complete 78-bp recognition sequence. The identified Gin recognition sites were distributed across several goat chromosomes but were absent from many chromosomes and unlocalized genomic scaffolds. The number of recognition sites varied among chromosomes, indicating a non-uniform genomic distribution (Figure 1). No Gin recognition sites were detected on the mitochondrial genome. The chromosome-wise distribution provides an overview of the genomic locations of naturally occurring Gin target sites in the goat genome. Approximately 450 recombinase target sites have been identified in human genome, revealing the extensive presence of naturally occurring recombinase recognition motifs across the genome [4]. Similarly, 426 potential RGR target sites were identified in buffalo genome for recombinase Beta with spacer length six [7]. Comparable genome-wide surveys in livestock have similarly demonstrated the widespread distribution of recombinase recognition sites across chromosomes. For instance, numerous genomic target sites for the hyperactivated Beta recombinase with five base spacers and an extensive distribution of hyperactivated Gin recombinase recognition sites with a six-base-pair spacer have been identified in the cattle genome [8, 9]. Although the number of recognition sites varied among chromosomes, such variation is expected because chromosome size, nucleotide composition, and sequence complexity differ across the genome. Larger chromosomes generally contain more opportunities for occurrence of degenerate DNA motifs, whereas smaller chromosomes harbour proportionally fewer target sites. The widespread chromosomal distribution observed in this study suggests that naturally occurring Gin recognition sequences are common genomic features rather than isolated loci. These findings expand the repertoire of naturally occurring recombinase target sites identified in livestock genomes and provide a genomic resource for the potential application of Gin recombinase in targeted genome engineering.

Further, in order to determine the genomic features of the identified recognition sites, all candidate loci were annotated using the corresponding NCBI RefSeq GFF3 genome annotation file. It was found that among 34 unique recognition sites, 15 (44.1%) overlapped annotated genes, whereas 19 (55.9%) were located within intergenic regions (Table 1). Similar results have been revealed in identification of RNA guided recombinase target sites for hyperactivated recombinase Gin and Beta in bovine genome. For example, out of 426 identified target sites in bovine genome for hyperactivated recombinase Beta with spacer length six, 207 RGR target sites got identified in genic region and 263 RGR target sites in intergenic region [7]. For hyperactivated recombinase beta with five base spacer it was found that 173 RGR target sites lies in genic region and 263 in intergenic region while for hyperactivated recombinase Gin with spacer length six, it was found that 247 RGR target sites lie in genic region and 430 in intergenic region [8, 9]. Chaikind *et al.* (2016) also searched ENSEMBL (release 81) to identify which predicted recCas9 target sites fall within annotated genes [10].

Feature-level annotation further classified gene-associated recognition sites into coding sequence (CDS), exon, or other gene-associated regions, while loci with no overlap to annotated genes were classified as intergenic. Analysis of genomic feature distribution revealed that intergenic regions constituted the largest proportion of identified Gin recognition sites, followed by gene-associated regions. A smaller number of recognition sites overlapped annotated exons and coding sequences (Figure 2). This distribution indicates that naturally occurring Gin recognition sites are preferentially enriched in non-coding regions of the goat genome.

Several Gin recognition sites were located within annotated genes, including MTHFD1L, SLC17A5, PACSIN2, and SMOC1, whereas the remaining sites were associated with additional annotated or predicted genes. The predominance of intergenic recognition sites suggests that naturally occurring Gin target sequences are more frequently located outside annotated protein-coding genes, thereby reducing the likelihood of disrupting endogenous coding regions during targeted genome engineering. The presence of recognition sites within intergenic regions could represent relatively safer targets for genome engineering by minimizing disruption of protein-coding genes. Likewise, recognition sites located within intronic regions may provide opportunities for targeted genome modification while preserving coding sequences, although functional validation would be required before considering any site for experimental applications. Location and features of target Sites identified in genic region of goat genome for hyperactivated recombinase gin with spacer length six has been shown in Table 2.

The genome-wide identification of endogenous Gin recombinase recognition sites provides a valuable resource for the development of RNA-guided recombinase (RGR)-based genome engineering in goats. The RGR platform integrates programmable CRISPR-based DNA targeting with the catalytic activity of serine recombinases to enable precise DNA integration and deletion [11]. Similarly, Standage-Beier, K. *et al.* demonstrated targeted genomic DNA deletion and integration using an RNA-guided recombinase-Cas9 fusion

[12]. The endogenous Gin recognition sites identified in this study represent potential candidate loci for future RGR-mediated genome engineering and functional genomics applications in goats.

Overall, the identified Gin recognition sites constitute a genome-wide catalogue of candidate substrates for hyperactivated Gin recombinase with six base spacer. The complete dataset, including genomic coordinates, strand orientation, recognition sequence, associated genes and genomic feature annotation provides a valuable resource for evaluating potential recombination targets for genome engineering applications in goat.

The widespread occurrence of naturally occurring Gin recognition sites is consistent with the degenerate nature of the recombinase recognition motif. Unlike recombinases that require highly conserved target sequences, hyperactivated Gin recombinase recognizes a flexible DNA motif containing degenerate nucleotide positions and variable spacer lengths. Such relaxed sequence requirements increase the probability of naturally occurring target sites throughout complex mammalian genomes. Consequently, computational genome-wide screening provides an efficient approach for identifying candidate recognition sites before experimental validation.

The computational pipeline developed in this study also demonstrates an efficient framework for large-scale identification of recombinase recognition sites. Automated motif searching combined with genome annotation enables rapid characterization of candidate target sites and can readily be adapted to other recombinases, recognition motifs, or reference genomes. Such genome-wide catalogs may facilitate the selection of suitable endogenous target sites for future genome engineering studies in livestock and other mammalian species.

Although this study provides a comprehensive in silico catalogue of candidate Gin recognition sites, several limitations should be acknowledged. The presence of a recognition motif does not necessarily imply efficient recombinase binding or recombination activity *in vivo*. Factors such as chromatin accessibility, DNA methylation, local sequence context, and three-dimensional genome organization may substantially influence recombination efficiency (Thurman *et al.*, 2012). Furthermore, the analyses were performed using a single reference genome assembly and therefore do not account for genetic variation among goat breeds or individuals. Experimental validation of selected recognition sites will be essential to determine their recombination efficiency and genome engineering potential.

4. Conclusion

In conclusion, this study presents the first comprehensive genome-wide identification and annotation of hyperactivated Gin recombinase with six base spacer recognition sites in the goat genome. The broad chromosomal distribution and diverse genomic localization of these recognition sites indicate that naturally occurring Gin target sequences are abundant throughout the goat genome. The dataset generated in this study provides a valuable genomic resource for future investigations of site-specific recombination and establishes a foundation for the development of recombinase-mediated genome engineering strategies in goats.

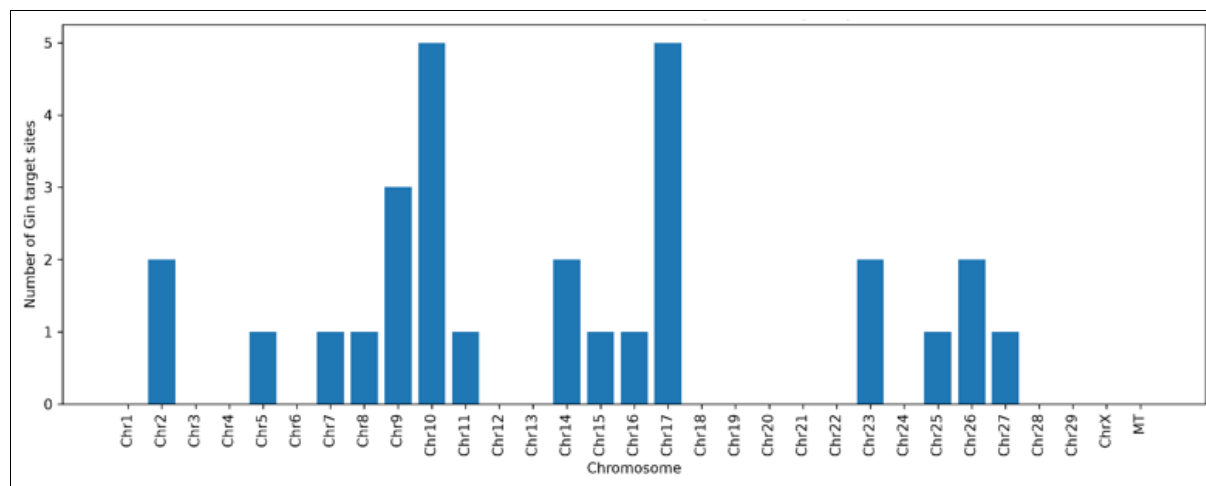


Fig 1: Chromosome-wise distribution of hyperactivated Gin recombinase with six base spacer recognition sites identified in the *Capra hircus* (ARS1.2) reference genome.

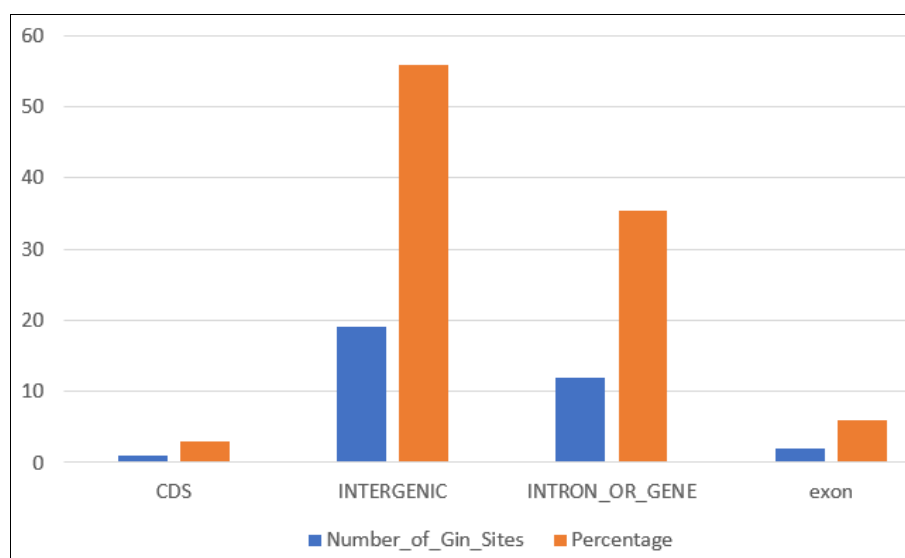


Fig 2: Distribution of hyperactivated Gin recombinase recognition sites across genomic features, including intergenic regions, gene-associated regions, exons, and coding sequences (CDS).

Table 1: Distribution of identified hyperactivated Gin recombinase recognition sites across major genomic feature categories

Genomic Feature	Number of Gin Sites	Percentage
CDS	1	2.94
INTERGENIC	19	55.88
INTRON OR GENE	12	35.29
EXON	2	5.88

Table 2. Genomic annotation of hyperactivated Gin recombinase recognition sites identified in the goat genome, including chromosome, genomic coordinates, strand orientation, associated gene, genomic feature, and recognition sequence

Chr. No.	Start	End	Gene	Feature	RGR target site sequence
Chr10	41383008	41383086	FBN1	Intron or gene	CCAGAGGTTCTTGGGCAAGAGAGTTATTTAATCAAAGCATCGTTTTAGGAAGATAAATCTAAGAAGGTGGATTGTAGG
Chr10	21400387	21400465	SMOC1	Intron or gene	CCTCAGAAACCATGCCAGCGTGTCTGAACAGCCAAACCTAGCTTTCCAGAAATCAAGCCAGCTTGAGGCACCCAGG
Chr5	112661689	112661767	PACSIN2	Intron or gene	CCAAGCTCAGAGCATCTGTCTCTGACTTGAAGAAACCAAGCTTTTCCAGACCCAGAGCGTGTGGCATGGTCTGG
Chr10	44415479	44415557	DMXL2	Intron or gene	CCCTCCTTGATGATTCAATGTGTATACAGAAATAAACCTTGCTTTACTTCAGTTCCTCTGTCCCTTGCCCTGAGTGG
Chr10	62722406	62722484	MDGA2	Intron or gene	CCCACTTTTTTATAGAATTTTTTTTAACTTAAAGGATGCTTTTATACAATGGTTATTGAGACTTGGGTATTTGG
Chr27	8155528	8155606	CHRN3	Intron or gene	CCTGTTTCTGTCTCCAGCCTGATGCTTTGAGAAAGCATGCTTTTTTTTTTCTTTCTGATTCTAGCAAAAGGAAGG
Chr26	35944863	35944941	PLCE1	Intron or gene	CCAACGTTATCCAGCTGCTCTTTCTCCTTCTCTAAACCTTCTTTCCAAGATCAATGCAATGATTTCTTATTCCAGG
Chr16	17834393	17834471	USH2A	Intron or gene	CCCAGTGGTAAAGAATCCAAGTCCCAATGCAAGAAACGTAGGTTTGATCCCCAGGTTGGGAAGATGCCCTGGAGAAGG

Chr17	60855255	60855333	ARHGAP10	Intron or gene	CCGCTGTCCTTTCCTGATTAGTAAGACACCATTAAAGGTTCC CTTTTCTGCTAAAGAATGGAGTGACTTTTGCTGTTGG
NW_017189517.1	17506736	17506814	TENM1	Intron or gene	CCAGCGTGCAGACACAAAAGGGGATGTGTGACGAAAGGTTCC TTTGCAAGACCCCCTGGGTATGGAGTGGGCTCTGGG
Chr9	75362975	75363053	MTHFD1L	Intron or gene	CCTGGCACTCTGTGTCCCAAGGAAGAGGAGCCAAAGGATGG TTTCTCCAGCACTGCTTATCACTACAGAGCTTAGGG
Chr9	270118	270196	SLC17A5	Intron or gene	CCCTGGGCCTCATGCTTGAACACAGGGATACAAAGGATC CTTTAGATCCCTGATGATCTAAAGACATAGATTCTGG
Chr17	60872259	60872337	NR3C2	exon	CCAAACTGTTTACATATAAAATTTTATTTTAAACCTACC TTTTAAAAATCTAAAAACCTCTAATTATCGTGCGG
Chr17	25847816	25847894	SFSWAP	exon	CCCCGGCACTCTGCTCAAGCATGCTGTACCTTTAAACCATGG TTTCTATTGAATGATGCTGCCTGAGTTGTCAGGCGG
Chr14	38172346	38172424	ZNF704	CDS	CCTGCTCTTTGATGAGCCCATCCCCAGGAAACGAAAGGTTGG TTTCAGGTTTCATTGGGGGGAACAGGCTTCCTAGG

References

- Henchion M, Moloney AP, Hyland J, Zimmermann J, McCarthy S. Review: Trends for meat, milk and egg consumption for the next decades and the role played by livestock systems in the global production of proteins. *Animal*. 2021;15(Suppl 1):100287.
- 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 Research*. 2013;41(6):3937-3946.
- 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 Research*. 2014;42(7):4755-4766.
- Chaikind B, Bessen LJ, Thompson BB, Hu JH, Liu RD. A programmable Cas9-serine recombinase fusion protein that operates on DNA sequences in mammalian cells. *Nucleic Acids Research*. 2016;44(20):9758-9770.
- Palgrave C, Gilmour L, Lowden C, Lillico S, Mellencamp M, Whitelaw B. Species-specific variation in RELA underlies differences in NF- κ B activity: A potential role in African swine fever pathogenesis. *Journal of Virology*. 2011;85(12):6008-6014.
- Ikeda M, Matsuyama S, Akagi S, Ohkoshi K, Nakamura S, Minabe S, *et al.* Correction of a disease mutation using CRISPR/Cas9-assisted genome Editing in Japanese black cattle. *Scientific Reports*. 2017;7:17827.
- Tarang M, Kumar S, Sonwane A, Sandeep A, Zafir I, Lodhi SS. RNA-guided hyperactivated Beta recombinase target sites identification for genome editing in buffalo. *The Pharma Innovation*. 2023;12(12):4092-4102.
- Pathak SK, Sonwane A, Kumar S. Exploration of RNA-guided recombinase target sites for hyperactivated recombinase Beta in the bovine genome. *Indian Journal of Animal Research*. 2019;54(9):1086-1097.
- Pathak SK, Sonwane A, Kumar S. Bovine genome analysis to unravel the location and feature of target sites of RNA-guided hyperactivated recombinase Gin with spacer length six. *Indian Journal of Animal Research*. 2024;58(12):2040-2045.
- Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, *et al.* Ensembl 2015. *Nucleic Acids Research*. 2015;43(D1):D662-D669.
- Sonwane AA. Development of RNA-guided recombinase (RGR) platform for targeted DNA integration. New Delhi: Department of Biotechnology, Government of India; 2021. Project Report.
- Standage-Beier K, Brookhouser N, Balachandran P, Zhang Q, Brafman DA, Wang X. RNA-guided recombinase-Cas9 fusion targets genomic DNA deletion

and integration. *The CRISPR Journal*. 2019;2(4):209-222.

How to Cite This Article

Pathak SK, Kumar S, Sonwane AA. Identification and genomic characterization of hyperactivated gin recombinase recognition sites with a six-base-pair spacer in the goat (*Capra hircus*) genome. *International Journal of Veterinary Sciences and Animal Husbandry*. 2026;11(7):31-35.

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.