Trends in Biotechnology

Article type: Forum

CRISPR-based directed evolution for crop improvement

Haroon Butt¹, Syed Shan-e-Ali Zaidi², Norhan Hassan¹, and Magdy Mahfouz^{1,*}

¹*Laboratory for Genome Engineering and Synthetic Biology, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

²*Plant Genetics Lab, TERRA Research and Teaching Centre, Gembloux Agro BioTech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium*

*Correspondence: Magdy.mahfouz@kaust.edu.sa (Magdy Mahfouz)

Key words: CRISPR-Cas9, NHEJ, HDR, CRISPR-directed evolution, trait engineering, crop breeding

Abstract:

Directed evolution involves generating diverse sequence variants of a gene of interest to produce variants with a desirable trait under selective pressure. CRISPR-Cas9 systems can be programmed to target any genomic locus and perform a targeted CRISPR-mediated directed evolution (CDE). Here, we discuss the opportunities and challenges of this emerging platform for targeted crop improvement.

Directed evolution

In plants, low genetic diversity limits crop improvement, particularly in crops that have undergone protracted selection, and hampers biotechnological approaches that seek to harness plants' autotrophic primary metabolism and versatile secondary metabolism. Rather than depending on natural genetic diversity, directed evolution accelerates evolution by inserting random mutations into a target gene in a host organism to generate modified proteins with potentially superior performance. Directed evolution requires 1) a method to make mutations in a target sequence and 2) a method to screen or select for proteins with superior performance.

For the first method, DNA mutations can be produced using chemical mutagens, radiation, or nonchemical methods that increase the error rate during DNA replication [1-3]. Error-prone PCR has been used to generate random mutagenesis libraries in prokaryotes [4]. However, the advent of genome-editing technologies such as CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR associated) has given researchers the ability to target mutations to a specific gene or region, as described in detail below.

For the second method, approaches to measure the performance of individual protein variants are often based on screening for a visible phenotype or selecting for survival of the cell or organism under specific conditions [1]. Various screening and selection methods have been used, including reporter gene activity assays based on staining or fluorescence [1, 5]. Other methods are based on affinity binding or organism survival [1].

CRISPR/Cas-mediated directed evolution

To ensure the successful performance of the evolved gene variant, it is essential to conduct directed evolution experiments in the original host within the same cellular context intended for its ultimate use. Unfortunately, most techniques were developed in prokaryotic systems, mainly due to the lack of efficient technologies for site-specific sequence diversification in eukaryotes. However, the advent of CRISPR/Cas technologies for eukaryotic genome editing and site-specific mutagenesis [6-8] has revolutionized our ability to generate specific changes, including double-strand breaks (DSBs), single-stranded nicks, or specific base changes via base editing; all of which can be used for CRISPR-mediated directed evolution (CDE).

CDE via imprecise DSB repair:

CRISPR-generated DSBs are repaired via the imprecise nonhomologous end joining (NHEJ) or the precise homology-directed repair (HDR) process. NHEJ generates small insertions, deletions, and base substitutions and is therefore ideal for targeted mutagenesis and the functional knockout of target genes. HDR requires a repair template, from which information is copied across the break site. NHEJ is the most common, efficient repair process in most cell types, including plant cells; HDR is inefficient and challenging due to the requirement for a repair template at the break site.

A recent study used NHEJ for CDE in rice [9] with an sgRNA library targeting *SPLICING FACTOR 3B SUBUNIT 1* (*OsSF3B1*). *OsSF3B1* is an essential gene that is conserved among eukaryotes and targeted by the splicing inhibitors pladienolide B and herboxidiene (GEX1A) [9]. For genome engineering applications, CRISPR specificity comes from a nucleic acid, the single-guide RNA (sgRNA) and to target *OsSF3B1*, a tiled sgRNA library was designed containing all possible sgRNAs on the forward and reverse DNA strands. CRISPR requires that the sgRNA include a three-nucleotide protospacer-adjacent motif (PAM) of NGG; this limits the number of sgRNAs possible for a specific gene. The herbicide GEX1A was used for CDE selection of *OsSF3B1* variants. The variants generated in this proof-of-concept study displayed different levels of GEX1A resistance, with plants harboring K1049R, K1050E, and G1051H mutations being almost insensitive to GEX1A treatment [9]. All of the observed transmitted mutations were either deletions or substitutions but were in frame. This pioneering study demonstrates the promise of CDE for applications in plant systems [9].

In mammalian cells, CDE was used to generate a wide variety of functional in-frame mutations in essential genes [10]. The first experiments used three cancer drugs for selection, each targeting a different protein. The sgRNAs were designed to target hotspots in each locus in a human cell line. After treatment, the resistant colonies were sequenced. Most of the mutations were in frame with known or novel protein variants. By the same strategy, sgRNA tiling libraries targeting multiple genes to obtain resistance against anticancer drug, bortezomib. Analysis of sgRNAs identified *PSMB5* as the sole target in bortezomib-resistant cells and sequencing of *PSMB5* revealed multiple in-frame mutations. These results highlight the feasibility of using CDE for large-scale target resolution. Using the same method, the authors identified a protein nicotinamide phosphoribosyl transferase (NAMPT) as the molecular target of the anticancer compound KPT-9274 [10].

CDE via EvolvR:

Instead of producing DSBs, CRISPR systems and sgRNAs can be used with other enzymes, such as a ‘nickase’ (nCas9) that cuts only one of the two DNA strands. This nick can be used as the starting point for DNA synthesis with an error-prone enzyme that introduces different nucleotides, thereby generating mutations and protein variants for CDE. For example, EvolvR uses CRISPR-guided nCas9 fused to an error-prone DNA polymerase I (PolI3M) [11] (Figure 1A). The sgRNA guides nCas9-PolI3M to the target genomic region, where nCas9 produces a nick in the target DNA and then dissociates from the nicked region. Next, PolI3M associates with the nicked region, where it produces a new DNA strand with errors, displacing and cleaving the old DNA strand and leaving a nick that can be repaired by the cell’s machinery [11]. EvolvR is a powerful tool for enhancing mutagenesis, but its use in eukaryotic systems remains to be demonstrated.

CDE via base editing:

CRISPR/Cas-mediated base editing does not generate DSBs, instead inducing chemical changes that convert one nucleotide to another (such as C to T, causing a G to C change in the opposite strand) near the PAM (Figure 1B). The cytidine deaminase base editors are fusions of nCas9 and the APOBEC deaminase, which converts C to T [12]. The nCas9 domain of the base editor generates a nick and restricts long-patch base-excision repair to using the deaminated strand for repair [12, 13]. In an alternative approach, human activation-induced cytidine deaminase (AID) was fused to MS2 (coat protein of the RNA bacteriophage) and recruited to the target locus via

two MS2 RNA hairpins inserted into the sgRNA backbone. The AID*Δ editor (with a deletion of the nuclear export signal, NES) has been used in the CRISPR-X system, conveying increased deamination activity, improved mutation frequency, and a larger window size of +50 to −50 relative to the PAM [13, 14] (Figure 1B). A modified approach was proposed termed CREATE, where one base editor is stably transformed with multiple sgRNA targeting a narrow region of genome and TILLING could be performed to for screening of new phenotypes [15]. The base editors can be used to develop herbicide resistance in plant, virus resistance crops by inducing stop codons in virus genomes and production of biotic and abiotic tolerance via precise editing of the crop genomes.

CDE for crop improvement

Current agricultural research aims to generate new herbicide tolerance traits, improve resistance to biotic and abiotic stress, reduce inputs (water and nitrogen), and improve output traits such as yield and quality. CDE can be used to accelerate this research, depending on the identification of appropriate target genes and the ability to devise informative screens and selections. For crop improvement, CDE can target EPSPS (5-enolpyruvylshikimate 3-phosphate synthase), for improved resistance to the herbicide glyphosate; Rubisco, for improved photosynthesis and growth; ascorbate peroxidase and isoprene synthase, for abiotic stress tolerance; nucleotide-binding leucine-rich repeat (NB-NLR)-containing proteins and FLAGELLIN-SENSING2, for increased pathogen resistance. Other noncoding elements that modulate transcriptional activity, including promoter sequences and *cis*-elements, can also be used for CDE. CDE can be used to evolve protein variants controlling responses to nutritional stress, thereby improving the plant's ability to adapt to this condition.

CDE: Challenges and future outlook

CDE generates variants with desired traits with a more focused approach than other genome-wide or single-gene directed evolution techniques in heterologous hosts (Figure 2). By harnessing nature's ability to evolve gene variants via NHEJ, protein variants could be generated that are extremely difficult, if not impossible, to be obtained using HDR approaches. Key challenges for

CDE approach include delivery of CRISPR-Cas reagents and selection of the mutant variant with desired trait of interest. Future applications of other CDE methods, such as EvolvR, will expand the possibilities. CDE will have myriad applications in basic plant biology research, crop improvement, and biotechnology. For example, CDE may be useful for molecular farming, in which plants are used as hosts to produce biologics or pharmaceutical compounds, to generate protein variants suitable for the overproduction of enzymes or catalytically active biologics [2].

However, enhancement of CDE technology will be required for widespread adoption. For example, improvements in transformation methods and CRISPR/Cas reagent delivery into the target cells will enable CDE in various tissues, including germline cells, and increase the number of plant species amenable to CDE. Refined, genome-wide functional studies will be required to select appropriate targets for sgRNA design.

Ongoing research in this emerging field holds great promise to help address important agricultural issues, including increasing population and climate change, advance basic research, and enable novel biotechnological applications.

References

1. Packer, M.S. and Liu, D.R. (2015) Methods for the directed evolution of proteins. *Nat Rev Genet* 16 (7), 379-94.
2. Porter, J.L. et al. (2016) Directed Evolution of Enzymes for Industrial Biocatalysis. *ChemBiochem* 17 (3), 197-203.
3. Brandenberg, O.F. et al. (2017) Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr Opin Biotechnol* 47, 102-111.
4. Dumas, L. et al. (2018) Structure-Function Analysis of Chloroplast Proteins via Random Mutagenesis Using Error-Prone PCR. *Plant Physiol* 177 (2), 465-475.
5. Sachsenhauser, V. and Bardwell, J.C. (2018) Directed evolution to improve protein folding in vivo. *Curr Opin Struct Biol* 48, 117-123.
6. Cho, S.W. et al. (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31 (3), 230-2.
7. Nekrasov, V. et al. (2013) Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31 (8), 691-3.
8. Eid, A. et al. (2016) High efficiency of targeted mutagenesis in *arabidopsis* via meiotic promoter-driven expression of Cas9 endonuclease. *Plant Cell Rep* 35 (7), 1555-8.
9. Butt, H. et al. (2019) CRISPR directed evolution of the spliceosome for resistance to splicing inhibitors. *Genome Biol* 20 (1), 73.

10. Neggers, J.E. et al. (2018) Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nat Commun* 9 (1), 502.
11. Halperin, S.O. et al. (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* 560 (7717), 248-252.
12. Kim, Y.B. et al. (2017) Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat Biotechnol* 35 (4), 371-376.
13. Eid, A. et al. (2018) CRISPR base editors: genome editing without double-stranded breaks. *Biochem J* 475 (11), 1955-1964.
14. Hess, G.T. et al. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat Methods* 13 (12), 1036-1042.
15. Jacob, P. et al. (2018) Translational Research: Exploring and Creating Genetic Diversity. *Trends Plant Sci* 23 (1), 42-52.

Figure legends:

Figure 1: CDE techniques

(A) The EvolvR system consists of nCas9 (nickase Cas9) and PolI3M (error-prone polymerase I). sgRNA guides the complex to genomic targets. nCas9 nicks the DNA and PolI3M synthesizes the new DNA strand, introducing mutations. The old strand is displaced and cleaved by flap endonuclease activity. PolI3M dissociates from the ligatable nick and can retarget the region, depending on the mutation. (B) Cytosine base editors with their editing windows. BE3 (based on APOBEC1) fused with nCas9 functions in the small region N4-N8. In the CRISPR-X system, AID*Δ is fused with sgRNA via MS2, which can function in the large region –N50 to +N50. Blue represents the PAM, red represents the mutation, and green represents the new strand synthesized by PolI3M.

Figure 2: CDE for trait engineering in eukaryotes

i. sgRNAs are designed to target a particular domain or the entire locus. sgRNAs can also be designed to target a particular hotspot if the function of a protein encoded nearby is known. ii. The CRISPR/Cas system can be delivered via plasmids. A single sgRNA cloned in the plasmid to target a single site. To target multiple sites at a time polycistronic tRNA-gRNA (PTG) system can be utilized, where up to eight sgRNAs are cloned under *U3* promoter. In an alternative approach, sgRNAs are in vitro transcribed (IVT), complexed with CRISPR/Cas proteins, and delivered as ribonucleoprotein complexes (RNPs). iii. The plasmids/RNPs are transformed in eukaryotic cells.

200 In plants, callus transformation can be performed via *Agrobacterium*-mediated transformation or
201 biolistics using a gene gun. iv. The cells/tissues are grown under selection pressure. In plants,
202 selection and regeneration can be performed under selection pressure. v. The resistant cells are
203 genotyped and further analyzed.

A. Mechanism of EvolvR

i. Target is nicked

ii. nCas9 dissociates

iii. DNA Polymerase associates

iv. Error-prone strand synthesis & old strand displacement

v. Old strand cleaved by flap endonuclease domain

vi. Poll3M dissociates from ligatable nick

B. Cytosine base editors with their editing window

BE3 fused to nCas9 converts C to T

sgRNA containing MS2 recruit AIDΔ* converts C to T