



OPEN ACCESS

EDITED BY

Vladimir Nekrasov,
Rothamsted Research, United Kingdom

REVIEWED BY

Ellen Slaman,
Flemish Institute for Biotechnology, Belgium
Anindya Bandyopadhyay,
International Maize and Wheat Improvement
Center, Mexico

*CORRESPONDENCE

Magdy Mahfouz

✉ magdy.mahfouz@kaust.edu.sa

RECEIVED 15 June 2024

ACCEPTED 12 August 2024

PUBLISHED 02 September 2024

CITATION

Kababji AM, Butt H and Mahfouz M (2024)
Synthetic directed evolution for targeted
engineering of plant traits.
Front. Plant Sci. 15:1449579.
doi: 10.3389/fpls.2024.1449579

COPYRIGHT

© 2024 Kababji, Butt and Mahfouz. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Synthetic directed evolution for targeted engineering of plant traits

Ahad Moussa Kababji, Haroon Butt and Magdy Mahfouz*

Laboratory for Genome Engineering and Synthetic Biology, Division of Biological Sciences, 4700 King
Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Improving crop traits requires genetic diversity, which allows breeders to select advantageous alleles of key genes. In species or loci that lack sufficient genetic diversity, synthetic directed evolution (SDE) can supplement natural variation, thus expanding the possibilities for trait engineering. In this review, we explore recent advances and applications of SDE for crop improvement, highlighting potential targets (coding sequences and *cis*-regulatory elements) and computational tools to enhance crop resilience and performance across diverse environments. Recent advancements in SDE approaches have streamlined the generation of variants and the selection processes; by leveraging these advanced technologies and principles, we can minimize concerns about host fitness and unintended effects, thus opening promising avenues for effectively enhancing crop traits.

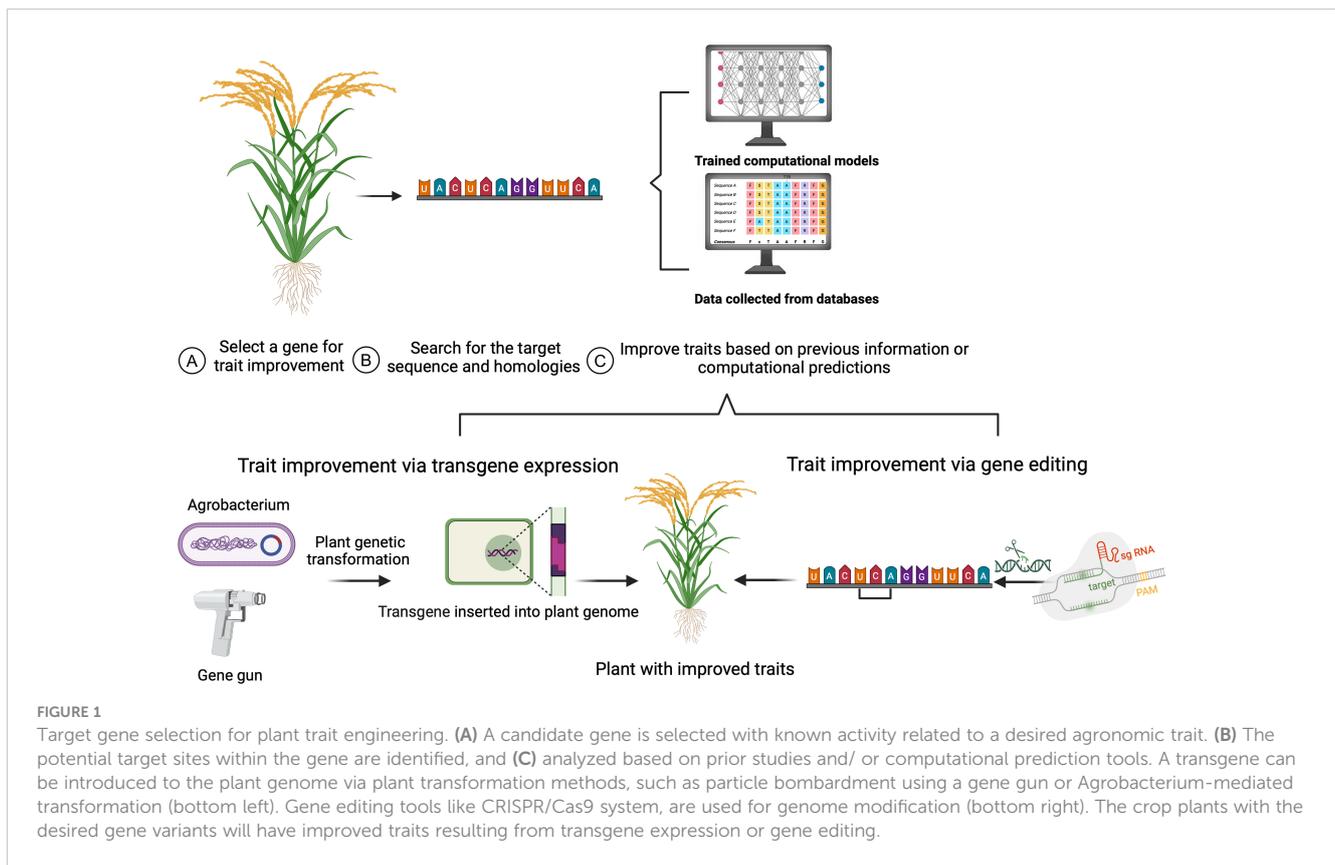
KEYWORDS

CRISPR-Cas9, synthetic directed evolution, CRISPR-directed evolution, rational protein design, trait engineering, crop breeding, noncoding DNA, regulatory elements

1 Introduction

Plant breeding introduces genetic diversity into breeding populations by intercrossing plants with outstanding traits and then selecting progeny with desired characteristics (Zsogon et al., 2022). However, many species and loci lack sufficient genetic diversity, meaning that breeders do not have germplasm with alleles that will produce the desired characteristics. Genome engineering technologies, including the CRISPR/Cas9 system, offer new avenues to improve crop traits (Das et al., 2022) (Figure 1). Although CRISPR/Cas systems have been primarily used to generate loss-of-function alleles, they can also be used as a powerful tool for generating genetic diversity at specific loci. This enables new, innovative approaches such as synthetic directed evolution (SDE).

SDE uses iterative cycles of artificial diversification of a gene sequence, followed by selection for a trait of interest (*in vitro* or *in vivo*) to produce desirable alleles. This versatile approach can generate genetic diversity and drive the evolution of organisms towards



desired traits (Simon et al., 2019). Recent advancements have automated all essential stages of SDE, significantly reducing the time required for variant generation and minimizing the effort demanded from researchers during the transfection, diversification, and screening stages (Packer and Liu, 2015). Modern SDE methods do not require prior knowledge of the gene background or structure as necessary for rational engineering. To generate variants, when specific genes are targeted via these modern SDE techniques, there are fewer concerns regarding host fitness and viability. Instead, variants are selected based on their performance under a range of selective pressures. However, SDE also presents some challenges, including designing selective assays, generating large, unbiased DNA libraries, avoiding false signals, and significant manual effort during mutagenesis and selection (Chen, 2001).

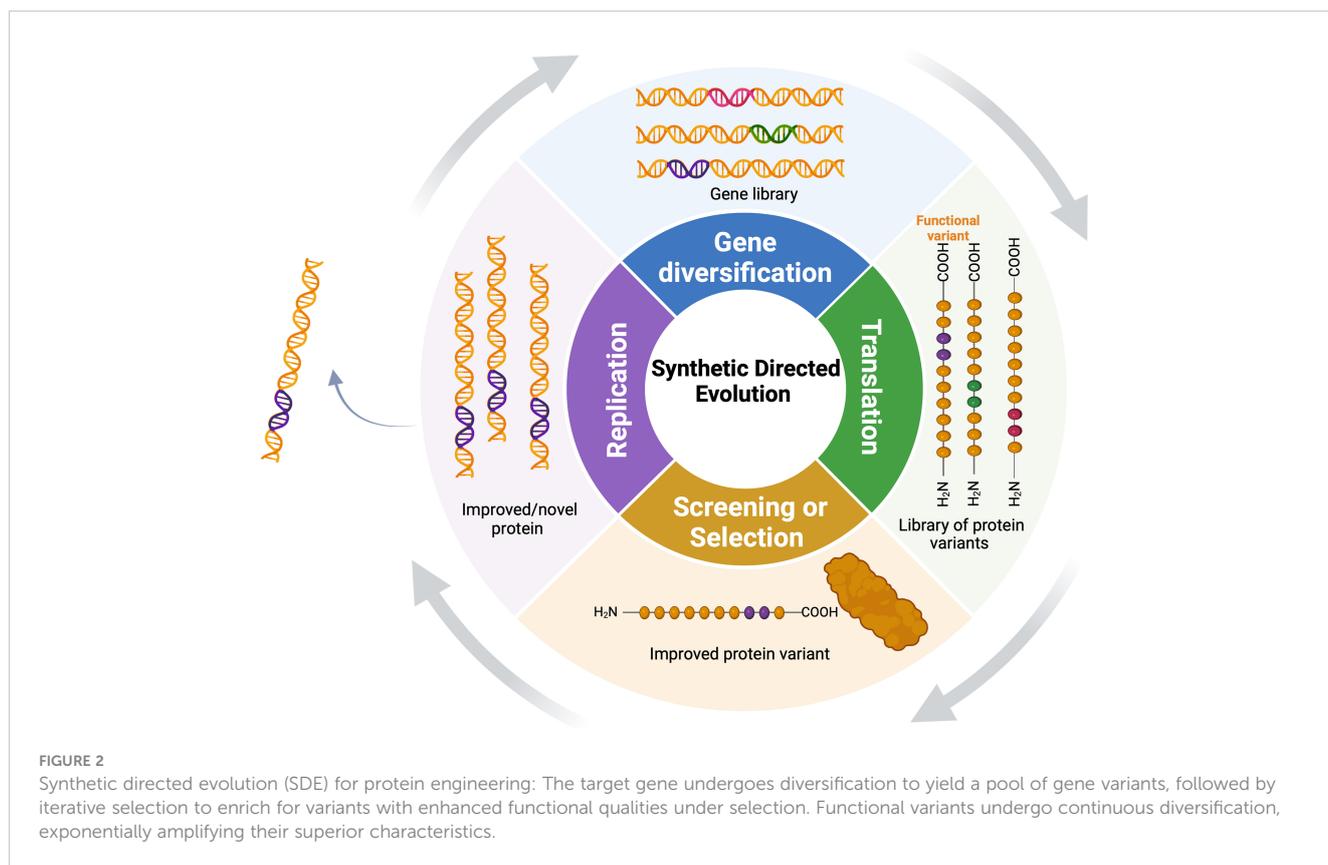
Here, we focus on SDE approaches for protein engineering to generate crop lines with improved agronomic and climate-resilient traits. We shed light on current efforts to harness SDE and expand genetic diversity for crop improvement. The SDE technologies have been utilized in plants to engineer and improve herbicide resistance. Later, we provide an overview of engineering *cis*-regulatory elements (CRE) and motifs in plant trait improvement. We discuss the computational tools and databases for predicting *cis*-regulatory elements in plants. At the end, we discuss with suitable examples how targeting promoter regions, upstream-open reading frames (uORF), enhancers, transcription factors, 5'-UTR, and 3'-UTR can improve plant traits. We discuss the constraints and prospects inherent in directed evolution for trait improvement in crops.

2 Synthetic directed evolution

SDE has two basic steps: in the diversification stage, researchers introduce changes in specific gene sequences, generating a library of gene variants encoding a diverse set of protein variants (Figure 2). The diversification step can involve mutagenesis at various levels, including single genes, multiple genes, genomes, or within complex pathways, employing direct (gene-focused) or random (whole-genome) platforms (Wang et al., 2021). Depending on the methods used for diversification, these variants can be handled *in vitro*, transformed into cells or organisms, or generated *in vivo*. In the selection stage, the host population is subjected to a specific selection pressure to identify individuals with the desired traits (Simon et al., 2019). Selection and screening enable the isolation and recovery of only those variants that exhibit the desired functional characteristics. Subsequently, selected variants undergo multiple rounds of SDE to further enrich and amplify desired traits.

As the field of SDE has advanced, there has been a transition from manipulating individual components and modifying products to address more complex pathways. This transition was highlighted by the awarding of the Nobel Prize in Chemistry to Frances H. Arnold, George P. Smith, and Sir Gregory P. Winter in 2018, for their work on development of new pharmaceutical enzymes through directed evolution (Arnold, 2019; Service, 2018).

Although SDE has shown great promise in bacteria, yeast, and mammalian cell lines, it has fewer success stories in plants and most of these have been proof-of-concept studies generating herbicide tolerance by mutating genes known to confer this phenotype.



Nevertheless, leveraging SDE has the potential to enable the discovery of novel traits, broadening the spectrum of useful traits, and accelerating trait development in targeted crops essential for food security (Rao et al., 2021; Butt et al., 2020; Garcia-Garcia et al., 2022; Wright et al., 2023). Moreover, the integration of SDE principles with biotechnology methodologies has revealed extensive potential for generating protein variants, enhancing natural enzyme catalytic activity, probing desired phenotypes, and introducing novel traits in diverse living organisms.

2.1 Methods and strategies for synthetic directed evolution

Diversification can be performed with various strategies, including random mutagenesis, DNA recombination and CRISPR/Cas-based methods (Wang et al., 2021). The groundwork for SDE was laid in the early 1970s when a chemical mutagens like ethyl methanesulfonate (EMS) or 1-methyl-3-nitro-1-nitrosoguanidine (NG) induced a new functional phenotype in *Aerobacter aerogenes* (Lerner et al., 1964). Use of a phage-expressing library as an *in vitro* diversification method yielded novel and desirable properties for various targeted genes (Levisohn and Spiegelman, 1969). Subsequent *in vitro* studies introduced mutations into specific genetic regions, followed by iterations of mutation and *in vivo* screening. This led to the gradual accumulation of advantageous mutations in a strategy termed combinatorial cassette mutagenesis (Reidhaar-Olson and Sauer, 1998; Mills et al., 1967).

Error-prone polymerase chain reaction (PCR) has emerged as a groundbreaking technique for *in vitro* random mutagenesis and diversifying sequences across multiple rounds, either segmentally or entirely. CasPER (Cas9-mediated protein evolution reaction) employ CRISPR/Cas9 for DSB at genomic targets to integrate mutagenized 300–600 bp linear DNA donors for the directed evolution of enzymes in native genomic contexts. The CasPER technique used error-prone PCR to generate combinatorial libraries serving as templates for introducing diversity into two essential enzymes within the mevalonate pathway (Jakociunas et al., 2018). However, random mutagenesis using error-prone PCR has limitations such as the low mutation rate, restricted amino acid variants along the target gene, and the low probability of encountering adjacent mutations (Chen and Arnold, 1993, 1991; Wen et al., 2007). One variation of epPCR is site saturation mutagenesis (SSM) which displayed substantial enhancement in functional performance. OmniChange, a multi-site-saturation method, uses PCR and synthetic DNA oligonucleotides for simultaneous and efficient saturation of five independent codons (Dennig et al., 2011).

In the mid-1990s, the recombination approach exemplified by DNA shuffling was established as an *in vitro* diversification strategy, exchanging larger fragments from similar genes to generate a library of functional variants (Stemmer, 1994a, b; Cramer et al., 1998; Jermutus, 2000). SDE through genome shuffling or inter-strain fragment shuffling has been widely employed in various bacterial strains and other organisms (Zhang et al., 2002; Patnaik et al., 2002; Schmidt-Dannert et al., 2000; Kumamaru et al., 1998; Biot-Pelletier

and Martin, 2014). The merging of enhanced variations via DNA recombination after selection enables the integration of functional and improved variants while removing harmful ones, resulting in significant trait improvements over the course of multiple generations.

Another approach uses oligo libraries suitable for recombineering and CRISPR-based genome editing. It generates unique collections of codons for each amino acid to be mutated, ensuring the desired level of redundancy necessary to efficiently achieve a diverse population. Excluding the wild-type amino acid helps reduce library size without compromising diversity. As an application, λ -red recombineering via CRISPR-based-selection was accomplished by creating an engineered cell library devoid of wild-type amino acids or stop codons, lacking redundant codons, and incorporating highly utilized codons in *E. coli* (Pines et al., 2015).

The Multiplex Automated Genome Engineering system (MAGE) was employed to create rapid continuous diversification and mutagenesis of specific sites within the bacterial genome, generating variants with combinatorial libraries *in vivo* (Wang et al., 2009). Because MAGE can introduce short oligonucleotides to create a library of the desired genomes only in *E. coli*, Co-Selection MAGE (CoS-MAGE) was modified to include targeting regulatory elements of the desired targeted gene (Wang et al., 2012). Similarly, the Directed Evolution with Random Genomic Mutations technique (DIvERGE) allows for the evolution of multiple loci in their native genomic context within a targeted coding gene and its promoter regions (Nyerges et al., 2018).

Exploring combinatorial sequence space experimentally can be resource-intensive. To overcome this issue, machine learning models were initially trained on tested variants to swiftly generate computational libraries. For example, a computational model was trained to predict the best combinatorial library for the human GB1 binding protein based on a sequence-function dataset, followed by using the predicted library in the evolution round for validation. As a proof of concept, an enzyme for stereodivergent carbon-silicon bond formation, a new-to-nature chemical transformation, was evolved to produce a pair of enantiomers. This method accurately predicted libraries with enriched enzyme activity, leading to the identification of variants with selective catalytic properties (Wu et al., 2019). A basic machine learning protocol was extensively reviewed and was supported by case studies in which models can guide protein engineering and enhance SDE in various species (Yang et al., 2019).

CRISPR/Cas-based techniques have revolutionized the field of SDE. Designed with precision, the CRISPR/Cas9 system not only can induce both random and localized diversification but also can serve as an effective guiding mechanism for other diversification agents like base editors and EvolvR (Doudna and Jiang, 2017). A notable example is the successful editing of gram-negative bacteria *Pseudomonas aeruginosa* using the methods pnCasPA-BEC and pCasPA/pACRISPR. In pnCasPA-BEC system, a cytidine deaminase APOBEC1 is fused with Cas9-nickase that enables highly efficient gene inactivation and point mutations in a variety of *Pseudomonas* species. The pCasPA/pACRISPR system harnessed the CRISPR/Cas9 and the phage λ -Red recombination systems to accelerate a wide variety of investigations, such as bacterial

physiology study, drug target exploration, and metabolic engineering (Chen et al., 2018). The EvolvR technology uses nCas9 (D10A) along with a bacterial error-prone DNA polymerase (PolI3M) and a thioredoxin-binding domain (TBD), in addition to a synthetically designed sgRNA, allowing ongoing localized sequence diversification within bacterial targets (Halperin et al., 2018). However, the editing window was limited to -13 to -18 bases upstream of the protospacer adjacent motif (PAM) sequence.

Producing a diversified gene sequence is only the first step; for SDE, the phenotypes resulting from each modification must undergo selection to identify variants that produce phenotypes of interest. Phage-assisted continuous evolution (PACE) is an M13 phage-based technique for the continuous directed evolution of proteins. PACE revealed various advantageous applications in diversifying peptides, including enzymes, antibodies, and receptors, through iterative phage propagation and selection cycles. PACE has the capacity to evolve any protein associated with the expression of an essential phage gene (Popa et al., 2020; Miller et al., 2020). A recent comparison between Phage-Assisted Continuous Evolution (PACE) and Phage-Assisted Non-Continuous Evolution (PANCE) revealed various advantages and applications in diversifying the selection cycles. PACE induces random mutations within the DNA sequence in a bioreactor, generating new variants at a much higher rate than natural evolution. The expressed proteins are then selected based on their fitness *in situ*. In contrast, the PANCE system allows for multiplexing of phage-based evolution and can enhance the evolution of desired targets with initially low starting activity. Hence, PANCE is slower and is more time-intensive compared to the PACE system (Popa et al., 2020; Miller et al., 2020).

To illustrate the diverse, complex approaches used for SDE in eukaryotes, we have described some selected examples in the following section.

2.2 Examples of synthetic directed evolution in eukaryotes

Among eukaryotes, a major success of SDE has been reported in yeast cells. The yeast OrthoRep system utilizes a highly error-prone orthogonal DNA polymerase (TP-DNAP1) for *in vivo* diversification of desired DNA segments independent of the remaining host genome sequence. These genes are located on orthogonal, non-genomic plasmids that replicate using error-prone DNA polymerases (Ravikumar et al., 2018). As a proof of principle, the OrthoRep system evolved primary metabolic enzymes, including thiazole synthase (THI4), aiming to modify plant enzymes for plant applications, and adapt enzymes from prokaryotes to function effectively in mild, plant-like conditions (Garcia-Garcia et al., 2022). A novel orthogonal replication system, TP-DNAP1^{Y427A}, was developed as a platform for *in vivo* continuous evolution in *Saccharomyces cerevisiae*. This system allows for independent manipulation of replicative properties compared to the host organism. The mutation rate per base of TP-DNAP1^{Y427A} is approximately 400-fold higher compared to the host genome (Ravikumar et al., 2014).

TRIDENT (TaRgeted *In vivo* Diversification ENabled by T7 RNAP) is another platform for continuous diversification of target genes at mutation rates one-million fold higher than natural genomic error rates. The TRIDENT system is designed to enable *in vivo* diversification by integrating an error-prone base editor with a DNA-dependent RNA polymerase derived from a bacteriophage (T7RNA polymerase) (Cravens et al., 2021). This fusion is under the control of a T7 phage promoter that is specifically inserted upstream of the desired target gene. The TRIDENT system introduces substitutions along the template strand during transcription, without creating double-strand breaks (DSBs). In this system, T7-RNA polymerase is fused to pmCDA1-CBE along with repair factors, synthetic inducible promoters, and a distributed UNG strain (*ungΔ* strain). Uracil DNA Glycosylase (UNG) encodes the primary enzyme responsible for DNA repair of uracil and disruption of the UNG1 gene improves C/G mutagenesis (Kurt et al., 2021). Mutational diversity is tunable within this system due to DNA repair factors that localize to the sites of deaminase targeting in the yeast cell. Using the TRIDENT system in yeast produced A:T and C:G mutations at similar rates across multiple kb of DNA fragments (Cravens et al., 2021).

SDE has also been developed and successfully employed in mammalian cell lines. The T7 RNA polymerase-driven Continuous Editing (TRACE) system has been harnessed for continuous mutagenesis applications. TRACE uses a T7 polymerase fusion with two naturally occurring cytosine deaminases, catalytic polypeptide (APOBEC1) and a hyperactive mutant of activation-induced cytidine deaminase (AID*Δ), which are independently used for continuous mutagenesis. The base editing window downstream of the T7 promoter reached 2 kb of the targeted DNA, and continuous nucleotide diversification throughout 20 cell generations was seen (Chen et al., 2020).

An example of a CRISPR-based screening tool is the Targeted AID-mediated Mutagenesis (TAM) system, which provides a genetic tool to generate a broad spectrum of diverse variants suitable for gain-of-function screening. Unlike other large-scale screening methods, which mainly rely on loss-of-gene function or changes in gene expression, the TAM tool uses a fusion of AID-BE with catalytically inactive dCas9 for high-throughput screening of functional variants. This led to the identification of known and novel variants conferring imatinib resistance in chronic myeloid leukemia cells (Ma et al., 2016). Unlike TAM where AID is fused with dCas9, a modified approach CRISPR-X was used, where a hyperactive AID fused to MS2 binding protein recruited through sgRNA containing MS2 hairpins. The dCas9 variant complexes with the sgRNA and guides the sgRNA-MS2-AID complex to the desired genomic location (Hess et al., 2016). The CRISPR-X extended editing window responding to the PAM sequence compared to TAM. Repurposing the somatic hypermutation machinery, CRISPR-X facilitates *in situ* protein engineering, generating diverse libraries at localized points through multiplexing or multiple rounds of sgRNA library induction. It can target multiple genomic locations simultaneously, with minimal off-target effects. By utilizing the CRISPR-X system, researchers successfully evolved the cancer therapeutic bortezomib by focusing

on *PSMB5*, identifying both known and novel mutations conferring bortezomib resistance (Hess et al., 2016).

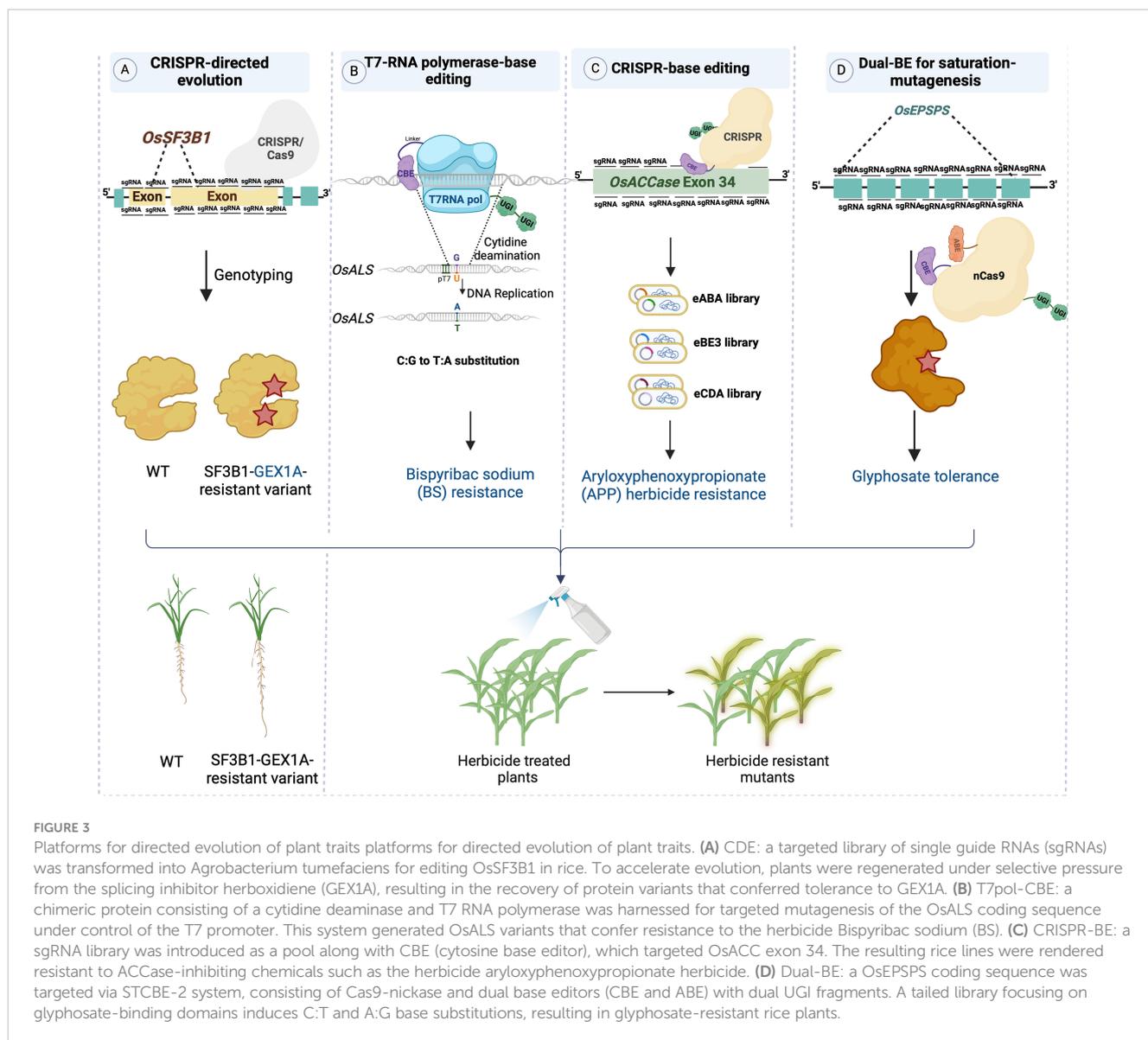
The recently developed CRISPRres (CRISPR-induced resistance) system facilitates swift diversification and identification of drug resistance mutations in vital genes. CRISPRres creates genetic variation via CRISPR-Cas-induced non-homologous end-joining (NHEJ) repair to generate a wide variety of functional in-frame mutations. One of the applications of CRISPRres is to target the anticancer agent *KPT-9274*, pinpointing nicotinamide phosphoribosyltransferase (*NAMPT*) as its primary target. Using tiling libraries of sgRNAs directed at multiple genes that conferred resistance to the anticancer drug bortezomib, CRISPRres produced multiple in-frame mutations within the *PSMB5* coding sequence, a key focal point in resistant-treated cells (Neggers et al., 2018).

Over the last decades, viruses have played pivotal roles in gene therapy and vaccine development within the emerging field of synthetic virology. Enhanced diversification by the naturally error-prone Sindbis viral RNA-dependent RNA polymerase, known as viral evolution of genetically actuating sequences (VEGAS), enables straightforward directed evolution in mammalian cells. VEGAS has been successfully applied to evolve TFs (English et al., 2019). VEGAS evolves within the signaling framework of the host cell, wholly dependent on the host cell, and selection is constant and highly mutagenic, enabling it to overcome many of the pitfalls inherent to complex fitness landscapes.

2.3 Synthetic directed evolution in plants: herbicide tolerance

Because of the ease of selection, many proof-of-concept studies in plants have used SDE to induce herbicide tolerance. Moreover, most of the plant genes tested were evolved in bacteria and later introduced into plant cells (Hendel and Shoulders, 2021; Rodrigues et al., 2021). The sequences evolved in bacterial systems were often unstable in plant cells due to the complex nature of internal pathways, codon usage patterns, RNA instability, as well as the unique anatomy and physiology of plants that ultimately hindered the broader adoption of the SDE tools to plant models. Devising technologies for high-efficiency gene sequence variation within plant cells coupled with selection pressure would enable the synthetic-directed evolution for trait discovery and engineering in plants. Developing, testing, optimizing, and establishing technologies for high-efficiency localized sequence variation coupled with selection pressure are indispensable to evolve gene variants for traits of value.

SDE was initially established by targeting a splicing factor called *SF3B1* in rice using CRISPR/Cas9 system (Butt et al., 2019; Zhang and Qi, 2019) (Figure 3A). *SF3B1* is a subunit of the SF3B spliceosome within the U2 small nuclear ribonucleoprotein (U2SnRNP) complex. A library of sgRNAs was employed to target *SF3B1* in rice. The edited sequences were generated randomly through NHEJ repair pathway. This CDE (CRISPR-directed Evolution) platform used to produce SF3B1-mutant



variants showing varying levels of resistance to the splicing inhibitor herbicide (*GEX1A*) (Table 1) (Butt et al., 2019).

To establish continuous evolution of target genes in plant cells, TRACE has been employed in plant systems (Figure 3B). Remarkably high C:T editing efficiency was achieved in transient assays conducted in *Nicotiana benthamiana* (Butt et al., 2022). Acetolactate synthase (ALS), which is responsible for initiating the biosynthesis of the branched-chain amino acids such as valine, leucine, and isoleucine, serves as the primary target for at least five structurally distinct classes of herbicides (Shimizu et al., 2011). Point mutations were produced via targeting *OsALS* in rice by utilizing a chimeric fusion of T7-RNA polymerase and CBE guided by the specificity of the T7 promoter. By employing a T7 promoter-driven targeting the *OsALS* sequence, which was stably integrated into the rice genome, C:T and G:A transitions were generated. Subsequent application of bispyribac sodium (BS) as the selection

pressure allowed herbicide-responsive residues within the *OsALS* sequence to be identified (Butt et al., 2022).

ALS has also been targeted with many SDE tools such as the Base Editing-Mediated Gene Evolution tool (BEMGE), which was constructed by combining nCas9, cytosine-adenine base editors, and an sgRNA library covering the full length of the *OsALS* coding region. Using this tool, base editors were guided by an sgRNA pool, resulting in a higher editing rate than when using samples transfected with a single gRNA. Additionally, four types of amino acid substitutions (R190H, P171F, P171L, and P171S) were obtained upon targeting, producing varying levels of BS herbicide resistance in plants (Kuang et al., 2020). As highlighted in another study, the Target-AID tool, which combines nCas9 with AID-BE, allows for precise base substitutions (specifically G:A or C:T) instead of random mutagenesis. The system was guided by sgRNAs directed toward the NGG PAM sequences located in the

TABLE 1 Herbicide-resistant point mutations engineered in Plants.

Herbicide gene	Type of mutation	Bioengineering Tool	Herbicide	Species	Reference
<i>SPLICING FACTOR 3B SUBUNIT 1 (OsSF3B1)</i>	K1049R+K1050E +G1051H/ H1048Q+A1064S	CRISPR- directed evolution	GEX1A/Pladienolide B	Rice	(Butt et al., 2019, 2021)
<i>PHD finger protein 5A (OsPHF5A)</i>	Y36C	Overexpression			
Acetolactate synthase (ALS)	A96V	Target-AID	Imazamox	Rice calli	(Shimatani et al., 2018)
	G95A	<i>In vitro</i> ALS activity assay and overexpression in rice using ubiquitin promoter	pyrimidinyl carboxy (PC): Bispyribac sodium (BS)	<i>Oryza sativa</i>	(Okuzaki et al., 2007)
	A121T P187S S652N	Site-directed mutagenesis	Londax (sulfonylurea) Cadre (imidazolinone)	<i>Nicotiana benthamiana</i>	(Chong and Choi, 2000)
	A122T A122 T/V A122V	Random mutagenesis Transgene expression	Sulfonylureas (SU): Chlorsulfuron Imazamox, bispyribac-sodium, and penoxsulam Pyrimidinylcarboxy(PCs)	<i>Raphanus raphanistrum</i> (Wild radish) Barnyardgrass (<i>Echinochloa crusgalli</i>) <i>Arabidopsis</i>	(Han et al., 2012) (Riar et al., 2013) (Shimizu et al., 2011)
	P171F	Base-editing mediated gene evolution (BEMGE)	Bispyribac sodium (BS)	<i>Oryza sativa</i>	(Kuang et al., 2020)
	P165 A/S/L/W	Base editing	Sulfonylureas (SU): Chlorsulfuron	<i>Maize</i>	(Li et al., 2020c)
	P171 S/A/Y/F G628E G629S Triple mutant (P171F +G628E+G629S)	Base editing (CBE)	Bispyribac sodium (BS) For Triple mutant: nicosulfuron, imazapic, pyroxsulam, BS, and flucarbazone	<i>Oryza sativa</i>	(Zhang et al., 2021)
	W548L/S627I (Double mutation) W548L/and S627I (Two point mutations)	Overexpression of dual mutated gene Site-specific modification	Bispyribac sodium (BS) Bispyribac sodium (BS)	<i>Oryza sativa</i> L. cv. <i>Nipponbare</i> <i>Oryza sativa</i> L. cv. Kinmaze	(Kawai et al., 2007) (Endo et al., 2007)
	P197-(H/T/A/L/S/A/I/M/L/W) P197 S/A D376E	Target-site mutation PCR-RFLP assay QuikChange site-directed mutagenesis system	Sulfonylureas (SU) Imidazolinones (IMI) Imidazolinones (IMI) Sulfonylureas (SU): (Tribenuron) Sulfonylureas (SU)	<i>Kochia</i> (<i>Kochia scoparia</i>) <i>Lolium rigidum</i> <i>Papaver rhoeas</i> L. (corn poppy) Weed <i>Monochoria vaginalis</i> (<i>syn. Pontederia vaginalis</i>)	(Warwick et al., 2008) (Anthimidoua et al., 2020) (Délyea et al., 2010) (Tanigaki et al., 2021)
	W574L	Target-Site Mutation	Imazamox	<i>Euphorbia heterophylla</i> L (Wild Poinsettia)	(Mendes et al., 2020)
	S653I	Transgene expression	Bispyribac sodium (BS) Pyrimidinylcarboxylates (PCs)	<i>Oryza sativa</i> L. subsp. <i>Japonica</i> <i>Arabidopsis</i>	(Shimizu et al., 2011)
	G654 E/D	Random mutagenesis	Imidazolinones (IMI)	<i>Setaria viridis</i>	(Laplante et al., 2009)
	S186P K416G L662P	Hybridization	Imazethapyr	Strawhull (SH) red rice	(Rajguru, 2005)
Acetyl-CoA carboxylase (ACCCase)	W2125L W2125Q W2125 S/R I2139N C2186H G2194S	Base editing-mediated targeted evolution	Haloxypfop-R- methyl (Gallant)	<i>Japonica</i> rice (<i>Nipponbare</i>)	(Wang et al., 2022)

(Continued)

TABLE 1 Continued

Herbicide gene	Type of mutation	Bioengineering Tool	Herbicide	Species	Reference
	G2194A C2186H				
	W2125S C2186R I1879V	CRISPR-mediated base editing tools (eABE, eBE3, and eCDA)	Haloxifop-R-methyl	<i>Oryza sativa</i>	(Liu et al., 2020)
	P1927F W2125C S1866F A1844P	STEME-1 STEME-NG	Haloxifop	Rice protoplasts	(Li et al., 2020a)
	D2078G I1781L W2027C I2041N D2078G C2088R	<i>In Vitro</i> inhibition of ACCase activity	Clethodim APP and CHD	<i>Lolium rigidum</i>	(Yu et al., 2007)
5-enolpyruvylshikimate-3-phosphate synthase (EPSPS)	D213N (glyphosate binding domain)	STCBE-2	Glyphosate	<i>Oryza sativa</i>	(Zhang et al., 2023)
	T102I P106S	CRISPR/Cas9 (Gene replacements and insertions) Should I include this here?	Glyphosate	Rice protoplast	(Li et al., 2016)
	T169I A170V P173S	Prime editing	Glyphosate	<i>Japonica rice</i>	(Li et al., 2020b)
	Triple substitution (T102I + A103V + P106S)	Screening naturally evolved variants via high-resolution melting analysis (HRMA)	Glyphosate	<i>Amaranthus hybridus</i>	(Perotti et al., 2019)
	G172A T173I P177S	Overexpression of resistant variants	Glyphosate	<i>Japonica rice (Nipponbare)</i>	(Achary et al., 2020)
Tubulin (<i>TubA2</i>)	M268T R243 M/K	rBE14 <i>Expression of mutant α-tubulin gene</i>	Dinitroanilines (Pendimethalin, trifluralin) Dinitroanilines (Trifluralin)	<i>Oryza sativa</i> <i>Lolium rigidum</i> (annual ryegrass)	(Liu et al., 2021) (Chu et al., 2018)
<i>OsPDS</i>	R304S R300S	Homologous recombination (HR)-mediated gene targeting (GT)/	Fluridone or Norflurazon	<i>Oryza sativa</i>	(Endo et al., 2021)
Glutamine synthetase (GS)	R295K R295K	DNA shuffling Transgenic overexpression	Glufosinate Glufosinate	<i>Oryza sativa</i> <i>Arabidopsis</i>	(Tian et al., 2015)
	A63V+P64S S59G+T60A	BEMGE	Gufosinate (PPT)	<i>Oryza sativa</i>	(Ren et al., 2023)
	-	Overexpression of OsGS1 or OsGS2 Co-overexpression of <i>OsGS1;1/OsGS2</i>	Gufosinate (PPT) limited tolerance to Gufosinate (PPT)	<i>Oryza sativa</i>	(James et al., 2018)
	S59G	Expression of naturally evolved variant	Gufosinate	Malaysian <i>E. indica</i>	(Zhang et al., 2022)

OsALS coding sequence, resulting in multiple mutations associated with herbicide resistance to imazamox in rice (Table 1) (Kuang et al., 2020).

Based on the concept of engineering allelic diversity, dCas9-BE can target the gene encoding acetyl-coenzyme A carboxylase, which catalyzes the first step of fatty acid biosynthesis in plants (Figure 3C). *OsACC*, which is the target of a major group of commercial herbicides, comprises 35 exons and encodes a protein consisting of 2327 amino acid residues. Herbicide-resistant-variants of this gene are located within the carboxyltransferase (CT) domain of *OsACC* (Jang et al., 2013). To comprehensively investigate *OsACC*, a pool of 141 sgRNAs was designed and introduced into rice callus, targeting the 1653-bp coding sequence of the CT domain situated in the 34th exon of *OsACC*. This was achieved using CRISPR-mediated base editing tools (eABE, eBE3, and eCDA), facilitating the identification of additional homozygous mutants that confer herbicide resistance in rice (Table 1) (Liu et al., 2020).

Another example of targeting herbicide tolerance genes *in planta* is the utilization of the STCBE-2 system, which facilitates near-saturated mutagenesis of *EPSPS*, which encodes 5-enolpyruvylshikimate-3-phosphate synthase, the target of glyphosate (Figure 3D) (Zhang et al., 2023). The STCBE-2 system consists of a dual cytosine and adenine base editor fused to nCas9-NG. Through multiple rounds of base editing and selections, researchers successfully obtained a novel allele of *OsEPSPS* that conveyed glyphosate tolerance to rice plants. This allele exhibited specific nucleotide changes that modified the coding sequence of *OsEPSPS*, particularly at position 213, where Asp was replaced by Asn (D213N). This alteration made the protein less susceptible to the inhibitory effects of glyphosate. The *OsEPSPS-D213N* modified allele was located in the predicted glyphosate-binding domain and provided robust glyphosate tolerance in rice (Table 1) (Zhang et al., 2023).

Other recent examples showcasing the efficacy of directed evolution involve the utilization of multiple versions of the CRISPR-based editing tool, such as SpCas9-NGv1, nSpCas9-NGv1 fused with APOBEC1, nSpCas9-NGv1-APOBEC-UGI, nSpCas9-NGv1-AID and nSpCas9-NGv1-AID-UGI. The SpCas9-NGv1-AID tool enabled the introduction of C:T substitutions into the *OsNGN*, *ALS*, *EPSPS* and the drooping leaf (*DL*) genes (Endo et al., 2019). The fusion construct efficiently induced mutations with variable efficiencies within the rice and *Arabidopsis* genomes (Endo et al., 2019).

3 Noncoding DNA and *cis*-regulatory elements

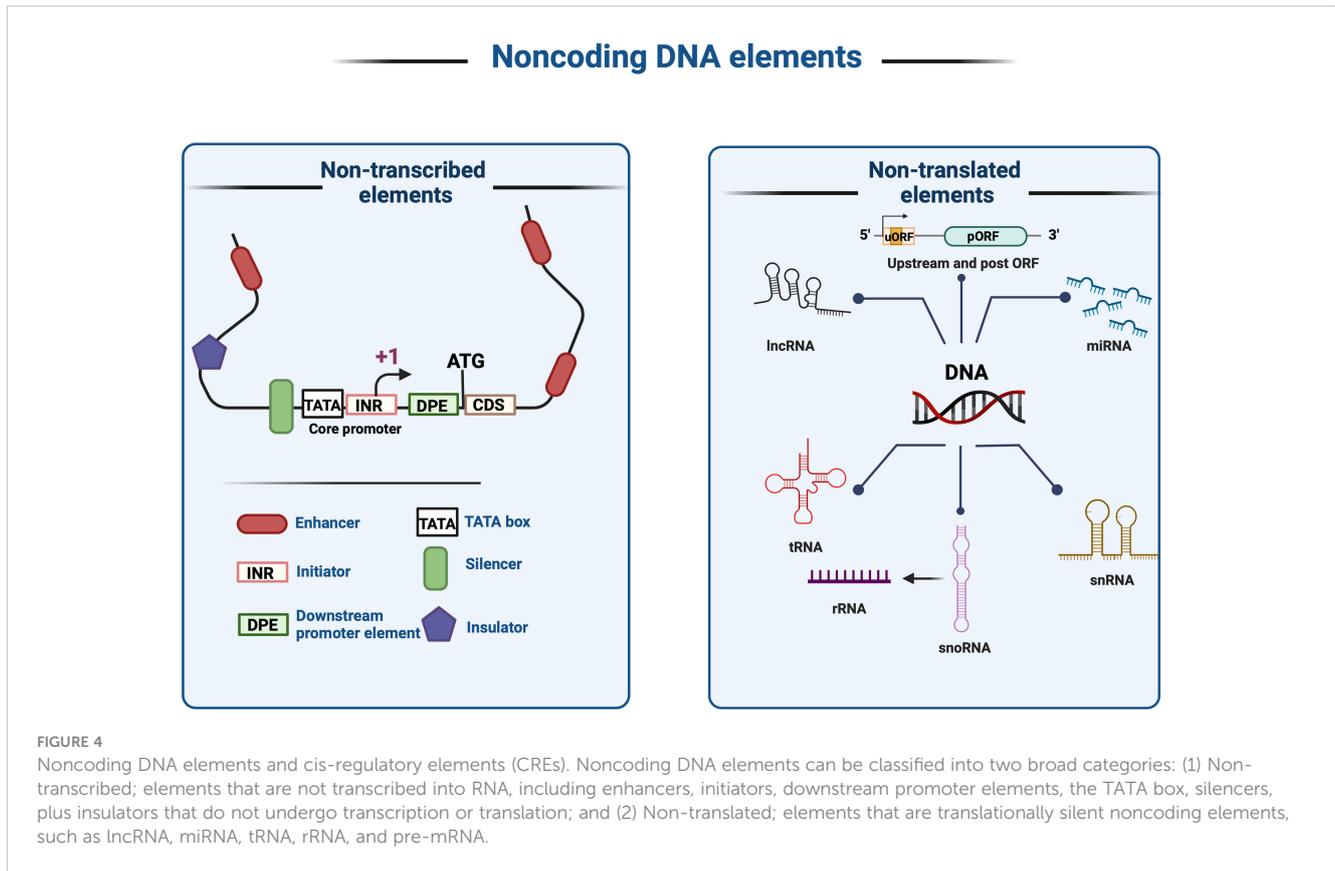
Although most approaches to date have targeted coding sequences for SDE of proteins, emerging technologies are beginning to examine the mechanisms regulating gene expression. Indeed, the diversity observed in crop characteristics arises from alteration not only in protein-coding genes but also in non-coding DNA elements (ncDNA). Modification in coding genes have the potential to disrupt protein structure and function, whereas alterations in non-coding DNA, specifically in *cis*-regulatory regions, regulate gene expression levels, thereby influencing a wide range of plant traits.

Non-coding DNA elements are segments of DNA which does not encode proteins. ncDNA include two types of sequences: 1) non-transcribed, that remain un-transcribed but serve regulatory roles such as core promoters, enhancers, insulators, silencers, and response elements, which also known as *cis*-regulatory elements (CRE); or 2) non-translated, that transcribed into RNA, such as ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), microRNAs (miRNAs), and long non-coding RNAs (lncRNAs) but not translated into proteins (Figure 4). CRE plays pivotal roles in regulating gene expression, modulating growth patterns, controlling spatial and temporal expression patterns, and regulating plant growth and development (Chekanova, 2015; Swinnen et al., 2016). In eukaryotic genomes, only 1% of the total DNA is responsible for encoding proteins (coding sequences), while the remaining 99% constitutes of non-coding DNA sequences. CREs are distinct DNA sequences recognized, for example, by transcription factors (TFs), which modulate transcription patterns. TFs bind to CREs, thereby activating or repressing the transcription of nearby genes. Regulatory elements ensure precise gene activation while preventing unwanted transcription. Coordinated actions of DNA-binding proteins and CRE interactions govern crucial cellular processes such as development and spatial patterning in planta (Jiang, 2015; Marand et al., 2023). Cells are tightly regulated by CREs to maintain proper gene expression and translation patterns. However, deciphering *cis*-regulatory roles and locations is challenging due to the vast diversity of regulatory sequences, epigenetic modifications, and environmental regulations (Schmitz et al., 2022; Bilas et al., 2016; Wittkopp and Kalay, 2011).

3.1 Computational tools and databases for predicting *cis*-regulatory elements in plants

To use SDE on CREs, researchers must first identify the CREs that regulate expression of the gene of interest in the relevant conditions and tissues. In recent years, there has been a notable focus on generating public databases and developing novel computational tools to study *cis*-regulatory roles and motifs *in planta*. These resources continually evolve, relying on systematic annotation with experimentally validated data to enable precise prediction of transcriptional networks. Integrating these predictions into plant applications is crucial both for conducting biologically significant analyses of regulatory sequence variation and for generating novel variants linked to desired phenotypes. This integration resolves issues such as annotation disparities, duplications, and standardized nomenclature, thereby enhancing the utility of these resources (Figure 5).

One of the earliest databases in plant research is PLACE (plant *cis*-acting regulatory DNA elements, which provides an accessible dataset of motifs within *cis*-acting regulatory DNA elements and their variants in vascular plants. These motifs are sourced from previous studies and are accompanied by concise descriptions. Additionally, variations of these motifs in different genes or plant species are documented in the PLACE database (Higo et al., 1998). Various studies have highlighted databases and prediction tools for predicting roles and locations of



CREs. For example, the Database of Rice Transcription Factors (DRTF) is a valuable resource for studying TFs in *Oryza sativa*, providing comprehensive information on putative TFs, chromosomal localization, and sequence alignments. The DRTF dataset is supported by manual selection and computational predictions (Gao et al., 2006). AGRIS TRANSFAC is a database resource that provides valuable data on *Arabidopsis* promoter sequences, transcription factors, and their binding sites. It consists of two minor databases: AtTFDB, which focuses on *Arabidopsis* transcription factors, and AtcisDB, which is dedicated to *Arabidopsis* cis-regulatory elements, making the database more generically useful (Davuluri et al., 2003).

PlantCARE is a generic database identifying plant cis-acting regulatory elements, enhancers, and repressors, offering links to various databases such as EMBL, TRANSFAC, and MEDLINE (Lescot et al., 2002). An additional database generated for *Arabidopsis* is the Database of Arabidopsis Transcription Factors (DATF), which is specifically tailored for *Arabidopsis* transcription factors and encompasses 1827 genes spanning 56 families. Notably, DATF offers distinctive information and other features including 3D structure templates, expression information derived from ESTs, transcription factor binding sites, and nuclear localization signals (Guo et al., 2005). The Database of Poplar Transcription Factors (DPTF) is a collection of transcription factors compiled through a combination of computational predictions and manual selection in rice and *Arabidopsis* (Zhu et al., 2007). The Plant Transcription Factor Database (Plant-TFDB) provides information for TFs from various plant species, integrating high-quality non-redundant TF binding motifs and diverse regulatory elements and their interactions (Jin

et al., 2017). Recently, Plant-TFDB, which has been developed to provide details concerning transcriptional regulation, expands its coverage across a diverse array of species for evolutionary annotation purposes, and highlights binding motifs along with TF functionality upon targeting (Tian et al., 2020). In the same study, they extended the development to generate the PlantRegMap tool, which focuses on functional regulatory mapping in plants. It assists in identification of transcriptional regulatory networks by predicting interactions between TFs and the target gene and by exploring TF binding sites as upstream regulators for the input gene (Tian et al., 2020). The recent PlantPAN 4.0 update is tailored to detect and analyze conserved non-coding sequences (CNSs) in plant promoters, providing valuable insights into the regulatory mechanisms underlying gene expression in a variety of plant species. Additionally, PlantPAN 4.0 can identify CNSs among related genes and explore different combinations and nucleotide variations of cis-regulatory elements, making it an essential tool for studying plant regulatory landscapes (Chow et al., 2024). As a proof of concept, seven bioinformatic models were employed to investigate cis-regulatory elements in soybean, *Arabidopsis*, maize, and rice (Zemlyanskaya et al., 2021; Ferebee and Buckler, 2023; Liu and Yan, 2019).

iCREPCP is a tool that employs deep learning techniques to identify CREs within plant core promoters, especially in maize and tobacco. By leveraging convolutional neural networks, iCREPCP accurately detects critical CREs that significantly contribute to promoter strength. This tool not only provides precise predictions of promoter strength but also identifies the position of each CRE with base-level resolution (Deng et al., 2023). A recent deep learning

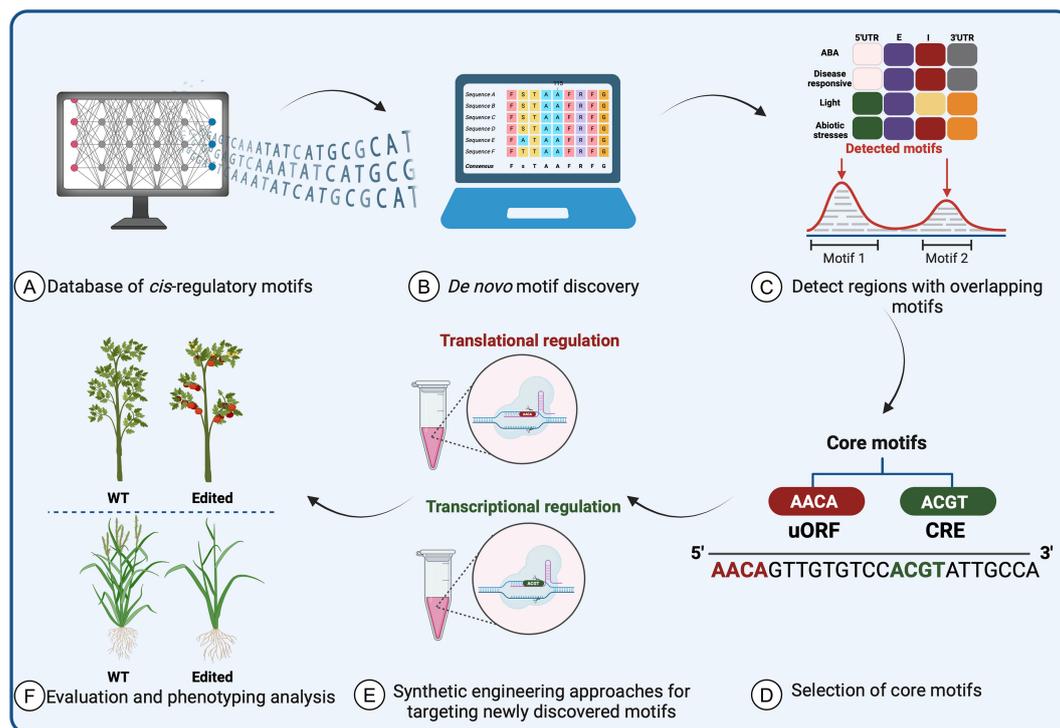


FIGURE 5

Prediction of *cis*-regulatory element motifs using computational approaches and databases in plants. (A) The required data are collected from validated databases and (B) converted into a matrix and structuring tool to form a baseline prediction tool. Following the discovery of motifs and *cis*-elements, (C) a search for homologs and overlapping motifs across species is performed. (D) After core motifs are identified and used in synthetic protein engineering methods to improve crops, (E) experimental validation and characterization are necessary to assess the accuracy of the data-driven machine-learning tools. (F) The data prediction tools will identify the most promising candidates for further testing as potential targets to improve crop traits.

model, the Deep Learning Important Features (DeepLIFT) algorithm, was employed for accurately predicting regulatory effects based on genetic variation. DeepLIFT connects gene sequence data with mRNA copy numbers in various plant species, facilitating the identification of specific sequence features associated with gene expression (Peleke et al., 2024) (Table 2).

In contrast, there are few computational methods to predict the presence of enhancer sequences in plants. These databases, prediction tools, and computational models play crucial roles in advancing our understanding of plant transcriptional regulation and regulatory sequences roles, offering valuable insights into the intricate regulatory networks controlling gene expression and deepening our comprehension of plant evolution. Enhancers are crucial *cis*-regulatory elements capable of exerting control over genes from a distance, whether upstream, downstream, embedded within intronic sequences, or in close proximity to the coding sequence. As a proof of concept, predicted enhancers in intergenic regions were experimentally validated using reporter assays (Zhu et al., 2015) (Table 3). The self-transcribing active regulatory region sequencing (STARR-seq) method was employed to map functional enhancers and measure the enhancer activity of candidate sequences in rice. Approximately 9642 potential enhancers were predicted and identified through protoplast transfection of a genomic library derived from rice (Sun et al., 2019). For trait discovery, the STARR-seq analysis of the ATAC-seq (assay for transposase-accessible chromatin sequencing) library in maize unveiled

enhancer activity in distal promoters linked with accessible chromatin regions and TF locations (Ricci et al., 2019). This signature-based enhancer prediction system utilized relative DNase-seq read enrichment in leaves compared to flowers to predict the tissue specificity of putative enhancers in *Arabidopsis*. For more precise and reproducible detection of enhancer activity, a variant of STARR-seq technology used for efficient identification of enhancers in transiently transformed *Nicotiana benthamiana* leaves. STARR-seq combined with site-saturation mutagenesis (SSM) to pinpoint functional regions within an enhancer, which recombined to create synthetic enhancers (Jores et al., 2020).

3.2 Targeting *cis*-regulatory elements for crop improvement

Although CREs harbor a wealth of untapped potential variation for crop improvement, this aspect has remained largely unexplored until recently. Inducing precise mutations or producing a library of variants of these elements and motifs can lead to altered gene expression and induce phenotypic diversity within crop species. Understanding the role of CREs is crucial for unraveling the molecular mechanisms underlying their expression and function. These elements provide insights into the complex regulatory networks involved in producing valuable crop traits.

TABLE 2 Summary of computational approaches and databases for identifying plant transcription factors, core promoters, and UTRs.

Computational Tool/Database	Model	<i>cis</i> -regulatory elements	Links	References
Database of plant <i>cis</i> -acting regulatory DNA elements (PLACE)	Vascular plants	<i>cis</i> -regulatory regions, motifs, and homologs involved in plant regulation	http://www.dna.affrc.go.jp/htdocs/PLACE/	(Higo et al., 1998)
Database of Rice Transcription Factor (DRTF)	<i>Oryza sativa</i>	Provides comprehensive information on putative TFs, chromosomal localization, and sequence alignments of rice TFs	http://drtf.cbi.pku.edu.cn	(Gao et al., 2006)
Arabidopsis Gene Regulatory Information Server (AGRIS)	<i>Arabidopsis thaliana</i>	<i>cis</i> -regulatory elements, TFs, and putative binding sites	http://arabidopsis.med.ohio-state.edu	(Davuluri et al., 2003)
PlantCARE	Monocots, dicots and higher plants	Database of plant <i>cis</i> -acting regulatory elements, enhancers, and repressors within plant core promoters	http://sphinx.rug.ac.be:8080/PlantCARE/	(Lescot et al., 2002)
Database of Arabidopsis Transcription Factors (DATF)	<i>Arabidopsis thaliana</i>	TF binding sites, nuclear location signals, and 3D structure templates of ArTFs	http://datf.cbi.pku.edu.cn	(Guo et al., 2005)
The database of poplar transcription factors (DPTF)	<i>Populus trichocarpa</i> , <i>Oryza sativa</i> and <i>Arabidopsis thaliana</i>	TF prediction	http://dptf.cbi.pku.edu.cn	(Zhu et al., 2007)
The Plant Transcription Factor Database and prediction tool (Plant TFDB)	<i>Oryza rufipogon</i> and <i>O. nivara</i> for <i>O. sativa</i> and green plants	Identify and analyze plant TFs, regulatory elements, and their interactions	http://planttfdb.cbi.pku.edu.cn/	(Jin et al., 2017)
Plant Regulatory data and analysis platform (PlantRegMap)	Green plants	Charting plant transcriptional regulatory maps, prediction of regulatory interactions between TFs and input gene, and exploring of TF binding sites upstream regulators for the input gene	http://plantregmap.cbi.pku.edu.cn/	(Tian et al., 2020)
PlantPAN 4.0	<i>Arabidopsis thaliana</i> , <i>Oryza sativa</i> , and <i>Zea mays</i>	A tool for constructing transcriptional regulatory networks for diverse plant species	http://PlantPAN.itps.ncku.edu.tw/	(Chow et al., 2024)
iCREPCP	<i>Zea mays</i> and tobacco	A deep learning-based web server for single base-resolution <i>cis</i> -regulatory elements within plant core promoters	https://github.com/kaixuanDeng95/iCREPCP	(Deng et al., 2023)
Deep Learning Important Features (DeepLIFT)	<i>Arabidopsis thaliana</i> , <i>Sorghum bicolor</i> , <i>Solanum lycopersicum</i> , and <i>Zea mays</i>	UTR's	https://github.com/JanZrimec/DeepExpression	(Peleke et al., 2024)

TABLE 3 Overview of Computational approaches and databases for identification of plant enhancers.

Computational Tool/Database	Model	<i>cis</i> -regulatory elements	Links	References
Global Quantitative Mapping of Enhancer Activity in Rice Genome by STARR-seq	<i>Oryza sativa L. ssp japonica</i>	Enhancers within the 5'UTR	Gene Expression Omnibus (GEO) as GEO: GSE121231	(Sun et al., 2019)
STARR-seq for genetic mapping	<i>Maize</i>	Transcriptional enhancer activity	https://github.com/schmitzlab/Widespread-Long-range-Cis-Regulatory-Elements-in-the-Maize-Genome/	(Ricci et al., 2019)
Signature-based enhancer prediction system	<i>Arabidopsis thaliana</i>	Enhancers in intergenic regions	BioProject/PRJNA252965	(Zhu et al., 2015)

By manipulating CREs, researchers can fine-tune gene expression and modify plant traits, leading to alterations in plant morphology and overall plant architecture. For example, targeting the conserved *cis*-regulatory elements in the *SIKLUH* promoter using CRISPR/Cas9 system significantly reduced the proportion of small tomatoes and enhanced the weight of all fruits along the inflorescence (Li et al., 2022). The Waxy (*Wx*) gene encodes a granule bound NDP-glucosestarch glucosyltransferase in rice. Varied activities of natural *Wx* alleles regulate different amylose contents (AC), gel consistency (GC) and pasting viscosity of grain starches thus ultimately influencing the grain appearance and cooking/eating quality. Rice grains with higher ACs and lower GC values have poor eating quality, while those with moderate ACs and higher GC values give better taste for most consumers (Zeng et al., 2020). Disrupting the region near the TATA-box in the core promoter region of the *Wx* allele led to a moderate decrease in the amylose content in rice (Huang et al., 2020; Butardo et al., 2017). Editing of the *Wx* promoter to disrupt the A-box, CAAT-box, and 5' UTR intronic splicing site (5' UISS) produced *Wx* variants with variable amylose content (Zeng et al., 2020). Editing CREs, such as the RY-element and 2S seed protein motif of the fatty acid desaturase 2 promoter (*FAD2*), in peanut seeds manipulated its expression and improved the fatty acid profile by increasing oleic acid levels (Neelakandan et al., 2022).

3.3 Engineering promoters for crop improvement

The core promoter region varies among species and genes and is comprised of multiple elements. In eukaryotes, the TATA-box is crucial for minimal promoter activity. Additionally, elements like the initiator (*Inr*) sequence, transcription factor binding sites, downstream promoter element (*DPE*), and *Y-Patch* are distributed in diverse combinations, playing essential roles in facilitating efficient transcription factor binding and supporting the regulation of gene expression.

Modifying CREs can selectively eliminate specific motifs, resulting in manipulating transcript abundance, controlling expression patterns, and generating a series of phenotypic variants (Figure 6). For example, regulating meristem maintenance and determinacy in tomato was achieved by inducing targeted modifications in the promoter of *SLCLV3* (Rodríguez-Leal et al., 2017) (Figure 6A). Moreover, CRISPR/Cas9-induced mutagenesis can introduce novel quantitative variation for diverse traits, leading to observable differences in fruit size and inflorescence branching (Rodríguez-Leal et al., 2017). In a similar way, the prediction of *cis*-regulatory elements in the *SLG7* promoter using Plant-CARE led to the generation of novel beneficial alleles for enhancing rice appearance quality. Targeting the AC II element-containing region of the *SLG7* promoter via CRISPR/Cas9 gene editing weakened the accessibility of MYB protein AH2 to the *SLG7* promoter, resulting in increased amylose content, decreased gel consistency, and improved chalkiness of milled rice (Figure 6B) (Tan et al., 2023).

Similarly, mutagenesis of the effector-binding element (EBE) within the promoter of *SUGAR WILL EVENTUALLY BE EXPORTED TRANSPORTER* (*OsSWEET14*) using CRISPR/Cas9 conferred resistance to bacterial blight (Duy et al., 2021). Recent

research used CRISPR/Cas12a mutagenesis to enhance resistance to blight bacteria. Rice leaf blight is caused by the bacterium *Xanthomonas oryzae* *pv.* *oryzae* (*Xoo*). The upregulated by transcription activator-like 1 (UPT) effector box in the promoter region of the rice *Xa13* gene plays a key role in *Xoo* pathogenicity (Yu et al., 2021). Specifically, mutating a critical bacterial protein-binding site within the UPT box of the *Xa13* promoter effectively abolishes bacterial expression, offering a promising CRISPR/Cas12a mutagenesis strategy to enhance rice resistance against bacterial pathogens (Yu et al., 2021).

The *TaVRN1-A1* promoter contains several regulatory sites namely, a VRN box, CA_RG box, and putative AG hybrid box associated with wheat vernalization. A tool termed A3A-PBE, comprising an enhanced human APOBEC3A-CBE fused to nCas9 (D10A), was developed to target *cis*-elements of the *TaVRN1-A1* promoter to produce variants with different vernalization requirements. A3A-PBE edited all cytidines in positions of the protospacer corresponding to the VRN-box site, thereby disrupting the binding site of bZIP TFs. Variants incorporating these modifications were generated in wheat, rice, and potato (Zong et al., 2018). Several studies have examined the pleiotropic rice gene *IDEAL PLANT ARCHITECTURE 1* (*IPA1*), which encodes the transcription factor *OsSPL14* and regulates various phenotypic traits such as tiller numbers, grain production, and panicle size. In this study, a tiling deletion-based CRISPR-Cas9 screen identified a 54-base pair *cis*-regulatory region in *IPA1* that is when deleted, leads to increased grain yield per plant. Further studies revealed that the 54-bp deleted fragment serves as a target site for the transcription factor An-1, which represses *IPA1* expression in panicles and roots (Song et al., 2022).

3.4 Engineering upstream open reading frames for crop improvement

Gene regulation extends beyond transcription to include posttranscriptional mechanisms that fine-tune gene expression. Elements operating at the translational level play a significant role in modulating the translation of downstream primary open reading frames (pORFs). The upstream open reading frame (uORF) is located upstream the coding gene and is not usually categorized as a non-coding DNA regulatory element. uORFs are recognized for their ability to influence translation, either positively or negatively, and sometimes initiate nonsense-mediated mRNA decay. In plants, approximately 30-40% of genes contain uORFs (von Arnim et al., 2014). In recent years, considerable attention has been given to studying the physiological roles of plant-specific uORFs, which have been extensively reported (van der Horst et al., 2020). Modifying uORFs through genetic editing tools can lead to the regulation of translation. This can be achieved either by introducing new uORFs or by extending the original uORFs through targeted modifications of their stop codons. For example, altering the uORFs in *OsDLT*, *OsTBI*, *OSBR1*, and *OsTCB19* via base editing and prime editing allowed control of gene expression and generation of a variant series exhibiting many phenotypical phenotypes in rice (Xue et al., 2023). Additionally, utilizing the CRISPR/Cas9 system for manipulating uORFs via genome editing

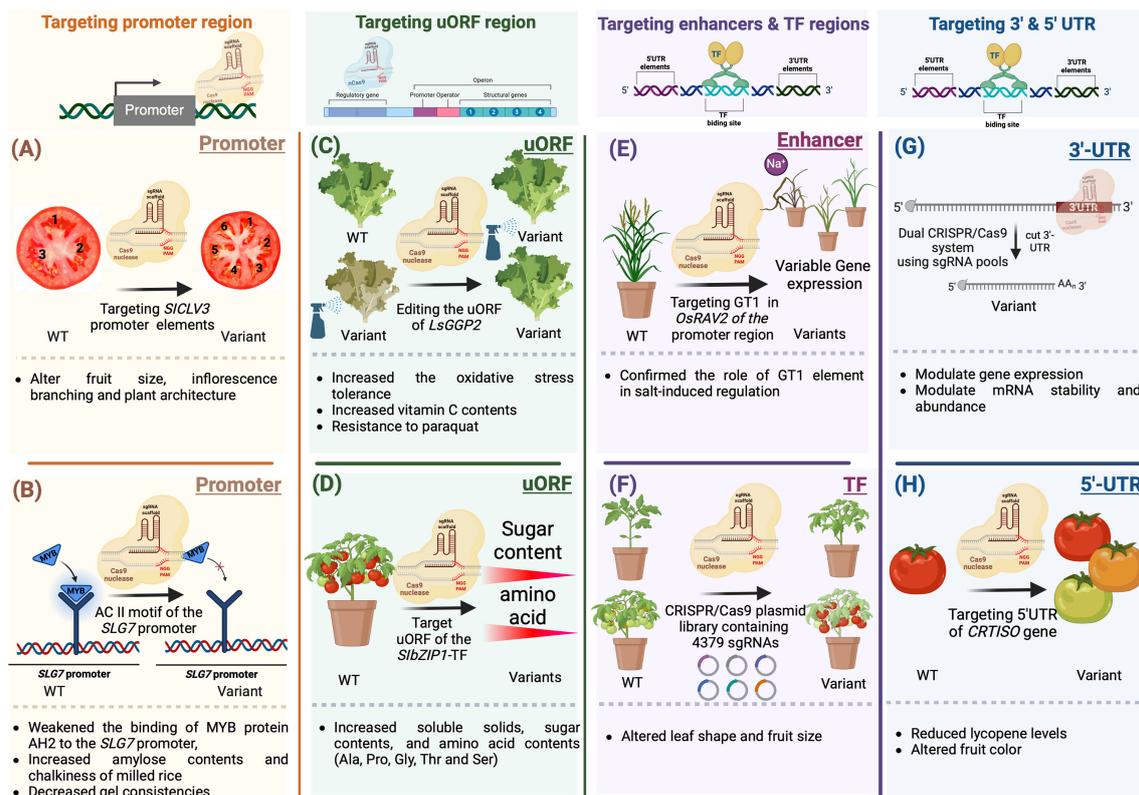


FIGURE 6

Gene editing of cis-regulatory elements and motifs for crop improvement. (A) CRISPR/Cas9-induced mutagenesis can introduce novel quantitative and qualitative variations for a variety of desirable traits, leading to observable differences in fruit characteristics (B) Prediction of cis-regulatory elements in the *SLG7* promoter using Plant-CARE led to the generation of novel beneficial alleles for enhancing rice appearance. The AC II element-containing region of the *SLG7* promoter was targeted via CRISPR/Cas9 gene editing. This weakened the accessibility of MYB protein AH2 to the *SLG7* promoter, resulting in increased amylose content, decreased gel consistency, and improved chalkiness of milled rice. (C) Induced mutations in the SC-uORF of the tomato transcription factor gene *SibZIP1* by the CRISPR/Cas9 system led to increased sugar and amino acid contents in tomato fruits. (D) A CRISPR/Cas9 system guided by dual gRNAs was employed to introduce targeted mutations in the uORF regions of *SibZIP1*, which encodes a protein responsible for sucrose-induced repression of translation in tomatoes. Targeting the SC-uORF led to significant alterations in sugar and amino acid concentrations in tomato fruits, with amino acid levels increasing by up to 132% compared to WT fruits. (E) The GT-1 element plays a crucial role in facilitating adaptive salt responses through the *OsRAV2* promoter. This was confirmed through *in situ* validation in plants with targeted mutations generated using the CRISPR/Cas9 system. This approach revealed the 65-bp region responsible for salt tolerance within the *OsRAV2* promoter. (F) CRISPR/Cas9 plasmid libraries were utilized to target transcription factors in the tomato genome. Mutant populations were obtained by employing a large-scale pooled library of *Agrobacterium* plasmids for genetic transformation. Phenotypic changes observed in field-grown plants included alterations in leaf shape or number, seedling yellowing, modifications in fruit morphology or size, and abnormal development of floral organs (G) Recombinant Cas9 and *in vitro* transcribed gRNA can be used to target 3'-UTRs. This high-throughput approach identified regions that regulate mRNA abundance and mRNA stability. (H) CRISPR/Cas9 system was utilized to modify the 5' UTR of *CRTISO*. This resulted in reduced levels of lycopene, a pigment responsible for the red color in tomatoes.

allows fine adjustments in mRNA translation, thereby influencing protein concentrations. For example, uORFGDP-1-galactose phosphorylase (*LsGGP2*) was targeted, resulting in an 80–140% increase in ascorbate (vitamin C) content in lettuce (*Lactuca sativa*) leaves, as well as enhanced tolerance to oxidative stress (Zhang et al., 2018). In a recent study, the CRISPR/Cas9 system guided by dual gRNAs was employed to introduce targeted mutations in the uORF regions of *SibZIP1*, which is responsible for sucrose-induced translation repression (SIRT) in tomatoes (Figure 6D). Targeting the SC-uORF led to significant alterations in sugar and amino acid concentrations in tomato fruits, with amino acid levels increasing up to 132% compared to WT plants. These induced mutations influenced the transcription levels of *SibZIP1* and related genes involved in sugar and amino acid biosynthesis, resulting in varying mRNA levels in mutant tomato lines. This highlights how

precise modifications of regulatory elements can manipulate the expression of key transcription factors, profoundly impacting vital fruit quality traits (Nguyen et al., 2023).

3.5 Engineering enhancers, transcription factors, 5'UTRs and 3'UTRs for crop improvement

Genetic variations in a transcription factor can influence TF binding, abundance, and activity. In a recent study, TFs linked to male sterility in maize were targeted. CRISPR/Cas9-mediated gene mutagenesis was employed to identify specific TFs regulating male sterility through genetic manipulation (Jiang et al., 2021). The GT-1 element in the *OsRAV2* promoter plays a crucial role in facilitating adaptive salt

responses. This was confirmed through *in situ* validation with targeted mutations generated using the CRISPR/Cas9 system in plants. Initially, regulatory elements in the *OsRAV2* promoter were identified using databases and computational tools such as PLACE and PlantCare (Figure 6E) (Lescot et al., 2002; Higo et al., 1998). Subsequently, *Agrobacterium*-mediated transformation was employed to induce site-directed mutagenesis in the promoter region of *RAV2*, specifically targeting the GT-1 element in rice. This approach ultimately revealed the 65-bp region responsible for salt tolerance within the *OsRAV2* promoter (Duan et al., 2016). For genome-wide analysis, use of a CRISPR/Cas9-sgRNA library identified 990 transcription factors in the tomato genome and synthesized 4379 plasmid-sgRNAs for targeted modification. This library enabled the creation of variations in multiple transcription factors using the CRISPR/Cas9-*Agrobacterium* transformation method. The efficacy of the CRISPR/Cas9-TF mutant library was validated through the observation of diverse phenotypic changes, such as modified leaf shape or number, seedling yellowing, alterations in fruit morphology or size, and abnormal floral organ development (Figure 6F) (Bi et al., 2023).

Regulatory elements within the 5' and 3' untranslated regions (UTRs) of genes are known to play crucial roles in controlling gene expression and posttranscriptional processes (Li et al., 2015). A recent study utilized a dual-CRISPR/Cas9 system to target modifications of regulatory elements within 3' UTRs (Figure 6G). This involved the use of a gRNA library for high-throughput localization of these elements, therefore targeting the full-length 3' UTRs. A high-throughput approach was then employed to identify regions regulating mRNA abundance. Individual gRNAs were subsequently used to screen the impact of targeting these regions on transcription and mRNA stability (Zhao et al., 2017). The CRISPR/Cas9 system was also utilized to modify the 5' UTR of *CRTISO*, which encodes carotenoid isomerase in tomato. By introducing genetic modifications in the 5' UTR region, researchers were able to manipulate the expression level of *CRTISO*. This resulted in reduced levels of lycopene, a pigment responsible for the red color in tomatoes (Figure 6H) (Lakshmi Jayaraj et al., 2021).

These studies emphasize the importance of utilizing SDE through genome engineering tools to investigate, understand, and characterize the functional significance of regulatory elements. Direct manipulation of these regions allows researchers to uncover the intricate mechanisms governing gene regulation and potentially devise novel strategies for crop enhancement and trait modification.

4 Conclusions

Genetic variation provides fuel for plant breeding and, by harnessing synthetic directed evolution (SDE), diverse alleles can

be generated for crop improvement. SDE tools not only modify the coding regions but also can be used to engineer non-coding regions to develop novel traits in crop plants. In this review, we have summarized the methodologies and strategies for protein engineering via SDE. These SDE tools were initially developed in bacteria but have been adapted to eukaryotic systems and are now used for plant trait engineering. Although SDE efforts have essentially focused on protein engineering and on recovering variants associated with desired traits, this represents only a small fraction of the possible variation that could be engineered under selective pressure. Other noncoding genetic elements that positively or negatively modulate transcriptional activity, including promoter sequences, uORF, Transcription Factors, 5' -UTR, and 3' -UTR, and *cis*-elements, could also be used for SDE. There is a huge potential to target and engineer the non-coding regions and *cis*-regulatory elements for trait improvement in plants, enabling us to tackle the current challenges of climate changes and global food production.

Author contributions

AK: Investigation, Writing – original draft, Writing – review & editing. HB: Supervision, Writing – original draft, Writing – review & editing. MM: Conceptualization, Project administration, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Achary, V. M. M., Sheri, V., Manna, M., Panditi, V., Borphukan, B., Ram, B., et al. (2020). Overexpression of improved EPSPS gene results in field level glyphosate tolerance and higher grain yield in rice. *Plant Biotechnol. J.* 18, 2504–2519. doi: 10.1111/pbi.13428

Anthimidou, E., Ntoaidou, S., Madesis, P., and Eleftherohorinos, I. (2020). Mechanisms of *Lolium rigidum* multiple resistance to ALS- and ACCase-inhibiting herbicides and their impact on plant fitness. *Pesticide Biochem. Physiol.* 164, 65–72. doi: 10.1016/j.pestbp.2019.12.010

- Arnold, F. H. (2019). Innovation by evolution: bringing new chemistry to life (Nobel lecture). *Angew Chem. Int. Ed Engl.* 58, 14420–14426. doi: 10.1002/anie.201907729
- Bi, M., Wang, Z., Cheng, K., Cui, Y., He, Y., Ma, J., et al. (2023). Construction of transcription factor mutagenesis population in tomato using a pooled CRISPR/Cas9 plasmid library. *Plant Physiol. Biochem.* 205, 108094. doi: 10.1016/j.plaphy.2023.108094
- Biot-Pelletier, D., and Martin, V. J. (2014). Evolutionary engineering by genome shuffling. *Appl. Microbiol. Biotechnol.* 98, 3877–3887. doi: 10.1007/s00253-014-5616-8
- Bilás, R., Szafran, K., Hnatuszko-Konka, K., and Kononowicz, A. K. (2016). Cis-regulatory elements used to control gene expression in plants. *Plant Cell Tissue Organ Cult.* 127, 269–287. doi: 10.1007/s11240-016-1057-7
- Butardo, V. M. Jr., Anacleto, R., Parween, S., Samson, I., de Guzman, K., Alhambra, C. M., et al. (2017). Systems genetics identifies a novel regulatory domain of amylose synthesis. *Plant Physiol.* 173, 887–906. doi: 10.1104/pp.16.01248
- Butt, H., Bazin, J., Alshareef, S., Eid, A., Benhamed, M., Reddy, A. S. N., et al. (2021). Overlapping roles of spliceosomal components SF3B1 and PHF5A in rice splicing regulation. *Commun. Biol.* 4, 529. doi: 10.1038/s42003-021-02051-y
- Butt, H., Eid, A., Momin, A. A., Bazin, J., Crespi, M., Arold, S. T., et al. (2019). CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. *Genome Biol.* 20, 73. doi: 10.1186/s13059-019-1680-9
- Butt, H., Ramirez, J. L. M., and Mahfouz, M. (2022). Synthetic evolution of herbicide resistance using a T7 RNAP-based random DNA base editor. *Life Sci. Alliance* 5, (12). doi: 10.26508/lsa.202201538
- Butt, H., Zaidi, S. S., Hassan, N., and Mahfouz, M. (2020). CRISPR-based directed evolution for crop improvement. *Trends Biotechnol.* 38, 236–240. doi: 10.1016/j.tibtech.2019.08.001
- Chekanova, J. A. (2015). Long non-coding RNAs and their functions in plants. *Curr. Opin. Plant Biol.* 27, 207–216. doi: 10.1016/j.pbi.2015.08.003
- Chen, R. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol.* 19 (1), 13–14. doi: 10.1016/S0167-7799(00)01522-5
- Chen, K. Q., and Arnold, F. H. (1991). Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnol. (N Y)* 9, 1073–1077. doi: 10.1038/nbt1191-1073
- Chen, K., and Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5618–5622. doi: 10.1073/pnas.90.12.5618
- Chen, H., Liu, S., Padula, S., Lesman, D., Griswold, K., Lin, A., et al. (2020). Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor. *Nat. Biotechnol.* 38, 165–168. doi: 10.1038/s41587-019-0331-8
- Chen, W., Zhang, Y., Zhang, Y., Pi, Y., Gu, T., Song, L., et al. (2018). CRISPR/cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine deaminase-mediated base editing in *Pseudomonas* species. *iScience* 6, 222–231. doi: 10.1016/j.isci.2018.07.024
- Chong, C. K., and Choi, J. D. (2000). Amino acid residues conferring herbicide tolerance in tobacco acetolactate synthase. *Biochem. Biophys. Res. Commun.* 279, 462–467. doi: 10.1006/bbrc.2000.3958
- Chow, C. N., Yang, C. W., Wu, N. Y., Wang, H. T., Tseng, K. C., Chiu, Y. H., et al. (2024). PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Res.* 52, D1569–D1578. doi: 10.1093/nar/gkad945
- Chu, Z., Chen, J., Nyporko, A., Han, H., Yu, Q., and Powles, S. (2018). Novel alpha-Tubulin Mutations Conferring Resistance to Dinitroaniline Herbicides in *Lolium rigidum*. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00097
- Cramer, A., Raillard, S. A., Bermudez, E., and Stemmer, W. P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* 391, 288–291. doi: 10.1038/34663
- Cravens, A., Jamil, O. K., Kong, D., Sockolosky, J. T., and Smolke, C. D. (2021). Polymerase-guided base editing enables *in vivo* mutagenesis and rapid protein engineering. *Nat. Commun.* 12, 1579. doi: 10.1038/s41467-021-21876-z
- Das, D., Singha, D. L., Paswan, R. R., Chowdhury, N., Sharma, M., Reddy, P. S., et al. (2022). Recent advancements in CRISPR/Cas technology for accelerated crop improvement. *Planta* 255, 109. doi: 10.1007/s00425-022-03894-3
- Davuluri, R. V., Sun, H., Palaniswamy, S. K., Matthews, N., Molina, C., Kurtz, M., et al. (2003). AGRIS- Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinf.* 4, 25. doi: 10.1186/1471-2105-4-25
- Délyea, C., Pernin, F., and Scarabel, L. (2010). Evolution and diversity of the mechanisms endowing resistance to herbicides inhibiting acetolactate-synthase (ALS) in corn poppy (*Papaver rhoeas* L.). *Plant Sci.* 180 (2), 333–342. doi: 10.1016/j.plantsci.2010.10.007
- Deng, K., Zhang, Q., Hong, Y., Yan, J., and Hu, X. (2023). iCREPCP: A deep learning-based web server for identifying base-resolution cis-regulatory elements within plant core promoters. *Plant Commun.* 4, 100455. doi: 10.1016/j.xplc.2022.100455
- Dennig, A., Shivange, A. V., Marienhagen, J., and Schwaneberg, U. (2011). OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One* 6, e26222. doi: 10.1371/journal.pone.0026222
- Doudna, J. A., and Jiang, F. (2017). CRISPR–cas9 structures and mechanisms. *Annu. Rev. Biophys.* 46, 505–529. doi: 10.1146/annurev-biophys-062215-010822
- Duan, Y. B., Li, J., Qin, R. Y., Xu, R. F., Li, H., Yang, Y. C., et al. (2016). Identification of a regulatory element responsible for salt induction of rice OsRAV2 through *ex situ* and *in situ* promoter analysis. *Plant Mol. Biol.* 90, 49–62. doi: 10.1007/s11103-015-0393-z
- Duy, P. N., Lan, D. T., Pham Thu, H., Thi Thu, H. P., Nguyen Thanh, H., Pham, N. P., et al. (2021). Improved bacterial leaf blight disease resistance in the major elite Vietnamese rice cultivar TBR225 via editing of the OsSWEET14 promoter. *PLoS One* 16, e0255470. doi: 10.1371/journal.pone.0255470
- Endo, M., Iwakami, S., and Toki, S. (2021). Precision genome editing in plants via gene targeting and subsequent break-induced single-strand annealing. *Plant Biotechnol. J.* 19, 563–574. doi: 10.1111/pbi.13485
- Endo, M., Mikami, M., Endo, A., Kaya, H., Itoh, T., Nishimasu, H., et al. (2019). Genome editing in plants by engineered CRISPR-Cas9 recognizing NG PAM. *Nat. Plants* 5, 14–17. doi: 10.1038/s41477-018-0321-8
- Endo, M., Osakabe, K., Ono, K., Handa, H., Shimizu, T., and Toki, S. (2007). The Plant Journal - 2007 - Endo - Molecular breeding of a novel herbicide-tolerant rice by gene targeting. *Plant J.* 52 (1), 157–166. doi: 10.1111/j.1365-313X.2007.03230.x
- English, J. G., Olsen, R. H. J., Lansu, K., Patel, M., White, K., Cockrell, A. S., et al. (2019). VEGAS as a platform for facile directed evolution in mammalian cells. *Cell* 178, 748–761 e717. doi: 10.1016/j.cell.2019.05.051
- Ferebee, T. H., and Buckler, E. S. (2023). Exploring the utility of regulatory network-based machine learning for gene expression prediction in maize. *bioRxiv*. doi: 10.1101/2023.05.11.540406
- Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W., et al. (2006). DRTF: a database of rice transcription factors. *Bioinformatics* 22, 1286–1287. doi: 10.1093/bioinformatics/btl107
- García-García, J. D., Van Gelder, K., Joshi, J., Bathe, U., Leong, B. J., Bruner, S. D., et al. (2022). Using continuous directed evolution to improve enzymes for plant applications. *Plant Physiol.* 188, 971–983. doi: 10.1093/plphys/kiab500
- Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., et al. (2005). DATF: a database of Arabidopsis transcription factors. *Bioinformatics* 21, 2568–2569. doi: 10.1093/bioinformatics/bti334
- Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., and Dueber, J. E. (2018). CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* 560, 248–252. doi: 10.1038/s41586-018-0384-8
- Han, H., Yu, Q., Purba, E., Li, M., Walsh, M., Friesen, S., et al. (2012). A novel amino acid substitution Ala-122-Tyr in ALS confers high-level and broad resistance across ALS-inhibiting herbicides. *Pest Manag. Sci.* 68, 1164–1170. doi: 10.1002/ps.3278
- Hendel, S. J., and Shoulders, M. D. (2021). Directed evolution in mammalian cells. *Nat. Methods* 18, 346–357. doi: 10.1038/s41592-021-01090-x
- Hess, G. T., Fresard, L., Han, K., Lee, C. H., Li, A., Cimprich, K. A., et al. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* 13, 1036–1042. doi: 10.1038/nmeth.4038
- Higo, K., Ugawa, Y., Iwamoto, M., and Higo, H. (1998). PLACE - A database of plant cis-acting regulatory DNA elements. *Nucleic Acids Res.* 26 (1), 358–359. doi: 10.1093/nar/26.1.358
- Huang, L., Li, Q., Zhang, C., Chu, R., Gu, Z., Tan, H., et al. (2020). Creating novel Wx alleles with fine-tuned amylose levels and improved grain quality in rice by promoter editing using CRISPR/Cas9 system. *Plant Biotechnol. J.* 18, 2164–2166. doi: 10.1111/pbi.13391
- Jakociunas, T., Pedersen, L. E., Lis, A. V., Jensen, M. K., and Keasling, J. D. (2018). CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metab. Eng.* 48, 288–296. doi: 10.1016/j.ymben.2018.07.001
- James, D., Borphan, B., Fartyal, D., Ram, B., Singh, J., Manna, M., et al. (2018). Concurrent overexpression of osGS1;1 and osGS2 genes in transgenic rice (*Oryza sativa* L.): impact on tolerance to abiotic stresses. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00786
- Jang, S., Marjanovic, J., and Gornicki, P. (2013). Resistance to herbicides caused by single amino acid mutations in acetyl-CoA carboxylase in resistant populations of grassy weeds. *New Phytol.* 197, 1110–1116. doi: 10.1111/nph.12117
- Jermutus, L. (2000). Tailoring *in vitro* evolution for protein affinity or stability. *Proc. Natl. Acad. Sci.* 98, 75–80. doi: 10.1073/pnas.98.1.75
- Jiang, J. (2015). The 'dark matter' in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin. *Curr. Opin. Plant Biol.* 24, 17–23. doi: 10.1016/j.pbi.2015.01.005
- Jiang, Y., An, X., Li, Z., Yan, T., Zhu, T., Xie, K., et al. (2021). CRISPR/Cas9-based discovery of maize transcription factors regulating male sterility and their functional conservation in plants. *Plant Biotechnol. J.* 19, 1769–1784. doi: 10.1111/pbi.13590
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, D1040–D1045. doi: 10.1093/nar/gkw982
- Jores, T., Tonnies, J., Dorrity, M. W., Cuperus, J. T., Fields, S., and Queitsch, C. (2020). Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. *Plant Cell* 32, 2120–2131. doi: 10.1105/tpc.20.00155

- Kawai, K., Kaku, K., Izawa, N., Fukuda, A., Tanaka, Y., and Shimizu, T. (2007). Functional analysis of transgenic rice plants expressing a novel mutated ALS gene of rice. *J. Pesticide Sci.* 32 (4), 385–392. doi: 10.1584/jpestics.G07-08
- Kuang, Y., Li, S., Ren, B., Yan, F., Spetz, C., Li, X., et al. (2020). Base-editing-mediated artificial evolution of osALS1 in planta to develop novel herbicide-tolerant rice germplasm. *Mol. Plant* 13, 565–572. doi: 10.1016/j.molp.2020.01.010
- Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T., and Furukawa, K. (1998). Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat. Biotechnol.* 16, 663–666. doi: 10.1038/nbt0798-663
- Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., et al. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* 39, 41–46. doi: 10.1038/s41587-020-0609-x
- Lakshmi Jayaraj, K., Thulasidharan, N., Antony, A., John, M., Augustine, R., Chakravartty, N., et al. (2021). Targeted editing of tomato carotenoid isomerase reveals the role of 5' UTR region in gene expression regulation. *Plant Cell Rep.* 40, 621–635. doi: 10.1007/s00299-020-02659-0
- Laplanche, J., Rajcan, I., and Tardif, F. J. (2009). Multiple allelic forms of acetohydroxyacid synthase are responsible for herbicide resistance in *Setaria viridis*. *Theor. Appl. Genet.* 119, 577–585. doi: 10.1007/s00122-009-1067-5
- Lerner, S. A., Wu, T. T., and Lin, E. C. (1964). Evolution of a catabolic pathway in bacteria. *Science* 146, 1313–1315. doi: 10.1126/science.146.3649.1313
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30 (1), 325–327. doi: 10.1093/nar/30.1.325
- Levisohn, R., and Spiegelman, S. (1969). Further extracellular darwinian experiments with replicating rna molecules - diverse variants isolated under different selective conditions. *P Natl. Acad. Sci. U.S.A.* 63, 805. doi: 10.1073/pnas.63.3.805
- Li, Q., Feng, Q., Snouffer, A., Zhang, B., Rodriguez, G. R., and van der Knaap, E. (2022). Increasing fruit weight by editing a cis-regulatory element in tomato KLUH promoter using CRISPR/Cas9. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.879642
- Li, H., Li, J., Chen, J., Yan, L., and Xia, L. (2020b). Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* 13, 671–674. doi: 10.1016/j.molp.2020.03.011
- Li, W., Ma, M., Feng, Y., Li, H., Wang, Y., Ma, Y., et al. (2015). EIN2-directed translational regulation of ethylene signaling in Arabidopsis. *Cell* 163, 670–683. doi: 10.1016/j.cell.2015.09.037
- Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., et al. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat. Plants* 2, 16139. doi: 10.1038/nplants.2016.139
- Li, C., Zhang, R., Meng, X., Chen, S., Zong, Y., Lu, C., et al. (2020a). Targeted, random mutagenesis of plant genes with dual cytosine and adenine base editors. *Nat. Biotechnol.* 38, 875–882. doi: 10.1038/s41587-019-0393-7
- Li, Y., Zhu, J., Wu, H., Liu, C., Huang, C., Lan, J., et al. (2020c). Precise base editing of non-allelic acetolactate synthase genes confers sulfonylurea herbicide resistance in maize. *Crop J.* 8, 449–456. doi: 10.1016/j.cj.2019.10.001
- Liu, L., Kuang, Y., Yan, F., Li, S., Ren, B., Gosavi, G., et al. (2021). Developing a novel artificial rice germplasm for dinitroaniline herbicide resistance by base editing of OsTubA2. *PINT Biotechnol. J.* 19 (1), 5–7. doi: 10.1111/pbi.13430
- Liu, X., Qin, R., Li, J., Liao, S., Shan, T., Xu, R., et al. (2020). A CRISPR-Cas9-mediated domain-specific base-editing screen enables functional assessment of ACCase variants in rice. *Plant Biotechnol. J.* 18, 1845–1847. doi: 10.1111/pbi.13348
- Liu, H. J., and Yan, J. (2019). Crop genome-wide association study: a harvest of biological relevance. *Plant J.* 97, 8–18. doi: 10.1111/tpj.14139
- Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* 13, 1029–1035. doi: 10.1038/nmeth.4027
- Marand, A. P., Eveland, A. L., Kaufmann, K., and Springer, N. M. (2023). cis-regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* 74, 111–137. doi: 10.1146/annurev-arplant-070122-030236
- Mendes, R. R., Takano, H. K., Oliveira, R. S., Adegas, F. S., Gaines, T. A., and Dayan, F. E. (2020). A trp574Leu target-site mutation confers imazamox resistance in multiple herbicide-resistant wild poinsettia populations from Brazil. *Agronomy* 10, (8). doi: 10.3390/agronomy10081057
- Miller, S. M., Wang, T., and Liu, D. R. (2020). Phage-assisted continuous and non-continuous evolution. *Nat. Protoc.* 15, 4101–4127. doi: 10.1038/s41596-020-00410-3
- Mills, D. R., Peterson, R. L., and Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.* 58 (1), 217–224. doi: 10.1073/pnas.58.1.217
- Neelakandan, A. K., Wright, D. A., Traore, S. M., Chen, X., Spalding, M. H., and He, G. (2022). CRISPR/Cas9 Based Site-Specific Modification of FAD2 cis-Regulatory Motifs in Peanut (*Arachis hypogaea* L.). *Front. Genet.* 13. doi: 10.3389/fgene.2022.849961
- Neggess, J. E., Kwanten, B., Dierckx, T., Noguchi, H., Voet, A., Bral, L., et al. (2018). Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nat. Commun.* 9, 502. doi: 10.1038/s41467-017-02349-8
- Nguyen, N. H., Bui, T. P., Le, N. T., Nguyen, C. X., Le, M. T. T., Dao, N. T., et al. (2023). Disrupting Sc-uORFs of a transcription factor bZIP1 using CRISPR/Cas9 enhances sugar and amino acid contents in tomato (*Solanum lycopersicum*). *Planta* 257, 57. doi: 10.1007/s00425-023-04089-0
- Nyerges, A., Csorgo, B., Draskovits, G., Kintses, B., Szili, P., Ferenc, G., et al. (2018). Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5726–E5735. doi: 10.1073/pnas.1801646115
- Okuzaki, A., Shimizu, T., Kaku, K., Kawai, K., and Toriyama, K. (2007). A novel mutated acetolactate synthase gene conferring specific resistance to pyrimidinyl carboxy herbicides in rice. *Plant Mol. Biol.* 64, 219–224. doi: 10.1007/s11103-007-9146-y
- Packer, M. S., and Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nat. Rev. Genet.* 16, 379–394. doi: 10.1038/nrg3927
- Patnaik, R., Louie, S., Gavrilovic, V., Perry, K., Stemmer, W. P., Ryan, C. M., et al. (2002). Genome shuffling of *Lactobacillus* for improved acid tolerance. *Nat. Biotechnol.* 20, 707–712. doi: 10.1038/nbt0702-707
- Peleke, F. F., Zumkeller, S. M., Gültas, M., Schmitt, A., and Szymański, J. J. (2024). Deep learning the cis-regulatory code for gene expression in selected model plants. *Nat. Commun.* 15 (1), 3488. doi: 10.1038/s41467-024-47744-0
- Perotti, V. E., Larran, A. S., Palmieri, V. E., Martinatto, A. K., Alvarez, C. E., Tunesca, D., et al. (2019). A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant *Amaranthus* hybrid population from Argentina. *Pest Manag. Sci.* 75, 1242–1251. doi: 10.1002/ps.5303
- Pines, G., Pines, A., Garst, A. D., Zeitoun, R. I., Lynch, S. A., and Gill, R. T. (2015). Codon compression algorithms for saturation mutagenesis. *ACS Synth. Biol.* 4, 604–614. doi: 10.1021/sb500282v
- Popa, S. C., Inamoto, I., Thuronyi, B. W., and Shin, J. A. (2020). Phage-assisted continuous evolution (PACE): A guide focused on evolving protein–DNA interactions. *ACS Omega* 5, 26957–26966. doi: 10.1021/acsomega.0c03508
- Rajguru, S. N. (2005). Mutations in the red rice gene associated with resistance to imazethapyr. *Weed Sci.* 53 (5), 567–577. doi: 10.1614/WS-04-111R1.1
- Rao, G. S., Jiang, W., and Mahfouz, M. (2021). Synthetic directed evolution in plants: unlocking trait engineering and improvement. *Synth. Biol. (Oxf)* 6, ysab025. doi: 10.1093/synbio/ysab025
- Ravikumar, A., Arrieta, A., and Liu, C. C. (2014). An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* 10, 175–177. doi: 10.1038/nchembio.1439
- Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., and Liu, C. C. (2018). Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* 175, 1946–1957 e1913. doi: 10.1016/j.cell.2018.10.021
- Reidhaar-Olson, J. F., and Sauer, R. T. (1998). Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science*. 241 (4861), 53–57. doi: 10.1126/science.3388019
- Ren, B., Kuang, Y., Xu, Z., Wu, X., Zhang, D., Yan, F., et al. (2023). Three novel alleles of OsGS1 developed by base-editing-mediated artificial evolution confer glufosinate tolerance in rice. *Crop J.* 11, 661–665. doi: 10.1016/j.cj.2022.10.003
- Riar, D. S., Norsworthy, J. K., Srivastava, V., Nandula, V., Bond, J. A., and Scott, R. C. (2013). Physiological and molecular basis of acetolactate synthase-inhibiting herbicide resistance in barnyardgrass (*Echinochloa crus-galli*). *J. Agric. Food Chem.* 61, 278–289. doi: 10.1021/jf304675j
- Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., et al. (2019). Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* 5, 1237–1249. doi: 10.1038/s41477-019-0547-0
- Rodrigues, S. D., Karimi, M., Impens, L., Van Lerberge, E., Coussens, G., Aesaert, S., et al. (2021). Efficient CRISPR-mediated base editing in *Agrobacterium* spp. *Proc. Natl. Acad. Sci. U.S.A.* 118, (2). doi: 10.1073/pnas.2013338118
- Rodriguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Plant J.* 171, 470–480 e478. doi: 10.1016/j.plj.2017.08.030
- Schmidt-Dannert, C., Umeno, D., and Arnold, F. H. (2000). Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.* 18, 750–753. doi: 10.1038/77319
- Schmitz, R. J., Grotewold, E., and Stam, M. (2022). Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. *Plant Cell* 34, 718–741. doi: 10.1093/plcell/koab281
- Service, R. F. (2018). Protein evolution earns chemistry Nobel. *Science* 362, 142. doi: 10.1126/science.362.6411.142
- Shimatani, Z., Fujikura, U., Ishii, H., Terada, R., Nishida, K., and Kondo, A. (2018). Herbicide tolerance-assisted multiplex targeted nucleotide substitution in rice. *Data Brief* 20, 1325–1331. doi: 10.1016/j.dib.2018.08.124
- Shimizu, M., Kawai, K., Kaku, K., Shimizu, T., and Kobayashi, H. (2011). Application of mutated acetolactate synthase genes to herbicide resistance and plant improvement. *Herbicides, Theory and Applications*. 10, 193–212. doi: 10.5772/13449
- Simon, A. J., d'Oelsnitz, S., and Ellington, A. D. (2019). Synthetic evolution. *Nat. Biotechnol.* 37, 730–743. doi: 10.1038/s41587-019-0157-4
- Song, X., Meng, X., Guo, H., Cheng, Q., Jing, Y., Chen, M., et al. (2022). Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat. Biotechnol.* 40, 1403–1411. doi: 10.1038/s41587-022-01281-7
- Stemmer, W. P. (1994a). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10747–10751. doi: 10.1073/pnas.91.22.10747

- Stemmer, W. P. C. (1994b). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* 370, 389–391. doi: 10.1038/370389a0
- Sun, J., He, N., Niu, L., Huang, Y., Shen, W., Zhang, Y., et al. (2019). Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics Proteomics Bioinf.* 17, 140–153. doi: 10.1016/j.gpb.2018.11.003
- Swinnen, G., Goossens, A., and Pauwels, L. (2016). Lessons from domestication: targeting-regulatory elements for crop improvement. *Trends Plant Sci.* 21, 506–515. doi: 10.1016/j.tplants.2016.01.014
- Tan, W. C., Miao, J., Xu, B., Zhou, C. T., Wang, Y. R., Gu, X. Q., et al. (2023). Rapid production of novel beneficial alleles for improving rice appearance quality by targeting a regulatory element of. *Plant Biotechnol. J.* 21, 1305–1307. doi: 10.1111/pbi.14041
- Tanigaki, S., Uchino, A., Okawa, S., Miura, C., Hamamura, K., Matsuo, M., et al. (2021). Gene expression shapes the patterns of parallel evolution of herbicide resistance in the agricultural weed *Monochoria vaginalis*. *New Phytol.* 232, 928–940. doi: 10.1111/nph.17624
- Tian, Y. S., Xu, J., Zhao, W., Xing, X. J., Fu, X. Y., Peng, R. H., et al. (2015). Identification of a phosphinothricin-resistant mutant of rice glutamine synthetase using DNA shuffling. *Sci. Rep.* 5, 15495. doi: 10.1038/srep15495
- Tian, F., Yang, D. C., Meng, Y. Q., Jin, J., and Gao, G. (2020). PlantRegMap: charting functional regulatory maps in plants. *Nucleic Acids Res.* 48, D1104–D1113. doi: 10.1093/nar/gkz1020
- van der Horst, S., Filipovska, T., Hanson, J., and Smeeckens, S. (2020). Metabolite control of translation by conserved peptide uORFs: the ribosome as a metabolite multisensor. *Plant Physiol.* 182, 110–122. doi: 10.1104/pp.19.00940
- von Arnim, A. G., Jia, Q., and Vaughn, J. N. (2014). Regulation of plant translation by upstream open reading frames. *Plant Sci.* 214, 1–12. doi: 10.1016/j.plantsci.2013.09.006
- Wang, H., He, Y., Wang, Y., Li, Z., Hao, J., Song, Y., et al. (2022). Base editing-mediated targeted evolution of ACCase for herbicide-resistant rice mutants. *J. Integr. Plant Biol.* 64, 2029–2032. doi: 10.1111/jipb.13352
- Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., et al. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460, 894–898. doi: 10.1038/nature08187
- Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. (2012). Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* 9, 591–593. doi: 10.1038/nmeth.1971
- Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S. T., and Zhao, H. (2021). Directed evolution: methodologies and applications. *Chem. Rev.* 121, 12384–12444. doi: 10.1021/acs.chemrev.1c00260
- Warwick, S. I., Xu, R., Sauder, C., and Beckie, H. J. (2008). Acetolactate synthase target-site mutations and single nucleotide polymorphism genotyping in ALS-resistant kochia (*Kochia scoparia*). *Weed Sci.* 56, 797–806. doi: 10.1614/ws-08-045.1
- Wen, F., McLachlan, M., and Zhao, H. (2007). Directed evolution: Novel and improved enzymes. *Wiley Encyclopedia of Chemical Biology.* 14, 1–10. doi: 10.1002/9780470048672.webc125
- Wittkopp, P. J., and Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69. doi: 10.1038/nrg3095
- Wright, C., Neres, D. F., Chaisupa, P., and Bryant, J. A. (2023). Protein engineering and plants- the evolution of sustainable agriculture. *Biochem.* 45 (1), 12–17. doi: 10.1042/bio_2023_101
- Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J., and Arnold, F. H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci.* 116, 8852–8858. doi: 10.1073/pnas.1901979116
- Xue, C., Qiu, F., Wang, Y., Li, B., Zhao, K. T., Chen, K., et al. (2023). Tuning plant phenotypes by precise, graded downregulation of gene expression. *Nat. Biotechnol.* 41, 1758–1764. doi: 10.1038/s41587-023-01707-w
- Yang, K. K., Wu, Z., and Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* 16, 687–694. doi: 10.1038/s41592-019-0496-6
- Yu, Q., Collavo, A., Zheng, M. Q., Owen, M., Sattin, M., and Powles, S. B. (2007). Diversity of acetyl-coenzyme A carboxylase mutations in resistant *Lolium* populations: evaluation using clethodim. *Plant Physiol.* 145, 547–558. doi: 10.1104/pp.107.105262
- Yu, K., Liu, Z., Gui, H., Geng, L., Wei, J., Liang, D., et al. (2021). Highly efficient generation of bacterial leaf blight-resistant and transgene-free rice using a genome editing and multiplexed selection system. *BMC Plant Biol.* 21, 197. doi: 10.1186/s12870-021-02979-7
- Zemlyanskaya, E. V., Dolgikh, V. A., Levitsky, V. G., and Mironova, V. (2021). Transcriptional regulation in plants: Using omics data to crack the cis-regulatory code. *Curr. Opin. Plant Biol.* 63, 102058. doi: 10.1016/j.pbi.2021.102058
- Zeng, D., Liu, T., Ma, X., Wang, B., Zheng, Z., Zhang, Y., et al. (2020). Quantitative regulation of Waxy expression by CRISPR/Cas9-based promoter and 5'UTR-intron editing improves grain quality in rice. *Plant Biotechnol. J.* 18, 2385–2387. doi: 10.1111/pbi.13427
- Zhang, R., Chen, S., Meng, X., Chai, Z., Wang, D., Yuan, Y., et al. (2021). Generating broad-spectrum tolerance to ALS-inhibiting herbicides in rice by base editing. *Sci. China Life Sci.* 64, 1624–1633. doi: 10.1007/s11427-020-1800-5
- Zhang, Y. X., Perry, K., Vinci, V. A., Powell, K., Stemmer, W. P., and del Cardayre, S. B. (2002). Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature* 415, 644–646. doi: 10.1038/415644a
- Zhang, Y., and Qi, Y. (2019). CRISPR enables directed evolution in plants. *Genome Biol.* 20, 83. doi: 10.1186/s13059-019-1693-4
- Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., et al. (2018). Genome editing of upstream open reading frames enables translational control in plants. *Nat. Biotechnol.* 36, 894–898. doi: 10.1038/nbt.4202
- Zhang, C., Yu, Q., Han, H., Yu, C., Nyporko, A., Tian, X., et al. (2022). A naturally evolved mutation (Ser59Gly) in glutamine synthetase confers glufosinate resistance in plants. *J. Exp. Bot.* 73, 2251–2262. doi: 10.1093/jxb/erac008
- Zhang, C., Zhong, X., Li, S., Yan, L., Li, J., He, Y., et al. (2023). Artificial evolution of OsEPSPS through an improved dual cytosine and adenine base editor generated a novel allele conferring rice glyphosate tolerance. *J. Integr. Plant Biol.* 65, 2194–2203. doi: 10.1111/jipb.13543
- Zhao, W., Siegel, D., Biton, A., Tonqueze, O. L., Zaitlen, N., Ahituv, N., et al. (2017). CRISPR-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Res.* 45, 10800–10810. doi: 10.1093/nar/gkx675
- Zhu, B., Zhang, W., Zhang, T., Liu, B., and Jiang, J. (2015). Genome-wide prediction and validation of intergenic enhancers in arabidopsis using open chromatin signatures. *Plant Cell* 27, 2415–2426. doi: 10.1105/tpc.15.00537
- Zhu, Q. H., Guo, A. Y., Gao, G., Zhong, Y. F., Xu, M., Huang, M., et al. (2007). DPTF: a database of poplar transcription factors. *Bioinformatics* 23, 1307–1308. doi: 10.1093/bioinformatics/btm113
- Zong, Y., Song, Q., Li, C., Jin, S., Zhang, D., Wang, Y., et al. (2018). Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* 36, 950–953. doi: 10.1038/nbt.4261
- Zsogon, A., Peres, L. E. P., Xiao, Y., Yan, J., and Fernie, A. R. (2022). Enhancing crop diversity for food security in the face of climate uncertainty. *Plant J.* 109, 402–414. doi: 10.1111/tjpl.15626