

Recent Advances in Soybean Genome Editing: CRISPR Technologies, Precision Editors and Artificial Intelligence–Assisted Gene Editing

MANJEET PUNTAMBEKAR^{1,2}, ARUNA TIWARI¹, GIRIRAJ KUMAWAT³, SANJAY GUPTA³, VIRAJ KAMBLE³, K. H. SINGH³ MILIND B. RATNAPARKHE^{3*}

¹Indian Institute of Technology, Indore

²Murdoch University, Australia

³ICAR-National Soybean Research Institute, Indore

*Corresponding Author: E-mail: milind.ratnaparkhe@gmail.com

ABSTRACT

Soybean (Glycine max) improvement is increasingly constrained by climate variability, limited genetic diversity and the long timelines of conventional breeding. Recent advances in genome editing, particularly expanded CRISPR–Cas systems, precision editing approaches such as base and prime editing, and DNA-free delivery strategies, now enable efficient and targeted modification of endogenous genes directly in elite soybean backgrounds. These developments have accelerated functional validation and trait engineering for seed quality, plant architecture and stress tolerance. In parallel, artificial intelligence and machine-learning approaches are emerging as important tools for candidate gene prioritisation, guide RNA design and off-target risk assessment. This review summarises recent technological developments in soybean genome editing, highlights major gene targets and traits, and discusses current editing platforms together with emerging AI-driven strategies that are shaping precision breeding in soybean.

Keywords: CRISPR, Soybean Genome editing, Prime editing, PAM-less editing, Artificial intelligence, Trait improvement

1 | INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) stands as the world's foremost source of vegetable protein and edible oil, underpinning global food security, animal feed systems, and emerging plant-based protein markets. Its strategic importance is particularly pronounced in rapidly transitioning economies where dietary shifts and agricultural intensification are widening the gap between domestic production and surging demand for protein and oil (Mangal et al., 2024). Global production and trade remain concentrated in Brazil and the United States, reflecting soybean's central role in international commodity markets and agricultural geopolitics (Mangal et al., 2024).

Conventional breeding has been the backbone of soybean improvement for decades and has enabled substantial gains in yield and adaptation. Nevertheless, traditional breeding depends on repeated hybridisation cycles spanning 8 to 12 years for variety development. The introgression of beneficial alleles from wild or exotic germplasm is frequently constrained by linkage drag, whereby favourable traits are inherited together with undesirable characteristics. In addition, the continued use of a narrow pool of elite parental lines contributes to genetic erosion and genetic plateaus, thereby restricting the availability of novel allelic variation required for sustained genetic improvement (Krishna et al., 2023; Tester and Langridge, 2010).

To overcome these limitations, contemporary soybean improvement increasingly integrates genomics, next-generation sequencing (NGS) and genome editing such as CRISPR-Cas. Large-scale genotyping and re-sequencing enable the identification of genes and single-nucleotide polymorphisms (SNPs) associated with yield, seed composition and stress tolerance, and support genome-wide association studies, haplotype analysis and genomic selection across large soybean populations. These genomics-driven approaches provide the primary entry point for genome editing by prioritising candidate genes and causal variants that can subsequently be validated and engineered using CRISPR-based technologies.

CRISPR-Cas genome editing enables precise, locus-specific modifications directly in elite genetic backgrounds. Unlike breeding or genomic selection, editing generates novel variation without linkage drag or prolonged back crossing, compressing variety development timelines from years to months. Critically, DNA-free ribonucleo-protein delivery and other transgene-free strategies streamline regulatory approval pathways and improve public acceptance compared with conventional transgenic approaches (He et al., 2022).

Artificial intelligence and machine learning increasingly interface with these editing workflows, extracting predictive patterns from high-dimensional datasets to support guide RNA

design, off-target risk assessment, and regulatory element prioritisation (Dixit et al., 2024). This convergence of computational prediction with experimental genome editing is shifting soybean improvement towards more rational and data-guided engineering strategies, enabling rational design of complex, multi-trait genetic modifications.

In this review, we examine recent advances in soybean genome editing, highlighting major genes and traits targeted for improvement, emerging technologies including PAM-less editing and prime editing systems that have recently overcome the dicot efficiency barrier, and the expanding role of machine learning in guiding precision breeding strategies.

2 | RECENT DEVELOPMENTS

2.1 | Expansion of CRISPR–Cas Systems and PAM Flexibility

The soybean genome editing toolbox has expanded beyond conventional SpCas9 (NGG PAM) to include near-PAM-less variants such as SpRY, which recognises relaxed NRN PAM sites and substantially increases the number of editable loci in the soybean genome (Ren et al., 2021). Because PAM sequences are short DNA motifs required for nuclease binding, relaxing PAM constraints enables access to regulatory regions that were previously inaccessible, supporting fine-scale modulation of gene expression rather than simple gene disruption. In soybean, GmUBI3 and GmM4-driven SpRY expression systems have been shown to support efficient editing across multiple genomic loci, achieving editing efficiencies of up to 57.7%, demonstrating the practical applicability of PAM-relaxed editing in soybean (Chen et al., 2024).

In parallel, the hypercompact Cas12j-8 system has been adapted for plant genome editing. Owing to its small size, approximately one-quarter that of Cas9, Cas12j-8 facilitates vector construction and delivery and allows incorporation of additional regulatory components within transformation constructs. The optimised en4Cas12j-8 editor achieves editing efficiencies comparable to SpCas9 at TTN PAM sites, and its base-editing derivative increases C-to-T conversion efficiency by 5.36- to 6.85-fold without introducing detectable indels (Bai et al., 2025b). The ability to minimise indel formation is particularly relevant for precise allele engineering in soybean.

2.2 | Multiplex Editing and Trait Stacking

Multiplex genome editing enables the simultaneous modification of multiple loci using several guide RNAs and is

particularly valuable in soybean because of its paleopolyploid genome and extensive gene duplication as a consequence of ancient whole-genome duplication events. Recent studies have demonstrated parallel knockout of four *GmSPL9* genes and three *GmPIN1* genes associated with shoot architecture (Li et al., 2025). However, increasing the number of target sites substantially reduces the frequency of biallelic mutations, with simultaneous editing of four to six genes yielding biallelic rates of only 4–17% (Li et al., 2025).

Despite these efficiency constraints, multiplex editing currently represents the most practical strategy for stacking traits controlled by gene families or duplicated loci and for generating higher-order loss-of-function combinations in soybean.

2.3 | Precision Editing: Base Editing and Prime Editing

Base editing enables precise nucleotide conversion without creating double-strand breaks. Cytosine and adenine base editors mediate C-to-T and A-to-G substitutions, respectively, and systematic screening identified ABE8e and A3A/Y130F-V04 as the most efficient adenine and cytosine base editors for both *Arabidopsis* and soybean within defined activity windows (Jeong et al., 2025).

Prime editing further enables precise base substitutions and small insertions or deletions without double-strand breaks, but its efficiency in dicot crops has been limited. A recently developed flanking-nicks prime editor (FLICK-PE) system markedly improved prime editing in soybean, achieving a 15.7-fold increase in intended editing efficiency over PE2 and enabling three amino-acid substitutions in *EPSPS1a* to confer glyphosate resistance with 21.1% editing efficiency (4/19 lines) and stable field performance, in contrast to PE3, which achieved an effective intended editing efficiency of 8.7% (2/23 lines) (Bai et al., 2025a).

2.4 | DNA-Free Delivery and Regulatory Streamlining

DNA-free genome editing based on delivery of preassembled ribonucleoprotein (RNP) complexes avoids the use of plasmid DNA and eliminates the risk of transgene integration. An *in planta* bombardment-RNP (iPBRNP) strategy targeting shoot apical meristem cells has achieved heritable edits at frequencies of 0.4–4.6% without tissue culture and without strong genotype dependence (Kuwabara et al., 2024).

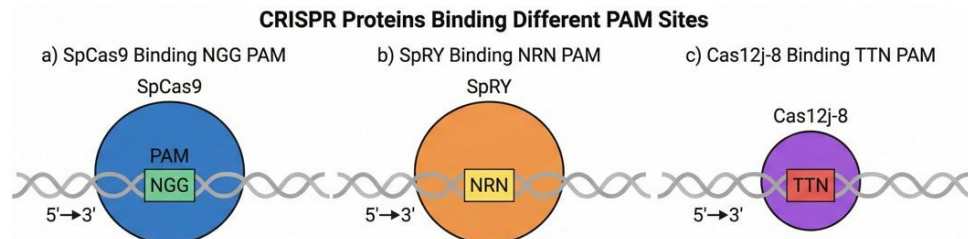


Figure 1: Comparison of CRISPR nucleases with different PAM requirements and protein sizes. DNA binding preferences of SpCas9 (NGG PAM), SpRY (NRN PAM) and Cas12j-8 (TTN PAM).

2.5 | AI-Driven Target Prioritisation and Design

Machine-learning models are increasingly used to predict guide RNA activity and off-target propensity using soybean-specific genomic features, thereby reducing empirical screening and enabling genome-scale target prioritisation in *Glycine max* (Das et al., 2023). In parallel, deep-learning frameworks are applied to infer regulatory and epigenomic features, including DNA and RNA modification sites and chromatin accessibility, supporting data-driven identification of regulatory elements and targets for genome and epigenome editing (Zhang et al., 2025). More recently, large language model-based agents, such as CRISPR-GPT, have been proposed to assist genome-editing experiment design by integrating guide RNA design, editing system selection, delivery strategy recommendation and validation planning into unified computational workflows (Zhang et al., 2025).

Beyond target discovery, structure-based deep-learning tools such as AlphaFold further enable rational, allele-level editing by guiding the design of functional variants, as demonstrated for *GmSWEET10b*-mediated oil improvement in soybean (Wang et al., 2025b).

3 | GENES TARGETED

In soybean, genome editing has been applied to improve disease resistance, drought and salinity tolerance, flowering time, yield, herbicide tolerance, plant architecture, seed oil composition, palatability and phytic acid content, and has also been widely used to support functional genomic studies and gene validation (Vargas-Almendra et al., 2024). Table 1 summarises representative genes that have been edited in soybean, the associated target traits, editing strategies and the main phenotypic outcomes reported in recent studies.

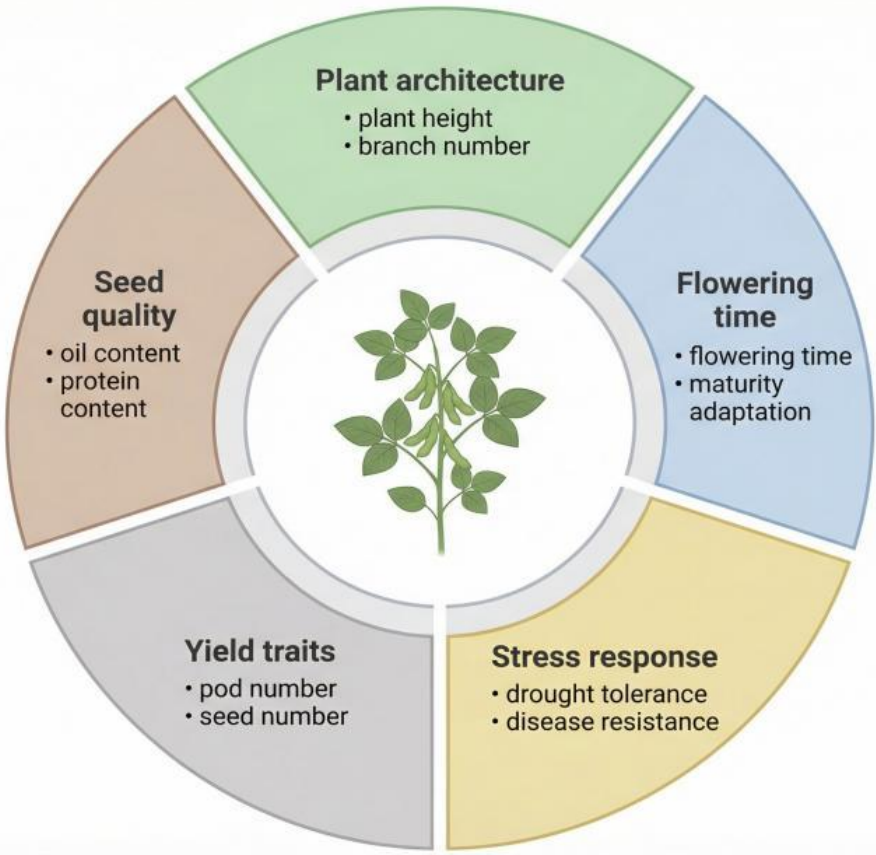


Figure 2: Conceptual overview of major functional categories of genes that have been targeted for genome editing in soybean. Example traits shown under each category illustrate typical biological outcomes reported in the literature.

Table 1: Genes targeted by genome editing in soybean.

Gene	Trait / biological role	Editing platform	Editing strategy	Main outcome	References
GmFAD2-1A/GmFAD2-1B	Fatty acid composition (oil quality; oleic and linolenic acid content)	TALEN	Knockout	Reduced linolenic acid (~2.5%) and increased oleic acid (~82.2%)	(Demorest et al., 2016)
LOX1, LOX2, LOX3	Beany flavour / lipid oxidation	CRISPR–Cas9 (pooled system)	Multiplex knockout	Reduced off-flavour; transgene-free line obtained	(Wang et al., 2020)
SPL9 family	Plant architecture/regulation of plastochron length	CRISPR–Cas9	Multiplex indel based knockout	Functional validation of <i>GmSPL9</i> genes; higher-order mutants showed increased node and branch number	(Bao et al., 2019)
GmSWEET10a, GmSWEET10b	Seed oil content	CRISPR–Cas9	AlphaFold- guided gene editing of C- terminal region	High-oil alleles (sw10b-ho-1, sw10bho-2) increased oil content without affecting seed yield	(Wang et al., 2025b)
PPD1, PPD2	Leaf and pod morphology	CRISPR–Cas9	Multiplex knockout	≥33% of T ₂ seeds carried mutations in both loci; ~19% of double mutants were transgene-free	(Kanazashi et al., 2018)
GmARM	Multiple stress tolerance (salt, alkali and <i>Phytophthora</i> root rot resistance)	CRISPR–Cas9	Knockout (dual-sgRNA, indel- based)	Improved salt and alkali tolerance and enhanced resistance to <i>Phytophthora</i> root rot	(Luo et al., 2024)
SMV HC-Pro (viral gene)	Resistance to soybean mosaic virus (SMV)	CRISPR–CasRx	Multiplex RNA targeting of viral HC-Pro gene	Up to 39.02% of T ₁ plants were highly resistant and 35.77% were resistant to SMV	(Gao et al., 2024)
GmAHAS4	Resistance To AHAS inhibiting herbicides	CRISPR/Cas9 cytosine base editor (BE3)	Precise C→T base substitution generating a P180S point mutation in GmAHAS4	Transgene-free, heritable herbicide resistant soybean with normal growth	(Wei et al., 2023)
GmFT2a, GmFT5a	Major flowering time regulators controlling floral transition and photoperiod response in soybean	CRISPR–Cas9	Dual-sgRNA mediated targeted deletions in <i>GmFT2a</i> and <i>GmFT5a</i> homologs	Stable mutant plants with altered flowering time; mutation frequencies of ~31– 42% and clear effects under different genetic backgrounds	(Freitas-Alves et al., 2025)
GmPRR3b, GmPRR37	Circadian clock– associated regulators involved in control of flowering time	CRISPR–Cas9	Targeted knockout of major functional alleles (including <i>GmPRR3b</i> H6 and <i>GmPRR37</i>)	<i>GmPRR3b</i> knockout delayed growth and floral transition, while <i>GmPRR37</i> mutants showed early flowering under long-day conditions	(Freitas-Alves et al., 2025)
GmLTK2 (homologous genes)	Light- and clock regulated genes influencing photoperiod dependent flowering	CRISPR–Cas9	Multiplex knockout of four homologous <i>GmLTK2</i> genes (quadruple mutants)	Quadruple mutants flowered significantly earlier under long-day conditions	(Freitas-Alves et al., 2025)

3.1 | Discussion

Table 1 shows that soybean genome-editing studies to modify agronomic and other traits. This trend is evident for genes controlling plant architecture and flowering, including the SPL9 family, PPD1/PPD2, GmFT2a, GmFT5a, GmPRR3b and LNK2 homologues, where loss-of-function mutations were used to both validate gene function and modify agronomic traits. Several studies in Table 1 also demonstrate trait improvement through precise allele engineering rather than simple gene disruption. For example, GmAHAS4 was edited by cytosine base editing to generate herbicide resistant plants, and GmSWEET10a/b were modified using an AlphaFold-guided CRISPR strategy to obtain high-oil alleles. Likewise, GmFAD2-1A and GmFAD2-1B were edited using TALENs to improve oil quality.

Multiplex editing is repeatedly applied when gene families or functionally redundant genes are targeted, such as LOX1, LOX2 and LOX3, as well as multiple flowering time regulators including GmFT2a, GmFT5a and LNK2, highlighting the importance of simultaneously editing several loci in soybean.

Table 1 also includes examples of both biotic and abiotic stress improvement, such as editing of GmARM for multi stress tolerance and the use of CRISPR/CasRx to target the soybean mosaic virus genome, illustrating that genome editing in soybean is being applied beyond developmental and quality-related traits.

Overall, Table 1 indicates that Cas9-based knockout and multiplex strategies remain the most commonly used approaches in soybean, whereas base editing and structure guided editing are increasingly adopted when precise and predictable trait modification is required.

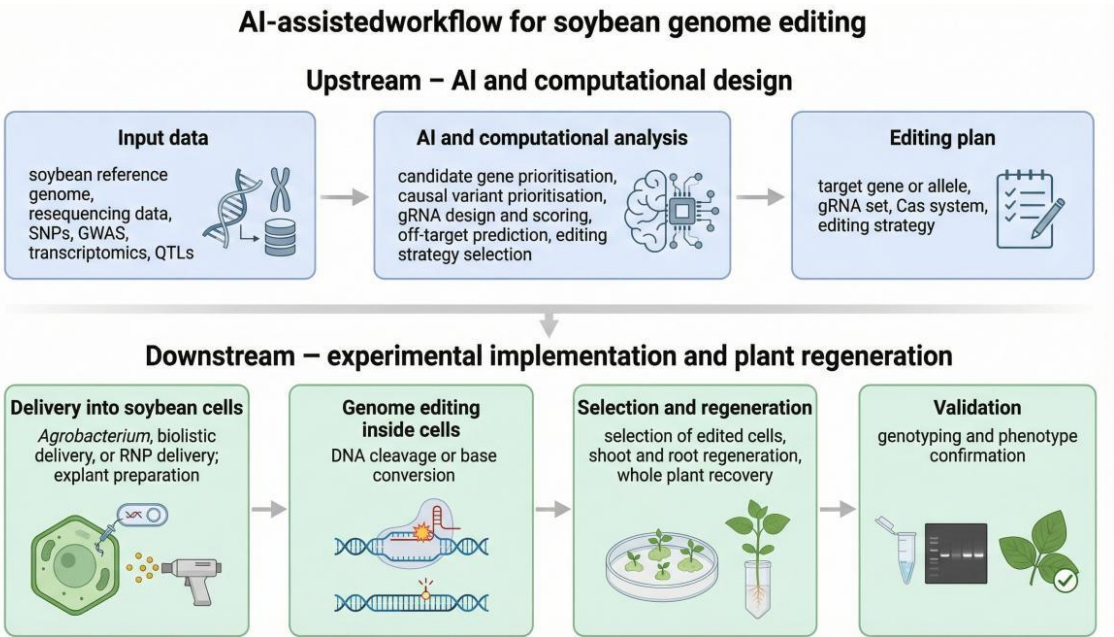
4 | GENOME EDITING TECHNOLOGIES AND THEIR PREFERRED APPLICATIONS IN SOYBEAN

Different genome editing technologies are suitable for different research and breeding goals in soybean. The choice of an editing approach mainly depends on whether the objective is to disrupt a gene, introduce a precise sequence

change, modify several genes at the same time, or rapidly test gene function before developing stable edited lines. Table 2 summarises the main genome editing platforms and technical strategies currently used in soybean and outlines the situations in which each approach is most practical and effective.

Table 2: Various genome editing technologies and their preferred applications in soybean.

Technology / platform	Main application in soybean	Key advantage	Primary limitation
CRISPR–Cas9 (SpCas9)	Routine targeted knockout of single genes in soybean when suitable NGG PAM sites are present near the target region	Well-established and easy-to-use system with reliable performance and broad community experience in soybean studies	Editing is restricted to genomic sites containing the required PAM sequence and mainly produces small insertions or deletions rather than precise changes
CRISPR–SpRY (PAM-relaxed)	Editing of genes or regulatory regions that cannot be accessed using standard PAM restricted CRISPR systems	Greatly expands the range of editable sites across the soybean genome, allowing targets that were previously inaccessible	The relaxed PAM requirement increases the need for careful guide design and off-target assessment
CRISPR–Cas12a (Cpf1)	Editing of AT-rich genomic regions and construction of multiplex editing systems in soybean	Uses a different PAM requirement and simplifies multiplex guide array design for targeting several loci	Editing efficiency can be inconsistent and remains genotype- and target dependent in soybean
Multiplex CRISPR editing	Simultaneous modification of multiple related genes or duplicated gene copies that control the same trait	Enables coordinated editing of gene families or redundant genes using a single transformation event	The probability of obtaining plants carrying complete edits at all target loci is lower than for single-gene editing
Base editing (CBE/ABE)	Introduction of precise single-base substitutions in genes controlling agronomic and quality traits	Allows predictable nucleotide conversion without creating double-strand breaks, reducing unintended large mutations	The editable positions are limited to a narrow window and nearby bases may be unintentionally modified
Prime editing	Precise base changes and small insertions or deletions for fine-scale trait improvement	Provides flexible and programmable sequence changes without generating double-strand breaks	The system is technically complex and editing efficiency in soybean is still variable and often low
DNA-free RNP delivery (e.g. iPB-RNP)	Production of transgene-free edited soybean plants suitable for downstream breeding and regulatory compliance	Avoids stable DNA integration and, in the iPB-RNP system, can reduce dependence on genotype-specific tissue culture	Stable and heritable editing frequencies are currently much lower than DNA-based transformation approaches
Hairy-root and protoplast systems	Rapid screening and functional testing of candidate genes before committing to stable plant transformation	Enables fast and low-cost validation of gene function without lengthy regeneration procedures	Hairy-root edits are not transmitted to the next generation, and regeneration of whole soybean plants from protoplasts remains technically difficult



Overall, there is no single best genome editing technology for soybean. The most suitable method depends on the specific objective of the study. Standard CRISPR–Cas9 remains the preferred choice for routine gene knockout and gene function studies because it is simple and well established in soybean.

Base editing and prime editing are more appropriate when precise sequence changes are required, such as modifying an existing allele without completely disrupting gene function. Multiplex genome editing is particularly useful in soybean because many genes occur as duplicated copies or belong to gene families, and several genes often need to be edited simultaneously to obtain clear phenotypic effects.

For rapid gene function testing, protoplast-based and hairy-root systems are valuable because results can be obtained quickly. However, these approaches cannot be used directly to generate stable breeding materials. Consequently, stable transformation and regeneration methods are still required when the goal is to develop improved soybean varieties.

5 | ARTIFICIAL INTELLIGENCE IN SOYBEAN GENOME EDITING

Recent advances in artificial intelligence (AI) have expanded its role in soybean research from large-scale phenotyping and genomic prediction to more direct support of genome editing workflows. Broad applications of AI in soybean breeding and production include phenomics, genomic selection, stress detection, yield prediction and cyber-agricultural systems for decision support and crop improvement (Singh et al., 2025). More recently, AI has also begun to contribute directly to the upstream design stage of genome editing, as demonstrated by an AlphaFold guided strategy to design superior *GmSWEET10b* protein variants that were subsequently introduced into soybean using precise genome editing and validated under field conditions (Wang et al., 2025b). Together, these studies indicate that AI is increasingly influencing both the computational design of editing targets and the broader breeding and evaluation pipeline, providing a basis for organising AI-assisted soybean genome editing into upstream computational design and downstream experimental implementation (Figure 3).

5.1 | AI-Assisted Design and Target Prioritisation for Upstream Genome-Editing

In this review, the term *upstream* refers to the computational and data-driven steps that support genome-editing design prior to experimental implementation, including candidate gene prioritisation, off-target prediction, guide RNA design and editing strategy selection (Figure 3).

Several artificial intelligence and machine learning approaches have already been developed and applied at this upstream stage in soybean and other crop species. For regulatory target discovery, the SoyTFtarget platform integrates motif information, co-expression analysis, family specific and global machine-learning models together with large-scale DAP-seq data to predict transcription factor–target relationships in soybean, thereby facilitating the identification

of regulatory genes for downstream editing applications (Zhu et al., 2025).

For trait-oriented target prioritisation, deep learning based genomic selection frameworks such as CropARNet and Cropformer have been developed to model complex genotype–phenotype relationships across large multimarker datasets. These models employ convolutional and attention-based architectures to improve prediction accuracy and robustness for complex agronomic traits and to assist in identifying genomic regions and variants associated with trait variation (Zhou et al., 2025; Wang et al., 2025a). Although these models are primarily used for breeding value prediction, their outputs provide valuable guidance for selecting candidate loci and pathways for subsequent genome editing.

In addition to target prioritisation, AI models have also been developed to predict editing outcomes. The BEEndo deep-learning framework incorporates both sequence features and endogenous genomic factors, such as chromatin state and transcriptional activity, to improve the prediction accuracy of adenine and cytosine base-editing efficiencies, providing useful guidance for the design of precise editing strategies (Yuan et al., 2024).

5.2 | AI Applications in Downstream Genome Editing and Plant Regeneration Pipelines

Following upstream computational target selection and editing design (Figure 3), genome editing enters a downstream experimental pipeline in which candidate edits are implemented and validated *in planta*. This stage typically involves delivery of editing reagents (e.g. *Agrobacterium*-mediated transformation, biolistic delivery or RNP delivery), induction of DNA cleavage or base conversion, selection of edited cells in tissue culture, shoot and root regeneration, whole-plant recovery, and final genotyping and phenotyping for edit confirmation.

Current AI applications in this downstream phase are mainly concentrated in plant regeneration and phenotyping rather than in transformation or molecular validation steps. Machine-learning models, particularly artificial neural networks and hybrid optimisation frameworks, have been widely surveyed and applied to optimise tissue culture and regeneration conditions, including culture media composition and growth regulator regimes, in order to improve regeneration efficiency and reduce experimental trial-and-error (Ibrahim et al., 2023).

For post-regeneration assessment, AI has been more actively adopted in phenotyping and field-level validation. An open-source platform, AirMeasurer, integrates computer vision and machine-learning algorithms with low-cost drone imagery to automatically estimate agronomic traits such as plant height, canopy coverage and growth stage, thereby enabling high-throughput evaluation of plant populations under field conditions (Sun et al., 2022). In addition, deep learning combined with feature selection based on random-forest feature importance and recurrent neural networks has been used to analyse multi-temporal satellite imagery for crop classification and monitoring, providing scalable phenotyping

and crop status assessment across large agricultural areas (Yuan et al., 2023).

Overall, downstream AI applications in crop genome editing remain largely focused on regeneration optimisation and phenotypic assessment after plant recovery, with limited use in transformation, editing activity monitoring, and molecular confirmation steps.

5.3 | Future Directions for AI-Enabled Genome Editing in Crops

While upstream stages of genome editing increasingly benefit from advanced AI-based target discovery and editing design (Section 4.1), the downstream experimental pipeline remains constrained by several fundamental biological and technical bottlenecks. A major limitation is the low and strongly genotype-dependent transformation efficiency, together with substantial variability in tissue culture and regeneration responses across crop species and cultivars (Lee and Wang, 2023). These factors significantly restrict the scalability and reproducibility of genome editing pipelines, particularly in elite and locally adapted germplasm.

In addition, the full potential of genome editing in crop improvement is currently limited by an incomplete understanding of complex biological processes involving genes, pathways and regulatory networks, as well as their interactions with environmental factors (Yang, 2020). This restricts the identification of robust and agronomically meaningful targets that are suitable for precise genome editing. Although CRISPR-based technologies have enabled high-throughput functional genomics in both model and non-model crops, the integration of multidisciplinary data required for effective target prioritisation remains a major challenge.

Another critical bottleneck lies in achieving efficient and precise base replacement in plants. While most genome editing applications still focus on targeted mutagenesis, the efficiency and applicability of base editing and precise nucleotide replacement remain low in many crop species, and are absent in several important crops (Yang, 2020).

Looking forward, advances in artificial intelligence and machine learning offer substantial opportunities to address these downstream challenges. AI-driven integration of multi-omics, phenomics and environmental datasets could improve the prioritisation of biologically meaningful and editable targets, support genotype-flexible transformation strategies, and guide the optimisation of tissue culture and regeneration protocols. Moreover, data-driven modelling of editing outcomes and repair processes may help accelerate the development of more efficient base replacement systems. Collectively, expanding the application of AI across downstream genome editing stages represents a critical direction for improving the reliability, precision and translational impact of crop genome editing pipelines.

ACKNOWLEDGEMENT

Authors acknowledge the support from Indian Council of Agricultural Research, Department of Agricultural Research and Education, Government of India. (ICAR) Genome Editing Project and MEITY, AgriHub Project (No.IITI/AgriHub/2025/PRJ-I/02).

REFERENCES

- Bai, M., Zhang, J., Lin, W., Zhou, Y., Jiang, M., Wu, H., Peng, C., Lin, J., He, F., Kuang, H., and Guan, Y. (2025a). A flanking-nicks prime editor (FLICK-PE) system to boost prime editing in dicots. *Nat. Commun.*, 17(1):337.
- Bai, S., Cao, X., Hu, L., Hu, D., Li, D., and Sun, Y. (2025b). Engineering an optimized hypercompact CRISPR/Cas12j-8 system for efficient genome editing in plants. *Plant Biotechnol. J.*, 23(4):1153–1164.
- Bao, A., Chen, H., Chen, L., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant Biol.*, 19:131.
- Chen, X., Zhong, Z., Tang, X., et al. (2024). Advancing PAM-less genome editing in soybean using CRISPR-SpRY. *Hortic. Res.*, 11(8):uhae160.
- Das, J., Kumar, S., Mishra, D. C., Chaturvedi, K. K., Paul, R. K., and Kairi, A. (2023). Machine learning in the estimation of CRISPR-Cas9 cleavage sites for plant system. *Front. Genet.*, 13:1085332.
- Demorest, Z. L., Coffman, A., Baltes, N. J., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol.*, 16(1):225.
- Dixit, S., Kumar, A., Srinivasan, K., Durai Raj, V. P. M., and Nadesh, R. K. (2024). Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions. *Front. Bioeng. Biotechnol.*, 11:1335901.
- Freitas-Alves, N. S., Moreira-Pinto, C. E., Távora, F. T.P. K., Paes-de Melo, B., Arraes, F. B. M., Lourenço-Tessutti, I. T., Moura, S. M., Oliveira, A. C., Morgante, C. V., Qi, Y., and Grossi-de Sa, M. F. (2025). CRISPR/Cas genome editing in soybean: challenges and new insights to overcome existing bottlenecks. *J. Adv. Res.*, 73:53–72.
- Gao, L., Xie, L., Xiao, Y., et al. (2024). CRISPR/CasRx-mediated resistance to soybean mosaic virus in soybean. *Crop J.*, 12(4):1093–1101.
- He, Y., Mudgett, M., and Zhao, Y. (2022). Advances in gene editing without residual transgenes in plants. *Plant Physiol.*, 188(4):1757–1768.
- Ibrahim, N., Adedamola, A. M., Ibrahim, B., Ahmed, R. T., Raji, I. D., and Bello-Salau, H. (2023). Survey of machine learning and optimization algorithms in plant tissue culture. *Eng. Proc.*, 56(1):122.
- Jeong, Y. Y., Han, J. H., Yu, J., Bae, S., and Seo, P. J. (2025). Identification of optimal adenine and cytosine base editors

- for genome editing in *Arabidopsis* and soybean. *BMB Rep.*, 58(7):288–292.
- Kanazashi, Y., Hirose, A., Takahashi, I., et al. (2018). Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. *Plant Cell Rep.*, 37(3):553–563.
- Krishna, T. P. A., Veeramuthu, D., Maharajan, T., and Soosaimanickam, M. (2023). The era of plant breeding: conventional breeding to genomics-assisted breeding for crop improvement. *Curr. Genomics*, 24(1):24–35.
- Kuwabara, C., Miki, R., Maruyama, N., et al. (2024). A DNA-free and genotype-independent CRISPR/Cas9 system in soybean. *Plant Physiol.*, 196(4):2320–2329.
- Lee, K. and Wang, K. (2023). Strategies for genotype-flexible plant transformation. *Curr. Opin. Biotechnol.*, 79:102848.
- Li, N., Yuan, X., Han, B., Guo, W., and Chen, H. (2025). CRISPR/Cas-mediated optimization of soybean shoot architecture for enhanced yield. *Int. J. Mol. Sci.*, 26(16):7925.
- Luo, T., Ma, C., Fan, Y., et al. (2024). CRISPR-Cas9-mediated editing of *GmARM* improves resistance to multiple stresses in soybean. *Plant Sci.*, 346:112147.
- Mangal, V., Verma, L. K., Singh, S. K., Saxena, K., Roy, A., Karn, A., et al. (2024). Triumphs of genomic-assisted breeding in crop improvement. *Heliyon*, 10(15):e35513.
- Ren, Q., Sretenovic, S., Liu, S., et al. (2021). PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat. Plants*, 7(1):25–33.
- Singh, A., Jones, S., Van der Laan, L., Ayanlade, T., Raigne, J., Saleem, N., et al. (2025). Use of artificial intelligence in soybean breeding and production. *Adv. Agron.*, 180:1–55.
- Sun, G., Lu, H., Zhao, Y., Zhou, J., Jackson, R., Wang, Y., et al. (2022). AirMeasurer: open-source software to quantify static and dynamic traits derived from multiseason aerial phenotyping to empower genetic mapping studies in rice. *New Phytol.*, 236(4):1584–1604.
- Tester, M. and Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science*, 327(5967):818–822.
- Vargas-Almendra, A., Ruiz-Medrano, R., Núñez-Muñoz, L. A., Ramírez-Pool, J. A., Calderón-Pérez, B., and Xoconostle-Cázares, B. (2024). Advances in soybean genetic improvement. *Plants (Basel)*, 13(21):3073.
- Wang, H., Yan, S., Wang, W., Chen, Y., Hong, J., He, Q., Diao, X., Lin, Y., Chen, Y., Cao, Y., Guo, W., and Fang, W. (2025a). Cropformer: an interpretable deep learning framework for crop genomic prediction. *Plant Commun.*, 6(3):101223.
- Wang, J., Kuang, H., Zhang, Z., et al. (2020). Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. *Crop J.*, 8(3):432–439.
- Wang, J., Zhang, L., Wang, S., et al. (2025b). AlphaFold-guided bespoke gene editing enhances field-grown soybean oil contents. *Adv. Sci. (Weinh.)*, 12(23):e2500290.
- Wei, T., Jiang, L., You, X., Ma, P., Xi, Z., and Wang, N. N. (2023). Generation of herbicide-resistant soybean by base editing. *Biology*, 12(5):741.
- Yang, B. (2020). Grand challenges in genome editing in plants. *Front. Genome Ed.*, 2:2.
- Yuan, T., Wu, L., Li, S., et al. (2024). Deep learning models incorporating endogenous factors beyond DNA sequences improve the prediction accuracy of base editing outcomes. *Cell Discov.*, 10:20.
- Yuan, X., Liu, S., Feng, W., and Dauphin, G. (2023). Feature importance ranking of random forest-based end-to-end learning algorithm. *Remote Sens.*, 15(21):5203.
- Zhang, D., Xu, F., Wang, F., Le, L., and Pu, L. (2025). Synthetic biology and artificial intelligence in crop improvement. *Plant Commun.*, 6(2):101220.
- Zhou, S., Cheng, K., Lv, L., Jiang, J., Zhou, S., Zhou, Y., Xu, Z., Huang, Q., Yang, H., Chen, L., Xu, Y., Yao, Z., and Zhao, T. (2025). CropARNet: a deep learning framework for crop genomic prediction with attention and residual modules. *Crop Des.*, 4(4):100118.
- Zhu, X., Zhang, B., Yang, L., Huang, J., Huang, F., Zhang, H., Song, Q., and Huang, J. (2025). A machine learning-based tool for predicting transcription factor–target gene interactions in soybean. *Plant Physiol.*, 198(4):kiaf318.