Creation of Hypoallergenic Mustard (*Brassica juncea*) through Genome Editing and Development of Precise Base Editing Tools for Plants

Von der Naturwissenschaftlichen Fakultät

der Gottfried Wilhelm Leibniz Universität Hannover

Zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Dingbo Zhang, Master of Engineering,

University of Chinese Academy of Science

2023

Referent: Prof. Dr. rer. nat. Jens Boch

Korreferentin: Prof. Dr. rer. hort. Traud Winkelmann

Tag der Promotion: 31.08.2023

Abstract

Improved agricultural safety through novel breeding techniques is urgently required to increase access to nutritious foods worldwide. Genome engineering using clustered regularly interspaced short palindromic repeats (CRISPR)-based or transcription activator-like effector (TALE)-based technologies provides a unique ability to modify targeted genes for precise breeding. This technology shows promise in various applications of allergy research. The major allergen Bra j I from brown mustard (*Brassica juncea*) is a seed storage protein that belongs to the 2S albumin family. One aim of this thesis was to create a hypoallergenic variety of mustard by utilizing genome editing techniques and a second aim was to develop novel base editing tools for plants.

Firstly, two CRISPR/Cas9 constructs with multiplex single guide RNAs were employed to induce large deletions or frameshift mutations in both *Bra j IA* and *Bra j IB* homoeologs in two brown mustard lines (Terratop and CR2664). High mutation efficiencies were observed in the T_0 transgenic mustard plants. The *Bra j IB* allele exhibited large deletions ranging from 566 to 790 bp in four lines. Additionally, nine out of 18 Terratop T_0 lines exhibited small indels in the targeted regions. Similarly, 14 out of 16 CR2664 T_0 lines analyzed had indels, while three lines exhibited mutations in all four *Bra j I* alleles. The mutations were stably inherited to the T_1 progeny. Moreover, immunoblotting results demonstrated a decrease or complete absence of the Bra j I protein in the seed extracts of selected T_1 lines. This work highlights the value of genome editing technologies in creating hypoallergenic food plants.

Secondly, two base editing tools: TALE-derived DddA-based cytosine base editors (TALE-DdCBEs) and TALE-derived adenine base editors (TALE-ABEs) were developed for precise C•G-to-T•A and A•T-to-G•C editing, respectively. TALE-DdCBEs containing evolved DddA variants (DddA6 or DddA11) showed a significant improvement in editing efficiency *in Nicotiana benthamiana* and rice protoplasts. TALE-DdCBEs containing DddA11 exhibited broader sequence compatibility for editing non-TC targets. Furthermore, a series of TALE-ABEs with different deaminase fusion architectures were tested in *N. benthamiana* and rice. The results showed that TALE-ABEs enable the conversion of A•T-to-G•C in rice protoplast. The application of TALE-base editors can result in a dramatic change because they can be deployed for nuclear genes or, alternatively, target the genomes of plastids or mitochondria by N-terminal targeting sequences.

Key words: genome editing, CRISPR/Cas, deaminase, TALE, *Brassica juncea*, base editors, rice.

Zusammenfassung

Eine verbesserte landwirtschaftliche Sicherheit durch neue Züchtungsverfahren ist dringend erforderlich, um den Zugang zu nahrhaften Lebensmitteln weltweit zu verbessern. Das Genom-Engineering mit Hilfe von CRISPR (clustered regularly interspaced short palindromic repeats)-Technologien oder TALE (transcription activator-like effector)-Technologien bietet die einzigartige Möglichkeit, gezielt Gene für eine präzise Züchtung zu verändern. Diese Technologie ist vielversprechend für verschiedene Anwendungen in der Allergieforschung. Das Hauptallergen Bra j I aus braunem Senf (*Brassica juncea*) ist ein Saatgut-Speicherprotein, das zur 2S-Albuminfamilie gehört. Ein Ziel dieser Arbeit war die Schaffung einer hypoallergenen Senfsorte durch den Einsatz von Genome Editing-Techniken und ein zweites Ziel war die Entwicklung neuer Baseneditoren für Pflanzen.

Zunächst wurden zwei CRISPR/Cas9-Konstrukte mit Multiplex-Single-Guide-RNAs eingesetzt, um große Deletionen oder Frameshift-Mutationen in den beiden Homöologen *Bra j IA* und *Bra j IB* in zwei Linien des braunen Senf (Terratop und CR2664) zu induzieren. In den transgenen T0-Senfpflanzen wurden hohe Mutationseffizienzen beobachtet. Das *Bra j IB*-Allel wies in vier Linien große Deletionen zwischen 566 und 790 bp auf. Außerdem wiesen neun von 18 Terratop-T0-Linien kleine Indels in den Zielregionen auf. In ähnlicher Weise wiesen 14 der 16 analysierten CR2664 T0-Linien Indels auf, während drei Linien Mutationen in allen vier *Bra j I*-Allelen aufwiesen. Die Mutationen wurden stabil an die T1-Nachkommen vererbt. Darüber hinaus zeigten Immunoblotting-Ergebnisse eine Abnahme oder ein vollständiges Fehlen des Bra j I-Proteins in den Samenextrakten ausgewählter T1-Linien. Diese Arbeit unterstreicht den Wert von Genom Editing Technologien für die Schaffung hypoallergener Lebensmittelpflanzen.

Zweitens wurden zwei Baseneditoren entwickelt: TALE-abgeleitete DddA-basierte Cytosin-Baseneditoren (TALE-DdCBEs) und TALE-abgeleitete Adenin-Baseneditoren (TALE-ABEs). Sie wurden entwickelt, um eine präzise Bearbeitung von C•G-to-T•A bzw. A•T-to-G•C zu erstellen. TALE-DdCBEs, die DddA-Varianten (DddA6 oder DddA11) enthielten, zeigten eine deutliche Verbesserung der Editierungseffizienz sowohl in *Nicotiana benthamiana* als auch in Reisprotoplasten. TALE-DdCBEs mit DddA11 wiesen eine bessere Sequenzkompatibilität für die Bearbeitung von Nicht-TC-Zielen auf. Darüber hinaus wurden verschiedene TALE-ABEs mit unterschiedlichen Desaminase-Fusionsarchitekturen in Reis und *N. benthamiana* getestet. Die Ergebnisse zeigten, dass TALE-ABEs A•T-to-G•C Umwandlungen im Reisprotoplasten ermöglichen. TALE-Base-Editoren können für nukleare Gene eingesetzt werden oder alternativ über N-terminale Targeting-Sequenzen die Genome von Plastiden oder Mitochondrien zum Ziel haben.

Schlüsselworte: genome editing, CRISPR/Cas, deaminase, TALE, *Brassica juncea*, base editors, rice.

Contents

1 General Introduction

Crops play a crucial role in human civilization by providing essential resources such as food, fodder, fuel, and other consumable resources. Global climate change, water shortage, limited arable land, and population growth are all raising serious problems for food security. Food allergy have posed significant challenges for the safety and health of many individuals. It is reported that the global prevalence of people allergic to various foods is around 5% among adults and 8% among children (Sicherer and Sampson, 2014). The development of genome editing techniques has enabled the introduction of precise and predictable genome modifications into plants, resulting in the acquisition of desirable traits. These techniques have paved the way for precision breeding methods, which are shaping the future of plant breeding (Gao, 2021). This thesis focused on the development of precise genome editing tools and the creation of hypoallergenic brown mustard (*Brassica juncea*) by modifying the allergen gene *Bra j I*.

1.1 Food allergy

An allergy is commonly defined as an adverse immune-mediated hypersensitivity to normally innocuous environmental substances known as allergens (Matsuo *et al*., 2015). Food allergies are characterized as a hypersensitive response to foods that is primarily mediated by an immune-globulin E (IgE) mechanism (Gould and Wu, 2018). In recent decades, the prevalence of food allergy has increased and is now considered as a significant public health issue in developed countries (Renz *et al*., 2018). It has been reported that approximately 5-8% of children under the age of three and 1-2% of adults suffer from a food allergy (Kagan, 2003). The typical symptoms of food allergy include disturbances to the skin, respiratory tract, and gastrointestinal tract, as well as cardiovascular aberrations. In the most extreme cases, immediate hypersensitivity to food can cause life-threatening hypovolemic shock (Figure 1) (Renz *et al*., 2018). A food allergy is typically developed during the initial two years of life, and sensitization takes place through the gastrointestinal tract (Lee and Burks, 2006). Moreover, exposure to an inhalant allergen such as pollen may lead to the generation of IgE antibodies

that have the potential to cross-react with proteins present in various fruits and vegetables (Vieths, 1997).

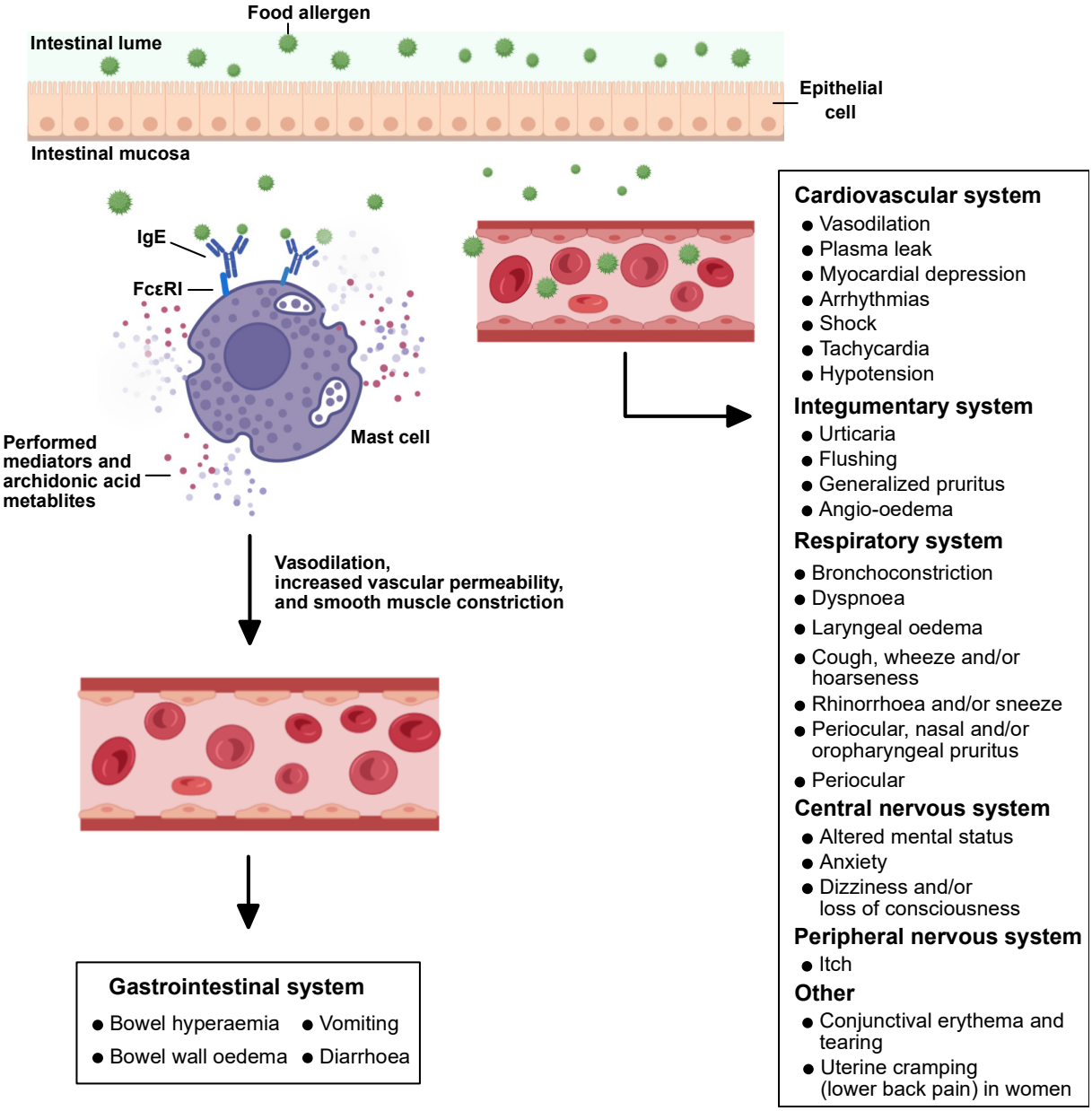


Figure 1. Food allergen responses mediated by IgE. Food allergens enter the intestinal epithelium through diverse mechanisms, such as transportation via epithelial cells, permeation through intercellular gaps, and absorption by microfold cells. These allergens subsequently come into contact with mast cells located in the mucosal layer. When IgE antibodies identify and attach to the allergen while bound to mast cells by the high-affinity immunoglobulin receptor FcεRI, receptor crosslinking occurs. This crosslinking induces the release of hypersensitive mediators and the stimulation of synthesis of arachidonic acid metabolites. The mediators facilitate the process of vasodilation and heightened vascular permeability, thereby leading to the swelling of the bowel wall. The smooth muscle contraction and mucus secretion also influenced by those mediators. When the allergen is dispersed throughout the body, the basophils circulating in the blood and the mast cells residing in the tissues, which have been sensitized by IgE, may interact, leading to the occurrence of anaphylaxis. IgE: immunoglobulin E. FcεRI: high-affinity immunoglobulin ε receptor. Figure modified from Renz *et al*., 2018.

Allergens originating from plants can be categorized based on their structure, function, biochemistry, and immunological features (Breiteneder and Radauer, 2004). The majority of allergens found in cereals and legumes are either pathogenesis-related (PR) proteins or seed storage proteins (Hoffmann-Sommergruber, 2002). PR-proteins are stimulated by pathogens or environmental factors, and categorized into 14 families (Breiteneder and Ebner, 2000). Numerous allergens found in plant-based foods are homologous to PR-proteins, such as chitinase (Pers a 1) from avocado (Sowka *et al*., 1998), thaumatin-like protein (Mal d2) from apple (Krebitz *et al*., 2003), and lipid transfer proteins (Pru p 3) from peach (Pastorello *et al*., 1999). Furthermore, a considerable number of allergenic proteins belong to prolamin, or cupin, or cysteine protease superfamily (Breiteneder and Radauer, 2004). There are three main groups of plant food allergens in the prolamin superfamily: the 2S albumins, lipid transfer proteins, and cereal alpha-amylase/trypsin inhibitors, which have related structures and are stable to thermal processing and proteolysis. They include major allergens found in Brazil nut, peanuts, peaches, and cereals like rice and wheat. The cupin superfamily comprises the major globulin storage proteins from a variety of plant species. The globulins have been found to be allergens in plant foods such as peanuts, soybean, and walnut. The cysteine protease superfamily contains the papain-like proteases found in microbes, plants, and animals. Those papain-like protease allergens can be found in kiwi fruit and soybean (Mills *et al*., 2004).

1.1.2 Bra j I — a major allergen from *Brassica juncea*

Brassicaceae is a diversified plant family that produces one of the most vast and diverse ranges of food products from a single plant genus. In Sanskrit and Sumerian documents from as early as 3,000 BC, mustard is described as a condiment, making *Brassica juncea* one of the first domesticated plants (Rahman *et al*., 2018). The study of plant genetics, along with findings from archaeology and written records, suggests that *B. juncea* likely originated from a single ancestor in West Asia between 8,000 and 14,000 years ago. There were at least three separate instances of domestication that occurred in the last 500 to 5,000 years, resulting in the development of different types of mustard (Kang et al., 2021b). Seed mustard was

developed in Central Asia, oilseed mustard in the Indian subcontinent, and root mustard in East Asia (Vaughan *et al*., 1963; Chen *et al*., 2013).

The *Brassica* species that have been cultivated comprise a vast array of vegetable and oil crops. *Brassica juncea* (AABB) with a genome size of approximately 920 Mb, is recognized as mustard and is a naturally occurring allopolyploid that results from the fusion of two diploid species: *Brassica rapa* (AA) and *Brassica nigra* (BB) (Figure 2) (He *et al*., 2021).

Mustard is a member of the *Brassicaceae* family and comprises two distinct species: *Sinapis alba* (yellow mustard) and *Brassica juncea* (oriental mustard). Mustard is considered to be the fourth most important dietary allergy for children, after eggs, peanuts, and cow milk (Morisset *et al*., 2003). Morisset *et al*. revealed that based on positive oral food challenges, 23.3% of individuals with positive skin prick tests are allergic to mustard. The mustard allergy typically occurs during early childhood, with a majority of affected children exhibiting symptoms before the age of three. Specifically, 8 out of 15 children with an allergy to mustard were found to have developed the condition before the age of three (Rancé, 2003). Moreover, mustard is regarded as a probable hidden allergen in the usual infant diet. This increased frequency of incidence at such young ages might be explained by sensitization *in utero,* and as well as the presence of mustard in infant foods (Rancé *et al*., 2000). Mustard is a common ingredient found in commercially available food products intended for consumption of infants and exposure to mustard is a frequent occurrence in infancy. According to the report, three infants between the ages of 12 and 18 months, who were breastfed until 11 months and had no prior exposure to mustard, exhibited a positive skin prick test for mustard (Niinimäi *et al*., 1989).

The major allergen of yellow mustard, named Sin a 1, is a thermostable protein that is resistant to digestion by trypsin and degradation by other proteolytic enzymes (Monsalve *et al*., 1993; De La Peña *et al*., 1996); Bra j I (Figure 3), the main allergen in oriental mustard, was shown to be structurally similar to Sin a 1 (Caballero *et al*., 1994). The structure prediction indicates that Bra j I comprises five α-helices (residues 7-16, 20-35, 47-58, 61-85, and 88-111) and four β-sheets (residues 36-40, 43-46, 118-121, and 124-127).

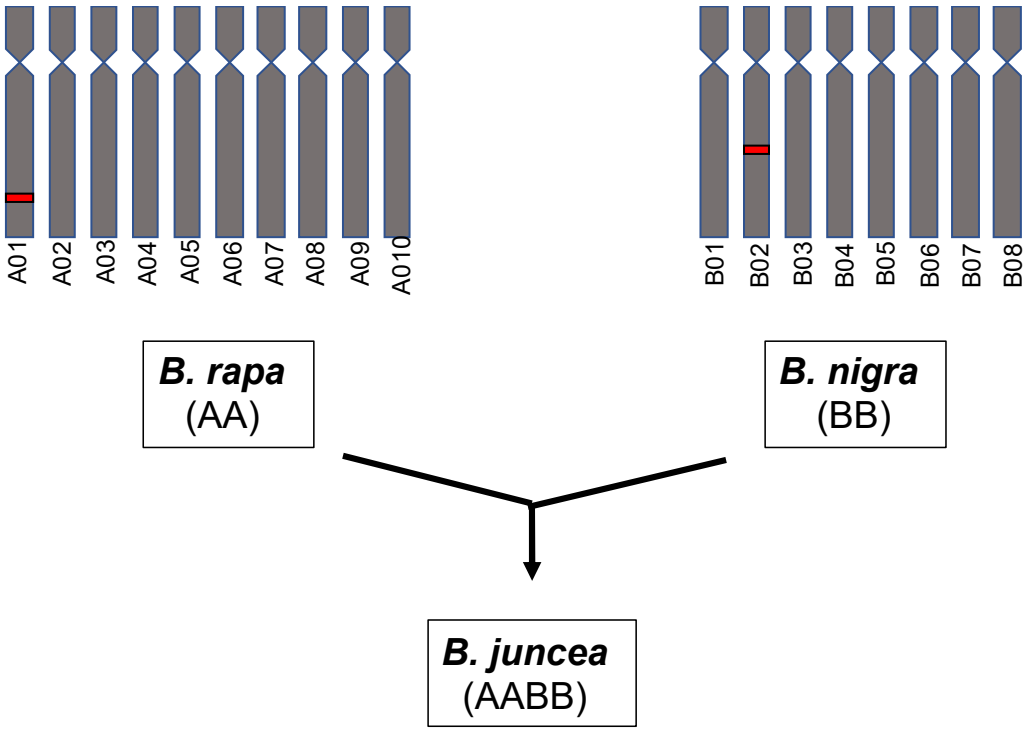


Figure 2. The hybridization between *B. rapa* (AA) and *B. nigra* (BB). *B. juncea* (AABB) formed by the hybridization between the diploid ancestors of *B. rapa* (AA) and *B. nigra* (BB), followed by spontaneous chromosome doubling. The localization of two Bra j I genes are indicated in red.

The Bra j I protein is a type of seed storage protein belonging to the 2S albumin family, and it has a molecular weight (MW) of approximately 16 kDa (Monsalve *et al*., 1993). The 2S albumins are a group of proteins that are classified as storage proteins in various dicotyledonous plant species, and are members of the prolamin superfamily (Shewry *et al*., 1995). The 2S albumins are small globular proteins rich in arginine, glutamine, asparagine, and often cysteine. These proteins are subjected to modifications after their synthesis. The majority of 2S albumins are split into a large and small subunit held together by a disulfide bond (Breiteneder and Ebner, 2000). Apart from Bra j I, several of the tree nut and seed allergens are 2S albumins, including Ber e 1 from Brazil nut (*Bertholletia excelsior*) (Nordlee *et al*., 1996), Jug r 1 from the English walnut (*Juglans regia*) (Teuber *et al*., 1999), and BnIII from oilseed rape (*Brassica napus*) (Monsalve *et al*., 1997).

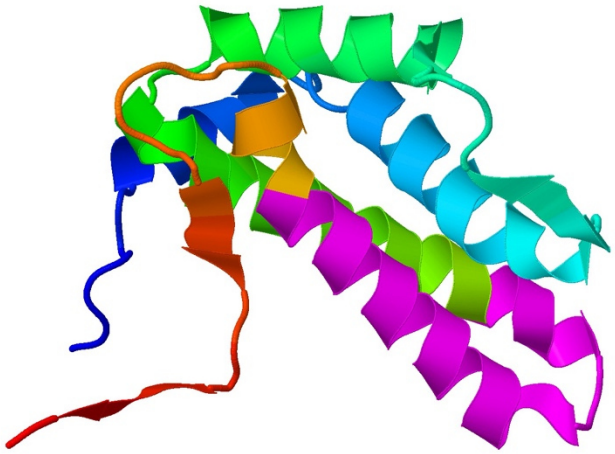


Figure 3. Structure of Bra j I. Predicted structure of Bra j I by AlphaFold (Jumper *et al*., 2021), UniProt accession number: P80207. Anti-Bra j I antibody against region is indicated in purple. The anti-Bra j I antibody was used for the detection of Bra j I from the mustard seed extracts in Chapter 2.

Until now, the best therapy for sensitive individuals to mustard would be to avoid all meals containing mustard. However, mustard is frequently used as a spice, particularly in mixtures such as curry, as well as a condiment and ingredient to season prepared foods, and, therefore, difficult to avoid in general foods. Hence, it is imperative to develop a mustard variant that does not elicit an allergic response.

1.2 Plant genome editing by sequence-specific nucleases

Genome editing by sequence-specific nucleases (SSNs) has revolutionized biology by allowing for targeted genomic alterations. SSNs including zinc finger nuclease (ZFN) (Figure 4A), transcription activator-like effector nucleases (TALENs) (Figure 4B), and the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated) system (Figure 4C) have been progressing at an unprecedented rate in gene therapy (Porteus, 2016) and crop breeding (Gao, 2021).

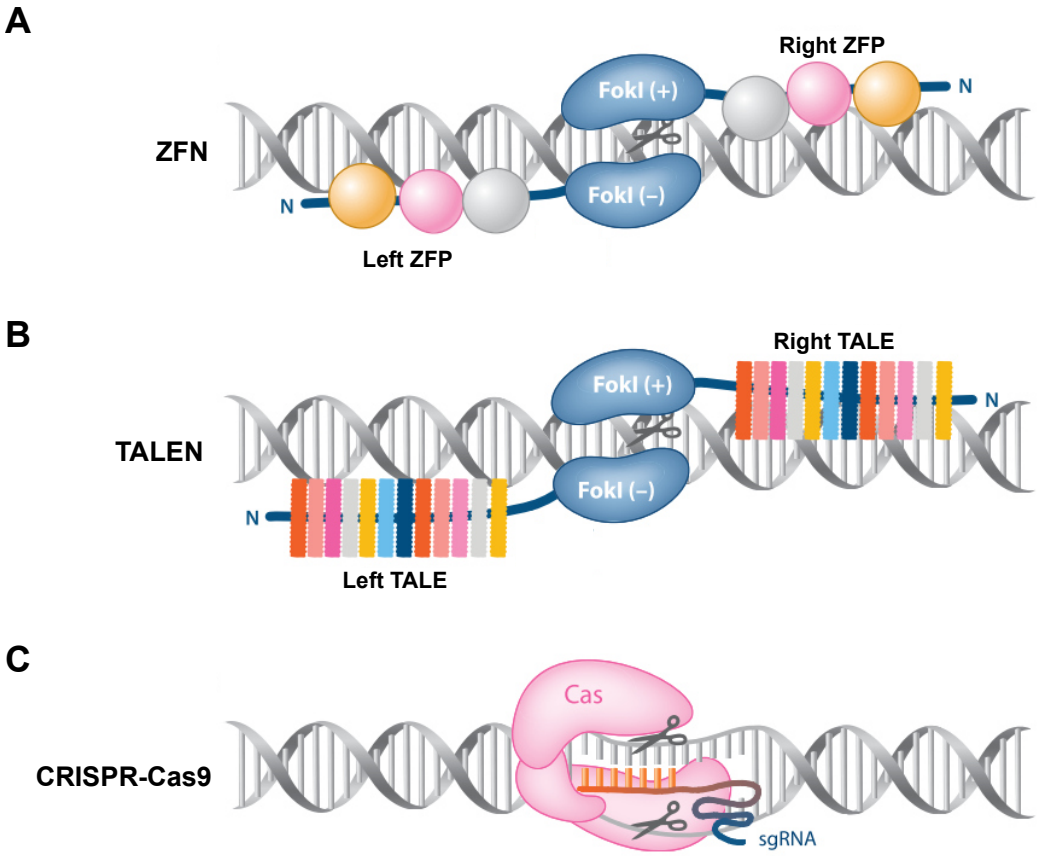


Figure 4. Schematic representation of three classes of programmable sequence-specific nucleases. (A) Representation of zinc finger nuclease (ZFN) pairs bound to DNA. Each ZFN is composed of the catalytic domain of *Fok*I linked to an array of three to six zinc finger that specifically recognize target sequences. The length of the spacer between two ZFNs binding sites is typically five to seven bases. **(B)** Illustration of a TALEN pair bound to DNA. Each TALEN consists of transcription-activator-like effector (TALE) DNA-binding domain and the *Fok*I catalytic domain. Each TALE domain contains 33 to 35 amino acid repeats, with one repeat recognizing one single DNA base. The amino acids at positions 12 and 13 confer DNA-binding specificity and are called repeat variable di-residue (RVD). **(C)** Schematic representation of CRISPR-Cas9 nuclease bound to a single guide RNA (sgRNA) complementary to a 20-bp target DNA sequence. Base pairing between a DNA sequence and sgRNA after protospacer adjacent motif (PAM) recognition allows DNA cleavage at the target site by the Cas9 nuclease domains (Langner *et al*., 2018).

The underlying principle of classical plant genome editing is the induction of DNA double-strand breaks (DSBs) by SSNs. The DSBs are subsequently repaired by by two major mechanisms: nonhomologous end-joining (NHEJ) and homology-directed repair (HDR) (Voytas, 2013; Puchta, 2017) and generated random or defined genetic outcomes.

1.2.1 Zinc finger nuclease

An essential requirement for gene targeting is the capacity to produce nucleases that are highly precise in targeting distinct genomic loci. ZFNs are composed of DNA recognition modules that consist of an array of Cys2–His2 zinc fingers and the catalytic domain of the type IIS restriction endonuclease *Fok*I (Figure 4A) (Urnov *et al*., 2005). The Cys2–His2 zinc finger was first reported in *Xenopus oocytes* (Miller *et al*., 1985), and each zinc finger recognizes a triplet of nucleotides (5′-G/ANN triplet, and some zinc fingers can also recognize 5′-C/TNN triplets, where N can be any nucleotide) via a single α-helix (Pavletich and Pabo, 1991; Wright *et al*., 2005). Multiple zinc fingers can be linked in tandem to form DNA-binding arrays that recognize a broad spectrum of DNA sequences with high specificity (Choo *et al*., 1994). Following the selection of a target site, the corresponding ZFNs are generated through combining the zinc-finger domains that bind to the particular set of triplets located at the target site. In principle, *Fok*I functions as a dimer. The fusion of each *Fok*I monomer with a distinct zinc-finger array that recognizes a unique DNA sequence enables the inactive monomers to form a functional enzyme only when the two recognition sites are in close proximity (Bibikova *et al*., 2001; Porteus and Baltimore, 2003). ZFNs have been employed for gene targeting through the induction of DSBs in a variety of plant species, such as tobacco (Wright *et al*., 2005), *Arabidopsis* (Lloyd *et al*., 2005), and *Zea mays* (Shukla *et al*., 2009).

1.2.2 Transcription activator-like effector nucleases

Transcription activator-like effectors (TALEs) are bacterial type III effector proteins derived from plant pathogenic *Xanthomonas* bacteria (Boch and Bonas, 2010). *Xanthomonas* bacteria employ a type III secretion system to deliver TALEs into plant cell. Once they enter the nucleus, those TALEs stimulate the transcription of specific genes to promote virulence (Van den Ackerveken *et al*., 1996; Boch and Bonas, 2010; Boch *et al*., 2014). TALE proteins harbor three

functional domains. The TALE N-terminal domain contains the bacterial secretion signal and a non-specific DNA-binding activity which is essential for the affinity of protein to DNA (Szurek *et al*., 2002; Kay *et al*., 2007; Gao *et al*., 2012). The TALE central tandem repeat domain consists of a variable number of tandem repeats, typically ranging from 33 to 35 amino acid in length. These repeats are crucial for the specific and programmable binding of DNA by TALEs, and the two amino acids residues known as repeat variable di-residue (RVD) located at position 12 and 13 of each repeat define the DNA binding specificity of a TALE (Boch *et al*., 2009; Moscou and Bogdanove, 2009). The specificities of all possible RVD combinations have been decoded. Among them HD, NG, NI, and NN are the most abundant naturally occurring RVDs, which specific binding to cytosine, thymine, adenine, and guanine/adenine, respectively (Boch *et al*., 2009; Yang *et al*., 2014; Miller *et al.*, 2015). The TALEs DNA binding specificities can be purposely altered by rearranging the repeats. The TALE C-terminal domain contains a transcription factor binding site, two nuclear localization signals, and an activation domain (Van den Ackerveken *et al*., 1996; Zhu *et al*., 1998). In vitro studies have indicated that TALE proteins adopt a loosely wrapped conformation around DNA templates during nonspecific search (Cuculis *et al*., 2016). Cuculis *et al*. employed single-molecule methodologies to observe of the dynamics of TALE search process along DNA templates (Cuculis *et al*., 2015). The results revealed a two-state model of TALE protein diffusion along DNA. TALE proteins undergo a conformational upon binding to the target DNA, transitioning from a "search" mode to a "bind" mode. In the "bind" mode, the TALE protein establishes robust interactions with the major groove of the DNA template through non-covalent bonds along the phosphate backbone. Moreover, the superhelical pitch of the TALE protein closely resembles that of B-form DNA. On the other hand, in the "search" mode, the protein adopts a more relaxed conformation with a larger superhelical pitch, similar to that of unbound TALE proteins (Cuculis *et al*., 2015; Cuculis *et al*., 2016).

Similar to ZFNs, TALENs are based on fusion of a TALE protein with the catalytic domain of the *Fok*I endonuclease (Figure 4B). Therefore, the utilization of TALENs enables the introduction of DSBs at specific genomic sites *(Huang et al., 2011; Cade et al., 2012; Miller et al*., 2015). TALENs have been utilized for crop improvement in various way. For instance,

TALENs have been used in the development of *Xanthomonas*-resistant rice (Li *et al*., 2012), powdery mildew-resistant wheat (Wang *et al*., 2014), high-quality lignocellulosic biomass sugarcane (Jung and Altpeter, 2016), and high oleic acid soybean, which is now available in the market as a genome-edited crop (Haun *et al*., 2014).

1.2.3 CRISPR/Cas systems for plant genome editing

The advent of genome-editing technologies has revolutionized plant research by enabling genetic modifications and modulating the function of DNA sequences in their endogenous genomic context in various organisms (Langner *et al*., 2018). Since the CRISPR/Cas9 system has been successfully implemented in rice, wheat (Shan *et al*., 2013), *Nicotiana benthamiana* (Nekrasov *et al*., 2013), and *Arabidopsis thaliana* (Li *et al*., 2013) in 2013, continuous improvements in CRISPR/Cas systems have resulted in the widespread adoption of genome editing as a cost-effective, user-friendly method for targeted genetic manipulation in crop improvement.

1.2.3.1 CRISPR/Cas systems

Beginning with a study of repetitive DNA sequences in a bacterial genome in 1987 *(Ishino et al*., 1987), the mystery of clustered regularly interspaced short palindromic repeats (CRISPRs) commonly found in microbial genomes together with genes encoding CRISPR-associated (Cas) proteins were uncovered by researchers. The CRISPR/Cas system is an RNA-mediated adaptive immunity system in bacteria and archaea that defends against phages and other invasive genetic elements by cleaving the invader's nucleic acid genome (Knott and Doudna, 2018). The CRISPR-Cas systems exhibit significant variation in protein composition, effector complex structure, genome locus architecture, and mechanisms of adaptation, CRISPR RNA (crRNA) processing, and interference. On the basis of such characteristics, the CRISPR/Cas systems are classified into two classes, which are further divided into six types depending on their typical Cas genes (Table 1). Class 1 CRISPR/Cas systems (types I, III, and IV) contain multi-subunit effector complexes, whereas class 2 CRISPR/Cas systems (types II, V, and VI) harbor single-protein effector modules (Koonin *et al*., 2017).

The *Streptococcus pyogenes* type II CRISPR/Cas9 system was the first to be proven to precisely cut DNA in vitro and in vivo (Jinek *et al*., 2012; Gasiunas *et al*., 2012; Cong *et al*., 2013). Three fundamental components are required for CRISPR-Cas9 genome editing: the CRISPR RNA (crRNA), trans-activating crRNA (tracrRNA), and the Cas9 nuclease. Generally, crRNA and tracrRNA can be combined to form a single guide RNA (sgRNA), which reducing the Cas9 programmable nuclease to only two components. The 5′-end of the sgRNA contains a sequence of 20 nucleotides that directs the Cas9-sgRNA complex to a specific target DNA site by Watson-Crick base-pairing (Figure 4C). The Cas9-sgRNA complex probes the DNA double helix for the canonical 5′-NGG protospacer-adjacent motifs (PAMs) and for potential guide RNA complementarity in the PAM flanking sequence. Base pairing of complementary nucleotides at the 8–12 bp seed region enables the formation of the sgRNA-DNA heteroduplex, followed by DNA cleavage by the Cas9 nuclease (Sternberg *et al*., 2014). RuvC and HNH are two distinctive nuclease domains within the Cas9 proteins. The RuvC domain cuts the complementary DNA strand, while the HNH domain cuts the noncomplementary DNA strand (Jinek *et al*., 2012). The mutation of catalytic residues in either domain results in the formation of a DNA nickase, known as nCas9 (D10A or H840A).

Table 1. Characteristics of CRISPR Cas nucleases

Nucleases	Cas9	Cas12	Cas13	Cascade-Cas3	Cas7-11
Type	type II	type V	type VI	type I	type III
Nuclease domain	RuvC and HNH	RuvC	HEPN	HD	RRM
Guide RNA	crRNA and tracrRNA	crRNA	crRNA	crRNA	crRNA
Substrate	dsDNA	dsDNA*	RNA	dsDNA	RNA
Cleavage pattern	mostly blunt end	sticky end	depends on local target sequence	nicks and degrades NTS DNA, then induces DSB	cleaves at the 3′-end of the binding site
Indel type	small indels	small indels	degraded RNA	large deletion	degraded RNA

*: mostly target dsDNA

It is noteworthy that inactivation of both domains produces a deactivated Cas9 protein (dCas9, D10A and H840A). Despite its inactivity, dCas9 retains the ability to selectively target specific genomic loci and act as a framework for recruiting effector proteins (Qi *et al*., 2013). Other bacteria-derived Cas9 enzymes, including those from *Neisseria meningitides* (NmCas9, requires a 5′-NNGGAA or 5′-NNGCTT PAM), *Staphylococcus aureus* (SaCas9, requires a 5′-NNGRRT PAM, with R representing A or G), and *Streptococcus thermophilus* (StCas9, requires a 5′-NNGGAA PAM), have also been created as tools for genome editing (Makarova *et al*., 2020). To expand the range of targetable genomic loci, Cas9 has been engineered to recognize different PAMs, such as VQR-Cas9 (requires a 5′-NGA PAM), EQR-Cas9 (requires a 5′-NGAG PAM), VRER- Cas9 (requires a 5′-NGCG PAM), SaKKH-Cas9 (requires a 5′-NNNRRT PAM) (Kleinstiver *et al*., 2015), xCas9 (requires a 5′-NG, or GAA, or GTA PAM) (Hu *et al*., 2018), and SpCas9-NG (requires a 5′-NG PAM) (Nishimasu *et al*., 2018). Furthermore, in contrast to the SpCas9 and StCas9 enzymes, the NmCas9 enzyme exhibits binding affinity towards a 24-nucleotide protospacer sequence on its intended DNA target, thereby imparting enhanced specificity beyond the previous 20-nucleotide protospacer of SpCas9 and StCas9 (Hou *et al*., 2013).

Similar to the CRISPR type II systems, the class 2 type V-A Cas enzyme Cas12a (also known as Cpf1) from *Francisella novicida* (FnCas12a), *Acidaminococcus sp.* BV3L6 (AsCas12a), and *Lachnospiraceae bacterium* (LbCas12a), *Eubacterium rectale* (ErCas12a, also named as MAD7), *Moraxella bovoculi* (Mb2Cas12a) have been developed to achieve targeted DNA modification (Zetsche *et al*., 2015; Zhang *et al*., 2021; Lin *et al*., 2021). Cas12a uses a T-rich PAM (5′-TTTN) sequence for target DNA recognition, and is guided by a 42-nt crRNA to induce DNA double-strand breaks that generates a sticky end (4- or 5-nt overhangs). Furthermore, it should be noted that Cas12a exhibits not only the ability to cleave DNA, but also possesses RNase III activity which facilitates the processing of pre-crRNA (crRNA that has not yet undergone processing). The endoribonuclease and endodeoxyribonuclease capabilities of Cas12a can be utilized to achieve multiplexing of diverse targets through the use of pre-crRNA expressing constructs that are arranged in tandem (Zetsche *et al*., 2017). To address the 5′-TTTN PAM limitations of Cas12a, Cas12a variants that can recognize various PAMs have been

developed, such as AsCpf1-RR (requires a 5′-TYCV PAM), AsCpf1-RVR (requires a 5′-TATVPAM), LbCpf1-RR (requires a 5′-CCCC or 5′-TYCV PAM), LbCpf1-RVR (requires a 5′-TATG PAM), and Mb2Cas12a-RVRR (can edit at 5′-TTTV, VTTV, TATV, TYCV, CCCV, and CTCV PAM sites) (Gao et al., 2017b; Li et al., 2018b; Zhang et al., 2021).

In contrast to Cas9 or Cas12a, which possess the ability to target DNA, Cas13 nucleases (formerly C2c2) from the class 2 type VI CRISPR system are guided by a single crRNA and enable precise editing of RNA via the HEPN domains (Abudayyeh *et al*., 2016). The utilization of Cas13-mediated RNA targeting presents a viable approach for RNA knockdown in RNA studies (Abudayyeh *et al*., 2017).

1.2.3.2 CRISPR-induced DNA double-strand breaks for genome editing

The most commonly utilized method for genome editing is the employment of SSNs to induce a site-specific DSB at the desired target locus, which is subsequently repaired by the endogenous repair mechanism (Figure 5A). DSBs are repaired by two major mechanisms: NHEJ and HDR (Figure 5B). NHEJ is the major pathway used to repair DSBs, and when DSBs are repaired by NHEJ, indels (insertions and deletions) may be introduced at the junctions of the rejoined chromosomes (Chen *et al*., 2019a). NHEJ processes are typically efficient in most mammalian and plant cells and do not require a homologous repair template, while the HDR process is typically present in cells that are actively dividing and relies on proteins that are mainly produced during the S and G2 phases of the cell cycle (Heyer *et al*., 2010; Lieber, 2010; Lin *et al*., 2014).

The CRISPR-Cas nucleases are frequently employed to effectively and specifically disrupt genes of interest. Generally, CRISPR-Cas nucleases induce DSBs are repaired by the NHEJ processes, and resulting in the insertion, deletion, or substitution of nucleotides around the DSB site that might avoid subsequent recognition and re-cutting by the Cas nuclease. Controlling the indels that occur from DSBs is challenging, but it can be predicted by using a machine-learning model (Shen *et al*., 2018; Chen *et al*., 2019b). Particularly when the break site is situated within a micro-homologous region (4-25 bp), this can trigger microhomology-mediated end joining (MMEJ) (Bosshard *et al*., 2019). The application of MMEJ-mediated

deletions provides a viable method for predicting and designing deletion fragments, making an effective technique for generating precise and predictable deletions in plant genome editing (Tan *et al*., 2020). When the Cas nucleases are directed towards open reading frames, the resulting indel products typically lead to frameshift mutations in coding sequences, which ultimately result in the loss of protein function. The Cas nucleases have the potential to disrupt *cis*-regulatory elements present in promoters (Korkmaz *et al*., 2016), as well as to characterize non-coding RNAs (Zhu *et al*., 2016). The 5'-untranslated regions of numerous eukaryotic mRNAs contain upstream open reading frames (uORFs), those uORFs function as translation regulators and finely adjusting the expression of protein encoded by primary ORF (van der Horst *et al*., 2020). The utilization of CRISPR-Cas9 for modifying the uORF of genes has been employed as a means of precisely adjusting the levels of protein expression in plants (Zhang et al., 2018b). Pairs of sgRNA that target adjacent regions of a chromosome sequence are generally used to delete larger segments (Cong *et al*., 2013). Besides, the type I Cas3 editing systems have also been exploited to create large deletions in mammalian cells (Dolan *et al*., 2019; Cameron *et al*., 2019). Moreover, the simultaneous introduction of two or more DSBs in a cell may result in chromosomal translocations or other additional rearrangements (inversions or deletions) (Huang and Puchta, 2019; Rönspies *et al*., 2022).

NHEJ is effective for extensive knockout research; however, it does not possess the accuracy needed for sophisticated genome engineering. While HDR-mediated genome editing, in theory, can be utilized to precisely incorporate particular point mutations and to insert or substitute desired sequences into the target DNA. HDR requires co-delivery of sgRNA, Cas nuclease, and a donor repair template with a sequence (double-stranded DNA or single-stranded DNA) homology to both borders of the cutting sites.

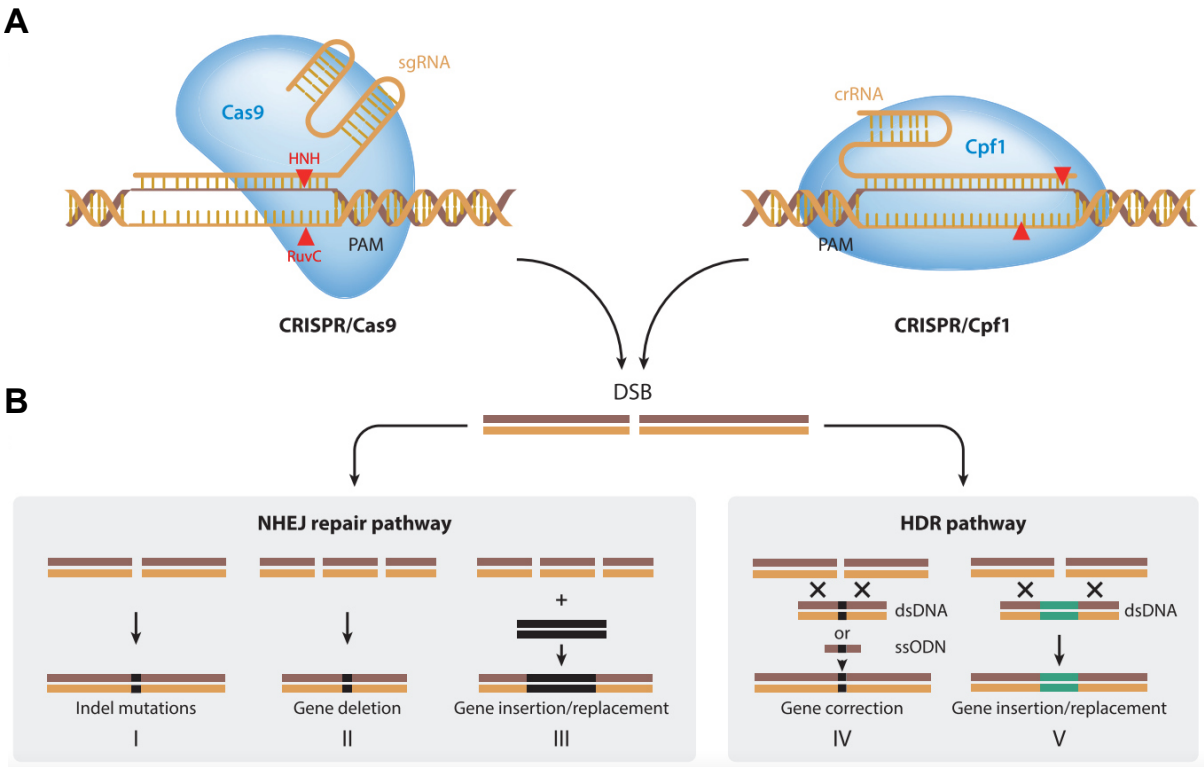


Figure 5. Schematic diagram of the NHEJ and HDR DNA repair pathways when DNA double-strand breaks (DSBs) are produced by sequence-specific nucleases (SSNs). (A) Two CRISPR/Cas systems Cas9 and Cpf1 used for generating DSBs. **(B)** Depending on the DSB repair pathways, genome editing with CRISPR/Cas systems can result in a variety of results: I, II, and III are outcomes of the dominant NHEJ repair pathway; IV and V are outcomes of the HDR pathway using an available DNA donor template. sgRNA: single guide RNA; crRNA: CRISPR RNA; DSB: double-strand break; dsDNA: double-strand DNA; HDR: homology-directed repair; NHEJ: nonhomologous end joining; PAM: protospacer-adjacent motif; ssODN: single-strand oligodeoxynucleotide (Chen *et al*., 2019a).

The ability to manipulate gene functions and combine multiple crop traits has been made possible through the precise insertion of DNA. However, it should be noted that the efficiency of HDR-mediated DNA insertion in plants is relatively low (Gao, 2021). The efficiency of HDR can be enhanced through the inhibition of proteins that facilitate nonhomologous end-joining (Yeh *et al*., 2019) or expressed proteins that can stimulate HDR (Rees *et al*., 2019; Nambiar *et al*., 2019). Geminivirus replicons based on bean yellow dwarf virus and wheat dwarf virus have been developed to increase the copy number of the repair donor template, which can boost the probability of HDR-mediated gene insertion in tomato and wheat (Čermák *et al*., 2015; Gil-Humanes *et al*., 2017).

1.3 Genome editing beyond DNA double-strand breaks

In addition to genome editing through DSBs, there exist base-editing systems that can facilitate targeted nucleotide substitutions without relying on HDR or donor DNA, and without the occurrence of DSBs. These systems offer a straightforward, efficient, and versatile approach to engineering nucleotide substitutions at desired sites. There are two main classes of base editors: cytosine base editors (CBEs) which catalyze the conversion of C•G base pairs to T•A base pairs (Figure 6A), and adenine base editors (ABEs) which catalyze A•T-to-G•C conversions (Figure 6B) (Anzalone *et al*., 2020).

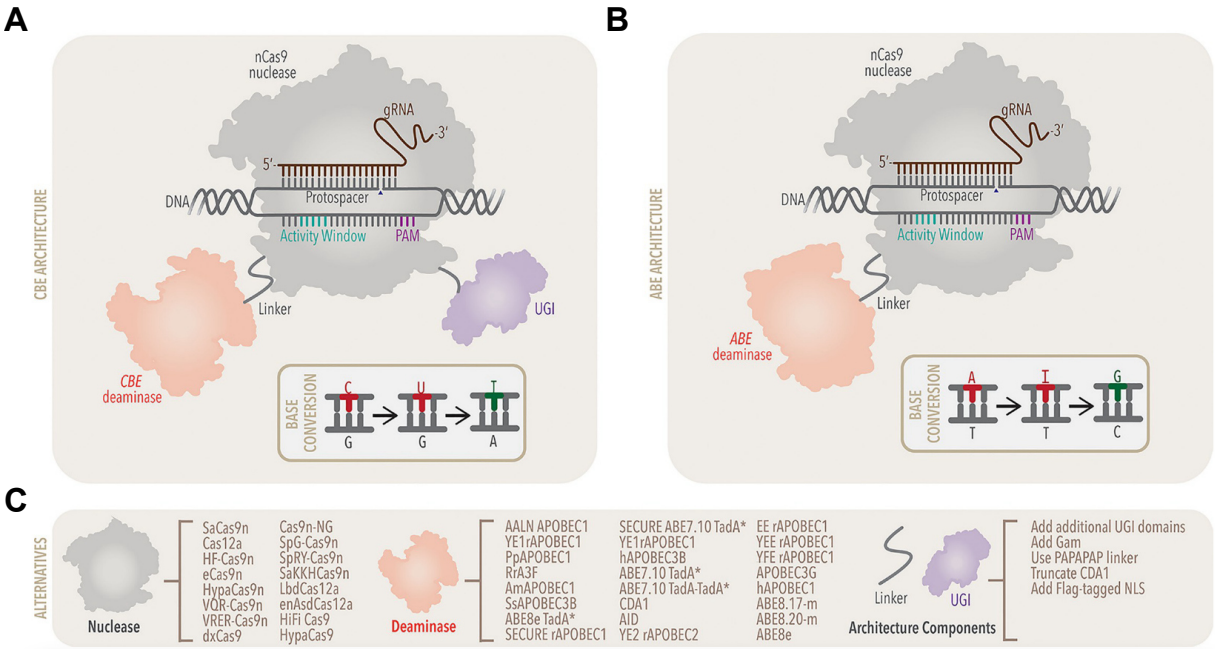


Figure 6. Overview of cytosine base editors and adenine base editors. (A) CBEs install C•G-to-T•A point mutations using nCas9 fused to cytidine deaminases and UGI. Deamination produces a U•G intermediate, which is processed by the cell to produce an overall C•G to T•A conversion. **(B)** ABEs install A•T-to-G•C mutations using a fusion of nCas9 and evolved TadA deoxyadenosine deaminase. **(C)** A selection of significant variances on CBE and ABE principal components are shown. These replacement components serve a variety of functions, including higher on-target editing, reduced off-target editing, and relaxed PAM requirements for increased utility (Porto and Komor, 2023).

1.3.1 Cytosine base editors

The cytosine base editor systems are usually comprised of a single-stranded DNA (ssDNA)-specific cytidine deaminase fused with a catalytically impaired Cas nuclease (nCas9, D10A) and an uracil glycosylase inhibitor (UGI) (Komor *et al*., 2016). These cytidine deaminases

include activation-induced cytidine deaminase (AID), and various apolipoprotein B mRNA-editing enzyme, catalytic polypeptide (APOBEC) enzymes, and cytosine deaminase 1 sourced from sea lamprey (pmCDA1) (Porto and Komor, 2023). In both CBEs and ABEs, the catalytically impaired Cas nuclease domain facilitates the localization of a single-stranded DNA deaminase enzyme to a specific target sequence within the genome. Upon binding of Cas to the target DNA strand, the sgRNA spacer hybridizes with the target DNA strand, leading to the displacement of the genomic DNA strand containing the PAM sequence, and resulting in the formation of an R-loop. This R-loop creates a localized region of accessible ssDNA, which is the natural substrate of these cytidine deaminases. Cytidines within this ssDNA region are deaminated by cytidine deaminases, which produce a C•G to U•G conversion. The deamination windows differ from deaminase to deaminase (Molla *et al*., 2021). The existence of the U•G mismatch intermediate initiates the base excision repair (BER) mechanism in the cells, which eliminates the uracil and restores the intermediate to its initial C•G base pair or generates small indels (Kunz *et al*., 2009). To avoid this, the nCas9 generates a DNA nick on the strand opposite the uracil-containing strand, consequently, the native DNA repair mechanisms of the cell are influenced to prioritize the replacement of the nicked strand and utilize the strand containing uracil as a template, thereby further enhancing the editing efficiency. The utilization of cytosine base editors has been swiftly adopted by plant biologists across various plant species, such as *Arabidopsis* (Chen et al., 2017a), wheat (Zong *et al*., 2017), maize (Zong *et al*., 2017), rice (Zong *et al*., 2018), watermelon (Tian *et al*., 2018), tomato (Veillet et al., 2019a), potato (Veillet *et al*., 2019b), cotton (Qin *et al*., 2020), soybean *(Cai et al*., 2020), strawberry (Xing *et al*., 2020), rapeseed (Wu *et al*., 2020), apple (Malabarba *et al*., 2021), pear (Malabarba *et al*., 2021), *Physcomitrium patens* (Guyon-Debast *et al*., 2021), and poplar (Li et al., 2021a).

The use of CRISPR/Cas-based base editors has been successful in accurately and effectively introducing single nucleotide variants (SNVs) into the nuclear genome. However, adapting this technology for editing DNA in organelles such as mitochondria and plastids is not possible because there are no established methods for delivering sgRNAs to these organelles. However, protein-only genome editing systems based on DNA-binding proteins, such as ZFN and TALEN,

can also be used as organellar genome editors. Therefore, the development of mitochondrially targeted ZFN (mtZFN) and mitochondrially targeted TALEN (mitoTALEN) have been successfully employed for heteroplasmic manipulation in various models of pathogenic mitochondrial DNA (mtDNA) variants through the induction of DSBs in mtDNA (Bacman *et al*., 2013; Gammage *et al*., 2014). The mitoTALEN has also been used to create cytoplasmic male sterility in rice (Kazama *et al*., 2019), and introduce point mutations into the tobacco mitochondrial genome (Forner *et al*., 2022). Remarkably, the discovery of cytosine deaminase DddAtox from *Burkholderia cenocepacia*, which deaminates cytosines in dsDNA enabled the C•G to T•A conversions in human nuclear and mitochondria DNA (Mok *et al*., 2020). To generate TALE-derived DddA-based cytosine base editors (TALE-DdCBEs), the toxic $DddA_{tox}$ had to be split into two inactive halves: DddA-N and DddA-C (Figure 7). Similar to the *Fok*I monomer assembly in TALE nucleases, the two DddA halves reconstitute the active enzyme when assembled by two adjacent tail-to-tail TALE arrays. Transit peptides (nuclear localization signal or mitochondrial targeting signal) guide the TALEs to the respective organelle, and the fusion of UGI to the C-terminus of the TALE-DdCBEs increases the editing efficiency and reduces byproducts by inhibiting uracil-DNA glycosylase (Mok *et al*., 2020). Instead of using the split DddA halves, non-toxic, full-length DddA variants were developed to make monomeric TALE-DdCBEs (DddA guided by one TALE protein) which also allow C•G to T•A editing in mitochondrial DNA (Mok et al., 2022a). $DddA_{tox}$-based CBEs showed an editing preference for 5′-T<u>C</u> contexts, this is comparable to the initial CBEs used for editing nuclear DNA. To improve the deaminate activity and address the rigid 5′-T<u>C</u> context limitation of DddA, evolved DddA variants DddA6 (Q1310R, S1330I, T1380I, and T1413I) with improved activity and DddA11 (S1330I, A1341V, N1342S, E1370K, T1380I, and T1413I) with an expanded targeting scope were created by protein engineering (Mok *et al*., 2022b).

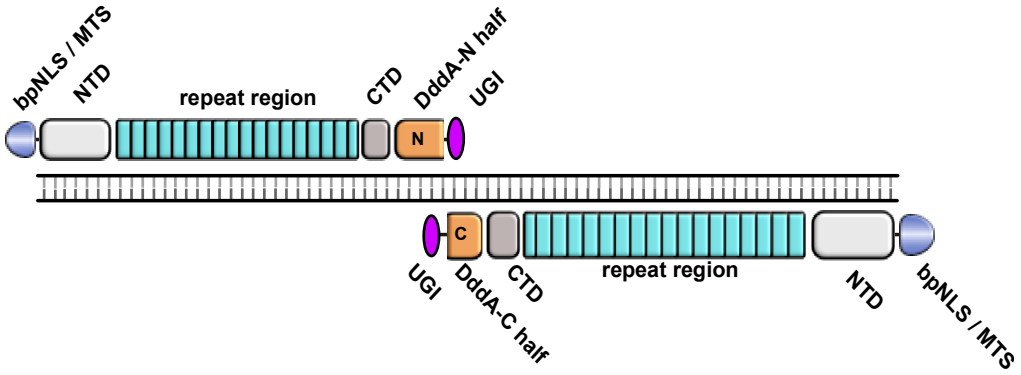


Figure 7. Schematic of TALE-DdCBEs. Illustration of a pair of TALE-DdCBEs interacting with double-stranded DNA. bpNLS: bipartite nuclear localization sequence; MTS: mitochondrial targeting signal; UGI: uracil glycosylase inhibitor; NTD: TALE N-terminal domain; CTD: TALE C-terminal domain.

In plants, TALE-DdCBEs were successfully used for editing the plastid genomes of *Arabidopsis thaliana* and rice (Nakazato *et al*., 2021; Li *et al*., 2021b), and the chloroplast and mitochondrial genomes of lettuce and rapeseed (Kang *et al*., 2021a).

1.3.2 Adenine base editors

Similar to the cytosine base editors, researchers sought to develop adenine base editors, which involve the utilization of adenosine deamination chemistry to install A•T to G•C base pair conversions. Although the idea of substituting cytidine deaminase with an adenosine deaminase enzyme that specifically targets ssDNA was straightforward, however, no such naturally occurring enzyme had been reported. Gaudelli *et al*. showed that, by directed evolution and protein engineering, adenosine deaminase variants that use ssDNA as a substrate were evolved from the *E. coli* transfer RNA adenosine deaminase (TadA) (Gaudelli *et al*., 2017). The ABEs consist of the evolved TadA (TadA*7.10) and nCas9, and they convert adenosines within the R-loop generated by Cas9 to inosines. Those inosines are read as guanines by polymerases which insert a cytosine in the opposite strand. Further directed evolution of TadA*7.10 resulted in TadA8e which has a higher adenine editing activity and specificity (Richter *et al*., 2020). The ABE system did not require any BER inhibition components (such as the UGI of the CBE system), possibly due to a lower efficiency of inosine excision than uracil excision by the BER glycosylase enzymes (Gaudelli *et al*., 2017). Like

CBEs, ABEs were rapidly applied in various plants, such as rice (Li *et al*., 2018a), wheat (Li *et al*., 2018a), *Arabidopsis* (Kang *et al*., 2018), *Brassica napus* (Kang *et al*., 2018), and poplar (Li *et al*., 2021a).

The same as TALE-DdCBEs, researchers utilized the TALE-DdCBE design but substituted the UGI domains with TadA8e to generate TALE-based ABEs (TALEDs) (Figure 8). Targeted A•T to G•C editing, but not C•G to T•A editing is achieved by simply omitting the UGI components from the TALE-DdCBE construct or by employing a catalytically inactivated DddA variant. TALEDs have recently also been developed to perform mitochondrial A•T to G•C base editing in mammalian mitochondria (Cho *et al*., 2022) and Arabidopsis chloroplast genes (Mok *et al*., 2022c).

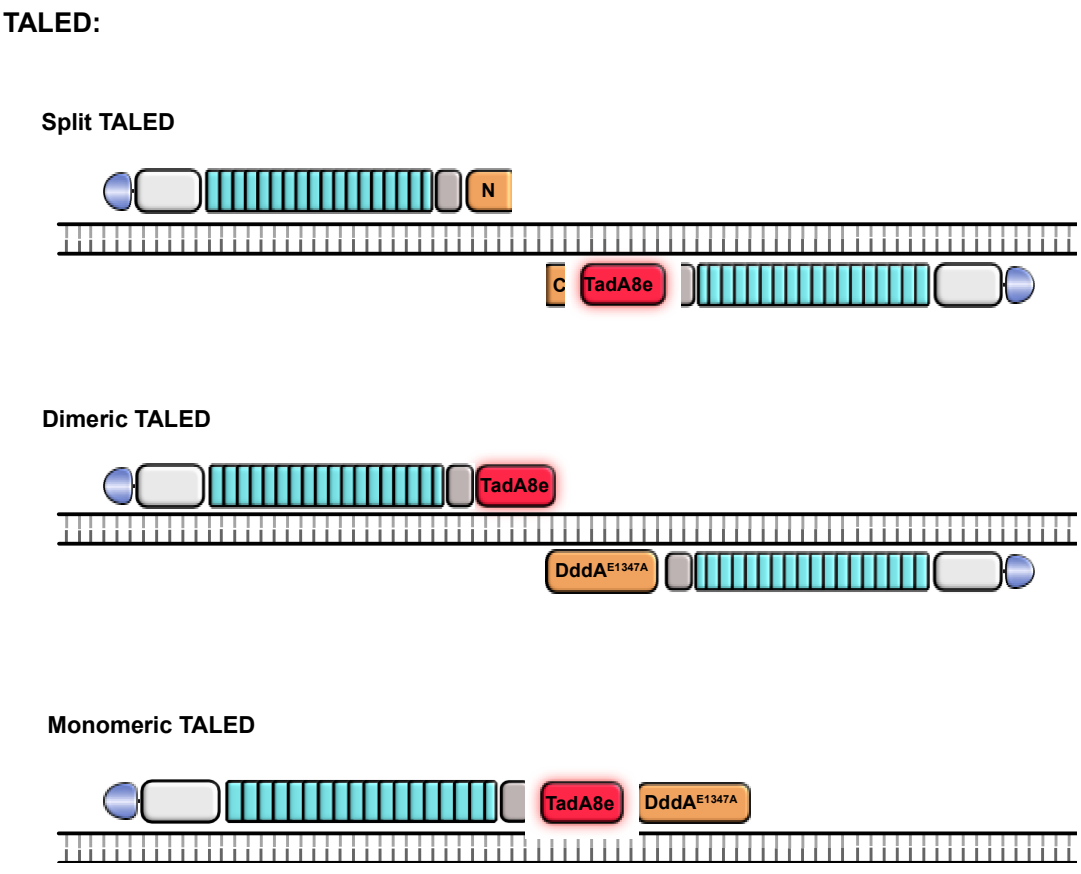


Figure 8. Schematic of TALE-based ABEs (TALEDs). Illustration of TALEDs interacting with double-stranded DNA. TALEDs have three different architectures: Split TALED, Dimeric TALED, and Monomeric TALED. ALL of them covalently fuse a located peptide, a TALE DNA-binding domain and a TadA8e. N: DddA-N half; C: DddA-C half; TadA8e: evolved adenine deaminase; DddA^{E1347A}: catalytically inactive DddA.

1.4 Aim of this thesis

Crop improvements can help us meet the challenge of feeding a population of 10 billion. This thesis focused on the development of precise genome editing tools, including the use of the CRISPR/Cas9 system in brown mustard (*Brassica juncea*) to create hypoallergenic mustard by modifying the allergen gene *Bra j I* (Chapters 2). Furthermore, instead of introducing the DSBs to the gene of interest, this thesis also focused on the development of TALE-based base editors which can install the C•G to T•A (Chapters 3) as well as A•T to G•C (Chapters 4) base editing in plants without introducing any nick or DSB to the genome.

2 Removing the major allergen Bra J I from brown mustard (*Brassica juncea*) by CRISPR/Cas9

Juvenal Assou[1†], **Dingbo Zhang**[2†], Kristian Roth[3], Stephan Steinke[3], Michael Hust[3], Thomas Reinard[2], Traud Winkelmann[1*], and Jens Boch[2*]

[1] Institute of Horticultural Production Systems, Leibniz Universität Hannover, Hannover, Germany

[2] Department of Plant Biotechnology, Institute of Plant Genetics, Leibniz Universität Hannover, Hannover, Germany

[3] Department of Biotechnology, Institute of Biochemistry, Biotechnology, and Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany

† these authors contributed equally

* corresponding authors

Journal: The Plant Journal

Submitted: 15. August 2021

Status: published 16. November 2021

Identifier: doi: 10.1111/tpj.15584

Author	Contributions
Dingbo Zhang	Designed and performed the experiments, which included cloning the *Bra j I* genes from mustard, designing and constructing all the plasmids, genotyping and phenotyping the T_0 and T_1 plants, and detecting the presence of Bra j I in T_0 and T_1 mutant seeds. Contributed to the writing of the corresponding introduction, results, discussion, and conclusion parts.
Juvenal Assou	Designed and performed the experiments, analyzed the data, contributed to the writing of the manuscript
Kristian Roth	Designed and performed the experiments, analyzed data, contributed to the writing of the manuscript
Stephan Steinke	Designed and performed the experiments, analyzed the data
Michael Hust	Conceived and designed the experiments and coordinated the research
Thomas Reinard	Coordinated the research, provided material, assisted with interpretations of results
Traud Winkelmann	Conceived and designed the experiments, coordinated the research, contributed to the writing of the manuscript
Jens Boch	Conceived and designed the experiments, coordinated the research, contributed to the writing of the manuscript

the plant journal

The Plant Journal (2021) doi: 10.1111/tpj.15584

Removing the major allergen Bra j I from brown mustard (*Brassica juncea*) by CRISPR/Cas9

Juvenal Assou[1,†], Dingbo Zhang[2,†], Kristian D. R. Roth[3], Stephan Steinke[3], Michael Hust[3], Thomas Reinard[2], Traud Winkelmann[1,*] and Jens Boch[2,*]

[1]*Institute of Horticultural Production Systems, Leibniz Universität Hannover, Hannover, Germany,*
[2]*Department of Plant Biotechnology, Institute of Plant Genetics, Leibniz Universität Hannover, Hannover, Germany, and*
[3]*Department of Biotechnology, Institute of Biochemistry, Biotechnology, and Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany*

Received 15 August 2021; revised 3 November 2021; accepted 8 November 2021.
*For correspondence (e-mails traud.winkelmann@zier.uni-hannover.de [T.W.]; jens.boch@genetik.uni-hannover.de [J.B.]).
†These authors contributed equally to this work.

SUMMARY

Food allergies are a major health issue worldwide. Modern breeding techniques such as genome editing via CRISPR/Cas9 have the potential to mitigate this by targeting allergens in plants. This study addressed the major allergen Bra j I, a seed storage protein of the 2S albumin class, in the allotetraploid brown mustard (*Brassica juncea*). Cotyledon explants of an Indian gene bank accession (CR2664) and the German variety Terratop were transformed using *Agrobacterium tumefaciens* harboring binary vectors with multiple single guide RNAs to induce either large deletions or frameshift mutations in both *Bra j I* homoeologs. A total of 49 T_0 lines were obtained with up to 3.8% transformation efficiency. Four lines had large deletions of 566 up to 790 bp in the *Bra j IB* allele. Among 18 Terratop T_0 lines, nine carried indels in the targeted regions. From 16 analyzed CR2664 T_0 lines, 14 held indels and three had all four *Bra j I* alleles mutated. The majority of the CRISPR/Cas9-induced mutations were heritable to T_1 progenies. In some edited lines, seed formation and viability were reduced and seeds showed a precocious development of the embryo leading to a rupture of the testa already in the siliques. Immunoblotting using newly developed Bra j I-specific antibodies revealed the amount of Bra j I protein to be reduced or absent in seed extracts of selected lines. Removing an allergenic determinant from mustard is an important first step towards the development of safer food crops.

Keywords: mustard, Bra j I, CRISPR/Cas, transformation, food allergen, seed storage protein.

INTRODUCTION

Food allergy is considered a public health issue and the prevalence has risen in recent decades. In sensitized persons, food allergy can elicit allergic symptoms such as urticaria, itching, wheezing, dyspnea, abdominal pain, and live-threatening anaphylactic shocks (Matsuo et al., 2015). Plant food allergenic proteins are classified according to their structural or functional properties. Most of these belong to the cupin superfamily (7S and 11S seed storage proteins) or the prolamin superfamily (2S albumins, nonspecific lipid transfer proteins, α-amylase/trypsin inhibitors, cereal prolamins) or are functional plant defense proteins (pathogenesis-related proteins, proteases, and protease inhibitors) (Breiteneder and Ebner, 2000; Breiteneder and Radauer, 2004). Most of the 2S albumins are heterodimeric proteins consisting of a large and a small subunit that are connected by a disulfide bond (Shewry et al., 1995).

The known allergens among the 2S albumins include Sin a I from yellow mustard seeds (*Sinapis alba*) (Menéndez-Arias et al., 1988), Bra j I from oriental or brown mustard seeds (*Brassica juncea*) (González de la Peña et al., 1991; Monsalve et al., 1993), Ber e 1 from Brazil nut (*Bertholletia excelsior*) (Nordlee et al., 1996), Jug r 1 from the English walnut (*Juglans regia*) (Teuber et al., 1998), and Ses i 2 from sesame (*Sesamum indicum*) (Beyer et al., 2002). Mustard is a significant elicitor of allergic reactions to food, with yellow and black mustard being mostly consumed in Europe, whereas the brown mustard flour (*B. juncea* L.) is most commonly used in mustard extracts in the USA and Japan (González de la Peña et al., 1991; Monsalve et al., 2001). Mustard allergy usually appears before the age of three, thus mustard is considered as a hidden allergen in the infant's diet (Rancé et al., 2000). Currently, the most effective treatment of mustard and other food

allergies is to strictly avoid ingesting allergen-containing food. However, accidental consumption of allergen-containing food is a constant issue and sensitized persons are significantly compromised in their daily life. Therefore, the development of allergen-free foods is a strategy that would make foods safe for sensitized persons and relieve them from a serious burden.

Using random mutagenesis or traditional breeding strategies to obtain allergen-free plants is difficult and time consuming, especially for polyploid crops with multiple gene copies. Genetic engineering technologies such as RNA interference (RNAi) have been used to alleviate peanut (*Arachis hypogaea*) allergy by silencing genes encoding allergens (Chandran et al., 2015; Dodo et al., 2008). Using programmable nucleases, like CRISPR/Cas9 or TALEN, targeted random mutations and deletions can be induced (Manghwar et al., 2019). This breeding method is currently the one with the greatest potential for crop improvements (Gao, 2021). Many countries have classified genome-edited crops as non-genetically modified organisms, relieving them from extensive regulation, and the first products from genome-edited plants have reached the market (Metje-Sprink et al., 2020). As a first example for a low-allergen food, CRISPR/Cas9 has been used to create low-gluten wheat (*Triticum aestivum*) for gluten-intolerant consumers (Sánchez-León et al., 2018).

Brassica juncea (L.) Czern is one of the major oilseed brassicas, and for this purpose it is mostly cultivated in semi-arid tropics of the Indian subcontinent, as well as in drier areas of Canada, Australia, China, Russia, and the northern USA (Labana and Gupta, 1993). *Brassica juncea* is a self-pollinated plant and an amphidiploid (allotetraploid) hybrid (AABB; 2n = 36), derived from the two diploid progenitors *Brassica rapa* (AA genome, 2n = 20) and *Brassica nigra* (BB genome, 2n = 16) followed by successive chromosome doubling thousands of years ago (Redden et al., 2009; Yang et al., 2016). *Brassica juncea* has been divided into four subspecies, with different morphology, quality characteristics, and usages: (i) the subspecies *B. juncea integrifolia*, used as a leafy vegetable in Asia, (ii) the subspecies *B. juncea juncea*, mostly produced for its seeds and sometimes as fodder, (iii) the subspecies *B. juncea napiformis*, mainly used as a root-tuber vegetable, and (iv) the subspecies *B. juncea taisai*, the stalks and leaves of which are used as vegetables in China (Spect and Diederichsen, 2001). The oilseed of *B. juncea* is especially rich in fatty acid and is an excellent source of natural antioxidants, but its high glucosinolate content and a fatty acid profile with a high level of erucic acid present a health issue for consumption. On the other hand, due to its high content of erucic acid with its low flash points and good combustion and lubrication qualities it can also be a valuable component in biodiesel (Premi et al., 2013). Moreover, the high glucosinolate content renders *B. juncea* an excellent product for biofumigation as a tool to alleviate the effects of soilborne pathogens and replant diseases (Hanschen and Winkelmann, 2020; Mattner et al., 2008; Yim et al., 2016).

To date, no study has reported the production of agronomically viable low-allergenic lines in *B. juncea*. However, several studies have reported the use of genetic engineering to develop varieties of *B. juncea* improved in oil and meal quality. Augustine et al. (2013) applied RNAi-based target suppression to develop low-glucosinolate *B. juncea* lines with significantly improved seed meal quality. Furthermore, Sinha et al. (2007) reported using hairpin-RNA to silence the *fatty acid elongase* gene in *B. juncea*, resulting in a decrease of erucic acid production. Yusuf and Sarin (2007) were successful in using genetic engineering methods to overexpress the γ-*TMT* gene to increase the vitamin E (α-tocopherol) level in the seeds of *B. juncea*.

To our knowledge, the use of CRISPR/Cas9 in brown mustard has not been reported yet. Here, we show that CRISPR/Cas9 can be used to precisely and efficiently mutate the *Bra j I* gene in this allotetraploid crop such that the allergenic Bra j I protein is absent from mutant seeds.

RESULTS

Design of CRISPR/Cas9 constructs with multiplex sgRNA expression cassettes targeting *Bra j I*

Two brown mustard (*B. juncea*) lines were chosen for this study, a European (Terratop) and an Indian (CR2664) one. The two *Bra j I* homoeologs in the published genome of brown mustard (*Bra j IA* and *Bra j IB*) have highly similar nucleotide sequences and are identical for both mustard lines (Figure S1). To modify the two *Bra j I* homoeologs, we designed eight single guide RNAs (sgRNAs) (sg1, sg2, sg3, sg4, sg5, sg6, sg7A, and sg7B) targeting conserved regions in the *Bra j I* exon or 5′/3′-untranslated region (UTR) regions (Figure 1a, Figure S1, and Table S1). All sgRNAs were individually driven by the *AtU6* promoter. Two final binary constructs (named pBraj1256 and pBraj3477) both carrying genes for *SpCas9*, hygromycin resistance, and sgRNA expression cassettes were generated (Figure 1b). pBraj1256 harbored four sgRNAs (sg1, sg2, sg5, and sg6) and could potentially lead to a complete deletion of *Bra j IA* and *Bra j IB* by simultaneously targeting their 5′- and 3′-UTR regions. pBraj3477 contained four sgRNAs (sg3, sg4, sg7A, and sg7B) targeting *Bra j I* coding sequences to introduce deletions or insertions at the cleavage sites that can lead to frameshift mutations of *Bra j IA* and *Bra j IB*.

Genetic transformation of mustard

Cotyledon explants of the *B. juncea* lines Terratop and CR2664 were transformed with the binary vectors containing either the reporter gene construct pEGFP (carrying an

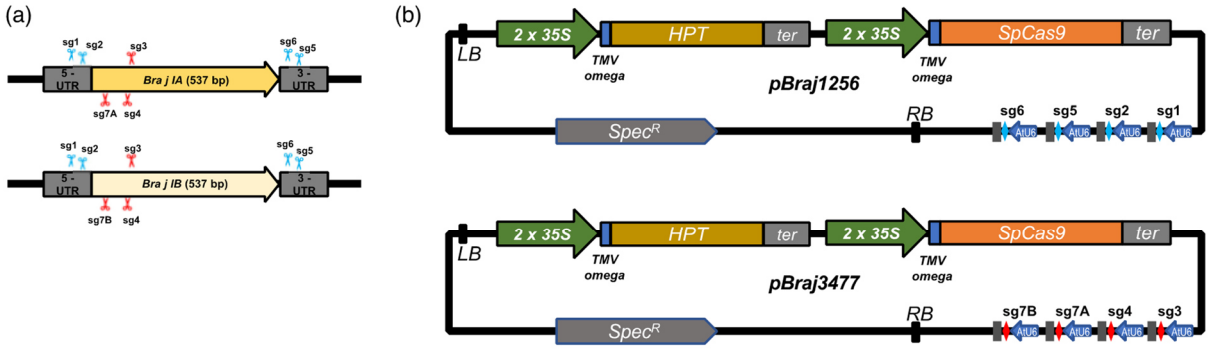


Figure 1. Constructs for targeted mutagenesis of *Bra j I* in *B. juncea* using CRISPR/Cas9. (a) Cartoon of the homoeologous *Bra j I* genes in the A and B subgenomes of tetraploid brown mustard (*B. juncea*) with sgRNA target sites (scissors). The four sgRNAs sg1, sg2, sg5, and sg6 (blue) target conserved 5′- and 3′-untranslated regions of *Bra j IA* and *Bra j IB*, whereas the four sgRNAs sg3, sg4, sg7A, and sg7B (red) target the *Bra j IA* and *Bra j IB* coding sequences. (b) Schematic of binary vectors designed for deleting (pBraj1256) or disrupting (pBraj3477) the coding sequences of *Bra j IA* and *Bra j IB*, respectively. *HPT*, *SpCas9*, and *Spec*^{R} encode hygromycin B phosphotransferase, *Streptococcus pyogenes* Cas9 endonuclease, and a spectinomycin resistance protein, respectively.

enhanced green fluorescence protein-encoding gene and a kanamycin selection marker) or one of two CRISPR/Cas9 constructs (pBraj1256 and pBraj3477, carrying a hygromycin selection marker) (Figure 2a,b). The regeneration of first putative transgenic shoots was observed 4 weeks after infection with *A. tumefaciens* (Figure 2c) and during the following three culture passages developed into plantlets (Figure 2d). Within the first 3 months, some weaker shoots died and these non-transgenic escapes were characterized by an albino-like phenotype. Only the shoots that survived after four culture passages on selection medium were considered to be transgenic shoots and were further tested by PCR. The pEGFP construct was used to establish the genetic transformation protocol (Table 1). To characterize the 57 putative transgenic plants obtained from the two CRISPR/Cas9 constructs, three vector-specific primer pairs targeting different regions of the transfer DNA (T-DNA) were used. Based on the PCR results, 19 out of 23 Terratop T_0 plants and 30 out of 34 CR2664 T_0 plants were PCR-positive plants (Figure S2). The transformation efficiency was line-dependent and also varied with the binary vectors used (Table 1) as well as among the independent experiments (Table S1). The transformation efficiencies in the line Terratop were 2 and 3.8% for pBraj1256 and pBraj3477, respectively, whereas in the line CR2664, transformation efficiencies of 3.8 and 3% were recorded with pBraj1256 and pBraj3477, respectively. In total, from the genetic transformation experiments 49 T_0 plants were transferred to the greenhouse. The surviving transgenic plants (36 T_0 plants) were grown until maturity and leaves and seeds were harvested for further characterization.

Characterization of *Bra j I* mutations in regenerated mustards

To identify CRISPR/Cas9-induced mutations in *Bra j I*, site-specific primer pairs were used to amplify the corresponding *Bra j IA* and *Bra j IB* regions from the transgenic T_0 mustard genomes. PCR on DNA from two T_0 lines (T0-1 from CR2664 and T0-47 from Terratop) which were transformed with the deletion construct pBraj1256 amplified a shorter *Bra j IB* fragment, potentially indicating a deletion (Figure S3). These DNA fragments were cloned and sequenced, revealing a 695-bp deletion of *Bra j IB* in line T0-1 (corresponding precisely to the region between the 5′- and 3′-flanking sgRNA target sites) and a 790-bp deletion of *Bra j IB* in line T0-47 (Figure S4). In addition, we found that two T_0 lines from Terratop (T0-21 and T0-55), which had been transformed with the frameshift mutation construct pBraj3477, also produced a PCR fragment indicating a deletion in *Bra j IB*. Sequencing revealed that line T0-21 has a 566-bp deletion in *Bra j IB* (the deleted region contained sg3 and sg4 target sites), and line T0-56 carries a 629-bp deletion (from the sg7 cutting site to the 3′-UTR) in *Bra j IB* (Figure S4).

Furthermore, we selected 16 transgenic T_0 lines transformed with pBraj3477 (10 from Terratop and six from CR2664) and 14 T_0 lines transformed with pBraj1256 (five from Terratop and nine from CR2664) for detailed *Bra j I* genotyping (Table 2 and Table S2). Among the 15 Terratop T_0 lines, Sanger sequencing indicated that two and five T_0 lines transformed with pBraj3477 and pBraj1256, respectively, carried indels in the targeted regions and the other eight lines did not show any mutation. For the 15 CR2664 T_0 lines, sequencing results showed that six and seven T_0 lines transformed with pBraj3477 and pBraj1256, respectively, carried indels in the targeted regions. Importantly, by specifically analyzing the *Bra j IA* and *Bra j IB* genes, we found that three CR2664 mutants (T0-22, T0-26, and T0-32) transformed with pBraj3477 had all four *Bra j I* alleles mutated (Table 2, Figure S4b).

Overall, for pBraj3477 the mutagenesis frequency of *Bra j IA*/*Bra j IB* was about 33.3 (4/12) and 100% (6/6) in

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2021), doi: 10.1111/tpj.15584

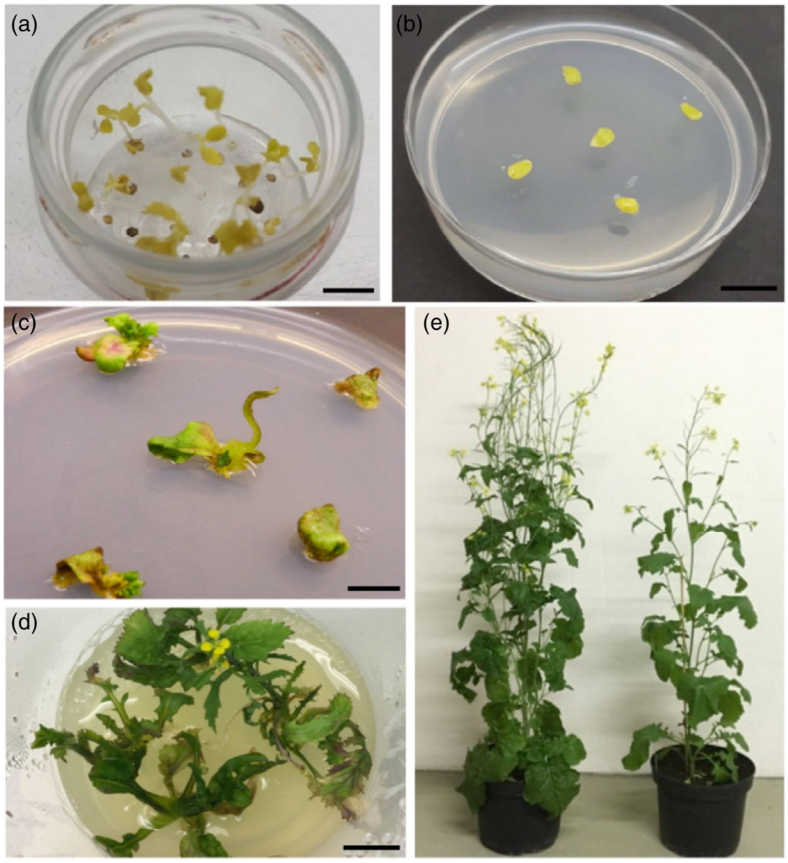


Figure 2. Generation of transgenic *B. juncea* plants (line CR2664) carrying the CRISPR/Cas9 constructs. (a) Five-day-old seedlings. (b) Cotyledon explants used for co-culture. (c) Shoot regeneration 4 weeks after co-culture with *A. tumefaciens* with pBraj1256. (d) Ten-week-old transgenic plants transformed with *A. tumefaciens* pBraj1256. (e) Flowering transgenic plants 10 weeks after acclimatization. Bars represent 1 cm.

Table 1 Transformation efficiencies for two *B. juncea* lines (Terratop and CR2664) using the three vectors pEGFP, pBraj1256, and pBraj3477

Vector[a]	Line	No. of explants[b]	No. of T_0 lines obtained	Transformation efficiency (%)[c]
pEGFP	Terratop	300	2	0.7
	CR2664	300	11	3.7
pBraj1256	Terratop	400	8	2
	CR2664	500	19	3.8
pBraj3477	Terratop	400	15	3.8
	CR2664	500	15	3

[a]Kanamycin selection for pEGFP and hygromycin selection for pBraj1256 and pBraj3477.
[b]For details see Table S1.
[c]Percentage of independent transgenic shoots per 100 explants. Transgenicity was analyzed by three different PCR reactions targeting the T-DNA.

Terratop and CR2664, respectively. Using pBraj1256, the mutagenesis frequency of *Bra j IA*/*Bra j IB* was 100 (6/6) and 80% (8/10) in Terratop and CR2664, respectively. Taken together, we successfully modified *Bra j I* in two brown mustard lines.

Inheritance of mutations in T_1 plants

To investigate the segregation of biallelic and heterozygous mutants, T_1 progenies were generated by self-pollination and analyzed. Sanger sequencing showed that 11 T_1 plants from two *bra j I* biallelic mutants (eight of them were from T0-22 and three from T0-32) were homozygous for *bra j IA* and *bra j IB* or again biallelic mutants (Table 3). For the heterozygous mutant T0-23 which harbored a biallelic mutation at the *Bra j IA* locus and a heterozygous mutation at the *Bra j IB* locus, 17 T_1 plants were analyzed, all of which were homozygous or biallelic mutations at the *Bra j IA* locus as expected. In contrast, only one offspring (line 23-7) carried a heterozygous mutation at the *Bra j IB* locus and the remaining 16 T_1 plants were homozygous wild type at the *Bra j IB* locus, which is not the expected 1:2:1 segregation ratio for a heterozygous locus after selfing (Table S3).

To assess whether the *Bra j IB* deletion identified in line T0-1 can be transmitted to the next generation, we tested 35 T_1 progeny plants, but found no PCR fragment indicating a deletion among them (Figure S5). We selected two T_1 progenies (1-1 and 1-2) from T0-1 for sequencing and new allelic mutations were detected in both *Bra j IA* and *Bra j*

Table 2 Genetic characterization of T_0 plants of the *B. juncea* lines Terratop and CR2664 regarding transgene presence and mutations in the target genes *Bra j IA* and *Bra j IB*, seed weight, and germination of seeds

Plant number	Plasmid	Line	Transgenicity assay[a]	Genotype[b] *Bra j IA* allele 1/allele 2	Genotype[b] *Bra j IB* allele 1/allele 2	Seed number obtained from free pollination	Weight of 100 seeds [g][c]	*In vitro* germination [%] (n = 10)	*Ex vitro* germination [%] (n = 20)
K2	-	Terratop	-	Wild type/wild type	Wild type/wild type	4477	0.257	100	95
K17	-	CR2664	-	Wild type/wild type	Wild type/wild type	4490	0.289	80	95
1	pBraj1256	CR2664	+	Wild type/wild type	Wild type/−695 bp	745	0.280[ns]	40	15
22	pBraj3477	CR2664	+	−14 bp (sg7)/+1 bp (sg7)	−1 bp (sg7)/+1 bp (sg7)	119	0.095	40	55
23	pBraj3477	CR2664	+	−14 bp (sg7)/+1 bp (sg7)	Wild type/+1 bp (sg7)	65	0.218	50	85
25	pBraj1256	CR2664	+	Wild type/wild type	Wild type/wild type	307	0.103***	20	20
26	pBraj3477	CR2664	+	+1 bp (sg4), +1 bp (sg7)/+1 bp (sg4), +1 bp (sg7)	+1 bp (sg7), −4 bp (sg4)/+1 bp (sg7), −1 bp (sg4)	482	0.133***	10	30
27	pBraj1256	CR2664	+	Wild type/wild type	Wild type/wild type	125	0.068	10	20
30	pBraj1256	CR2664	+	−1bp (sg1)/−1bp (sg1)	Wild type/−8 bp (sg1)	203	0.157	50	90
31	pBraj1256	Terratop	+	−1bp (sg1)/−1bp (sg1)	−4 bp (sg1)/−1 bp (sg1)	1264	0.183***	60	85
32	pBraj3477	CR2664	+	+1 bp (sg4); +1 bp (sg7)/+1 bp (sg4); +1 bp (sg7)	+1 (sg7), +1 (sg4)/−2 (sg7)	182	0.195	0	15
33	pBraj3477	CR2664	+	Wild type/−3 bp (sg7)	Wild type/+1 bp (sg7)	1303	0.198***	0	30
34	pBraj3477	CR2664	+	Wild type/−3 bp (sg7)	Wild type/−1 bp (sg7)	155	0.163	10	0
37	pBraj1256	CR2664	+	−1 bp (sg1)/−1 bp (sg1)	+1 bp (sg1)/+1 bp (sg1)	128	0.153	80	50

[a]'+' detected, '-' not detected.
[b]The sequencing chromatograms were decoded by ICE Analysis (https://ice.synthego.com/#/).
[c]For lines with less than 300 seeds no statistical analysis was conducted. For the remaining lines the 100-seed weight of three to four subsamples was compared against the wild type by Dunnett's test. ***P < 0.0001; ns, not significant.

IB alleles. T_1 line 1-1 and T_1 line 1-2 displayed homozygous 1-bp deletions (-C) at the *Bra j IA* sg1 site and a biallelic 1-bp insertion (+T/+A) at the *Bra j IB* sg1 target site (Table S3). It is possible that the T_0 plant was mosaic and the deleted allele was not present in the germline, whereas the detected new mutations occurred either in the T_0 or the T_1 generation. Nevertheless, these results show that the majority of our CRISPR/Cas9-induced mutations in mustard were heritable. We analyzed whether off-target mutations occurred during our editing approach. For this, 24 predicted off-target sites in the two edited T_1 lines 22-1 and 31-5 were PCR-amplified and analyzed by sequencing. No off-target mutation was detected (Figure S8).

Phenotyping of transgenic mustard seeds

The T_0 plants of both lines that clearly showed homozygous mutations (either monoallelic or biallelic) and the wild-type plants were analyzed for yield and viability traits (Table 2 and Table S2). Seed production was significantly reduced in some of the transgenic lines. This was especially obvious in the controlled crosses for which lower percentages of silique formation and much lower numbers of seeds per silique were recorded for most transgenic lines (Table S4). After open pollination, the wild-type plants of both lines produced approximately 4500 seeds per plant with a 100-seed weight of 0.257 g for Terratop and 0.289 g for CR2664, while only two out of 57

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2021), doi: 10.1111/tpj.15584

Table 3 Genetic characterization of T_1 plants of the *B. juncea* lines Terratop and CR2664 regarding transgene presence and mutations in the target gene *Bra j I*, seed weight, and germination of seeds

Plant number	Plasmid	Line	Transgenicity assay[a]	Genotype[b] *Bra j I A* allele 1/allele 2	Genotype[b] *Bra j I B* allele 1/allele 2	Seed number obtained from free pollination	Weight of 100-seed weight [g][c]	*In vitro* germination [%] (n = 10)	*Ex vitro* germination [%] (n = 20)
K17	Wild type	CR2664	-	Wild type/wild type	Wild type/wild type	3224	0.303	100	100
K2	Wild type	Terratop	-	Wild type/wild type	Wild type/wild type	2178	0.273	100	95
22-1	pBraj3477	CR2664	+	−14 bp (sg7)/ +1 bp (sg7)	−1 bp (sg7)/+1bp (sg7)	182	0.222	30	65
22-2	pBraj3477	CR2664	+	−14 bp (sg7)/ +1 bp (sg7)	+1 bp (sg7)/+1bp (sg7)	85	0.125	20	30
22-3	pBraj3477	CR2664	+	−14 bp (sg7)/ −14 bp (sg7)	−1 bp (sg7)/ −1 bp (sg7)	0	n.a.	n.a.	n.a.
22-4	pBraj3477	CR2664	+	+1 bp (sg7)/ +1 bp (sg7)	−1 bp (sg7)/+1bp (sg7)	612	0.154***	50	55
22-5	pBraj3477	CR2664	+	−14 bp (sg7)/ +1 bp (sg7)	−1 bp (sg7)/+1bp (sg7)	118	0.208	80	75
22-6	pBraj3477	CR2664	+	−14 bp (sg7)/ +1 bp (sg7)	+1 bp (sg7)/+1bp (sg7)	256	0.120	40	30
22-7	pBraj3477	CR2664	+	−14 bp (sg7)/ +1 bp (sg7)	−1 bp (sg7)/−1bp (sg7)	10	0.211	20	n.a.
22-8	pBraj3477	CR2664	+	−14 bp (sg7)/ −14 bp (sg7)	−1 bp (sg7)/+1bp (sg7)	521	0.181***	40	70
28-1	pBraj1256	CR2664	+	+1 bp (sg1)/ +1 bp (sg1)	−58 bp (sg1)/ −58 bp (sg1)	1199	0.381[ns]	100	100
31-2	pBraj1256	Terratop	-	−1 bp (sg1)/ −1 bp (sg1)	−4 bp (sg1)/ −1 bp (sg1)	1421	0.216***	100	95
31-5	pBraj1256	Terratop	+	−1 bp (sg1)/ −1 bp (sg1)	−1 bp (sg1)/ −1 bp (sg1)	3298	0.230***	80	90
32-1	pBraj3477	CR2664	+	+1 bp (sg4); +1 bp (sg7)/ +1 bp (sg4); +1 bp (sg7)	−2 bp (sg7)/ −2 bp (sg7)	1067	0.219***	40	40
32-2	pBraj3477	CR2664	+	+1 bp (sg4); +1 bp (sg7)/ +1 bp (sg4); +1 bp (sg7)	−2 bp (sg7)/ −2 bp (sg7)	700	0.430***	60	45
32-3	pBraj3477	CR2664	+	+1 bp (sg4); +1 bp (sg7)/ +1 bp (sg4); +1 bp (sg7)	+1 (sg7); +1 (sg4)/−2 bp (sg7)	1201	0.204***	70	80
35-3	pBraj1256	CR2664	+	+1 (sg1), −1 (sg 5)/+1 (sg1), −1 (sg 5)	−83 bp (sg5), −32 bp (sg2)/ −83 bp (sg5), −32 bp (sg2)	1037	0.156***	80	85
35-10	pBraj1256	CR2664	-	+1 (sg1)/+1 (sg1)	−83 bp (sg5), −32 bp (sg2)/ −83 bp (sg5), −32 bp (sg2)	47	0.095	n.a.	n.a.

[a]PCR with three T-DNA-specific primer pairs. +, fragment detected; -, no fragment detected.
[b]The sequencing chromatograms were decoded by ICE Analysis (https://ice.synthego.com/#/).
[c]For lines with less than 300 seeds no statistical analysis was conducted. For the remaining lines the 100-seed weight of three to four subsamples was compared against the wild type by Dunnett's test. ***P < 0.0001; ns, not significant; n.a., not analyzed.

transgenic lines produced more than 1000 seeds, i.e., lines 31 and 33 with 1264 seeds and a 100-seed weight of 0.183 g and 1303 seeds and a 100-seed weight of 0.198 g, respectively. Most of the remaining transgenic lines had a reduced seed production of less than 500 seeds per plant. The transgenic line 23 showed the lowest seed production

with only 65 seeds and a 100-seed weight of 0.218 g. A significantly reduced 100-seed weight was observed for four T_0 plants (25, 26, 31, and 33), and three further T_0 plants produced seeds of severely reduced weight (22, 27, and 36) but due to their low seed number could not be included in the statistical comparison (Table 2 and Table S2). In contrast, seeds harvested from plant 28 were significantly heavier than those of the wild type (Table S2). The seeds from the T_0 plants were germinated *in vitro* and *ex vitro* to evaluate seed viability. *In vitro*, the T_0 lines 35 and 37 showed the highest germination rates of 100% and 80%, respectively, whereas the seeds of T_0 lines 29, 32, and 33 did not germinate at all. *Ex vitro*, the highest germination rate of 90% was observed with line 30, followed by lines 23 and 31, with a germination rate of 85%. No germination was recorded for seeds of lines 24, 29, and 34.

The seeds from those T_1 lines that carried CRISPR-induced homozygous mutations were harvested and the 100-seed weight was as well determined (Table 3). Seed production was again much lower in most T_1 lines compared to wild-type plants, and the seed weight was reduced and strongly varied from plant to plant. Seed production of the T_1 plants of line 22 ranged from 0 to 612 seeds with a 100-seed weight of 0.120 g to 0.208 g. However, the T_1 plant 31-5 of line 31 showed a production of 3298 seeds, which was even higher than that from the wild type. When testing the germination ability, the *in vitro* germination rate of the seeds of the T_1 plants of line 22 ranged from 20 to 80%, and *ex vitro* germination was comparable (from 30 to 75%). For all remaining T_1 plants, the *in vitro* germination rate varied from 30 to 100% and the *ex vitro* germination rate varied from 40 to 100%, with most T_1 lines germinating at lower percentages than wild type.

In some transgenic lines, the seed phenotype clearly differed from that of wild-type plants (Figure 3). Some T_0 and

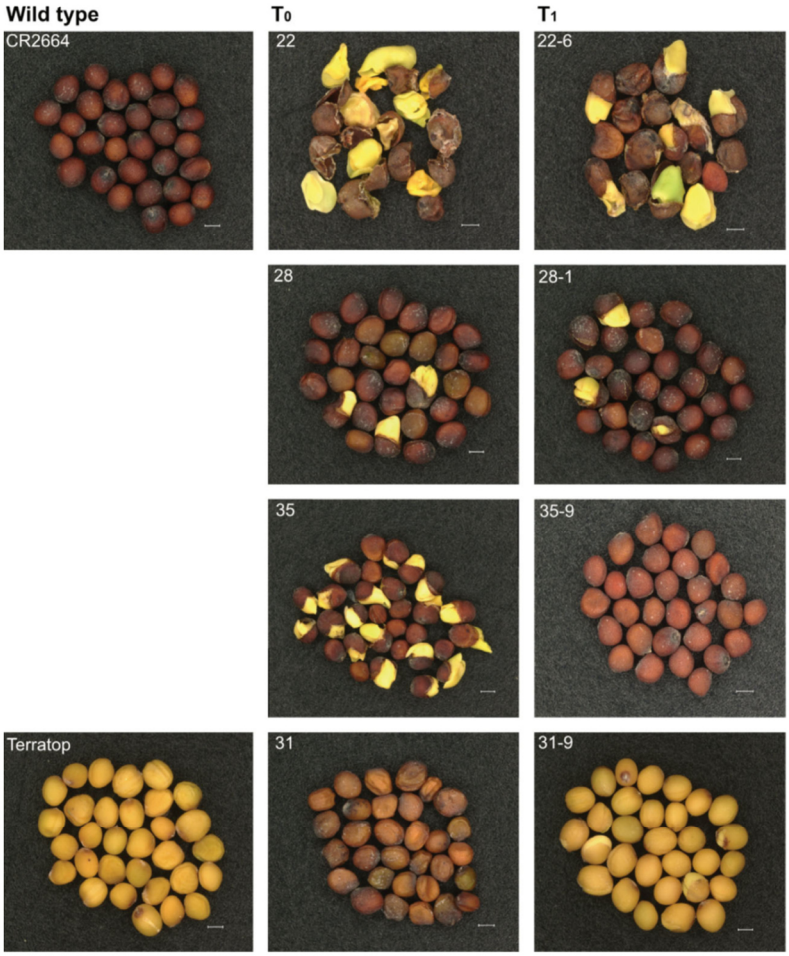


Figure 3. Seeds from wild-type, T_0, and T_1 lines of *B. juncea* (line CR2664 and Terratop) carrying the CRISPR/Cas9 constructs. Numbers 22, 28, 35, and 31 refer to T_0 lines (see Table 2 and Table S2); 22-6, 28-1, 35-9, and 31-9 refer to T_1 lines (see Table 3 and Table S3). Bar = 500 μm.

The Plant Journal, (2021), doi: 10.1111/tpj.15584

T_1 seeds showed a precocious development of the embryo leading to a rupture of the testa and a breakthrough of the embryo through the seed coat already in the siliques. T_0 and T_1 plants of line 22 were heavily affected with more than 70% of the seeds showing this abnormality. The T_0 and T_1 plants of line 28 only rarely expressed this phenotype, whereas in line 35 the T_0 was heavily affected while the seeds of the T_1 plant lines were less affected. For the T_0 line 31 of cultivar Terratop, a change in the seed coat color was observed from yellow to brown, which reverted to yellow in some of the T_1 lines (line 31-9) (Figure 3).

Immunodetection of Bra j I in T_1 mustard seeds

The seed storage protein Bra j I is a 2S albumin with a molecular mass of about 22 kDa which is processed into two subunits of 9.5 kDa and 12 kDa (Breiteneder and Ebner, 2000; L'Hocine et al., 2019). Two bands of corresponding sizes are visible in Coomassie-stained protein profiles of seed extracts from wild-type and EGFP-transgenic plants (Figure 4a–d). These protein bands are not visible in seed extracts from T_1 lines with frameshift mutations in all *bra j I* alleles (different T1-22 lines; Figure 4a), indicating that the induced mutations successfully block Bra j I protein production. In contrast, the Bra j I protein bands are reduced in their abundance but not completely absent in the T_1 lines T1-32-1 and T1-35-3 (Figure 4b,d). Line T1-32-1 carries frameshift mutations in both *Bra j IA* and *Bra j IB*, and line T1-35-3 carries mutations in the 5′- and 3′-UTRs of *Bra j I*, but not within the coding region (Table 3).

To analyze whether the identified protein bands correspond to the Bra j I protein, two specific antibodies (KRO58-A3 and STE2-G2) were generated by antibody phage display and used in immunoblotting with seed extracts (Figure 4e–h). Both antibodies were selected against the linear immunogenic epitope of Bra j I (Figure S6) (Monsalve et al., 1993). The antibodies were cloned into the scFv-Fc format (with the human IgG1 Fc fragment) and produced in EXPI293F cells. The binding of both antibodies to complete Bra j I was validated by ELISA on recombinant Bra j I. To determine the EC_{50} values, a titration ELISA was performed (Figure S7), resulting in subnanomolar EC_{50} values for both antibodies: 0.22 nM for KRO58-A3 and 0.32 nM for STE2-G2 (KRO58-A3 was chosen for the immunoblotting). As expected, the anti-Bra j I antibody efficiently detected a protein in wild-type and EGFP-transgenic seed extracts corresponding to the heavy chain (12 kDa) of Bra j I (Figure 4e–h). No anti-Bra j I antibody binding was detected in lines T1-22 and T1-32-1, verifying that the Bra j I protein was absent in seeds of these T_1 *bra j I* mutants (Figure 4e,f).

To further estimate whether mutation of *Bra j IA*, but not *Bra j IB* results in reduced Bra j I protein accumulation within the seeds, we selected four T_1 *bra j I* mutants from T_0 line 23: T1-23-8, T1-23-12, T1-23-15, and T1-23-16, with biallelic or homozygous mutations at *Bra j IA* but no editing at *Bra j IB* (Table S3). Surprisingly, practically no Bra j I protein accumulated in these four T1-23 lines (Figure 4c,g), suggesting that the intact *Bra j IB* alleles in these T_1 lines did not result in a partial protein accumulation.

In contrast, the T_1 lines T1-31-5 and T1-35-3 showed a reduced accumulation of Bra j I protein. Lines 31-5 and 35-3 contain mutations at the 5′- and/or 3′-UTR regions of *Bra j IA* and *Bra j IB*, but not the coding regions (Table 3). This suggests that mutations in these regions might also have an impact on the level of Bra j I production. Taken together, we successfully demonstrated that several of our edited mustard lines have no or a reduced accumulation of Bra j I protein.

DISCUSSION

Transformation efficiency differs between genotypes

The genetic transformation of both mustard lines (Terratop and CR2664) using CRISPR/Cas9 editing technology was successfully achieved. In this study, we focused on the two mustard lines CR2664, originating from India with a brown testa, and Terratop, derived from a German breeder and selected for being used as a catch crop with a yellow testa. The Indian *B. juncea* lines contain higher levels of erucic acid and butenyl and propenyl glucosinolates, while the European lines were selected for a low erucic acid level and only contain propenyl glucosinolates (Lionneton et al., 2004; Vaughan et al., 1963). Fazekas et al. (1986) reported good amenability to tissue culture of the Indian lines, with greater potential for adventitious shoot regeneration than the European lines. Since the regeneration capacity is the precondition for *A. tumefaciens*-mediated transformation, most of the transformation experiments in *B. juncea* reported so far involved Indian lines (Ahmed et al., 2017; Bhuiyan et al., 2011; Dutta et al., 2008; Mondal et al., 2007; Rani et al., 2017; Sharma et al., 2004; Singh et al., 2009; Thakur et al., 2020). This paper is one of the first reporting successful *A. tumefaciens*-mediated transformation of a European line of *B. juncea*. The transformation efficiencies obtained in this study from both lines (Terratop and CR2664) varied between 0.7 and 3.8%, thus being similar to the efficiency of 0.5 to 1.5% reported by Mondal et al. (2007) but lower than the previously reported transformation efficiencies of 6–7 (Sharma et al., 2004), 16.2 (Singh et al., 2009), or 16.4 to 19.7% (Rani et al., 2017). Factors that could improve the transformation efficiencies in future experiments could involve a delayed start of selection for 3–5 days (as shown by Bhuiyan et al., 2011) or a preculture of 2 days as suggested by Thakur et al. (2020).

Successful genome editing in polyploid species

Polyploidization in flowering plants has promoted genetic variation by genome duplication during their adaptive

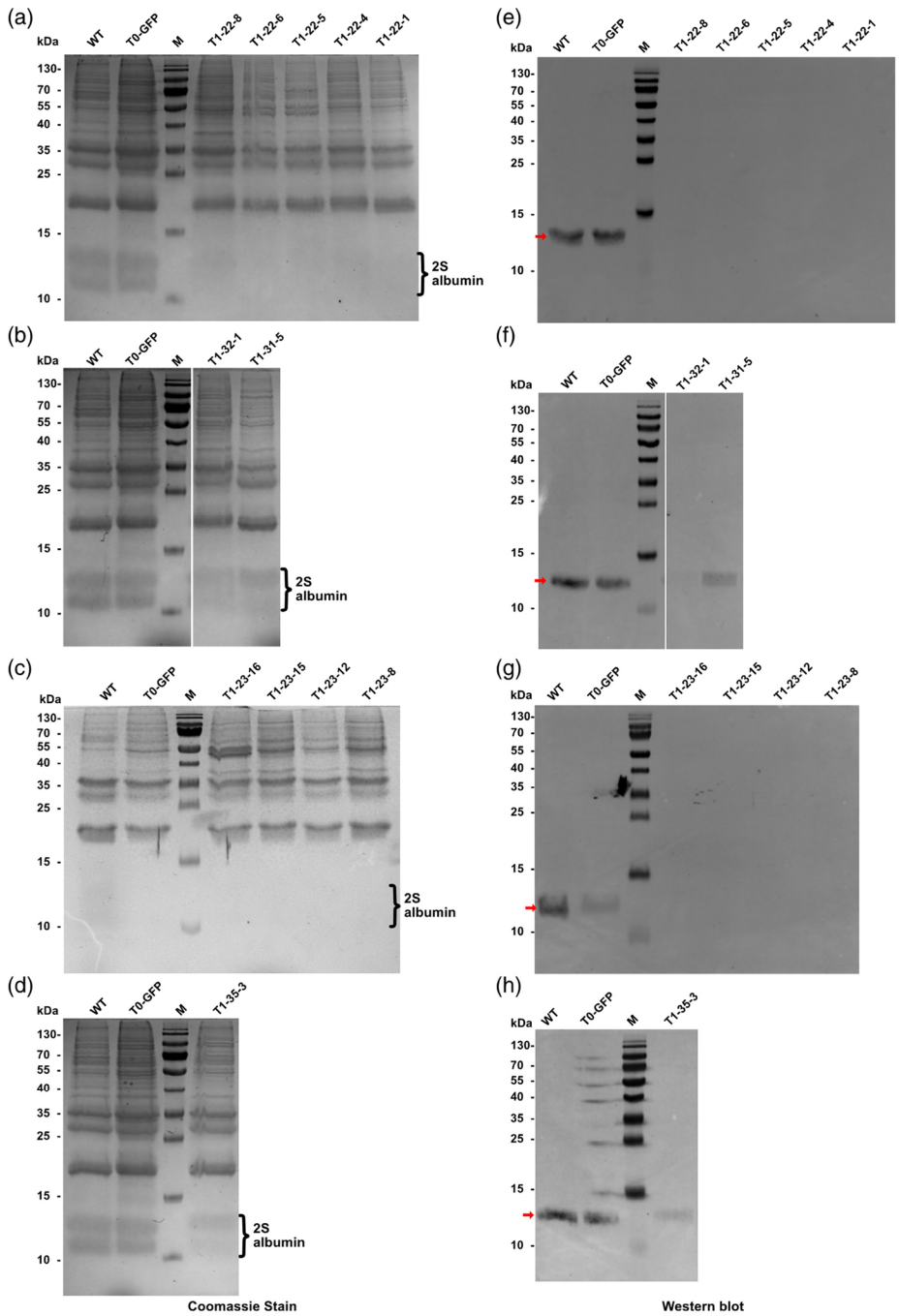


Figure 4. Loss of Bra j I protein in genome-edited mustard seeds. (a–d) Electrophoretic profiles of seed proteins from T_1 plants. The concentrations of protein extracts were normalized and about 40 μg was loaded per lane. The proteins were separated by SDS-PAGE and stained with Coomassie Blue R-250. (e–h) Immunoblot analysis of Bra j I protein in seeds from T_1 mustard plants. Following SDS-PAGE, proteins were transferred to a nitrocellulose membrane. The Bra j I -specific monoclonal antibody KRO58-A3 was used for Bra j I (red arrow, heavy chain, approximately 12 kDa) detection. WT, wild-type CR2664 seeds; T0-GFP, EGFP-transgenic seeds (non-edited control); M, protein marker.

evolution. A large number of crops are alloploids (e.g., rapeseed [*B. napus*], wheat, and cotton [*Gossypium hirsutum*]) combining genomes from different species (Pelé et al., 2018). Random mutagenesis technologies to improve traits in polyploid crops are usually inefficient; however, targeted genome editing has become a promising

The Plant Journal, (2021), doi: 10.1111/tpj.15584

alternative. The CRISPR/Cas9 system is the most widely used genome editing technology for precise plant genome modifications (Chen et al., 2019). Until now, CRISPR/Cas9-mediated genome editing has been successfully reported in some polyploid crops: In allotetraploid crops, the editing efficiencies varied; for example, the mutagenesis efficiency ranged from 96.8 to 100% in oilseed rape (*B. napus*) (Li et al., 2018) and from 66.7 to 100% in cotton (*G. hirsutum*) (Wang et al., 2017). A previous report showed that the Arabidopsis U6 small nuclear RNA (snRNA) promoter *AtU6-26* has a higher transcriptional activity than other *AtU6* snRNA promoters (Li et al., 2007), and in allotetraploid sweet basil (*Ocimum basilicum*) the mutation efficiency was 92.6% by using such an *AtU6-26*-sgRNA expression cassette (Navet and Tian, 2020). In this study, we utilized a multiplex CRISPR/Cas9 system (harboring the *AtU6-26*-sgRNA expression cassettes) to simultaneously edit four alleles of *Bra j I* in mustard, and the mutation efficiencies for *Bra j IA* and *Bra j IB* were 47.1 (8/17) and 50% (9/18) in Terratop and 81.3 (13/16) and 87.5% (14/16) in CR2664. These data implied that CRISPR/Cas9 is a suitable approach and our results confirmed that high editing efficiencies can be obtained also in mustard using a *35S* and *AtU6-26* promoter for *SpCas9* and sgRNA expression, respectively.

Off-target mutations are a concern for the applied use of CRISPR/Cas9. Such off-target events can be limited by using one of the high-fidelity SpCas9 variants instead of the wild-type SpCas9 (Zhang et al., 2017) or by choosing sgRNAs with few predicted off-targets (Liu et al., 2017). In plant research, possible off-target events can sometimes be removed by crossing or segregation in the next generation while keeping the desired mutations. In this study, we analyzed 24 predicted off-target sites in two edited T_1 lines and did not find any off-target mutation. Because these lines still carry the CRISPR/Cas9 transgene, it is possible that mutations will occur at a later time point or at a locus that was not analyzed. Before commercial application, a more thorough analysis might be advisable.

Altered seed phenotype in some transgenic lines

We noted that the seeds of several *bra j I* mutants were aberrant in shape (Figure 3), weight, and germination efficiency (Table 2, Table 3, Table S2, and Table S3). Those phenotypes indicate that Bra j I may have an influence on seed development. The seed yield and related traits revealed wide phenotypic variation between the wild-type plants and the transgenic T_0 and T_1. In this study, the wild types of both lines produced approximately 4500 seeds per plants with a 100-seed weight of 0.257 g for Terratop and 0.289 g for CR2664 (Table 2). The number of seeds is similar to the numbers reported by Stevens (1932, 1957) and the weight is within the range of 2.63 g and 1.84 g (1000-seed weight) reported by Stevens (1932, 1957), whereas the seed weight was higher than that recorded by Yoshimura et al. (2016) (1000-seed weight: 1.57 g). In contrast, Ramana and Ghildiyal (1997) reported much higher seed production of 3825 to 7075 per plant with a 1000-seed weight ranging from 4.49 to 4.64 g. In this study, a much lower seed production was observed for some of the transgenic lines, and this reduction was also found for the 100-seed weight and seed viability. The lower seed number might be caused by the isolation bags that had to be used for the transgenic plants as soon as flowers started to open. However, the significant reduction of seed weight and viability could also be the result of modified Bra j I levels in edited *B. juncea* seeds. Rolletschek et al. (2020) reported a seed weight reduction in transgenic lines when using RNAi to suppress seed storage protein synthesis in *B. napus*. Moreover, they observed an irregular shape of the cotyledons of mature transgenic seeds with a volume of void spaces significantly greater than in the wild-type seeds and the transgenic embryos being 7 to 9% smaller than the wild-type ones. For T_0 line 25, we could not detect a mutation in the *Bra j I* homoeologs, but observed a significantly reduced seed weight. Since this plant was proven to contain the editing construct, this fact could be explained by chimerism and a random selection of a leaf that did not carry mutations. A negative effect due to the integration site of the transgene could be an alternative explanation.

The allergen Bra j I belongs to the 2S albumin class, one of the most abundant seed storage protein classes in mustard (Monsalve et al., 1993). The seed storage proteins are essential for the seed germination vigor and provide an efficient scavenging system of the reactive oxygen species (ROS) actively generated during seed germination and, therefore, protect the other proteins that are essential for germination from oxidation (El-Maarouf-Bouteau et al., 2013; Muntz et al., 2001; Nguyen et al., 2015). Some of our T_0 and T_1 seeds showed a precocious development of the embryo leading to a rupture of the testa already in the siliques (Figure 3). Similar abnormalities were reported in F_1 hybrid seeds derived from a cross of transgenic *B. napus* and *B. juncea* that showed cracked seed coats (Tang et al., 2018). The phenotype is similar to that of mutants lacking a growth arrest during seed development and being deficient in dormancy resulting in vivipary. Vivipary occurs in mutants lacking genes that suppress germination and here especially adequate abscisic acid (ABA) levels are important to prevent sprouting (Bewley et al., 2012). The level of dormancy in a seed, i.e., its capacity to repress precocious germination, is mostly related to the capacity of the seed to synthesize ABA in the endosperm and in the embryo, which is an essential contributor to maintain dormancy (Lee et al., 2010). The reduction in seed storage protein levels in seed development often results in a low dormancy level of the seeds (Debeaujon et al., 2000; Nguyen et al., 2015; Sugliani et al., 2009). A possible link to ABA

can be the levels of free amino acids, which can be assumed to be altered in the *Bra j I* mutant lines. Scuffi et al. (2014) were able to show that hydrogen sulfide acts as a component of the ABA signaling pathway that is involved in stomatal closure. Hydrogen sulfide can be enzymatically produced from cysteine via L-cysteine desulfhydrase (Scuffi et al., 2014).

Mutations led to the intended decreased Bra j I content in seeds

The allergen Bra j I from brown mustard was initially found in the 2S albumin fraction and recognized by the IgE of sensitive individuals (L'Hocine et al., 2019). To create Bra j I-free mustard, the *Bra j I* gene was inactivated in the genome using CRISPR/Cas9. Three regenerated T_0 *bra j I* mutant lines (lines 22, 26, and 32) were identified to contain frameshift mutations in all four *Bra j I* alleles. The corresponding seeds of T_1 offsprings of T_0 lines 22 and 32 showed significantly less 2S albumin compared to wild-type seeds. Moreover, the immunoblotting results with an anti-Bra j I antibody showed no Bra j I protein in all six of these lines, confirming that we successfully obtained Bra j I-free mustard lines. These lines are a valuable start to develop low-allergenic mustard. We noticed increased abundance of some proteins in the seed extracts of T_1 lines, e.g., T1-23-16 or T1-22-5 (Figure 4a,c), possibly indicating that the loss of Bra j I caused an accumulation of other seed storage proteins. This is in agreement with recent findings in *Camelina sativa* cruciferin C knockout lines (Lyzenga et al., 2019).

CRISPR/Cas9-induced DNA double-strand breaks (DSBs) are mainly repaired by non-homologous end joining (NHEJ) and usually result in small insertions or deletions (indels). Indels can alter the gene-coding frame and create a premature stop codon. However, indels also produce pseudo-mRNAs that do not encode functional proteins and lead to unexpected effects in CRISRP-based gene knockout studies (Tuladhar et al., 2019). Besides, such pseudo-mRNAs from the truncated reading frame of a targeted gene could produce proteins with altered amino acid sequences. Thus, a clean genetic mutation would require to completely delete a gene using two or more sgRNAs simultaneously. In our study, we expressed four sgRNAs in tandem to delete the whole *Bra j I* coding region, and Sanger sequencing confirmed that the *Bra j IB* allele was deleted by the joining of the sg1 and sg5 cleavage sites in the T_0 plant line 1 (Figure S4). However, in another line, line 47, the *Bra j IB* deletion was an outcome of an NHEJ repair with DSBs occurring in sg1 or sg2 targeted sites (Figure S4). Similar outcomes were also found in line 21 and line 56, which have large fragment deletions at cleavage sites (Figure S4). We noticed that the efficiencies of complete deletions of *Bra j IA* and *Bra j IB* were lower than we expected, and neither a deletion of *Bra j IA* nor a complete *Bra j IB* deletion line was obtained. Several studies already showed that utilizing microhomology-mediated end joining (MMEJ)-assisted CRISPR/Cas9 editing, the deletion efficiency can be increased (Owens et al., 2019; Tan et al., 2020). An alternative approach might thus be to design sgRNAs targeting microhomologous sequences to induce a complete deletion of *Bra j IA* and *Bra j IB* by the MMEJ DNA repair mechanism.

CONCLUSIONS AND OUTLOOK

The Bra j I mutation lines that were generated in this work now serve as a key starting point to further study the impact of changes in seed storage protein composition in mustard. Future work will correlate specific changes in seeds with their allergenic potential for patients and their food quality. In principle, it might even be possible to precisely change epitopes in proteins that trigger IgE binding using precision genome editing (base editors or prime editors; Anzalone et al., 2020) without removing the protein altogether and thereby avoiding altered seed phenotypes.

The protocols established for genetic transformation and CRISPR/Cas9-mediated genome editing in *B. juncea* pave the way for application in modern breeding programs of this allotetraploid crop. Compared to traditional breeding, genome editing-based new breeding technologies accelerate crop improvement since they can generate precise genetic changes and desired traits in complex genetic backgrounds in a relatively short time (Gao, 2021). Our work demonstrates the removal of an allergenic protein from brown mustard potentially improving the safety of mustard-derived food products for sensitized individuals. This work also highlights the possibility for creating hypoallergenic food plants in general to enhance their safety for human consumption.

EXPERIMENTAL PROCEDURES

Plant material and explant preparation

Seeds of two brown mustard (*B. juncea* L.) lines were used, the European cultivar Terratop (kindly provided by the breeder P.H. Petersen Saatzucht Lundsgaard GmbH, Grundhof, Germany) and an Indian line of *B. juncea* (L.) Czern. with accession number CR2664 obtained from the gene bank of the IPK Gatersleben, Germany. The mature seeds of both lines were surface-disinfected with 1% NaOCl with 0.1% Tween 20 for 5 min followed by three washing steps with sterile deionized water under aseptic conditions. The treated seeds were germinated on half-strength Murashige and Skoog (MS) medium (Murashige and Skoog, 1962) for 5 days in darkness at 24°C. The 5-day-old seedlings were collected under aseptic conditions, and the cotyledons were vertically divided into two halves and used as explants for the genetic transformation experiments.

Vector constructs

All the vectors used in this work were assembled based on Modular Cloning-compatible vectors (Table S5) (Weber et al., 2011). In

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2021), doi: 10.1111/tpj.15584

short, to construct pEGFP, 2x*35S*::*nptII* and 2x*35S*::*EGFP* transcriptional units were ligated at positions 1 and 2, respectively, into pAGM8031 by Golden Gate assembly. Suitable target sequences of sgRNAs (Table S6) were chosen using the online web tool CRISPR-P (Liu et al., 2017). Complementary oligos with target sequences were synthesized by Microsynth AG (Balgach, Switzerland). Oligonucleotide pairs were annealed and cloned into an intermediate vector (pDI1, pDI2, pDI3, and pDI4E) with *Bpi*I, yielding *AtU6-26*::*sgRNA* expression cassettes. The 2x*35S*::*HPT* transcriptional unit and the 2x*35S*::*SpCas9* transcriptional unit together with four sgRNA (sg1, sg2, sg5, sg6) expression cassettes were ligated into pAGM8031 by Golden Gate assembly, yielding the final construct pBraj1256. The 2x*35S*::*HPT* transcriptional unit and the 2x*35S*::*SpCas9* transcriptional unit and four sgRNA (sg3, sg4, sg7A, sg7B) expression cassettes were ligated into pAGM8031 by Golden Gate assembly, resulting in the final construct pBraj3477. The final constructs were transformed into *A. tumefaciens* strain LBA4404 for transformation of plants. All oligonucleotides used in this work are listed in Table S7.

Transformation of mustard plants

Glycerol stocks of the *A. tumefaciens* strain LBA4404 harboring the binary vectors pEGFP, pBrj1256, and pBrj3477 were plated on YEB medium (0.5% [w/v] sucrose, 0.5% [w/v] tryptone, 0.5% [w/v] yeast extract, 0.5% [w/v] beef extract, 0.05% [w/v] magnesium sulfate, 1.4% [w/v] micro agar, pH 7.2) supplemented with 100 mg L^{-1} rifampicin and 100 mg L^{-1} spectinomycin and incubated for 24 h at 28°C. The bacteria were then collected from the plates and dissolved in simplified induction medium (Alt-Moerbe et al., 1988) containing 20 mM sodium citrate and 2% sucrose (pH 5.2) supplemented with 100 μM acetosyringone and 1 mM betaine hydrochloride (James et al., 1993) and incubated at room temperature for 4 h until an OD_{600} of 1.8 to 2.0 in the 1:10 dilution was reached. Medium without bacteria was used for control variants. The cotyledon explants were immersed in the different solutions supplemented with 0.03% Silwet L-77 for 10 min, blotted dry on sterile paper towel, and co-cultivated on shoot induction medium (full-strength MS medium supplemented with 3% [w/v] sucrose, 0.8% [w/v] agar, 8.88 μM benzyladenine, 5.37 μM naphthalene acetic acid, and 100 μM acetosyringone) for 3 days. After co-cultivation, the explants were washed in sterile deionized water and placed onto regeneration medium (shoot induction medium + 10 μM $AgNO_3$) and the different selection media depending on the construct (shoot induction medium + 10 μM $AgNO_3$ + 15 mg L^{-1} hygromycin for Terratop or 20 mg L^{-1} hygromycin for CR2664 with the CRISPR/Cas9 constructs and 50 mg L^{-1} kanamycin for both lines with the EGFP construct) for shoot organogenesis at 24 ± 1°C with a 16/8 h day/night cycle (35–40 μmol m^{-2} sec^{-1}, fluorescent tubes). After 28 days, the obtained shoots were separated and cultivated on the same media without $AgNO_3$ for 28 days. Surviving shoots were subcultured every 4 weeks for another 56 days. Four months after co-culture, putative transformants (T_0) were obtained. The plants from the non-transformed control (wild type) and the T_0 transgenic plants were acclimatized and grown in 5-L pots containing growing medium (3:1 mixture of peat and lime-free clay and fertilized with 1 ‰ Ferty3 Mega produced by Einheitserdewerke Patzer, Patzer Erden GmbH, Germany) under a 16/8 h day/night cycle using a Philips SON-T Agro 400 W at 22°C in the greenhouse for 4 months. Before anthesis of the first flowers, the plants were covered with isolation bags (Crispac-Beutel SM 330 × 750 mm, Baumann Saatzuchtbedarf GmbH, Germany) and shaken from time to time to improve self-pollination. Where possible, 20 flowers per plants were emasculated before anthesis and the anthers with mature pollen grains were collected from other flowers of the same plant for pollination (Table S4). At maturity, seeds were harvested to obtain the seeds for the T_1 generation.

Seed germination

The germination rates of seeds from wild-type, T_0, and T_1 transgenic plants were determined *in vitro* and in the greenhouse. For *in vitro* germination, 10 seeds from each plant were randomly chosen, subjected to surface disinfection, and incubated in darkness for 7 days at 24°C. A seed was scored as germinated 7 days after sowing when the radicle and the hypocotyl had emerged. For *ex vitro* germination, 20 seeds from wild-type, T_0, and T_1 transgenic plants were directly sown into multipot trays or 8-cm pots containing growing medium (3:1 mixture of peat and lime-free clay and fertilized with 1 ‰ Ferty3 Mega produced by Einheitserdewerke Patzer, Patzer Erden GmbH, Germany) under a 16/8 h day-night cycle using a Philips SON-T Agro 400 W, at 22°C in the greenhouse. The germination was scored after 21 days when the cotyledons had emerged above the substrate.

Seed characterization

The harvested seeds from each plant were counted with a counting machine (Contador, Pfeuffer, Kitzingen, Germany) to quantify the seed set per plant. Where possible, four times 100 seeds from each individual plant were randomly selected and the weight was determined using a precision balance (Denver Instrument SI-203.1, Cole-Parmer GmbH, Germany). Later, 15 to 30 seeds were randomly selected and photographed with a digital microscope (Figure 3) (VHX-S750E, Keyence, Japan).

Statistical analysis was conducted on the 100-seed weight using R (R Core Team, 2021) with the help of the packages tidyverse (Wickham et al., 2019) and dplyr (Hadley et al., 2021). The R package emmeans was also used to conduct pairwise comparisons (Dunnett test) (Lenth, 2021).

Identification of mutants

Frozen leaf samples were disrupted with a TissueLyser (QIAGEN GmbH, Germany) and genomic DNA was extracted using the innuPREP Plant DNA Kit (Analytik Jena, Germany) according to the instructions of the manufacturer. Three pairs of primers (hyg-F/hyg-R, F1/R1, and F2/R2) were used to detect transgenic mustards with T-DNA integrated into the genome. Subsequently, *Bra j IA* and *Bra j IB* were amplified by Q5 High-Fidelity Polymerase (NEB) from genomic DNA with site-specific primer pairs BrajA-F/BrajA-R and BrajB-F/BrajB-R (Table S7), respectively. The PCR amplicons were purified from the agarose gels using the GeneJET Gel Extraction Kit (Thermo Scientific Fermentas, USA) and sequenced (Microsynth Seqlab, Göttingen, Germany) or subcloned into pUC57 vector before sequencing. The Sanger sequencing chromatograms were decoded manually and using the ICE Analysis tool from SYNTHEGO (https://ice.synthego.com).

Generation of a monoclonal Bra j I-specific antibody

Antibody selection was performed as described previously (Russo et al., 2018) with modifications. In brief, antibodies were selected by phage display using the naïve antibody gene libraries HAL9 and HAL10 (Kügler et al., 2015) on a biotinylated peptide (Monsalve et al., 1993) (Peps4LS GmbH, Heidelberg, Germany) in microtiter plates (Corning, New York, USA). Selected antibodies were cloned into scFv-Fc format and were produced in EXPI293F cells as previously described (Wenzel et al., 2020). Titration ELISA

was performed with the selected antibodies. For a more detailed description see Supplementary Material.

Immunoblotting

Seeds were ground into a fine powder with 25% (w/v) extraction buffer (0.2 M Tris-Cl, pH 8.0, containing 0.1 M NaCl, 10 mM EDTA, 0.1% $MgCl_2$, 10% glycerol, 1 mM PMSF, and 10 nM β-mercaptoethanol) and centrifuged at 14 000 ***g*** for 10 min at 4°C. The supernatants were transferred to new tubes and centrifuged again at 11 304 ***g*** for 10 min at 4°C. These supernatants were used for further analysis. Total protein concentrations were measured using Bradford reagent (Carl Roth GmbH, Karlsruhe, Germany). Samples were boiled with loading dye at 95°C for 10 min. Protein samples were separated by 15% SDS-PAGE and visualized by staining with Coomassie Brilliant Blue R-250. For Bra j I detection, proteins were transferred onto 0.45 μm nitrocellulose membranes using a wet blotting system. Membranes were blocked overnight at 4°C in blocking buffer (TBST buffer with 5% defatted milk and 3% bovine serum albumin). Subsequently, membranes were incubated with 100 ng ml^{-1} primary antibody for 3 to 4 h at room temperature. After washing the membranes three times for 5 min each in TBST buffer, they were incubated with 1:20 000 diluted alkaline phosphatase-conjugated goat anti-human IgG secondary antibody (Dianova 109-055-98) for 1 h at room temperature. Then, the membranes were washed three times with TBST buffer for 5 min each and equilibrated with substrate buffer (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) for 10 min. Thereafter, the membranes were stained with NBT/BCIP substrate staining solution and the staining reaction was stopped by washing three times with H_2O after the signal developed. Images were taken by a ChemiDoc™ Touch Imaging System (Bio-Rad).

Off-target detection

Off-target sites were predicted by the online tool CRISPR-P (Liu et al., 2017). Based on the off-target score, the top three predicted off-target sites for each of the eight used sgRNAs were selected as potential off-target sites (Figure S8). Those sites were amplified from the genome of two selected T_1 plants by PCR and the PCR product was sequenced. Primer pairs for off-target amplification are added to Table S7.

ACKNOWLEDGEMENTS

This work was supported by funding from the Federal Ministry of Education and Research (BMBF LACOP consortium; grant number 031B0348). The authors thank Ewa Schneider for excellent technical assistance, David Wamhoff for help in statistics, and Jana Streubel for MoClo-compatible genome editing vectors. Open access funding enabled and organized by ProjektDEAL.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

TW, JB, and MH conceived the study and its design, coordinated the research together with TR, and assisted with interpretations of results. JA, DZ, KR, and SS designed and performed the experiments and analyzed the data. JA, DZ, TW, JB, and KR wrote the manuscript. TR provided the recombinant Bra j I protein.

DATA AVAILABILITY STATEMENT

All relevant data can be found within the article and supporting materials.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Sequence alignment of *Bra j I* homoeologs in *B. juncea* varieties Terratop and CR2664.

Figure S2. Outcomes of the PCR test for transgenicity of all obtained mustard lines.

Figure S3. PCR amplification of both *Bra j I* homoeologs from genomic DNA of the T_0 lines.

Figure S4. Analysis of deletions in *Bra j I* in mutated T_0 lines.

Figure S5. Analysis of the *Bra j I* homoeologs of T_1 offsprings to screen for inheritance of the deletion from the T0-1 line.

Figure S6. Amino acid sequence alignment of Bra j I homoeologs in *B. juncea* varieties Terratop and CR2664.

Figure S7. Titration-ELISA for EC_{50} determination on Bra j I peptide.

Figure S8. Off-target detection.

Table S1. Transformation efficiencies for two *B. juncea* lines (Terratop and CR2664) in different experiments using the three vectors pEGFP, pBraj1256, and pBraj3477.

Table S2. Genetic characterization of T_0 plants regarding transgene presence, seed weight, and germination of seeds of the *B. juncea* lines Terratop and CR2664.

Table S3. Genetic characterization of T_1 plants regarding transgene presence, seed weight, and germination of seeds of the *B. juncea* lines Terratop and CR2664.

Table S4. Controlled pollinated flowers of T_0 plants of *B. juncea* lines Terratop and CR2664 with the number of seeds and seed weight.

Table S5. MoClo-compatible vectors used in this study.

Table S6. sgRNA sequences.

Table S7. Oligonucleotides used in this study.

Supplementary Material. Antibody selection in microtiter plate.

REFERENCES

Ahmed, I., Yadav, D., Shukla, P., Vineeth, T.V., Sharma, P.C. & Kirti, P.B. (2017) Constitutive expression of *Brassica juncea* annexin, *AnnBj2* confers salt tolerance and glucose and ABA insensitivity in mustard transgenic plants. *Plant Science*, **265**, 12–28.

Alt-Moerbe, J., Neddermann, P., von Lintig, J., Weiler, E.W. & Schröder, J. (1988) Temperature-sensitive step in Ti plasmid *vir*-region induction and correlation with cytokinin secretion by *Agrobacteria*. *Molecular and General Genetics*, **213**, 1–8.

Anzalone, A.V., Koblan, L.W. & Liu, D.R. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature Biotechnology*, **38**, 824–844.

Augustine, R., Mukhopadhyay, A. & Bisht, N.C. (2013) Targeted silencing of *BjMYB28* transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*. *Plant Biotechnology Journal*, **11**, 855–866.

Bewley, J.D., Bradford, K. & Hilhorst, H. (2012) *Seeds: physiology of development, germination and dormancy*. Springer Science & Business Media.

Beyer, K., Bardina, L., Grishina, G. & Sampson, H.A. (2002) Identification of sesame seed allergens by 2-dimensional proteomics and Edman sequencing: seed storage proteins as common food allergens. *The Journal of Allergy and Clinical Immunology*, **110**, 154–159.

The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd.,
The Plant Journal, (2021), doi: 10.1111/tpj.15584

Bhuiyan, M.S.U., Min, S.R., Jeong, W.J., Sultana, S., Choi, K.S., Lim, Y.P. *et al.* (2011) An improved method for *Agrobacterium*-mediated genetic transformation from cotyledon explants of *Brassica juncea*. *Plant Biotechnology*, **28**, 17–23.

Breiteneder, H. & Ebner, C. (2000) Molecular and biochemical classification of plant-derived food allergens. *The Journal of Allergy and Clinical Immunology*, **106**, 27–36.

Breiteneder, H. & Radauer, C. (2004) A classification of plant food allergens. *The Journal of Allergy and Clinical Immunology*, **113**, 821–830.

Chandran, M., Chu, Y., Maleki, S.J. & Ozias-Akins, P. (2015) Stability of transgene expression in reduced allergen peanut (*Arachis hypogaea* L.) across multiple generations and at different soil sulfur levels. *Journal of Agriculture and Food Chemistry*, **63**, 1788–1797.

Chen, K., Wang, Y., Zhang, R., Zhang, H. & Gao, C. (2019) CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annual Review of Plant Biology*, **70**, 667–697.

Debeaujon, I., Leon-Kloosterziel, K.M. & Koornneef, M. (2000) Influence of the testa on seed dormancy, germination, and longevity in Arabidopsis. *Plant Physiology*, **122**, 403–414.

Dodo, H.W., Konan, K.N., Chen, F.C., Egnin, M. & Viquez, O.M. (2008) Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen *Ara h 2* leads to its significant reduction and a decrease in peanut allergenicity. *Plant Biotechnology Journal*, **6**, 135–145.

Dutta, I., Saha, P. & Das, S. (2008) Efficient *Agrobacterium*-mediated genetic transformation of oilseed mustard [*Brassica juncea* (L.) Czern.] using leaf piece explants. *In Vitro Cellular & Developmental Biology—Plant*, **44**, 401–411.

El-Maarouf-Bouteau, H., Meimoun, P., Job, C., Job, D. & Bailly, C. (2013) Role of protein and mRNA oxidation in seed dormancy and germination. *Frontiers in Plant Science*, **4**, 77.

Fazekas, G.A., Sedmach, P.A. & Palmer, M.V. (1986) Genetic and environmental effects on in vitro shoot regeneration from cotyledon explants of *Brassica juncea*. *Plant Cell Tissue Organ Culture*, **6**, 177–180.

Gao, C. (2021) Genome engineering for crop improvement and future agriculture. *Cell*, **184**, 1621–1635.

González de la Peña, M.A., Menéndez-Arias, L., Monsalve, R.I. & Rodríguez, R. (1991) Isolation and characterization of a major allergen from oriental mustard seeds, Bra jI. *International Archives of Allergy and Applied Immunology*, **96**, 263–270.

Hadley, W., Romain, F., Lionel, H. & Kirill, M. (2021) dplyr: *A Grammar of Data Manipulation*. R package version 1.0.7. CRAN—Package dplyr.

Hanschen, F.S. & Winkelmann, T. (2020) Biofumigation for fighting replant disease-A Review. *Agronomy*, **10**, 425.

James, D.J., Uratsu, S., Cheng, J., Negri, P., Viss, P. & Dandekar, A.M. (1993) Acetosyringone and osmoprotectants like betaine or proline synergistically enhance *Agrobacterium*-mediated transformation of apple. *Plant Cell Reports*, **12**, 559–563.

Kügler, J., Wilke, S., Meier, D., Tomszak, F., Frenzel, A., Schirrmann, T. *et al.* (2015) Generation and analysis of the improved human HAL9/10 antibody phage display libraries. *BMC Biotechnology*, **15**, 10.

L'Hocine, L., Pitre, M. & Achouri, A. (2019) Detection and identification of allergens from Canadian mustard varieties of *Sinapis alba* and *Brassica juncea*. *Biomolecules*, **9**, 489.

Labana, K.S. & Gupta, M.L. (1993) *Importance and Origin. In Breeding Oilseed Brassicas*. Berlin, Heidelberg: Springer, pp. 1–7. https://doi.org/10.1007/978-3-662-06166-4_1

Lee, K.P., Piskurewicz, U., Turečková, V., Strnad, M. & Lopez-Molina, L. (2010) A seed coat bedding assay shows that RGL2-dependent release of abscisic acid by the endosperm controls embryo growth in Arabidopsis dormant seeds. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 19108–19113.

Lenth, R.V. (2021) emmeans: estimated marginal means, aka Least-Squares Means. R Package Version, 1.6.2-1. https://CRAN.R-project.org/package=emmeans

Li, C., Hao, M., Wang, W., Wang, H., Chen, F., Chu, W. *et al.* (2018) An efficient CRISPR/Cas9 platform for rapidly generating simultaneous mutagenesis of multiple gene homoeologs in allotetraploid oilseed rape. *Frontiers in Plant Science*, **9**, 442.

Li, X., Jiang, D.H., Yong, K. & Zhang, D.B. (2007) Varied transcriptional efficiencies of multiple *Arabidopsis* U6 small nuclear RNA genes. *Journal of Integrative Plant Biology*, **49**, 222–229.

Lionneton, E., Aubert, G., Ochatt, S. & Merah, O. (2004) Genetic analysis of agronomic and quality traits in mustard (*Brassica juncea*). *Theoretical and Applied Genetics.*, **109**, 792–799.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K. & Chen, L.L. (2017) CRISPR-P 2.0: An improved CRISPR-Cas9 tool for genome editing in plants. *Molecular Plant*, **10**, 530–532.

Lyzenga, W.J., Harrington, M., Bekkaoui, D., Wigness, M., Hegedus, D.D. & Rozwadowski, K.L. (2019) CRISPR/Cas9 editing of three CRUCIFERIN C homoeologues alters the seed protein profile in Camelina sativa. *BMC Plant Biology*, **19**, 1–16.

Manghwar, H., Lindsey, K., Zhang, X. & Jin, S. (2019) CRISPR/Cas system: recent advances and future prospects for genome editing. *Trends in Plant Science*, **24**, 1102–1125.

Matsuo, H., Yokooji, T. & Taogoshi, T. (2015) Common food allergens and their IgE-binding epitopes. *Allergology International*, **64**, 332–343.

Mattner, S.W., Porter, I.J., Gounder, R.K., Shanks, A.L., Wren, D.J. & Allen, D. (2008) Factors that impact on the ability of biofumigants to suppress fungal pathogens and weeds of strawberry. *Crop Protection*, **27**, 1165–1173.

Menéndez-Arias, L.I., Dominguez, J. & Rodriguez, R. (1988) Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, Sin a I. *European Journal of Biochemistry*, **177**, 159–166.

Metje-Sprink, J., Sprink, T. & Hartung, F. (2020) Genome-edited plants in the field. *Current Opinion in Biotechnology*, **61**, 1–6.

Mondal, K.K., Bhattacharya, R.C., Koundal, K.R. & Chatterjee, S.C. (2007) Transgenic Indian mustard (*Brassica juncea*) expressing tomato glucanase leads to arrested growth of *Alternaria brassicae*. *Plant Cell Reports*, **26**, 247–252.

Monsalve, R.I., Gonzalez De La Peña, M.A., Menendez-Arias, L., Lopez-Otin, C., Villalba, M. & Rodriguez, R. (1993) Characterization of a new orientalmustard (*Brassica juncea*) allergen, Bra j IE: detection of an allergenic epitope. *The Biochemical Journal*, **293**, 625–632.

Monsalve, R.I., Villalba, M. & Rodríguez, R. (2001) Allergy to mustard seeds: the importance of 2S albumins as food allergens. *Internet Symposium Food Allergens*, **3**, 57–69.

Muntz, K., Belozersky, M.A., Dunaevsky, Y.E., Schlereth, A. & Tiedemann, J. (2001) Stored proteinases and the initiation of storage protein mobilization in seeds during germination and seedling growth. *Journal of Experimental Botany*, **52**, 1741–1752.

Murashige, T. & Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum*, **15**, 473–497.

Navet, N. & Tian, M. (2020) Efficient targeted mutagenesis in allotetraploid sweet basil by CRISPR/Cas9. *Plant Direct*, **4**, e00233.

Nguyen, T.P., Cueff, G., Hegedus, D.D., Rajjou, L. & Bentsink, L. (2015) A role for seed storage proteins in *Arabidopsis* seed longevity. *Journal of Experimental Botany*, **66**, 6399–6413.

Nordlee, J.A., Taylor, S.L., Townsend, J.A., Thomas, L.A. & Bush, R.K. (1996) Identification of a Brazil-nut allergen in transgenic soybeans. *New England Journal of Medicine*, **334**, 688–692.

Owens, D.D.G., Caulder, A., Frontera, V., Harman, J.R., Allan, A.J., Bucakci, A. *et al.* (2019) Microhomologies are prevalent at Cas9-induced larger deletions. *Nucleic Acids Research*, **47**, 7402–7417.

Pelé, A., Rousseau-Gueutin, M. & Chèvre, A.M. (2018) Speciation success of polyploid plants closely relates to the regulation of meiotic recombination. *Frontiers in Plant Science*, **9**, 907.

Premi, O.P., Kandpal, B.K., Rathore, S.S., Shekhawat, K. & Chauhan, J.S. (2013) Green manuring, mustard residue recycling and fertilizer application affects productivity and sustainability of Indian mustard (*Brassica juncea* L.) in Indian semi-arid tropics. *Industrial Crops and Products*, **41**, 423–429.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Ramana, S. & Ghildiyal, M.C. (1997) Contribution of leaf photosynthesis towards seed yield in *Brassica* species. *Journal of Agronomy and Crop Science*, **178**, 185–187.

Rancé, F., Dutau, G. & Abbal, M.M. (2000) Mustard allergy in children. *Allergy*, **55**, 496–500.

Rani, S., Sharma, V., Hada, A., Bhattacharya, R.C. & Koundal, K.R. (2017) Fusion gene construct preparation with lectin and protease inhibitor

genes against aphids and efficient genetic transformation of *Brassica juncea* using cotyledons explants. *Acta Physiologiae Plantarum*, **39**, 115.

Redden, R., Vardy, M., Edwards, D., Raman, H. & Batley, J. (2009) *Genetic and morphological diversity in the Brassicas and wild relatives*. 16th Australian Research Assembly on Brassicas. Ballarat, Victoria.

Rolletschek, H., Schwender, J., König, C., Chapman, K.D., Romsdahl, T., Lorenz, C. *et al.* (2020) Cellular plasticity in response to suppression of storage proteins in the *Brassica napus* embryo. *The Plant Cell*, **32**, 2383–2401.

Russo, G., Meier, D., Helmsing, S., Wenzel, E., Oberle, F., Frenzel, A. *et al.* (2018) Parallelized antibody selection in microtiter plates. *Methods in Molecular Biology*, **1701**, 273–284.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C.V., Giménez, M.J., Sousa, C., Voytas, D.F. *et al.* (2018) Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnology Journal*, **16**, 902–910.

Scuffi, D., Álvarez, C., Laspina, N., Gotor, C., Lamattina, L. & García-Mata, C. (2014) Hydrogen sulfide generated by L-cysteine desulfhydrase acts upstream of nitric oxide to modulate abscisic acid-dependent stomatal closure. *Plant Physiology*, **166**, 2065–2076.

Sharma, M., Sahni, R., Kansa, I.R. & Kaundal, K.R. (2004) Transformation of oilseed mustard *Brassica juncea* (L.) Czern & Coss cv. Pusajaikisan with snowdrop lectin gene. *Indian Journal of Biotechnology*, **3**, 97–102.

Shewry, P.R., Napier, J.A. & Tatham, A.S. (1995) Seed storage proteins: structures and biosynthesis. *The Plant Cell*, **7**, 945.

Singh, V.V., Pareek, A.K., Mathur, M., Yadav, R., Goyal, P., Thakur, A.K. *et al.* (2009) Optimization and development of regeneration and transformation protocol in Indian mustard using lectin gene from chickpea [*Cicer arietinum* (L.)]. *Journal of Plant Breeding and Crop Science*, **1**, 306–310.

Sinha, S., Jha, J.K., Maiti, M.K., Basu, A., Mukhopadhyay, U.K. & Sen, S.K. (2007) Metabolic engineering of fatty acid biosynthesis in Indian mustard (*Brassica juncea*) improves nutritional quality of seed oil. *Plant Biotechnology Reports*, **1**, 185–197.

Spect, C.E. & Diederichsen, A. (2001) Brassica. In: *Mansfeld's Encyclopedia of Agricultural and Horticultural Crops*. Vol **6** (Hanelt, P., Ed). Berlin, Heidelberg, New York: Springer-Verlag, pp. 1453–1456.

Stevens, O.A. (1932) The number and weight of seeds produced by weeds. *American Journal of Botany*, **19**, 784–794.

Stevens, O.A. (1957) Weights of seeds and numbers per plant. *Weeds*, **5**, 46–55.

Sugliani, M., Rajjou, L., Clerkx, E.J.M., Koornneef, M. & Soppe, W.J.J. (2009) Natural modifiers of seed longevity in the *Arabidopsis* mutants *abscisic acid insensitive3-5* (*abi3-5*) and *leafy cotyledon1-3* (lec1-3). *New Phytologist*, **184**, 898–908.

Tan, J., Zhao, Y., Wang, B., Hao, Y., Wang, Y., Li, Y. *et al.* (2020) Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. *Plant Biotechnology Journal*, **18**, 2161–2163.

Tang, T., Chen, G., Bu, C., Liu, F., Liu, L. & Zhao, X. (2018) Transgene introgression from *Brassica napus* to different varieties of *Brassica juncea*. *Plant Breed*, **137**, 171–180.

Teuber, S.S., Dandekar, A.M., Peterson, W.R. & Sellers, C.L. (1998) Cloning and sequencing of a gene encoding a 2S albumin seed storage protein precursor from English walnut (*Juglans regia*), a major food allergen. *The Journal of Allergy and Clinical Immunology*, **101**, 807–814.

Thakur, A.K., Parmar, N., Singh, K.H. & Nanjundan, J. (2020) Current achievements and future prospects of genetic engineering in Indian mustard (*Brassica juncea* L. Czern & Coss.). *Planta*, **252**, 56.

Tuladhar, R., Yeu, Y., Piazza, J.T., Tan, Z., Clemenceau, J.R. & Xiafeng, W. et al. (2019) CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nature Communications*, **10**, 4056.

Vaughan, J.G., Hemingway, J.S. & Schofield, H.J. (1963) Contributions to a study of variation in *Brassica juncea* Coss. & Czern. *Botanical Journal of the Linnean Society*, **58**, 435–447.

Wang, P., Zhang, J., Sun, L., Ma, Y., Xu, J., Liang, S. *et al.* (2017) High efficient multisites genome editing in allotetraploid cotton (*Gossypium hirsutum*) using CRISPR/Cas9 system. *Plant Biotechnology Journal*, **16**, 137–150.

Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS One*, **6**, e16765.

Wenzel, E.V., Bosnak, M., Tierney, R., Schubert, M., Brown, J., Dübel, S. *et al.* (2020) Human antibodies neutralizing diphtheria toxin *in vitro* and *in vivo*. *Scientific Reports*, **10**, 571.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R. *et al.* (2019) Welcome to the Tidyverse. *Journal of Open Source Software*, **4**(43), 1686.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B. *et al.* (2016) The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics*, **48**, 1225–1232.

Yim, B., Hanschen, F.S., Wrede, A., Nitt, H., Schreiner, M., Smalla, K. *et al.* (2016) Effects of biofumigation using *Brassica juncea* and *Raphanus sativus* in comparison to disinfection using Basamid on apple plant growth and soil microbial communities at three field sites with replant disease. *Plant and Soil*, **406**, 389–408.

Yoshimura, Y., Tomizono, S. & Matsuo, K. (2016) Seed production of wild *Brassica juncea* on riversides in Japan. *Japan Agricultural Research Quarterly*, **50**, 335–343.

Yusuf, M.A. & Sarin, N.B. (2007) Antioxidant value addition in human diets: genetic transformation of *Brassica juncea* with *γ-TMT* gene for increased α-tocopherol content. *Transgenic Research*, **16**, 109–113.

Zhang, D., Zhang, H., Li, T., Chen, K., Qiu, J.L. & Gao, C. (2017) Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biology*, **18**, 191.

The Plant Journal, (2021), doi: 10.1111/tpj.15584

3 Targeted C•G-to-T•A base editing with TALE cytosine deaminases in plants

Dingbo Zhang, Vanessa Pries, Jens Boch*

Leibniz Universität Hannover, Institute of Plant Genetics, Herrenhäuser Str. 2, 30419

Hannover, Germany

* corresponding authors

Journal: Plant Biotechnology Journal

Submitted: Submitted

Accepted: -

Available online: -

Author	Contributions
Dingbo Zhang	Designed and performed all the experiments, analyzed all the data, contributed to the writing of the manuscript
Vanessa Pries	Assisted with the plasmids cloning
Jens Boch	Conceived the studies and designed experiments

Targeted C•G-to-T•A base editing with TALE cytosine deaminases in plants

Dingbo Zhang, Vanessa Pries, Jens Boch

Leibniz Universität Hannover, Institute of Plant Genetics, Herrenhäuser Str. 2, 30419 Hannover, Germany

Abstract

TALE-derived DddA-based cytosine base editors (TALE-DdCBEs) use programmable DNA-binding transcription activator-like effector (TALE) arrays, split double-strand DNA cytidine deaminase (DddA), and an uracil glycosylase inhibitor to catalyze C•G-to-T•A editing in organelle and nuclear DNA. To investigate the editing rules of TALE-DdCBEs, we constructed a series of TALE arrays flanking the target cytosine in β-glucuronidase (GUS) reporter constructs. Compared to canonical DddA, TALE-DdCBEs containing evolved DddA variants (DddA6 or DddA11) showed a significant improvement in editing efficiency in *Nicotiana benthamiana* and rice, moreover, TALE-DdCBEs containing DddA11 have broader sequence compatibility for non-TC targets editing. We also found that the spontaneous assembly of split DddA halves can cause off-target editing by TALE-DdCBEs in plants. Our study expands the base editing toolbox in plants. The further development of high-fidelity double-strand DNA cytidine deaminase is highly beneficial for gene therapy and crop improvement.

Keywords: DddA, genome editing, deaminase, base editors

Introduction

Genome-editing technologies are rapidly revolutionizing plant breeding. Base editor is a significant innovation in the genome editing field. Instead of generating double-stranded DNA (dsDNA) breaks, base editors utilize DNA deaminases to precisely incorporate single nucleotide variants (SNVs) into the genome (Gu *et al*., 2021). Current base editors generally contain a catalytically impaired CRISPR-Cas nuclease linked to a single-stranded DNA (ssDNA) deaminase enzyme. Cytosine base editors (CBEs) catalyze the transition mutation of C•G-to-T•A through the deamination of deoxycytidine to deoxyuridine, the U•G mismatch can be repaired to U•A and results in a T•A base pair. Conjugating the uracil glycosylase inhibitor (UGI) to CBEs increases the editing efficiency and purity (Komor *et al*., 2016; Nishida *et al*., 2016; Anzalone *et al*., 2020). In addition, C•G-to-G•C base editors (CGBEs) have been developed by replacing the UGI with uracil N-glycosylase (UNG) in the CBE architecture (Kurt *et al*., 2021; Zhao *et al*., 2021). Adenine base editors (ABEs) catalyze the transition mutation of A•T-to-G•C through the deamination of deoxyadenosine to deoxyinosine, which has the base-pairing specificity of guanines (G) (Gaudelli *et al*., 2017).

Although the CRISPR/Cas-derived base editors have proven to provide efficient and precise introduction of SNVs in the nuclear genome, repurposing them for organellar (mitochondria and plastids) DNA editing is challenging due to the lack of methods for delivering sgRNAs to the mitochondria and plastids. However, protein-only genome editing systems based on DNA-binding proteins such as zinc fingers (ZFs) and transcription activator-like effectors (TALEs) can be used as organellar genome editors (Bacman *et al*., 2013; Gammage *et al*., 2014). Remarkably, the discovery of the cytosine deaminase $DddA_{tox}$ from *Burkholderia cenocepacia* that deaminates cytosines in dsDNA enabled the C•G-to-T•A conversions in human nuclear and mitochondrial DNA (Mok *et al*., 2020). To generate TALE-derived DddA-based cytosine base editors (TALE-DdCBEs), the toxic $DddA_{tox}$ was split into two inactive halves DddA-N and DddA-C. Similar to the *Fok*I monomer assembly in TALE nucleases (TALENs), the two DddA halves reconstitutes the active enzyme when assembled by two adjacent tail-to-tail TALE array. Transit peptides like nuclear localization signal or mitochondrial targeting signal guide them to their sub-cellular target locus and the fusion of UGI to the C terminus of the TALE-DdCBEs

increases their editing efficiency and reduces byproducts by inhibiting uracil-DNA glycosylase. Instead of using the split DddA halves, non-toxic, full-length DddA variants were developed to make monomeric TALE-DdCBEs (DddA guided by one TALE protein) which also allow C•G-to-T•A editing in mitochondrial DNA (Mok *et al*., 2022a). To improve the deamination activity and expand the strict 5′-TC context limitation of DddA, evolved DddA variants (DddA6 and DddA11) with improved activity and expanded targeting scope were created by protein engineering (Mok *et al*., 2022b). In plants, TALE-DdCBEs were successfully used for editing the plastid genomes of Arabidopsis thaliana and rice chloroplasts (Nakazato *et al*., 2021; Li *et al*., 2021), as well as the chloroplast and mitochondrial DNA of lettuce and rapeseed (Kang *et al*., 2021). Besides CBEs, TALE-based ABEs (TALEDs) have recently also been developed to perform mitochondrial A•T-to-G•C base editing in mammalian mitochondria (Cho *et al*., 2022) and *Arabidopsis* chloroplast genes (Mok *et al*., 2022c).

In this study, we developed a modular cloning (MoClo) pipeline for TALE-DdCBEs assembly, and established a simple β-glucuronidase (GUS) reporter assay in *Nicotiana benthamiana* for C•G-to-T•A editing efficiency evaluation to evaluate the dsDNA editing efficiency and specificity of TALE-DdCBEs in plants.

Results

An optimized architecture for TALE-DdCBEs

We first developed a β-glucuronidase (GUS) reporter system in the mode plant *Nicotiana benthamiana* to quickly assess the C•G-to-T•A conversion efficiencies of various TALE-DdCBEs architectural designs. GUS^{G537} was constructed to contain a missense mutation of glutamic acid (G<u>A</u>A) to glycine (G<u>G</u>A) in the active center (Islam *et al*., 1999). A C•G-to-T•A conversion could change the glycine back to glutamic acid and restore the GUS enzymatic activity (Figure 1A and Figure S1).

TALEs are often used as genome editing tools with a truncation of the N-terminal domain by 152 amino acids (retaining 135 amino acids), because this truncation is tolerated without sacrifcing the general DNA-binding capacity of the N-terminal domain (Kay *et al*., 2007). Nevertheless, the full-length N-terminal TALE domain (288 amino acids) confers stronger DNA-

binding (Zhang *et al*., 2011), and Schreiber *et al.* propose an N-terminal domain of 196 amino acids as optimal (Schreiber *et al*., 2015; Schreiber *et* al., 2019). Hence, we employed three different lengths of N-terminals and six different lengths of C-terminals domian to develop a series of TALE-DdCBEs containing DddA halves split at G1397 (Figure 1B). The activities of these TALE-DdCBEs were examined using the GUS^{G537} reporter via Agrobacterium-mediated transient expression in *N. benthamiana* leaves. These TALE-DdCBEs showed different GUS activities after normalization to a constitutively expressed functional GUS (WT GUS) (Figure 1C). Three different N-terminal architectures (N288, N196, and N135) showed comparable GUS activity when coupled with the same C-terminal domain (C47). Therefore, N196 (196 amino acids length with a deletion of 93 amino acids from the full-length N-terminus) was chosen for all subsequent studies. Similar truncation analysis of the C terminal domain demonstrated that N196/C17 (C17: C-terminal of 17 amino acids length) and N196/C96 (C96: C-terminal of 96 amino acids length) have the highest GUS activity with an average activity of 73.9% and 80.6%, respectively. Whereas N196/C247 (C247: 247 amino acids length of full-length C terminal domian without the native TALE activation domain) showed the lowest GUS activity. To minimize the size of TALE-DdCBEs and maintain its high editing efficiency, we selected N196/C17 as the TALE N-terminal and C-terminal domian combination for the following experiments.

Characterization of the TALE-DdCBEs editing window

To comprehensively investigate how a pair of TALE-DdCBEs should be positioned to modify a specific target base, we constructed TALE arrays of different lengths flanking the targeted cytosine in the GUS^{G537} reporter (Figure 2A). The different combinations of left TALE and right TALE enable the evaluation of varied widths of spacer regions ranging from 1-nt to 16-nt in length, as well as varying the position of the targeted cytosine within the spacer from position 1 to position 8 (C1 to C8) (Figure 2B). The DddA-C half was fused to the left TALE (left TALE-DddA-C) and combined with the DddA-N half fused to the right TALE (right TALE-DddA-N) or the opposite way around (left TALE-DddA-N and right TALE-DddA-C). We tested the editing efficiencies of those TALE-DdCBEs in *N. benthamiana* with the GUS^{G537} reporter. When using

the left TALE-DddA-C / right TALE-DddA-N architecture, the highest efficiency was achieved at the targeted cytosine positioned at C4, C5, or C6 across the 1-nt to 16-nt spacer (Figure 2C). For the left TALE-DddA-N / right TALE-DddA-C architecture, the highest editing efficiency was yielded at the targeted cytosine located at C5 or C6 across the 1-nt to 16-nt spacer (Figure 2D). When the spacer length was less than 4-nt or the targeted cytosine at C7 or C8, the editing efficiencies were dramatically decreased using either one of the combinations. These results indicated that TALE-DdCBEs prefer the target cytosine located at C5 or C6.

Likewise, to further characterize the optimal spacer length for TALE-DdCBEs, we analyzed the editing efficiencies of TALE-DdCBEs targeting the cytosine located at C4 or C5 or C6 within different lengths of spacing regions. The GUS activities showed that, for the left TALE-DddA-C / right TALE-DddA-N architecture with the target cytosine located at C5, the highest editing efficiency was achieved at a 16-nt spacer length (Figure 3A). The editing efficiencies are slightly lower with a spacer of 5-nt to 10-nt length. When the target cytosine is located at C4, it has a comparable high efficiency with spacer lengths from 5-nt to 9-nt and 15-nt (Figure 3A). For the left TALE-DddA-N / right TALE-DddA-C architecture, the highest editing efficiencies were obtained when the spacer length is 9-nt with the target cytosine located at C5 or C6 (Figuer 3B). Together, these findings reveal that when the TALE-DddA-C half is binding to the DNA strand harboring the target cytosine and the TALE-DddA-N half is binding to the other strand, the optimal editing efficiency can be achieved at C5 or C6 in a 9-nt spacer region.

DddA variants exhibited high activities in plant cells

It has been reported that fusing a single-strand DNA-binding domain from RADIATION SENSITIVE 51 (Rad51) to a CRISPR/Cas-based cytosine base editor (Zhang *et al*., 2020) or adenine base editor (Tan *et al*., 2022; Xue *et al*., 2023) could enhance the base editing capability in mammalian cells and rice. We wondered if the fusion of Rad51 could influence the editing efficiency of TALE-DdCBEs, therefore, we inserted the Rad51 domain before the DddA halves or before the TALE N-terminal domain (Figure S2). GUS activity showed that by using the N196/C17 or N196/C96 TALE N/C-terminal architectures, the additional fusion of Rad51 could not increase or even decreases the editing efficiencies of TALE-DdCBEs.

Recently, Mok *et al*. used rapid phage-assisted continuous evolution (PACE) and related phage-assisted non-continuous evolution (PANCE) methods to evolve the DddA protein (Mok *et al*., 2022b). They showed that, compared to wild-type DddA, the evolved DddA6 (Q1310R, S1330I, T1380I, T1413I) and DddA11 (S1330I, A1341V, N1342S, E1370K, T1380I, T1413I) variants had an average four-fold improvement of C•G-to-T•A editing efficiencies (Mok *et al*., 2022b). Moreover, DddA11 enabled editing non-TC targets whereas the wild-type DddA showed nearly no editing at those non-TC targets (Mok *et al*., 2020: Mok *et* al., 2022b). To test these variants in plants, we replaced the DddA-C/N halves with DddA6-C/N halves or DddA11-C/N halves and tested them in the GUS^{G537} reporter. Furthermore, we introduced the point mutations from DddA6 into DddA11 and generated a new DddA variant named DddA611, and tested it in the GUS^{G537} reporter as well. The reporter assays showed that the editing activities of DddA6, DddA11 and DddA611 are dramatically increased compared to canonical DddA (Figure 4).

To further characterize the targeting capabilities of DddA, DddA6, and DddA11, we targeted a tobacco gene (NB-T1) and the rice genes (OsALS-T1, OsALS-T2, and *OsPDS*) in protoplasts. Amplicons sequencing showed that DddA achieved C•G-to-T•A editing efficiency of approximately 2.4% of C10, while DddA6 yielded an editing efficiency of around 3.1% (Figure 5A). At the OsALS-T1 target site, both DddA and DddA11 showed similar editing efficiencies at C11 in a TC context (Figure 5B). For *OsPDS* and OsALS-T2, DddA showed no editing efficiency within the spacer regions, whereas DddA11 showed high editing efficiencies in multiple cytosines in *OsPDS* (C3, C5, C7, C11) (Figure 5C), as well as in OsALS-T2 (C9, C10) (Figure 5D). Notably, DddA11 exhibited non-TC targets editing activities of GC (C11, 1.1%), CC (C9, 2.6%), and AC (C10, 2.5%) at these two target sites. These results indicate that the three DddA variants DddA6, DddA11, and DddA611 substantially increase the C•G-to-T•A editing efficiency in plant cells, and DddA11 enables the editing of non-TC targets.

Specificity assay of DddA variants

TALE-DdCBEs-mediated base editing in mitochondria can generate off-target editing in mitochondria and even in nuclear chromosomes (Wei *et al*., 2022; Lei *et al*., 2022). To profile

the off-target activities of TALE-DdCBEs (harboring DddA or DddA6 or DddA11) in plants, we determined whether editing activity can be measured if only one TALE-array is binding to the target region. For this, we targeted the GUS^{G537} reporter by using a pair of specific TALE arrays to guide the DddA halves (left DddA-N / right DddA-C; the positive control) or lacking the right TALE array (left DddA-N / TALE-free DddA-C), or using one TALE array that containing non-target RVDs repeats that do not bind to the GUS^{G537} reporter (left DddA-N / unspecific-TALE DddA-C) (Figure 6A). The GUS activity results in *N. benthamiana* showed that for DddA6, the left DddA6-N / right DddA6-C, left DddA6-N / TALE-free DddA6-C, and left DddA6-N / unspecific TALE DddA6-C architectures showed editing efficiencies of 94.1%, 55.7%, and 75.6%, respectively (Figure 6B). For DddA11, the left DddA11-N / right DddA11-C, left DddA11-N / TALE-free DddA11-C, and left DddA11-N / unspecific TALE DddA11-C architectures showed editing efficiencies of 92%, 71%, and 79%, respectively (Figure 6C). These data suggest that while guided by one TALE array, the split DddA halves can spontaneously assembe to trigger C•G-to-T•A editing at loci where only one TALE array is binding.

Recently, to avoid the spontaneous assembly of split DddA halves, Lee developed a high-fidelity TALE-DdCBEs by substituting amino acid residues at the interface between the split DddA halves with alanine (K1389A or T1391A) (Lee *et al*., 2022). We wondered whether these high-fidelity TALE-DdCBEs could disrupt the formation of functional DddA in the absence of TALE-DNA interaction in plants, hence the K1389A or T1391A mutation was introduced into DddA6 (yielded DddA6^{K1389A} or DddA6^{T1391A}) and DddA11 (yielded DddA11^{K1389A} or DddA11^{T1391A}). We analyzed the off-target activity of these high-fidelity DddA variants in the *N. benthamiana* GUS^{G537} reporter (setup described above). Unfortunately, both DddA6^{K1389A} and DddA6^{T1391A}, showed similar editing efficiencies compared to DddA6 with left DddA6^{K1389A}-N or DddA6^{T1391A}-N / right DddA-C, left DddA6^{K1389A}-N or DddA6^{T1391A}-N / TALE-free DddA6-C, and left DddA6^{K1389A}-N or DddA6^{T1391A}-N / unspecific TALE DddA6-C architectures (Figure 6B). For DddA11^{K1389A} and DddA11^{T1391A}, they also showed comparable editing efficiencies to DddA11 in the three different TALE-DdCBEs architectures. Together, these results indicate that unlike in mitochondria (Lee *et al*., 2022), the high-fidelity DddA mutants could not reduce the spontaneous assembly of split DddA halves as well as off-target editing in plants.

Discussion

TALE cytosine base editors (TALE-DdCBEs) employ programmable DNA-binding TALE repeat arrays, a split dsDNA cytosine deaminase (DddA), and a uracil glycosylase inhibitor to mediate C•G-to-T•A editing in nuclear and organellar DNA. In this study, we developed a Modular Cloning (MoClo) system to assemble the TALE-DdCBEs using Golden Gate Cloning (Weber *et al*., 2011), and established a GUS reporter to evaluate the TALE-DdCBEs C•G-to-T•A editing efficiency in the plant nucleus. Our experiments using N- and C-terminal TALE truncations showed that the architecture influences the TALE-DdCBE editing capabilities. Previous studies already showed that TALEs require a portion of their N-terminal domain for binding to DNA (Herbers *et al*., 1992; Kay *et al*., 2007). Later, it was shown that within the full-length N-terminal region, the residues 1 to 152 contribute to the TALE DNA-binding affinity (Schreiber *et al*., 2015) and the residues from 62 to 92 contribute to transcriptional activation (Schreiber *et al*., 2019). The C-terminal domain does not contribute to DNA-binding (Kay *et al*., 2007), and parts as short as 17 amino acids that could function as linker sequences have been used in functional TALE nuclease (TALEN) designs (Mussolino *et al*., 2011). TALEs are large proteins which can contain more than 800 amino acids. To minimize the protein size of TALE-DdCBEs without sacrificing its activity, we recommend the use of N196/C17 N- and C-terminal combinations in the TALE-DdCBEs architectures.

Defining the appropriate editing window, the spacer length between two TALE-binding sites, and the optimal position of the target cytosine are crucial parameters to consider when employing TALE-DdCBEs. In mammalian cells, the TALE-DdCBEs (G1397-split DddA) preferentially edited T<u>C</u>s that were located approximately 4-7 (C4-C7) nucleotides upstream of the 3′ end of 14 to 18-nt spacer regions on either strand of the mitochondrial DNA (Mok *et al*., 2020) or at nuclear targets (Boyne *et al*., 2022). TALE-DdCBEs containing canonical DddA, DddA6 or DddA11 showed generally similar editing windows (Mok *et al*., 2022b). In our GUS reporter experiments, we found that the orientations of G1397-split DddA halves (left TALE-DddA-C / right TALE-DddA-N or left TALE-DddA-N / right TALE-DddA-C architecture) could affect the editing window. In general, TALE-DdCBEs prefer to edit T<u>C</u>s that were positioned at C5-C6 upstream of the 3′ end of the DNA strand in the GUS experiments. Moreover, the spacer

length can be shortened to 8-nt while maintaining the high editing efficiency at C5 or C6. Previous studies showed that the TALE-DdCBEs containing 41 amino acids long TALE C-terminal domains yielded the highest efficiencies in 14-nt to 16-nt spacer (Mok *et al*., 2020), whereas we used 17 amino acids long TALE C-terminal domains in our TALE-DdCBEs. Further experiments are needed to determine the relation between the lengths of the TALE C-terminal domain and the TALE-DdCBEs editing windows.

The TALE-DdCBEs containing canonical DddA have a strong preference for editing T<u>C</u> contexts (Mok *et al*., 2020) which might limit the application of TALE-DdCBEs in genome editing. Consistent with the previous study (Mok *et al*., 2022b), the utilization of DddA variants (DddA6, DddA11) can improve the TALE-DdCBEs editing activity at T<u>C</u> targets also in plants. Furthermore, DddA11 enables the editing of non-TC targets which expands the targeting scope of TALE-DdCBEs. Recently, Mi *et al*. identified a DddA homolog from *Simiaoa sunii* (Ddd_Ss) (Mi *et al*., 2023) and Guo *et al.* identified a DddA homolog from Roseburia intestinalis (riDddA$_{tox}$) (Guo *et al*., 2023), that can efficiently deaminate cytosines at non-T<u>C</u> targets. It would be helpful to compare the editing capabilities of the DddA variant and DddA homologs for further improvement of TALE-DdCBEs in plants. TALE-DdCBEs directed to the mitochondria have been reported to result in off-target editing in mitochondria and even the nuclear chromosomes. We found that the unspecific editing is at least partially caused by spontaneous assembly of DddA halves in the absence of TALE-DNA interactions. DddA and its variants DddA6 and DddA11 containing either the K1389A or T1391A point mutation can restrict the spontaneous assembly of DddA halves in human mitochondrial DNA editing (Lee *et al*., 2022). In contrast, the DddA6^{K1389A}, DddA6^{T1391A}, DddA11^{K1389A}, or DddA11^{T1391A} variants did not show a significant decrease of unspecific editing in our GUS^{G537} reporter assays. This might be due to a different cellular context between editing the human mitochondrial DNA and using the plant transient reporter assay. The GUS reporter might efficiently detect even low editing activities, because of the expected high copy number of transferred T-DNA molecules while at the same time limiting the maximal detectable activity due to saturation of the assay. Further studies targeting endogenous genes using high-fidelity DddA variants to characterize the TALE-DdCBE specificity in plants will be interesting. Additional protein engineering of DddA or

characterization of novel DddA homologs could further improve the efficiency and specificity of TALE-cytosine base editors. Notably, we found that while the DddA-N half was guided by one TALE array and the DddA-C half itself flowing non-targeted in the cell (without fusion to a TALE) C•G-to-T•A editing could be induced. Hence, we speculate that the TALE-DdCBEs architecture can possibly be simplified using one TALE to guide the DddA deaminase (in a structure of TALE - DddA-N - 2A peptide - DddA-C). This single transcription unit of the TALE-DdCBE architecture has a smaller protein size that could possibly be package into AAV or other viral vectors with limited cargo space facilitating gene therapy.

In summary, we successfully determined the architecture of TALE-DdCBEs with high editing efficiencies and tested their targeting scope in rice and tobacco nuclear DNA. The protein-only base editing tools TALE-DdCBEs broaden the plant genome editing toolbox and provide a valuable resource for organellar or nuclear DNA editing.

Methods

Plasmid Construction

The TALE-DdCBEs plasmids were generated by using the the modular cloning (MoClo) syntax (Weber *et al*., 2011; Geiβler *et al*., 2011; Grützner and Marillonnet, 2020). All the components were subcloned in individual modules that can be assembled using Golden Gate Cloning (Engler *et al*., 2008). The details of the cloning procedures are listed Figure S3. The plasmid modules used in this study were listed in Table S1.

Nicotiana benthamiana infiltration and GUS reporter assay

Assays for GUS reporter activities were performed as previously described (Boch *et al*., 2009). Briefly, *A. tumefaciens* GV3101 strains containing a TALE-CBE construct, the GUS reporter construct, respectively, were mixed 1:1:1 with OD_{600} of 0.8 and inoculated into *N. benthamiana* leaves. After two to three days inoculation, two leaf discs (diameter 0.8 cm) were harvested from the inoculation spot. Leaf tissues were homogenized and incubated with 4-methyl-umbelliferyl-β-D-glucuronide. GUS activities were measured using a TECAN reader (360 nm

excitation and 465 nm emission). Proteins were quantified by NanoDrop™ One (Thermo Fisher Scientific).

Protoplast isolation and transformation

Rice cultivar Kitaake leaves were used to prepare protoplasts. Rice protoplast and *N. benthamiana* protoplast isolation and transformation were performed as previously described (Li *et al*., 2013; Shan *et al*., 2014). 10 µg plasmid DNA per construct were introduced into protoplasts by PEG-mediated transfection. The transfected protoplasts were incubated at room temperature. After 48 h, the protoplasts were collected and the genomic DNA extracted.

DNA extraction and amplicon sequencing

We used innuPREP Plant DNA Kit (Analytik Jena) to extract plant genomic DNA. The targeted sequences were amplified with specific primers (see in Table S2), and the amplicons were purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) then quantified using Qubit™ 1X dsDNA High Sensitivity Kits (Thermo Fisher Scientific). Oligos used in this study were list in Table S2. Equal amounts of PCR products were pooled and sequenced (GENEWIZ, AMPLICON-EZ). Amplicon sequencing was performed three times for each target location using genomic DNA isolated from three different protoplasts transformation experiments. The target sites in the sequenced reads were analyzed for mutations using CRISPResso2 (Clement *et al*., 2019).

Plant growth condition

N. benthamiana plants were grown in a greenhouse with 16 hours of light, a relative humidity of 40-60%, and temperatures of 23°C and 19°C during the day and night, respectively. Four to six weeks old plants were used for *A. tumefaciens* inoculation experiments.

Statistical analysis

All values are shown as means ± SEM (standard error of the mean). Statistical differences between the values were tested using two-tailed unpaired Student's t-tests by GraphPad (Prism; www.graphpad.com).

Data availability

The amplicon sequencing data have been deposited in an NCBI BioProject database: PRJNA950930.

Funding

The work was conducted using university core funding.

Author contributions

D.Z. and J.B. designed the experiments. D.Z. and V.P. performed the experiments and analysed the data. D.Z. and J.B. wrote the manuscript.

References

Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38**:824–844.

Bacman, S. R., Williams, S. L., Pinto, M., Peralta, S., and Moraes, C. T. (2013). Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat Med* **19**:1111–1113.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of dna binding specificity of TAL-Type III effectors. *Science* **326**:1509–1512.

Boyne, A., Yang, M., Pulicani, S., Feola, M., Tkach, D., Hong, R., Duclert, A., Duchateau, P., and Juillerat, A. (2022). Efficient multitool/multiplex gene engineering with TALE-BE. *Front Bio Biotechnol* **10**:1940.

Cho, S.-I., Lee, S., Mok, Y. G., Lim, K., Lee, J., Lee, J. M., Chung, E., and Kim, J.-S. (2022). Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell* **185**:1764-1776.e12.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A. J., and Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**:757–761.

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., *et al*. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* **37**:224–226.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot, One Step, Precision cloning method with high throughput capability. *PloS One* **3**:e3647.

Gammage, P. A., Rorbach, J., Vincent, A. I., Rebar, E. J., and Minczuk, M. (2014). Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. *EMBO Mol Med* **6**:458–466.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**:464–471.

Geiβler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.-E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-specificity. *PloS One* **6**:e19509.

Grützner, R., and Marillonnet, S. (2020). Generation of MoClo standard parts using Golden Gate cloning. In *DNA Cloning and Assembly: Methods and Protocols*, pp. 107–123. New York, NY: Springer US.

Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: Expanding the types of DNA damage products harnessed for genome editing. *Gene Genome Edi* **1**:100005.

Guo, J., Yu, W., Li, M., Chen, H., Liu, J., Xue, X., Lin, J., Huang, S., Shu, W., Huang, X., *et al*. (2023). A DddA ortholog-based and transactivator-assisted nuclear and mitochondrial cytosine base editors with expanded target compatibility. *Mol Cell* **83**:1710-1724.e7.

Herbers, K., Conrads-Strauch, J., and Bonas, U. (1992). Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* **356**:172–174.

Islam, M. R., Tomatsu, S., Shah, G. N., Grubb, J. H., Jain, S., and Sly, W. S. (1999). Active site residues of human β-glucuronidase: Evidence for glu540 as the nucleophile and glu451 as the acid-base residue. *J Biol Chem* **274**:23451–23455.

Kang, B.-C., Bae, S.-J., Lee, S., Lee, J. S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., and Kim, J.-S. (2021). Chloroplast and mitochondrial DNA editing in plants. *Nat Plants* **7**:899–905.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). A Bacterial Effector Acts as a Plant Transcription Factor and Induces a Cell Size Regulator. *Science* **318**:648–651.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**:420–424.

Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., Grünewald, J., and Joung, J. K. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol* **39**:41–46.

Lee, S., Lee, H., Baek, G., and Kim, J.-S. (2022). Precision mitochondrial DNA editing with high-fidelity DddA-derived base editors. *Nat Biotechnol* Advance Access published October 13, 2022, doi:10.1038/s41587-022-01486-w.

Lei, Z., Meng, H., Liu, L., Zhao, H., Rao, X., Yan, Y., Wu, H., Liu, M., He, A., and Yi, C. (2022). Mitochondrial base editor induces substantial nuclear off-target mutations. *Nature* **606**:804–811.

Li, J.-F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G. M., and Sheen, J. (2013). Multiplex and homologous recombination–mediated genome editing

in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* **31**:688–691.

Li, R., Char, S. N., Liu, B., Liu, H., Li, X., and Yang, B. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. *Mol Plant* **14**:1412–1414.

Mi, L., Shi, M., Li, Y.-X., Xie, G., Rao, X., Wu, D., Cheng, A., Niu, M., Xu, F., Yu, Y., *et al*. (2023). DddA homolog search and engineering expand sequence compatibility of mitochondrial base editing. *Nat Commun* **14**:874.

Mussolino, C., Morbitzer, R., Lütge, F., Dannemann, N., Lahaye, T., and Cathomen, T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**:9283–9293.

Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., Hsu, F., Radey, M. C., Peterson, S. B., Mootha, V. K., *et al*. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**:631–637.

Mok, Y. G., Lee, J. M., Chung, E., Lee, J., Lim, K., Cho, S.-I., and Kim, J.-S. (2022a). Base editing in human cells with monomeric DddA-TALE fusion deaminases. *Nat Commun* **13**:4038.

Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., and Liu, D. R. (2022b). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat Biotechnol* **40**:1378–1387.

Mok, Y. G., Hong, S., Bae, S.-J., Cho, S.-I., and Kim, J.-S. (2022c). Targeted A-to-G base editing of chloroplast DNA in plants. *Nat Plants* Advance Access published December 1, 2022, doi:10.1038/s41477-022-01279-8.

Nakazato, I., Okuno, M., Yamamoto, H., Tamura, Y., Itoh, T., Shikanai, T., Takanashi, H., Tsutsumi, N., and Arimura, S. (2021). Targeted base editing in the plastid genome of *Arabidopsis thaliana*. *Nat Plants* **7**:906–913.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., *et al*. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**:aaf8729.

Schreiber, T., Sorgatz, A., List, F., Blüher, D., Thieme, S., Wilmanns, M., and Bonas, U. (2015). Refined requirements for protein regions important for activity of the TALE AvrBs3. *PloS One* **10**:e0120214.

Schreiber, T., Prange, A., Hoppe, T., and Tissier, A. (2019). Split-TALE: A TALE-based two-component system for synthetic biology applications in planta. *Plant Physiol* **179**:1001–1012.

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc* **9**:2395–2410.

Tan, J., Zeng, D., Zhao, Y., Wang, Y., Liu, T., Li, S., Xue, Y., Luo, Y., Xie, X., Chen, L., *et al*. (2022). PhieABEs: a PAM-less/free high-efficiency adenine base editor toolbox with wide target scope in plants. *Plant Biotechnol J***20**:934–943.

Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**:636–646.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PloS One* **6**:e16765.

Wei, Y., Xu, C., Feng, H., Xu, K., Li, Z., Hu, J., Zhou, L., Wei, Y., Zuo, Z., Zuo, E., *et al*. (2022). Human cleaving embryos enable efficient mitochondrial base-editing with DdCBE. *Cell Discov* **8**:1–4.

Xue, N., Liu, X., Zhang, D., Wu, Y., Zhong, Y., Wang, J., Fan, W., Jiang, H., Zhu, B., Ge, X., *et al*. (2023). Improving adenine and dual base editors through introduction of TadA-8e and Rad51DBD. *Nat Commun* **14**:1224.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**:149–153.

Zhang, X., Chen, L., Zhu, B., Wang, L., Chen, C., Hong, M., Huang, Y., Li, H., Han, H., Cai, B., *et al*. (2020). Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. *Nat Cell Biol* **22**:740–750.

Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., Rosser, S. J., Bi, C., and Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol* **39**:35–40.

Figure 1. Establishment of TALE-DdCBEs in *N. benthamiana*.

(A) Schematic of the GUS^{G537} cytosine base editing reporter. The C•G-to-T•A (highlight in red) editing in GUS^{G537} can alter the glycine (GGA) to glutamic acid (GAA) and restore GUS activity. TALE binding sites are in gray background, spacer region in cyan background, respectively. **(B)** Schematic of TALE-DdCBEs containing different lengths of N- and C-terminal. N288: 288 amino acids full-length N-terminus. N196: 196 amino acids length of truncated N-terminal. N135: 135 amino acids length of truncated N-terminal. Five truncated C-terminal: C247 has 247 amino acids, C96 has 96 amino acids, C63 has 63 amino acids, C47 has 47 amino acids, C28 has 28 amino acids, C17 has 17 amino acids. bpNLS: bipartite nuclear localization sequence. UGI: uracil glycosylase inhibitor. **(C)** C•G-to-T•A editing efficiencies of different TALE-DdCBEs architectures in GUS^{G537}. GUS activities were measured and normalized to 2x35S::GUS (WT GUS, positive control). Values and error bars indicate the mean ± SEM, n = 4, n.s. (not significant) using Student's two-tailed unpaired t-test.

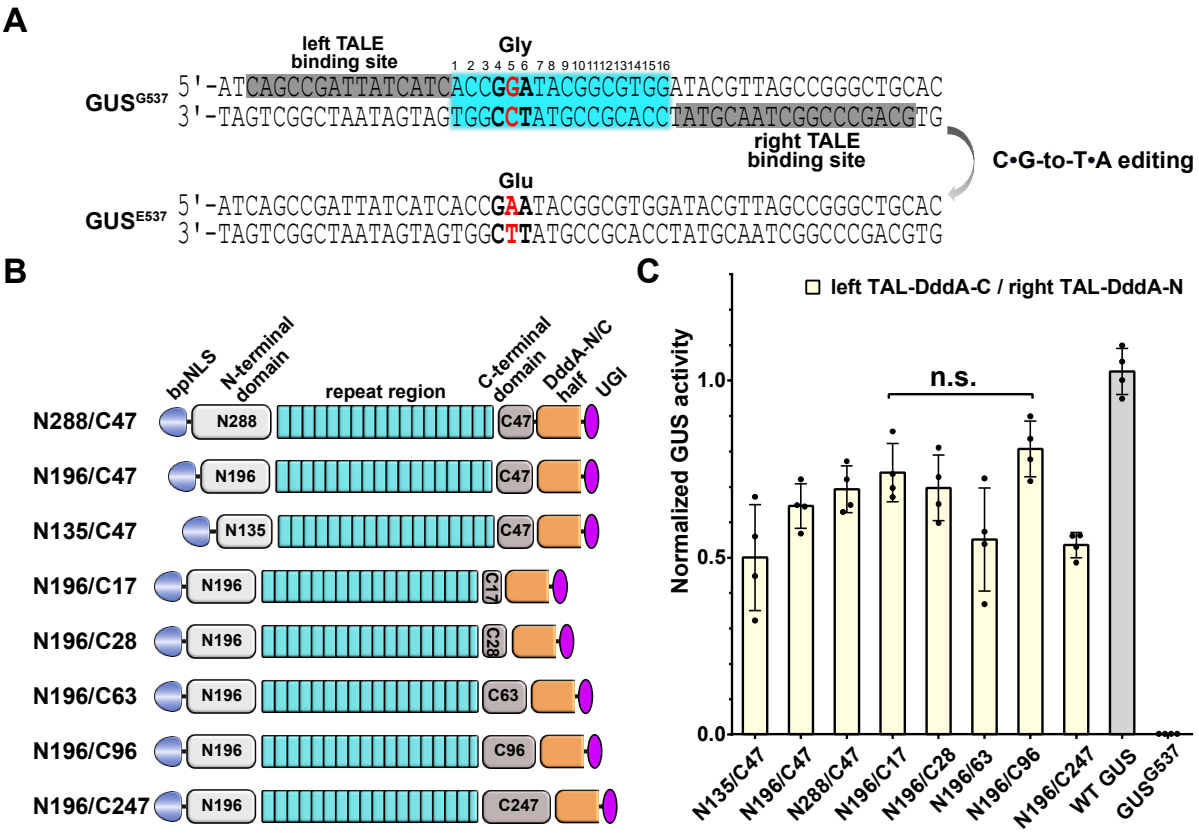
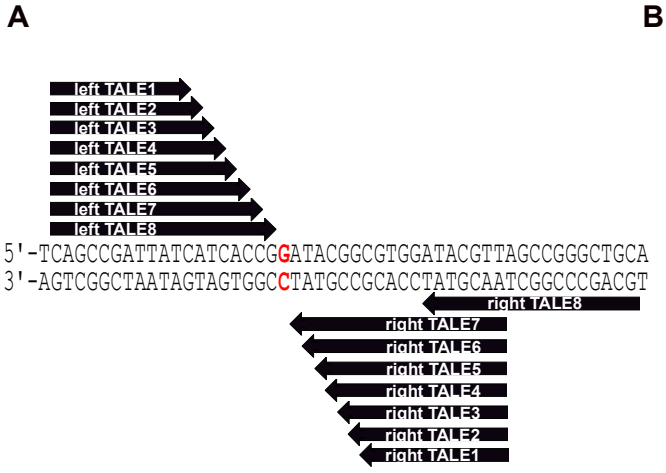


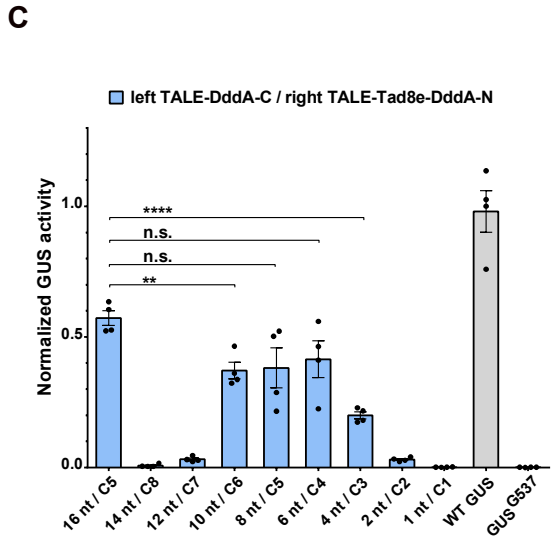
Figure 2. Analyzing the editing windows of pair TALE-DdCBEs.

(A) Schematic of shifting the spacer region of pair TALE-DdCBEs and the position of the target cytosine by using TALE arrays of different length, the binding sites of eight left TALEs and eight right TALEs are show by arrows. The targeted C•G base pair is in red. **(B)** Different spacer regions (from 1-nt to 16-nt) flanked by different left and right TALE combinations. The targeted cytosine is in red and bold. **(C)** and **(D)** C•G-to-T•A editing efficiencies of TALE-DdCBEs using left TALE-DddA-C / right TALE-DddA-N architecture (C) or left TALE-DddA-N / right TALE-DddA-C architecture (D) in the in GUS^{G537} reporter. GUS activities were measured and normalized to 2x35S::GUS (WT GUS, positive control). Values and error bars indicate the mean ± SEM, n = 4. Student's unpaired two-tailed t-test was applied. **:P < 0.01; ****:P < 0.0001; n.s.: not significant (P > 0.05).



B

	Spacer sequences	length	position of C
left TALE4 + right TALE8	ACCG**G**ATACGGCGTGG TGGC**C**TATGCCGCACC	16-nt	C5
left TALE1 + right TALE1	ATCACCG**G**ATACGG TAGTGGC**C**TATGCC	14-nt	C8
left TALE2 + right TALE2	TCACCG**G**ATACG AGTGGC**C**TATGC	12-nt	C7
left TALE3 + right TALE3	CACCG**G**ATAC GTGGC**C**TATG	10-nt	C6
left TALE4 + right TALE4	ACCG**G**ATA TGGC**C**TAT	8-nt	C5
left TALE5 + right TALE5	CCG**G**AT GGC**C**TA	6-nt	C4
left TALE6 + right TALE6	CG**G**A GC**C**T	4-nt	C3
left TALE7 + right TALE7	**G**G **C**C	2-nt	C2
left TALE8 + right TALE7	**G** **C**	1-nt	C1



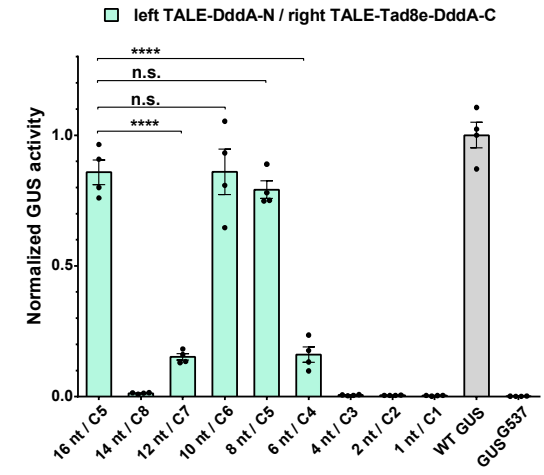


Figure 3. Analyzing the editing spacer length of pair TALE-DdCBEs.

(A) C•G-to-T•A editing efficiencies of the targeted cytosine at C4 or C5 within different length of spacer using left TALE-DddA-C / right TALE-DddA-N architecture in the in GUS^{G537} reporter. **(B)** C•G-to-T•A editing efficiencies of the targeted cytosine at C5 or C6 within different length of spacer using left TALE-DddA-N / right TALE-DddA-C architecture in the in GUS^{G537} reporter. GUS activities were measured and normalized to 2x35S::GUS (WT GUS, positive control). Values and error bars indicate the mean ± SEM, n = 4.

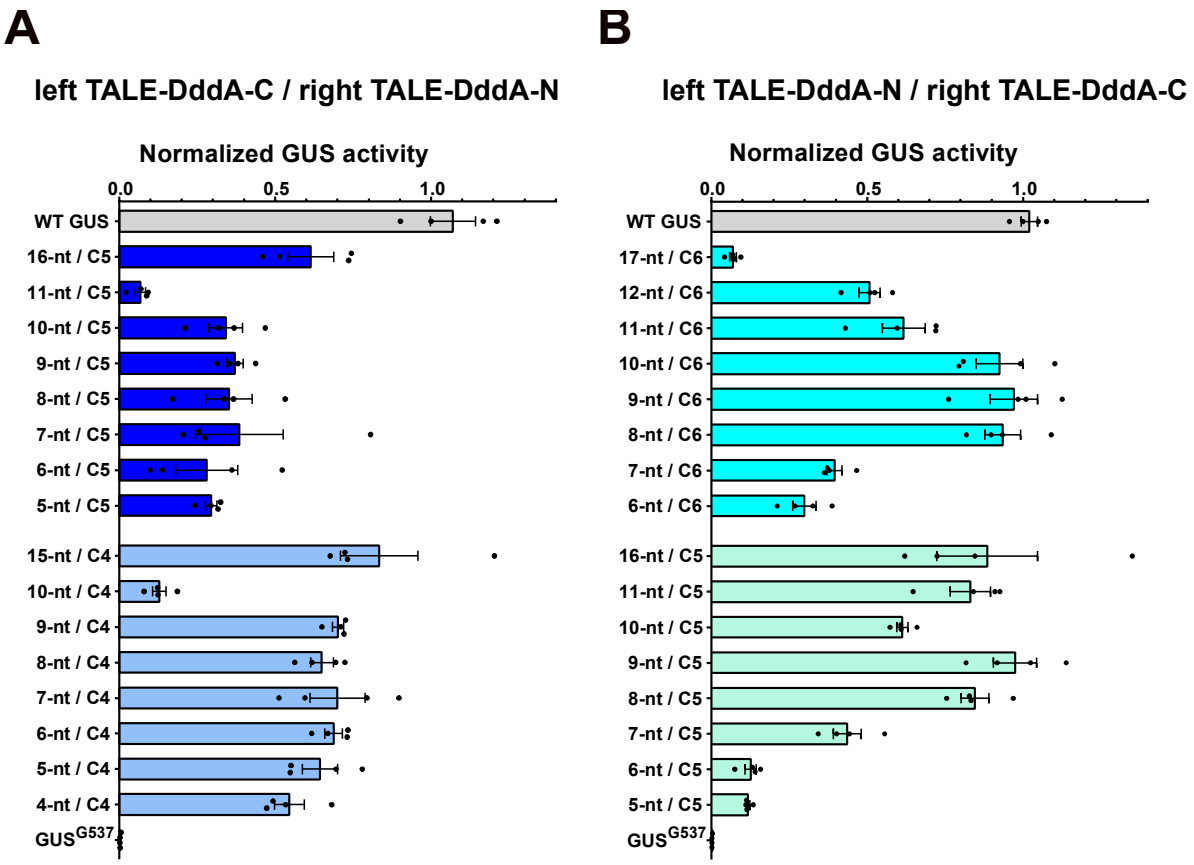


Figure 4. DddA variants showed enhanced editing in GUS^{G537} reporter.

C•G-to-T•A editing efficiencies of TALE-DdCBEs harboring canonical DddA or DddA6 or DddA11 or DddA611 in GUS^{G537} reporter. Top: TALE-DdCBEs using left TALE-DddA-C / right TALE-DddA-N architecture, targeted cytosine located at C5 within the 16-nt spacer. Bottom: values and error bars indicate the mean ± SEM, n=3.

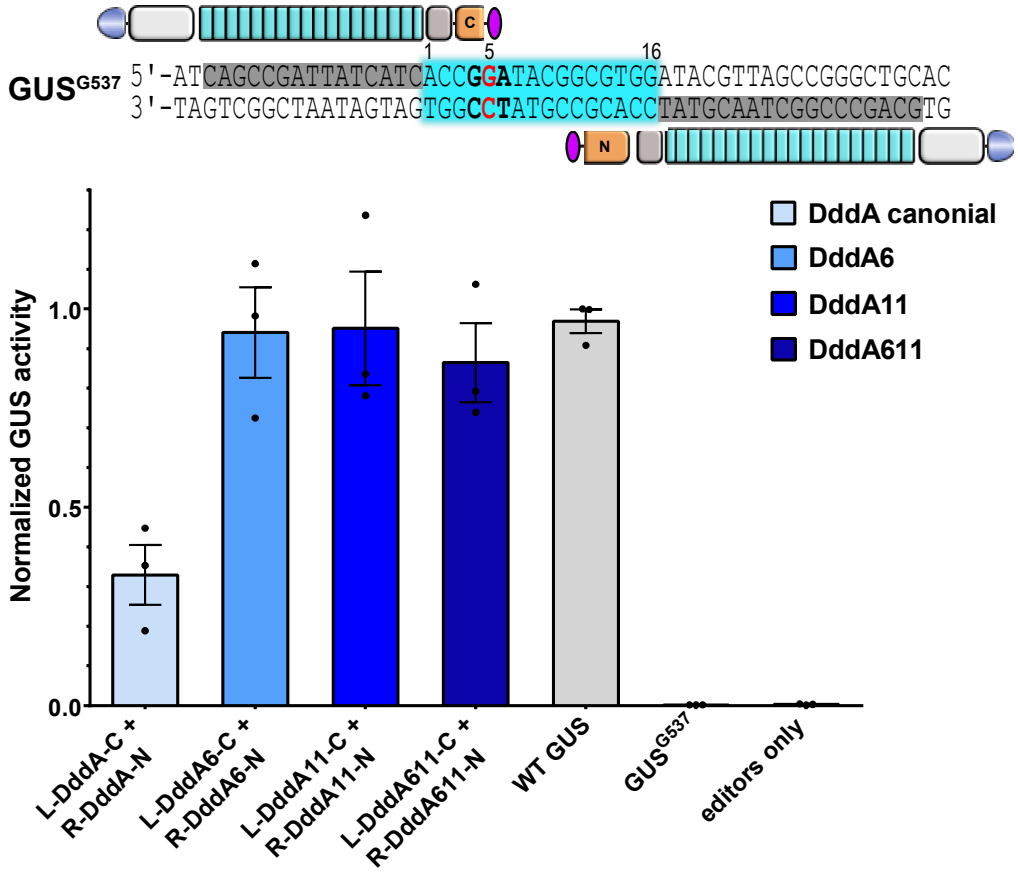


Figure 5. Editing efficiency of TALE-DdCBEs in rice and *N. benthamiana*.

(A-D) C•G-to-T•A editing efficiencies of TALE-DdCBEs were determined by amplicon sequencing of target regions from *N. benthamiana* (A) and rice protoplasts (B-D). Targeted sequences are listed above the panels. Values and error bars indicate the mean ± SEM, n=3 independent experiments.

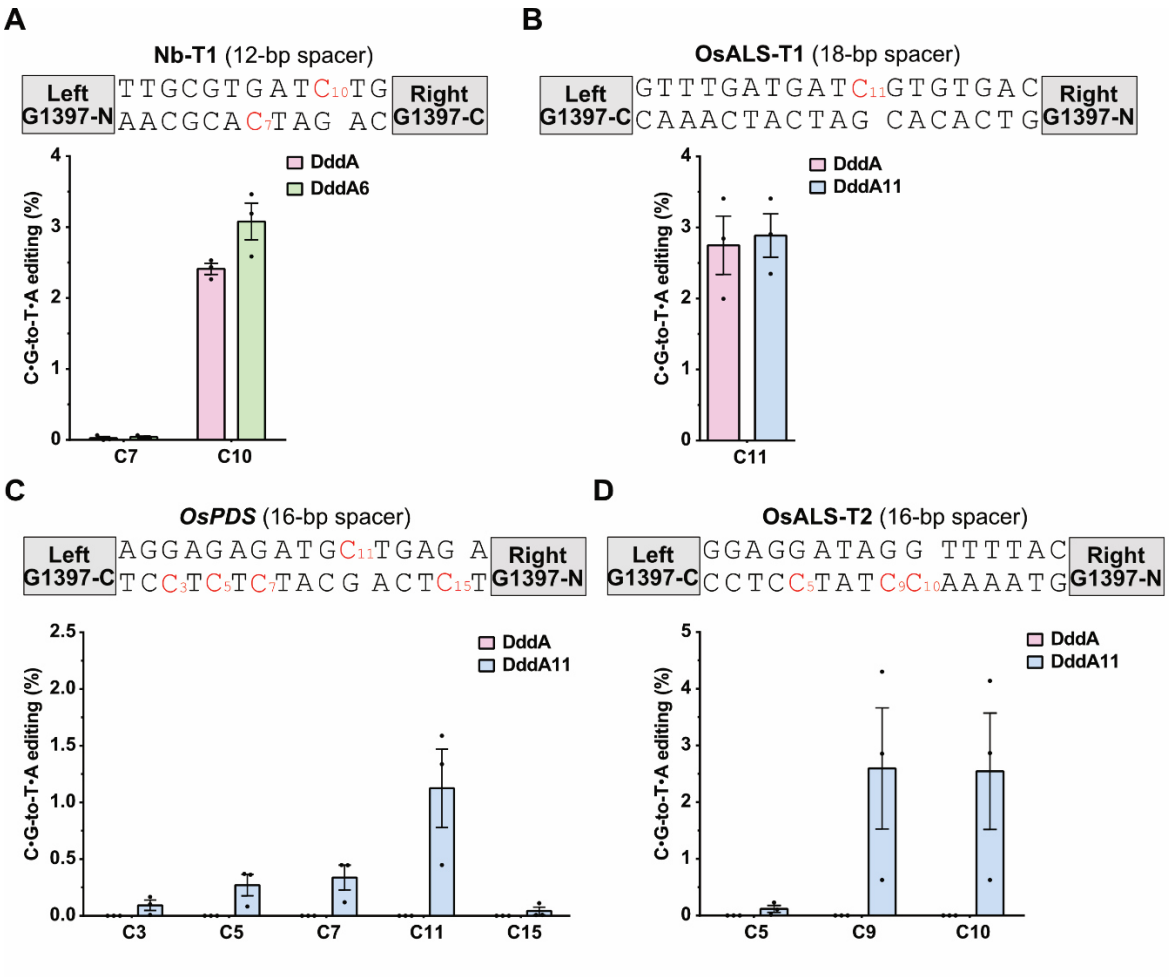
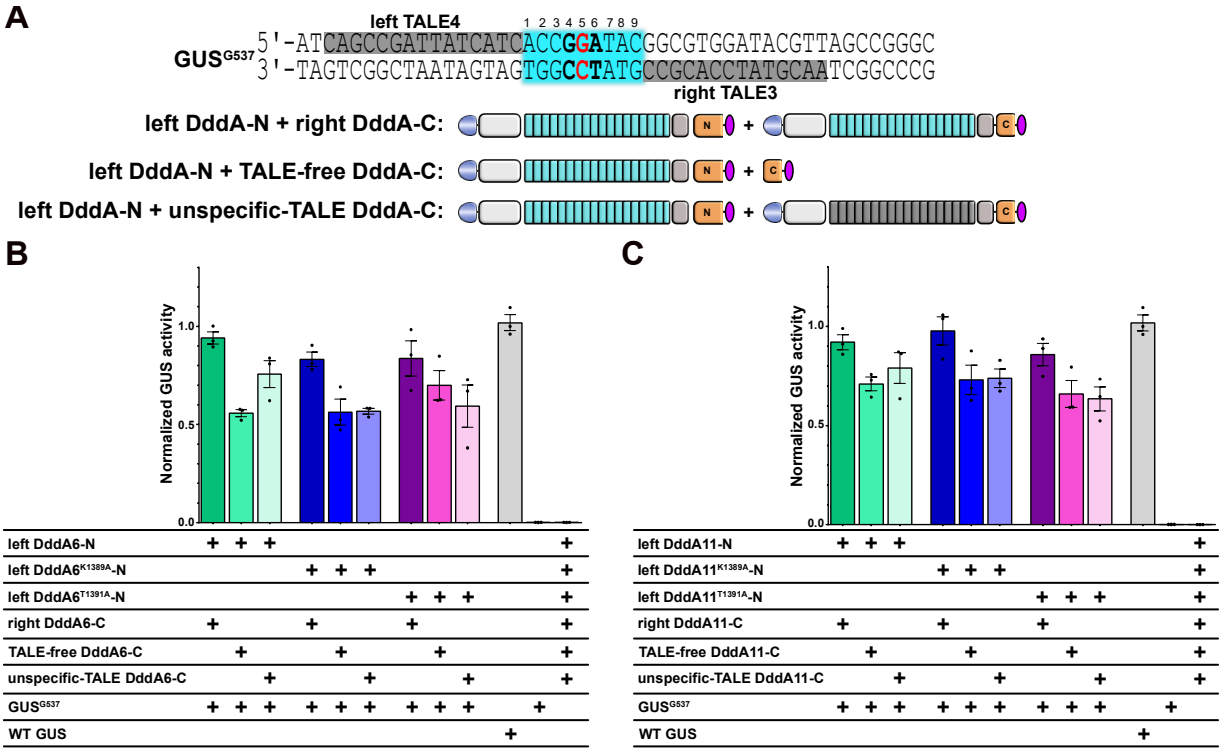


Figure 6. Off-target editing of TALE-DdCBEs in GUS^{G537} reporter.

(A) TALE-DdCBEs constructs were used to target GUS^{G537} reporter. Targeted cytosine located at C5 within the 9-nt spacer. **(B)** C•G-to-T•A editing efficiencies of TALE-DdCBEs that containing DddA6 or its variants DddA6^{K1389A} or DddA6^{T1391A}. **(C)** C•G-to-T•A editing efficiencies of TALE-DdCBEs that containing DddA11 or its variants DddA11^{K1389A} or DddA11^{T1391A}. GUS activities were measured and normalized to 2x35S::GUS (WT GUS, positive control). Values and error bars indicate the mean ± SEM, n = 3.



4 Development of TALE adenine base editors in plants

Dingbo Zhang, Jens Boch*

Leibniz Universität Hannover, Institute of Plant Genetics, Herrenhäuser Str. 2, 30419

Hannover, Germany

* corresponding authors

Journal: Plant Biotechnology Journal
Submitted: 01-Apr-2022
Accepted: -
Available online: -

Author	Contributions
Dingbo Zhang	Designed and performed all the experiments, analyzed the data, contributed to the writing of the manuscript
Jens Boch	Conceived the studies and designed experiments

Development of TALE adenine base editors in plants

Dingbo Zhang, Jens Boch

Leibniz Universität Hannover, Institute of Plant Genetics, Herrenhäuser Str. 2, 30419

Hannover, Germany

Abstract

Base editors enable precise nucleotide changes at targeted genomic loci without requiring double-stranded DNA breaks or repair templates. TALE-adenine base editors (TALE-ABEs) are genome editing tools, composed of a DNA-binding domain from transcription activator-like effectors (TALEs), an engineered adenosine deaminase (TadA8e), and a cytosine deaminase domain (DddA) and allow A•T-to-G•C editing in human mitochondrial DNA. However, the editing ability of TALE-ABEs in plants has not been well described so far, and the role of DddA is still unclear. We tested a series of TALE-ABEs with different deaminase fusion architectures in *Nicotiana benthamiana* and rice. The results indicate that the double-stranded DNA-specific cytosine deaminase DddA can boost the activities of single-stranded DNA-specific deaminases (TadA8e or APOBEC3A) on double-stranded DNA. We analyzed A•T-to-G•C editing efficiencies in a β-glucuronidase reporter system and show precise adenines editing in genomic regions with high product purity in rice protoplasts. Consequently, TALE-adenine base editors provide alternatives for crop improvement and gene therapy by editing nuclear or organellar genomes.

Keywords: genome editing, plant breeding, rice, adenine deaminase.

Introduction

The growing global population and the effects of climate change are challenging agricultural productivity. Genome editing by sequence-specific nucleases such as meganucleases, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the CRISPR-Cas system have revolutionized genetic studies and crop breeding by enabling precise modifications of genomes (Gao, 2021). Such nucleases can induce double-stranded DNA breaks (DSBs). In plants, the DSBs are predominantly repaired by non-homologous end joining, which can generate random nucleotide insertions or deletions (indels) (Chen *et al*., 2019).

Many quantitative trait loci (QTLs) that drive crop production and stress tolerance are linked to single nucleotide polymorphisms (SNPs) (Huang and Han, 2014). Thus, the development of tools that can effectively cause single nucleotide variants instead of random indels mutations is essential. Base editors are genome editing technologies that can convert targeted base pairs without requiring DSBs or donor DNA templates (Anzalone *et al*., 2020). Two types of base editors have been developed: cytosine base editors (CBEs) convert C•G base pairs to T•A base pairs (Komor *et al*., 2016), and adenine base editors (ABEs) catalyze A•T-to-G•C conversions (Gaudelli *et al*., 2017). Typically, CBEs are composed of a CRISPR/Cas nickase, a cytosine deaminase, and an uracil glycosylase inhibitor (UGI). ABEs are comprised of a CRISPR-Cas nickase and an adenosine deaminase (Liu *et al*., 2022). An earlier version of ABE (ABE7.10) used a heterodimer of a wild-type *E. coli* tRNA adenosine deaminase (TadA) and a synthetically evolved TadA (TadA7.10) to act on single-stranded DNA (ssDNA) (Gaudelli *et al*., 2017). A further evolved TadA variant (TadA8e) exhibits improved editing efficiency and targeting scope in mammalian cells and rice (Richter *et al*., 2020; Yan *et al*., 2021; Wei *et al*., 2021). *In vitro* studies showed that TadA8e catalyzes DNA deamination more than 1000-fold faster than TadA7.10 (Lapinaite *et al*., 2020). When the Cas protein binds to its target DNA sequence, the sgRNA hybridizes to the complementary DNA sequence and causes a ssDNA R-loop (Jiang and Doudna, 2017). This ssDNA exposure allows the ssDNA-specific CBE and ABE deaminases to chemically modify their target bases within a window at the PAM-distal end (Gu *et al*., 2021).

Recently, Mok *et al*. reported that the cytosine deaminase $DddA_{tox}$ from *Burkholderia cenocepacia* enables targeted C•G-to-T•A conversions in human nuclear and mitochondrial double-stranded DNA (dsDNA) (Mok *et al*., 2020). This enabled to use zinc finger or TALEs which don't cause ssDNA formation as targeting devices for the development of novel base editors (Lim *et al*., 2022; Mok *et al*., 2020; Mok *et al*., 2022a). TALEs can be placed more precisely than Cas nucleases, because they don't require the presence of a PAM sequence in a given distance to the target cytosine. DddA-derived cytosine base editors (DdCBEs) use two split halves of $DddA_{tox}$ (DddA-N and DddA-C) which are fused to two adjacent tail-to-tail TALE DNA-binding arrays, respectively. The assembly of the two DddA halves reconstitutes the active enzyme which triggers deamination of target cytosines within the spacer region between the TALE binding sites (Mok *et al*., 2020). Such DdCBEs and ZF-deaminases (ZFD) enabled base editing in organellar genomes, because they can be imported into organelles which is inefficient for CRISPR-based systems (Mok *et al*., 2020; Lim *et al*., 2022). Accordingly, it is highly relevant to further expand and optimize such genome editing tools. In plants, DdCBEs were successfully used for editing the plastid genome of *Arabidopsis thaliana* (Nakazato *et al*., 2021), the chloroplast and mitochondrial DNA of lettuce, rapeseed (Kang *et al*., 2021), and rice chloroplasts (Li *et al*., 2021). Besides CBEs, TALE-based ABEs (TALEDs) have recently also been developed to perform mitochondrial A•T-to-G•C base editing (Cho *et al*., 2022). These TALEDs are comprised of TALE DNA-binding arrays, a full-length DddA variant or split DddA, and an engineered deoxyadenosine deaminase (TadA8e) which catalyzes the base conversion. When tested in 17 human mitochondrial target sites, TALEDs exhibited high editing efficiencies of up to 49% (Cho *et al*., 2022). Recently, it was reported that TALEDs could generate A•T-to-G•C base conversions in chloroplast DNA in lettuce protoplasts and *Arabidopsis* (Mok *et al*., 2022b). The DddA domain is essential for this editing system, but how DddA promotes TadA8e activity is still unknown. Neither TALE-CBEs nor TALE-ABEs have been applied for nuclear targets in plants, yet.

To easily compare different TALE-base editor designs in plants, we developed a modular cloning (MoClo) pipeline for these tools and established a simple β-glucuronidase (GUS) reporter assay in *Nicotiana benthamiana*. We present a series of TALE adenine base editors

(TALE-ABEs), compared their A•T-to-G•C editing activities and demonstrate A•T-to-G•C editing in genomic loci in rice. Our experiments show that DddA enhances not only TadA8e, but also other strictly ssDNA-specific deaminases, which suggests that DddA somehow provides access to single-stranded DNA.

Results

Establishing a base editor reporter system with single TALE-ABEs

We first developed a β-glucuronidase (GUS) reporter system in *N. benthamiana* for evaluating A•T-to-G•C editing efficiencies of different base editors. For this, an inactivated GUS variant (GUS^{*424}) was constructed containing a stop-codon (TAA). An A•T-to-G•C conversion in the non-coding strand can revert the stop codon to a codon encoding glutamine (CAA) and facilitates the production of functional GUS protein (Figure 1A). The GUS activity can then be used as approximation for the efficiency of base editing.

Next, we employed two separate strategies for the construction of TALE-ABEs (Figure 1B). One is using a single TALE array containing a bipartite nuclear localization sequence (bpNLS), an N-terminal TALE domain, the RVD repeat region, the C-terminal TALE domain, and functional domains to perform the base editing, termed single TALE-ABE (sTABE). The other is using a pair of TALE-ABEs and named pair TALE-ABE (pTABE). To enable a simple construction of different base editor designs, individual parts were built as modular cloning (MoClo) modules and assembled using Golden Gate Cloning (Supplemental Figure 1). First, TadA8e or a TadA8e-dimer was fused to the TALE array to generate sTABE_v1 and sTABE_v2. The activity of these sTABEs was tested using the GUS^{*424} reporter by Agrobacterium-mediated transient expression in *Nicotiana benthamiana* leaves (Supplemental Figure 2). After normalization to a constitutively expressed functional GUS (WT GUS), both sTABEs conferred very low GUS activity (0.2-0.6%) with the sTABE-binding either left (non-coding strand, shown as L) or right (coding strand, shown as R) of the target adenine in the GUS^{*424} reporter (Figure 1A, C).

Next, we examined whether fusing additional domains to the sTABE can increase their A•T-to-G•C editing efficiency. Cho *et al*. reported that the dsDNA-specific cytosine deaminase

domain of DddA as catalytic inactive (DddA^{E1347A}) or active version in different designs drastically enhances the activity of TALE adenine base editors on human mitochondrial DNA (Cho *et al*., 2022). Accordingly, DddA^{E1347A} was fused to the single TadA8e and yielded sTABE_v3 (Figure 1B). sTABE_v4 harbors an engineered human AID (AID*Δ) which exhibited high deaminase activity in rice (Ren *et al*., 2018). sTABE_v5 contains the single-strand DNA-binding domain from RAD51 protein (Rad51DBD) which conferred increased activity in cytosine base editors (Zhang *et al*., 2020) and adenine base editors (Tan *et al*., 2022). sTABE_v6 contains the non-specific double-strand DNA-binding protein Sso7d from *Sulfolobus solfataricus* (Baumann *et al*., 1995). Overall, all setups with the exception of sTABE_v4 showed comparable GUS activity above background with slight preferences for binding either left or right of the target site (Fig. 1C). These results show that TadA8e can catalyze A•T-to-G•C editing in dsDNA in sTABE architectures, but the editing efficiencies are very low. Fusion of DddA^{E1347A}, AID*Δ, Rad51DBD or Sso7d could not increase the editing efficiency in our reporter assays.

Improving the A•T-to-G•C editing efficiency with pair TALE-ABEs

As an alternative design, we tested pair TALE-ABE (pTABE) architectures to combine TadA8e and DddA. The pTABEs (pTABE_v1 to pTABE_v6) are composed of a pair of TALE arrays in a tail-to-tail arrangement (Figure 1B). For pTABE_v1, TadA8e was fused to one TALE array and the catalytically inactive DddA^{E1347A} was fused to the other. The editing efficiencies of pTABE_v1 was comparable to the single TALE-ABE (sTABE_v3) (Figure 1D). The activity of the single sTABE_v3 varied between 0.7% and 1.6% when positioned on the left or right side of the target adenine, respectively. The activity of the pair pTABE_v1 was similar (1%) for both orientations, but remained low (Figure 1D). This indicates that the fusion of a full length DddA^{E1347A} in sTABE or pTABE could not increase the TALE-ABE A•T-to-G•C editing efficiency in our *N. benthamiana* assay.

Next, we tested split DddA designs with the DddA-N and DddA-C halves fused separately to a pair of TALE arrays with or without the addition of an uracil-glycosylase inhibitor (UGI) (pTABE_v2, pTABE_v3, pTABE_v4). Remarkably, two of these designs showed significantly

increased base editing activity. When DddA-C and TadA8e were fused to one TALE array and DddA-N to the other (pTABE_v2) the activities varied between 5.3% and 19.3% depending on the orientation of the pTABE pair (Figure 1E). If the position of the two DddA-halves in the fusion constructs was switched (pTABE_v3) the activities ranged from 7.6% to 10.3% (Figure 1E). In contrast, the activity of pTABE_v4 which has the same architecture as pTABE_v2, but harbors additional UGIs following the DddA-C and DddA-N halves was significantly lower (1.5%).

We speculated that the spatial position of TadA8e to the target adenine might influence editing efficiency and that a single adenine deaminase might not be optimal. To alleviate this, we added TadA8e to both TALE arrays in addition to one of the DddA halves (pTABE_v5). Indeed, pTABE_v5 exhibited a significantly higher activity (32.9%) than pTABE_v2, irrespective of the orientation of the pair (Figure 1F). Finally, we used the DddA variant DddA6 which has been reported to exhibit an increased catalytic activity (Mok *et al*., 2022a) as DddA6-N and DddA6-C halves (pTABE_v6). pTABE_v6 showed a slightly higher activity (37.5%) in one orientation, but a slightly lower one (29.6%) in the other orientation which was comparable to the normal DddA halves.

To compare the efficiencies of the TALE-base editors with CRISPR/Cas9 editors, we designed a sgRNA for the target sequence (Figure 1A) and used a catalytically dead dSpCas9-adenine base editor (TadA8e-dSpCas9) targeting the GUS^{*424} reporter. We chose dCas9 instead of a Cas9-nickase for a fair comparison, because TALE-base editors are also not able to guide repair via nicking of the non-edited DNA strand. The activity of this CRISPR/Cas9 base editor was slightly higher (42.7%), but in a similar range than the two TALE-base editors (pTABE_v5, pTABE_v6) (Figure 1F).

These results indicate that split-DddA but not full-length DddA^{E1347A} could increase A•T-to-G•C editing in TALE-ABEs in our *N. benthiamiana* reporter system. Moreover, pair TALE arrays containing TadA8e on both TALEs and split-DddA or DddA6 halves can further increase editing efficiency.

pTABE_v4 allows C•G-to-T•A editing

Because some pTABEs contain split DddA halves which are in principle catalytically active, we wondered whether these pTABEs can also edit cytosines. Hence, we developed a GUS reporter (GUS^{G537}) with a missense mutation that converts an amino acid in the active center of the enzyme from glutamate to glycine to abrogate its activity (Islam *et al*., 1999). Cytosine base editing can revert this mutation to restore the GUS activity. The target site contains a TC motif which is required for DddA activity (Figure 2A).

We then constructed TALE-CBEs (DdCBEs) that resembled the original design of Mok *et al*. with split $DddA_{tox}$ halves and UGI fused to each TALE array (Mok *et al*., 2020) as a positive control (Figure 2B). By utilizing the same left and right TALE arrays, the DdCBEs and pTABE_v4 (containing UGI) showed similar GUS activity of average 35.5% and 38.6%, whereas pTABE_v6 (without UGI) showed a very low GUS activity of 2.9% (Figure 2C). This indicates that the addition of UGI to a pair of TALE-ABEs (pTABE_v4) enables efficient C•G-to-T•A conversion, however, the absence of UGI (pTABE_v6) largely prohibits cytosine targeting.

$DddA^{E1347A}$ makes dsDNA accessible for ssDNA-specific deaminases

It was puzzling why the ssDNA-specific TadA8e can efficiently use a dsDNA substrate when fused to DddA-N and DddA-C halves. Our hypothesis was that the DddA acts on dsDNA, e.g. by partially unwinding the double strand and revealing ssDNA locally. To address this, we tested the activity of another highly ssDNA-specific deaminase, the cytosine deaminase human APOBEC3A (A3A), using our cytosine GUS^{G537} reporter (Figure 3A). For this, $DddA^{E1347A}$ was fused to the left TALE array and the highly active A3A variant $A3A^{Y130F}$ (Zhou *et al*., 2019; Ren *et al*., 2021) was fused to the right TALE array. A3A only exhibits cytosine deaminase activity against ssDNA (Moraes *et al*., 2021) while $DddA^{E1347A}$ has no cytosine deaminase activity against both, dsDNA and ssDNA (Mok *et al*., 2020). When targeting the GUS^{G537} reporter with $DddA^{E1347A}$ / $A3A^{Y130F}$ CBEs in *N. benthamiana* GUS assays, the combination of $DddA^{E1347A}$ and $A3A^{Y130F}$ CBEs exhibited a very high GUS activity of 48.7%, while $A3A^{Y130F}$ and $DddA^{E1347A}$ alone only show 10.3% and 3.6% GUS activity, respectively

(Figure 3B). This suggests that DddA^{E1347A} generally makes target bases in dsDNA more accessible for local ssDNA-specific deaminases, possibly by partially unwinding the dsDNA.

The spatial requirement of pair TALE-adenine base editors

To apply base editors, it is crucial to know the editing window, i.e., the target region where the deaminase is acting, relative to the DNA-binding site of the tool. Previously, this has been studied for TALE-base editors by amplicon sequencing of target regions which revealed the editing efficiencies of different possible bases in the regions (Mok *et al*., 2020; Mok *et al*., 2022a). In contrast, we aimed to understand how a pair of TALE-ABEs should be positioned to modify a specific target base. Therefore, we constructed TALE arrays of different length flanking the target adenine in the GUS^{*424} reporter (Figure 4A). The different combinations of left and right TALEs allow to test different sizes of spacer regions (from 4 to 16 nucleotides; 4-nt to 16-nt), and to vary the position of the targeted adenine in the spacer (from position 2 to 6; A2-A6) (Figure 4B). We tested the editing efficiencies of pTABE_v2 (with DddA-halves; Figure 4C) and pTABE_v6 (with DddA6-halves; Figure 4D) with different TALE combinations separately in *N. benthamiana* GUS assays. Across the 6-nt to 16-nt spacers, pTABE_v2 achieved the highest editing efficiency at position A4 in the 10-nt and 12-nt spacer regions, and the A4 editing activities are significantly reduced when the spacer is extended to 14-nt or shortened to 8-nt. When the targeted adenine was located at A2 or A6, pTABE_v2 showed poor editing activities. Similarly, pTABE_v6 also showed high activity at A4 in the 10-nt to 12-nt spacer regions. In addition, pTABE_v6 still showed significant activities at A6 in the 14-nt spacer and A2 in the 10-nt spacer. These results indicate that both pTABE_v2 and pTABE_v6 prefer target adenines located at A4 in a spacer region of 10-12 nucleotides.

Refining the editing range in the spacer region of pair TALE-base editors in plant protoplasts

More than one adenine might be edited in the spacer region of pair TALE-base editors, in particular, if both DNA strands could potentially be targeted. To reveal the editing range of pTABE_v6 in comparison to TadA8e-dSpCas9, we amplified the target region in the GUS^{*424}

reporter from DNA of *N. benthamiana* leaves infiltrated with *Agrobacterium* strains co-delivering the GUS reporter and base editor components and sequenced the amplicons by next-generation sequencing. Both, pTABE_v6 and TadA8e-dSpCas9 showed the highest adenine editing activity at position A4 (which restored the stop-codon to glutamine) with average 0.2% and 0.9% editing efficiencies, respectively (Figure 5A). In addition, TadA8e-dSpCas9 showed 0.1% and 0.7% efficiencies at A2 and A8, respectively, while pTABE_v6 showed very low editing (less than 0.1%) at these sites.

To further characterize the targeting capabilities of TALE-ABEs on plant nuclear loci, we used pTABE_v6 to target three chromosomal rice genes (*OsALS*, *OsSWEET14*, and *OsPDS*) and one chromosomal *N. benthamiana* gene (NB-T1) in protoplasts. Amplicon sequencing showed that, in the 16-nt spacer region of *OsALS* three A•Ts were edited (T9, T11, and T14) with efficiencies from 0.2% to 0.8% (Figure 5B). In *OsSWEET14* two A•Ts were edited with an efficiency of 1% for A6 and 0.1% for A15 (Figure 5C). In *OsPDS* four A•Ts were edited with the highest editing efficiency of 1.5% at A8 (Figure 5D). In NB-T1 only very low editing was detectable at T10 (Figure 5E). The average editing product purity of pTABE_v6 reached 97.6% with 2.4% transversions to C or G, and we did not identify any C•G-to-T•A editing in those five target sites (Figure 5F). These results show that pTABE_v6 generates A•T-to-G•C conversions with high product purity and can target chromosomal loci in plant cells.

To demonstrate targeted A•T-to-G•C editing in rice organelles, we chose the chloroplast gene *psaA* for the TALE-ABE editing (Figure 6A). Mutations in the *psaA* gene resulted in an albino phenotype because of decreased chlorophyll production (Li *et al*., 2021). As expected, we found that several regenerated lines exhibited the albino phenotype (Figure 6B). Sanger sequencing chromatogram from line 7 displays a desired conversion of A•T-to-G•C in the spacer region (Figure 6C). We noted that undesired editing was also detected at the TALE binding regions. This bystander editing, caused by TALE-ABE, has also been reported in mammalian cells (Cho *et al*., 2022). Additional research is required in the future to explore the use of TALE-DdCBEs and TALE-ABEs for producing genetically modified plants and examining their specificity *in vivo*.

Off-target editing by pair TALE-ABEs

TALE-CBEs directed to the mitochondria have been reported to result in off-target editing in mitochondria and even the nuclear chromosomes (Lei *et al*., 2022; Mok *et al*., 2022a; Lee *et al*., 2022). This off-target editing appears puzzling, given that DNA-recognition by TALE arrays is considerably specific and the requirement for two neighboring binding sites makes TALEN pairs explicitly specific. One possibility is, that the interaction between the two DddA halves is strong enough to enable reconstitution of the functional deaminase even if only one TALE array is bound to an off-target site.

To profile the off-target editing of our pair TALE-ABEs, we designed two pairs of TALE arrays based on pTABE_v6 with one TALE array binding to the target site in the GUS^{*424} reporter (shown as L-TadA8e-6N or R-TadA8e-6C, Figure 7) and the other one containing non-targeted RVD repeats that can not bind to the target sites (shown as NT-TadA8e-6N or NT-TadA8e-6C, Figure 7). Determining base editing activity in *N. benthamiana* GUS assays, we found that the combination of L-TadA8e-6N / NT-TadA8e-6C showed an average of 8.3% GUS activity (positive control is 62.4%, L-TadA8e-6N / R-TadA8e-6C) while L-TadA8e-6N alone (without the NT-TadA8e-6C) showed a background activity of only 0.4%. This indicated that there is a considerable editing if only one of the two pair TALE-ABEs is bound to the target site. If the DddA-halves are switched, the combination of NT-TadA8e-6N / R-TadA8e-6C or the single R-TadA8e-6C alone both lead to about 3% editing activity (positive control is 62.4%). This shows that one of the pair TALE-ABEs can trigger low-level off-target editing when bound to a DNA location.

Discussion

In the present study, we tested two different designs of TALE-adenine base editors, single TALE-ABE editors and paired TALE-ABE editors, and refined how to apply them in plant cells on nuclear target sequences. This work extends the initial description of efficient A•T-to-G•C editing via TALE-based genome editing tools (Cho *et al*., 2022). In comparison to the previous work which was focused on mitochondrial DNA in human cells, we established GUS reporter

to score cytosine and adenine base editing in the plant nucleus. To accomplish a quick assembly of different tool designs, we based all components on a modular MoClo design.

Our results confirm that fusion of DddA to the adenine deaminase is crucial to achieve a high adenine editing efficiency, also in plant cells. For this, we found that fusion of split halves of DddA or the enhanced DddA6 variant (Mok *et al*., 2022a) to the TadA8e adenine deaminase variant (Richter *et al*., 2020) is most effective compared to other protein fusion strategies. A spacer length of 10-12 nt is optimal, with the targeted adenine at position 4. Previous studies have shown that TadA8e has a higher DNA deaminating ability than other TadA variants. Although TadA8e can not access dsDNA, it is *in vitro* capable of rapidly deaminating transiently generated single-strand DNA that might occur during the search process of CRISPR/Cas-systems for target sites (Lapinaite *et al*., 2020). Accordingly, TadA8e fused to a TALE array alone only resulted in very low adenine editing in our assay which confirms that Tad8e itself can not act efficiently on dsDNA. In contrast, our experiments now support a model that the dsDNA-specific DddA provides access to ssDNA. A TALE-DddA^{E1347A} fusion enabled efficient base editing by the ssDNA-specific cytosine deaminase APOBEC3A (A3A^{Y130F}) when fused to a TALE array. A possible explanation is that DddA unwinds dsDNA locally to facilitate its own cytosine deaminase activity. Similarly, the catalytically dead DddA^{E1347A} variant transiently provides ssDNA as a substrate for A3A^{Y130F} and the split DddA variants enabled Tad8e activity in our reporter assays. Furthermore, DddA and DddA6 are limited to a 5'-TC context for cytosine base editing (Mok *et al*., 2020; Mok *et al*., 2022a), however, split DddA and DddA6 in our pTABEs do not require a 5'-TC motif to facilitate adenine base editing of Tad8e, which is consistent with the previous study in mammalian cell lines (Cho *et al*., 2022). Taken together, this opens up interesting questions regarding the mechanistic details whether and how dsDNA-specific cytosine deaminases like DddA possibly target DNA bases in a two-step process of unwinding DNA and subsequent base deamination.

Cho *et al*. showed an efficient A•T-to-G•C conversion in human mitochondrial DNA via monomeric TALEDs (similar architecture as sTABE_v3), dimeric TALEDs (similar architecture as pTABE_v1), and split TALEDs (similar architecture as pTABE_v2) (Cho *et al*., 2022). In contrast, in our GUS reporter assays, both sTABE_v3 and pTABE_v1 show only low activity

compared to pTABE_v2. It is worth noticing that there is a difference how the base modification is fixed in human mitochondria and the *N. benthamiana* transient expression system, respectively. The circular mitochondrial DNA in a multiplying cell culture is replicating quickly, which fixes heteroduplex situations into mutations in one of the daughter molecules without the need for a DNA repair process. This enhances mutational changes by TALE-base editors which don't nick DNA and which in contrast to CRISPR/Cas9 base editors (Komor *et al*., 2016) are unable to guide the replacement of the non-edited DNA strand. Furthermore, amplicon sequencing can overestimate base editing mutation rates, because also transient, non-resolved heteroduplexs are amplified and scored as mutations. In our GUS assay, the base deaminases target a base in the template strand which is directly transcribed into the desired modification even if a heteroduplex still exists, which also potentially enhances the apparent editing. The target adenine in our GUS reporter is part of a TA motif which is favored by TadA8e (Wu *et al*., 2022). At the same time, we can only detect TAA to CAA edits in the GUS^{*424} reporter and other base edits are not detected which might result in an underestimation of total editing rates.

So far, TALE-base editors have achieved near-homoplasmic editing rates in full organisms, namely Arabidopsis, lettuce, and rice (Kang *et al*., 2021; Nakazato *et al*., 2021; Li *et al*., 2021; Mok *et al*., 2022b). At present no targeted editing in the nuclear chromosomes by TALE-base editors have been reported. Although we achieved high editing rates in our reporter system which in principle is nuclear localized, and we detected editing of rice and *N. benthamiana* chromosomal loci, we did not accomplish to regenerate edited T0 plants, yet. Either, the base changes are not fixed efficiently in the nuclear chromosome without nicking of the non-edited strand or TALE-base editors have an overall deleterious effect on the cell, e.g., by high off-target rates, which hinders plant regeneration. Substantial nuclear off-target editing of TALE-CBEs have been reported even when the tool was directed to the mitochondria (Lei *et al*., 2022). The two DddA halves were speculated to associate to a functional enzyme even at sites where only one TALE array is bound. We could confirm that the activity of our pTABE_v6 editor can be detected when only one TALE array is binding the target locus. Apparently, spontaneous re-association of the DddA halves promotes this. The use of an engineered low-

off target DddA (Lei *et al*., 2022; Lee *et al*., 2022) might alleviate this. Also, in particular the TadA8e variant causes elevated off-target editing in plant genomes when used in CRISPR/Cas9 base editors (Wu *et al*., 2022). Possibly, this editing is linked to transient ssDNA areas caused by the Cas9 target search and would not appear by TALE-TadA8e tools.

In summary, we have refined the optimal architecture of the TALE-adenine base editing system in plant cells. This system can now be applied for mitochondrial and chloroplast genome editing to accelerate crop improvement. Future work will address the efficiency for nuclear chromosomal editing.

Methods

Plasmid Construction

We used the modular cloning (MoClo) syntax (Weber *et al*., 2011; Geiβler *et al*., 2011; Grützner and Marillonnet, 2020) to generate the TALE-ABE plasmids. For this, the components were subcloned in individual modules that can be assembled using Golden Gate Cloning (Engler *et al*., 2008). The details of the cloning procedures are listed in Supplemental Methods and Supplemental Figure 1. The plasmid modules used in this study were listed in Supplemental Table 1.

Plant growth condition

Nicotiana benthamiana plants were grown in a greenhouse with 16 hours of light, a relative humidity of 40-60%, and temperatures of 23°C and 19°C during the day and night, respectively. Four to six weeks old plants were used for *Agrobacterium tumefaciens* inoculation experiments.

Nicotiana benthamiana infiltration and GUS reporter assay

GUS reporter assays were performed as previously described (Boch *et al*., 2009). Briefly, *A. tumefaciens* GV3101 strains containing a TALE-ABE construct, the GUS reporter construct, and a p19 silencing inhibitor, respectively, were mixed 1:1:1 with OD_{600} of 0.8 and inoculated into *N. benthamiana* leaves. Two to three days after inoculation, two leaf discs (diameter 0.8 cm) were harvested from each inoculation spot. Leaf tissues were homogenized and incubated with 4-methyl-umbelliferyl-β-D-glucuronide. GUS activities were measured using a TECAN reader (360 nm excitation and 465 nm emission). For details see Supplemental Figure 2. Proteins were quantified by NanoDrop™ One (Thermo Fisher Scientific). Leaf disks were stained in X-Gluc solution and de-stained in ethanol.

Protoplast isolation and transformation

We used rice cultivar Kitaake leaves to prepare rice protoplasts. Protoplast isolation and transformation were performed as previously described (Shan *et al*., 2014). 10 µg plasmid DNA per construct were introduced into protoplasts by PEG-mediated transfection. The

transfected protoplasts were incubated at room temperature. After 48 h, the protoplasts were collected and the genomic DNA extracted.

DNA extraction and amplicon sequencing

Plant genomic DNA was extracted with the innuPREP Plant DNA Kit (Analytik Jena). The targeted sequences were amplified with specific primers, and the amplicons were purified with the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and quantified using Qubit™ 1X dsDNA High Sensitivity Kits (Thermo Fisher Scientific). Oligos used in this study were list in Supplemental Table 2. Equal amounts of PCR products were pooled and sequenced (GENEWIZ, AMPLICON-EZ). Amplicon sequencing was performed three times for each target location using genomic DNA isolated from three different protoplasts transformation experiments. The target sites in the sequenced reads were analyzed for mutations using CRISPResso2 (crispresso2.pinellolab.org; Clement *et al*., 2019).

Statistical analysis

All values are shown as means ± SEM (standard error of the mean). Statistical differences between the values were tested using two-tailed unpaired Student's t-tests by GraphPad (Prism; www.graphpad.com).

Data availability

The amplicon sequencing data have been deposited in an NCBI BioProject database (PRJNA909199). Plasmids used in this study will be made available through Addgene.

Funding

The work was conducted using university core funding.

Author contributions

D.Z. and J.B. designed the experiments. D.Z. performed the experiments and analysed the data. D.Z. and J.B. wrote the manuscript.

Acknowledgments

The authors thank Sebastian Becker and Jana Streubel for assistance. No conflict of interest is declared.

References

Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. **38**:824–844.

Baumann, H., Knapp, S., Karshikoff, A., Ladenstein, R., and Härd, T. (1995). DNA-binding surface of the Sso7d protein from *Sulfolobus solfataricus*. J. Mol. Biol. **247**:840–846.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326**:1509–1512.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. Annu. Rev. Plant Biol. **70**:667–697.

Cho, S.-I., Lee, S., Mok, Y. G., Lim, K., Lee, J., Lee, J. M., Chung, E., and Kim, J.-S. (2022). Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. Cell **185**:1764-1776.e12.

Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., Cole, M. A., Liu, D. R., Joung, J. K., Bauer, D. E., *et al*. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. **37**:224–226.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One **3**:e3647.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. Cell **184**:1621–1635.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. Nature **551**:464–471.

Geißler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.-E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-Specificity. PLoS One **6**:e19509.

Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: Expanding the types of DNA damage products harnessed for genome editing. Gene Genome Ed. **1**:100005.

Grützner, R., and Marillonnet, S. (2020). Generation of MoClo standard parts using Golden Gate cloning. In DNA Cloning and Assembly: Methods and Protocols (ed. Chandran, S. and George, K. W.), pp. 107–123. New York, NY: Springer US.

Huang, X., and Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **65**:531–551.

Islam, M. R., Tomatsu, S., Shah, G. N., Grubb, J. H., Jain, S., and Sly, W. S. (1999). Active site residues of human beta-glucuronidase: Evidence for Glu(540) as the nucleophile and Glu(451) as the acid-base residue. J. Biol. Chem. **274**:23451–23455.

Jiang, F., and Doudna, J. A. (2017). CRISPR–Cas9 structures and mechanisms. Annu. Rev. Biophys. **46**:505–529.

Kang, B.-C., Bae, S.-J., Lee, S., Lee, J. S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., and Kim, J.-S. (2021). Chloroplast and mitochondrial DNA editing in plants. Nat. Plants **7**:899–905.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature **533**:420–424.

Lapinaite, A., Knott, G. J., Palumbo, C. M., Lin-Shiao, E., Richter, M. F., Zhao, K. T., Beal, P. A., Liu, D. R., and Doudna, J. A. (2020). DNA capture by a CRISPR-Cas9–guided adenine base editor. Science **369**:566–571.

Li, R., Char, S. N., Liu, B., Liu, H., Li, X., and Yang, B. (2021). High-efficiency plastome base editing in rice with TAL cytosine deaminase. Mol. Plant **14**:1412–1414.

Lim, K., Cho, S.-I., and Kim, J.-S. (2022). Nuclear and mitochondrial DNA editing in human cells with zinc finger deaminases. Nat. Commun. **13**:366.

Liu, G., Lin, Q., Jin, S., and Gao, C. (2022). The CRISPR-Cas toolbox and gene editing technologies. Mol. Cell **82**:333–347.

Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., Hsu, F., Radey, M. C., Peterson, S. B., Mootha, V. K., *et al*. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature **583**:631–637.

Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., and Liu, D. R. (2022a). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nat. Biotechnol. **40**:1378–1387.

Mok, Y. G., Hong, S., Bae, S.-J., Cho, S.-I., and Kim, J.-S. (2022b). Targeted A-to-G base editing of chloroplast DNA in plants. Nat. Plants Advance Access published December 1, 2022, doi:10.1038/s41477-022-01279-8.

Moraes, M. H., Hsu, F., Huang, D., Bosch, D. E., Zeng, J., Radey, M. C., Simon, N., Ledvina, H. E., Frick, J. P., Wiggins, P. A., *et al*. (2021). An interbacterial DNA deaminase toxin directly mutagenizes surviving target populations. eLife **10**:e62967.

Nakazato, I., Okuno, M., Yamamoto, H., Tamura, Y., Itoh, T., Shikanai, T., Takanashi, H., Tsutsumi, N., and Arimura, S. (2021). Targeted base editing in the plastid genome of Arabidopsis thaliana. Nat. Plants **7**:906–913.

Ren, B., Yan, F., Kuang, Y., Li, N., Zhang, D., Zhou, X., Lin, H., and Zhou, H. (2018). Improved base editor for efficiently inducing genetic variations in rice with CRISPR/Cas9-Guided hyperactive hAID Mutant. Mol. Plant **11**:623–626.

Ren, Q., Sretenovic, S., Liu, G., Zhong, Z., Wang, J., Huang, L., Tang, X., Guo, Y., Liu, L., Wu, Y., *et al*. (2021). Improved plant cytosine base editors with high editing activity, purity, and specificity. Plant Biotechnol. J. **19**:2052–2068.

Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. Nat. Protoc. **9**:2395–2410.

Tan, J., Zeng, D., Zhao, Y., Wang, Y., Liu, T., Li, S., Xue, Y., Luo, Y., Xie, X., Chen, L., *et al*. (2022). PhieABEs: a PAM-less/free high-efficiency adenine base editor toolbox with wide target scope in plants. Plant Biotechnol. J. **20**:934–943.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene Constructs. PLoS One **6**:e16765.

Wu, Y., Ren, Q., Zhong, Z., Liu, G., Han, Y., Bao, Y., Liu, L., Xiang, S., Liu, S., Tang, X., *et al*. (2022). Genome-wide analyses of PAM-relaxed Cas9 genome editors reveal substantial off-target effects by ABE8e in rice. Plant Biotechnol. J. **20**:1670–1682.

Zhang, X., Chen, L., Zhu, B., Wang, L., Chen, C., Hong, M., Huang, Y., Li, H., Han, H., Cai, B., *et al*. (2020). Increasing the efficiency and targeting range of cytidine base editors through fusion of a single-stranded DNA-binding protein domain. Nat. Cell Biol. **22**:740–750.

FIGURE LEGENDS

Figure 1. Establishment of TALE-ABEs in *N. benthamiana*.

(A) Schematic of the GUS^{*424} adenine base editing reporter. The A•T-to-G•C (highlight in red) editing in GUS^{*424} can alter the stop codon (TAA) to Gln (CAA) and restore GUS activity. TALE binding sites are in gray background, spacer region in cyan background, sgRNA targeting sequence and PAM are indicated by a red and blue line, respectively. **(B)** Architectures of six single TALE-ABEs (sTABE_v1 - sTABE_v6) and six pair TALE-TABEs (pTABE_v1 - pTABE_v6). bpNLS: bipartite nuclear localization signal; N / C: DddA-N / DddA-C halves split at G1397; 6N / 6C: Ddd6A-N / Ddd6A-C halves split at G1397. **(C-F)** *A. tumefaciens* strains delivering constructs were mixed and infiltrated into *N. benthamiana* leaves. GUS activities were measured and normalized to 35S::GUS (WT GUS, positive control). Values are confirmed in independent experiments. **(C)** A•T-to-G•C editing efficiencies of six sTABEs binding to the left (left TALE) or right (right TALE) site in GUS^{*424}, n=8. **(D)** A•T-to-G•C editing efficiencies of sTABEs_v3 and pTABE_v1 binding to the left (L) or right (R) site in GUS^{*424}, n=4. **(E)** A•T-to-G•C editing efficiencies of pTABEs_v2, pTABE_v3, and pTABE_v4 at GUS^{*424} targeting sites, n=3. **(F)** A•T-to-G•C editing efficiencies of three pTABEs (pTABEs_v2, pTABE_v5, and pTABE_v6) and Tad8e-dSpCas9 at GUS^{*424} targeting sites, n=4. GUS^{*424}: negative control, w/o GUS (without GUS^{*424}, pTABEs only): negative control. Values and error bars indicate the mean ± SEM, *p < 0.05; **p < 0.01; n.s. (not significant) using Student's two-tailed unpaired t-test.

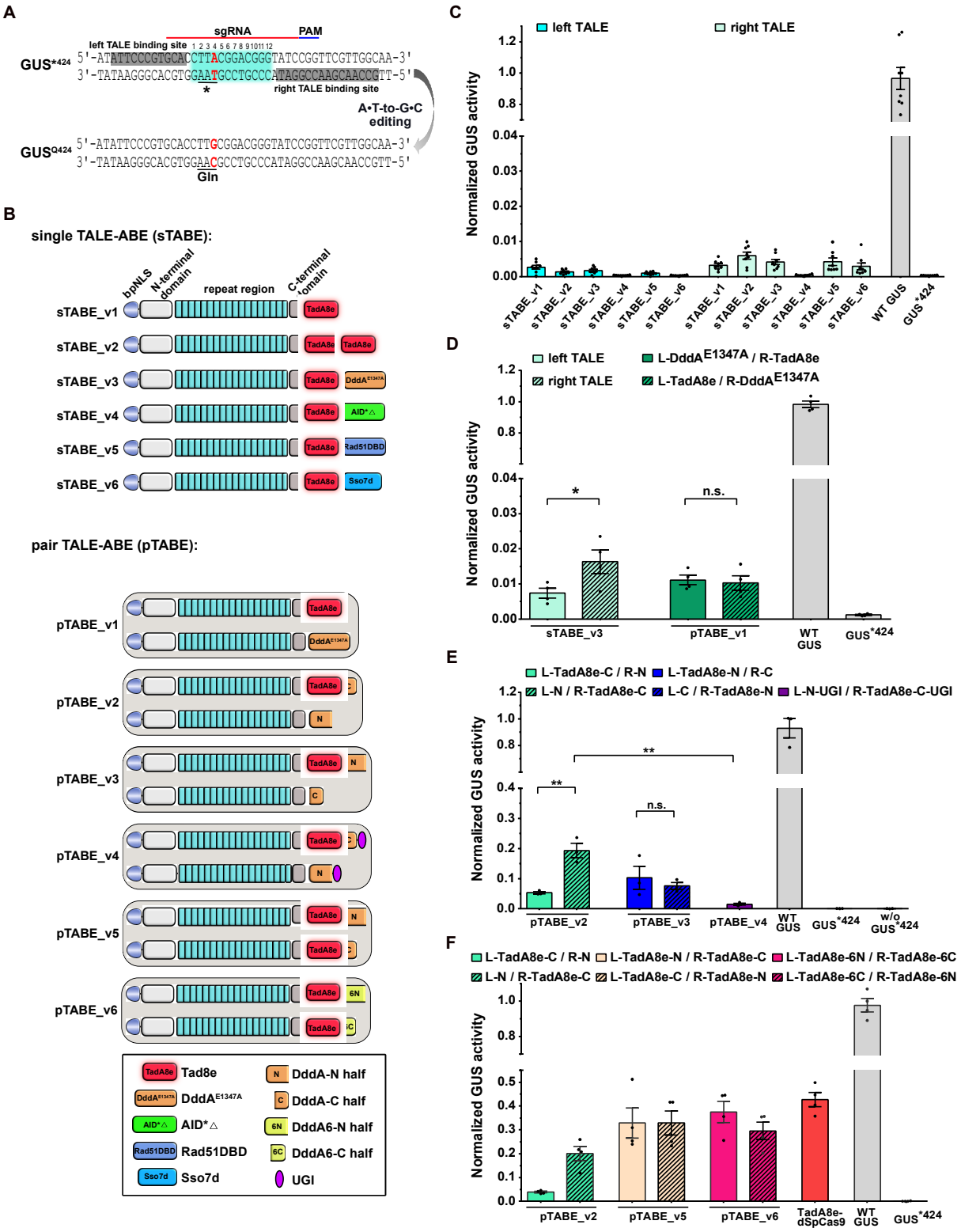

A
sgRNA
PAM
left TALE binding site
GUS*424
5'-ATATTCCCGTGCACCTTACGGACGGGTATCCGGTTCGTTGGCAA-3'
3'-TATAAGGGCACGTGGAATGCCTGCCCATAGGCCAAGCAACCGTT-5'
right TALE binding site
A•T-to-G•C editing
GUSQ424
5'-ATATTCCCGTGCACCTTGCGGACGGGTATCCGGTTCGTTGGCAA-3'
3'-TATAAGGGCACGTGGAACGCCTGCCCATAGGCCAAGCAACCGTT-5'
Gln
B
single TALE-ABE (sTABE):
bpNLS
N-terminal domain
repeat region
C-terminal domain
sTABE_v1
sTABE_v2
sTABE_v3
sTABE_v4
sTABE_v5
sTABE_v6
TadA8e
DddA^E1347A
AID*△
Rad51DBD
Sso7d
pair TALE-ABE (pTABE):
pTABE_v1
pTABE_v2
pTABE_v3
pTABE_v4
pTABE_v5
pTABE_v6
N
C
6N
6C
Tad8e
DddA^E1347A
AID*△
Rad51DBD
Sso7d
DddA-N half
DddA-C half
DddA6-N half
DddA6-C half
UGI
C
left TALE
right TALE
Normalized GUS activity
WT GUS
GUS*424
D
left TALE
right TALE
L-DddA^E1347A / R-TadA8e
L-TadA8e / R-DddA^E1347A
*
n.s.
sTABE_v3
pTABE_v1
E
L-TadA8e-C / R-N
L-TadA8e-N / R-C
L-N / R-TadA8e-C
L-C / R-TadA8e-N
L-N-UGI / R-TadA8e-C-UGI
**
n.s.
pTABE_v2
pTABE_v3
pTABE_v4
w/o GUS*424
F
L-TadA8e-C / R-N
L-TadA8e-N / R-TadA8e-C
L-TadA8e-6N / R-TadA8e-6C
L-N / R-TadA8e-C
L-TadA8e-C / R-TadA8e-N
L-TadA8e-6C / R-TadA8e-6N
pTABE_v5
pTABE_v6
TadA8e-dSpCas9

Figure 2. Efficient C•G-to-T•A editing occurs only in the presence of UGI.

(A) Schematic of the GUS^{G537} cytosine base editing reporter. The inactive GUS^{G537} contains Gly (GGA) at position 537. C•G-to-T•A (highlight in red) editing of GUS^{G537} can change the Gly to Glu (GAA) and restore GUS activity. Left and right TALE binding sites in gray background, spacer region in orange background. **(B)** Architectures of TALE-split DddA$_{tox}$ (DdCBE) to target GUS^{G537}. **(C)** C•G-to-T•A editing efficiencies of the cytosine base editor DdCBE, pair adenine base editors with (pTABE_v4) and without (pTABE_v6) UGI. GUS activities were determined from *A. tumefaciens*-infiltrated *N. benthamiana* leaves and normalized to 35S::GUS (WT GUS). Values and error bars indicate the mean ± SEM, n=3.

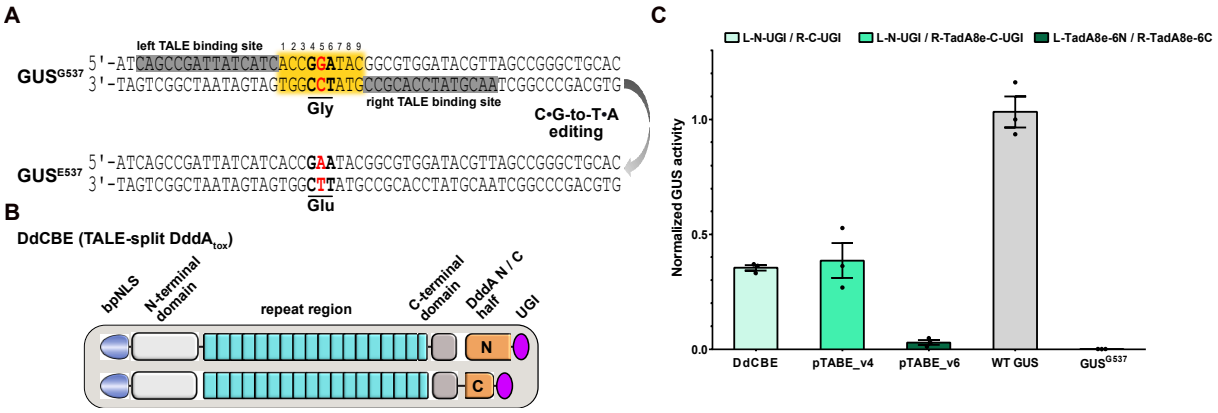


Figure 3. DddA enable efficient base editing of APOBEC3A.

(A) Schematic of the DddA^{E1347A} / A3A^{Y130F} cytosine base editing system targeting the GUS^{G537} cytosine base editing reporter. Left TALE fused with DddA^{E1347A}, right TALE fused with APOBEC3A (A3A^{Y130F}). **(B)** C•G-to-T•A editing efficiencies of cytosine base editors. One representative stained leaf disk of the qualitative assay is shown. GUS activities were determined from *A. tumefaciens*-infiltrated *N. benthamiana* leaves and normalized to 35S::GUS (WT GUS). Values and error bars indicate the mean ± SEM, n=3. ***p < 0.001 using Student's two-tailed unpaired t-test.

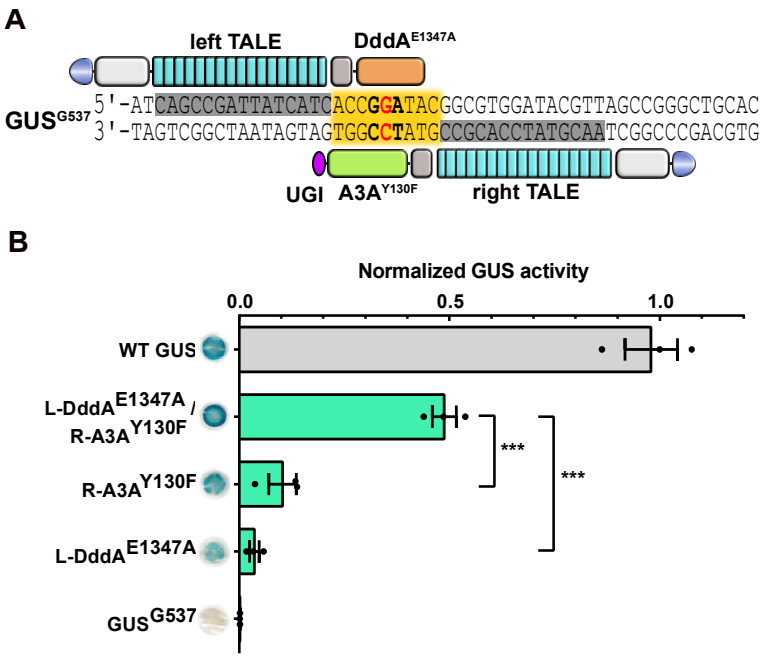
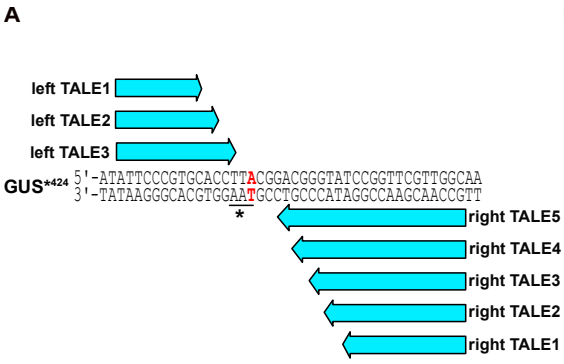


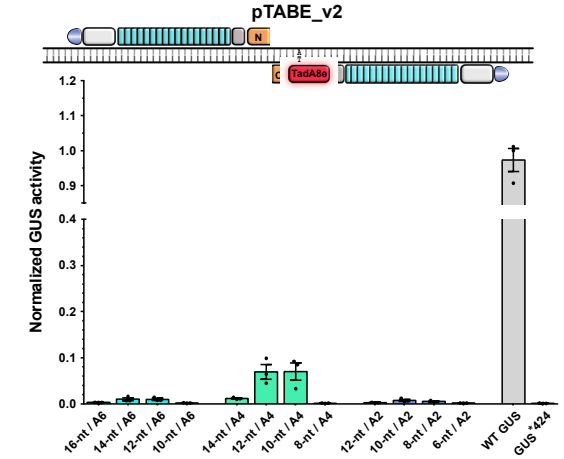
Figure 4. Analyzing the editing windows of pair TALE-ABEs.

(A) Schematic of shifting the editing window of pair TALE-ABEs (pTABEs) and the position of the target adenine by using TALE arrays of different length, the binding sites of three left TALEs and five right TALEs are show by blue arrows. The targeted A•T base pair is in red. **(B)** Different spacer regions (from 4 to 16) flanked by different left and right TALE combinations. The targeted adenine is in red and bold. **(C)** and **(D)** A•T-to-G•C editing efficiencies of pTABE_v2 **(C)** and pTABE_v6 **(D)** in the GUS^{*424} reporter. Top, architecture of pTABE_v2 or pTABE_v6, left TALE binding to the non-coding strand of GUS^{*424} and right TALE binding to the coding strand. Bottom, values and error bars indicating the mean ± SEM, n=3. GUS activities were determined from *A. tumefaciens* infiltrated *N. benthamiana* leaves and normalized to 35S::GUS (WT GUS).



B

pTABE	Spacer sequences	length	position of A
left TALE1 + right TALE1	ACCTTACGGACGGGTA TGGAATGCCTGCCCAT	16-nt	A6
left TALE1 + right TALE2	ACCTTACGGACGGG TGGAATGCCTGCCC	14-nt	A6
left TALE1 + right TALE3	ACCTTACGGACG TGGAATGCCTGC	12-nt	A6
left TALE1 + right TALE4	ACCTTACGGA TGGAATGCCT	10-nt	A6
left TALE1 + right TALE5	ACCTTACG TGGAATGC	8-nt	A6
left TALE2 + right TALE1	CTTACGGACGGGTA GAATGCCTGCCCAT	14-nt	A4
left TALE2 + right TALE2	CTTACGGACGGG GAATGCCTGCCC	12-nt	A4
left TALE2 + right TALE3	CTTACGGACG GAATGCCTGC	10-nt	A4
left TALE2 + right TALE4	CTTACGGA GAATGCCT	8-nt	A4
left TALE2 + right TALE5	CTTACG GAATGC	6-nt	A4
left TALE3 + right TALE1	TACGGACGGGTA ATGCCTGCCCAT	12-nt	A2
left TALE3 + right TALE2	TACGGACGGG ATGCCTGCCC	10-nt	A2
left TALE3 + right TALE3	TACGGACG ATGCCTGC	8-nt	A2
left TALE3 + right TALE4	TACGGA ATGCCT	6-nt	A2
left TALE3 + right TALE5	TACG ATGC	4-nt	A2



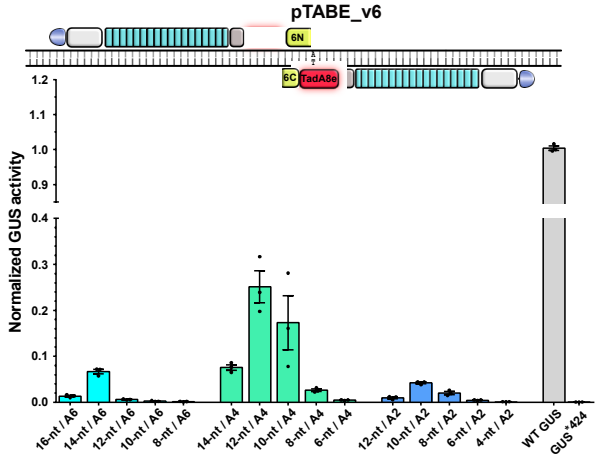


Figure 5. Editing efficiency of pTABE_v6 in rice and *N. benthamiana*.

(A-E) A•T-to-G•C editing efficiencies were determined by amplicon sequencing of target regions from the *A. tumefaciens*-infiltrated GUS*424 ABE reporter **(A)** or transformed rice **(B-D)** and *N. benthamiana* **(E)** protoplasts. Targeted sequences are listed above the panels. Spacer sequences are in bold. sgRNA for TadA8e-dSpCas9 is indicated by a rectangle. Blue: pTABE_v6, red: TadA8e-dSpCas9. **(F)** Product purities of pTABE_v6 from (B-E). Values and error bars indicate the mean ± SEM, n=3.

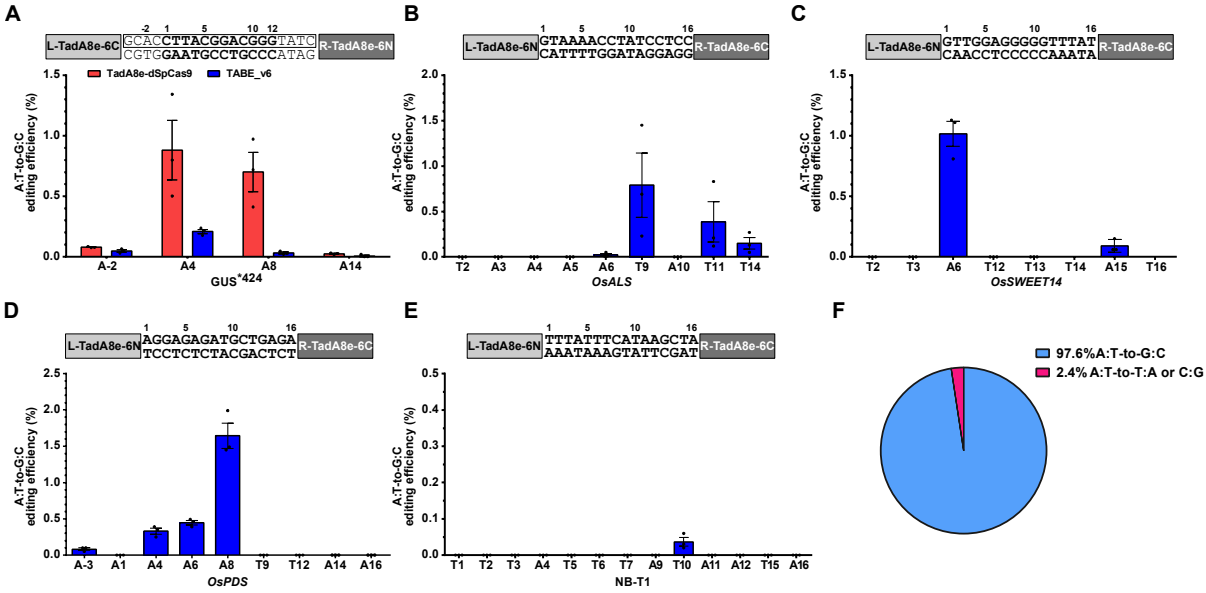


Figure 6. pTABE_v6 induce chloroplast genome editing in rice.

(A) Schematic of A•T-to-G•C editing in rice plants mediated by TALE-ABEs. A pair of pTABE_v6 plasmids, left pTABE_v6 and right pTABE_v6 are separately transformed into *A. tumefaciens* strain EHA105. Two *A. tumefaciens* strains, each containing one of the two pTABE_v6 plasmids, are mixed before transforming rice calli. The regenerated plants are genotyped after 6-7 weeks of selection on medium with 50 mg/L hygromycin. Hyg: Hygromycin. **(B)** Phenotypes of two transgenic lines, line 6 and Line 7. Bar = 1 cm. **(C)** The genotypes of line 6 and Line 7 are shown through the sequencing chromatograms. A•T-to-G•C editing bases are highlighted in red. R represents A and G, while Y represents T and C. WT: wild-type. TALE binding sites are in gray background, spacer region in cyan background.

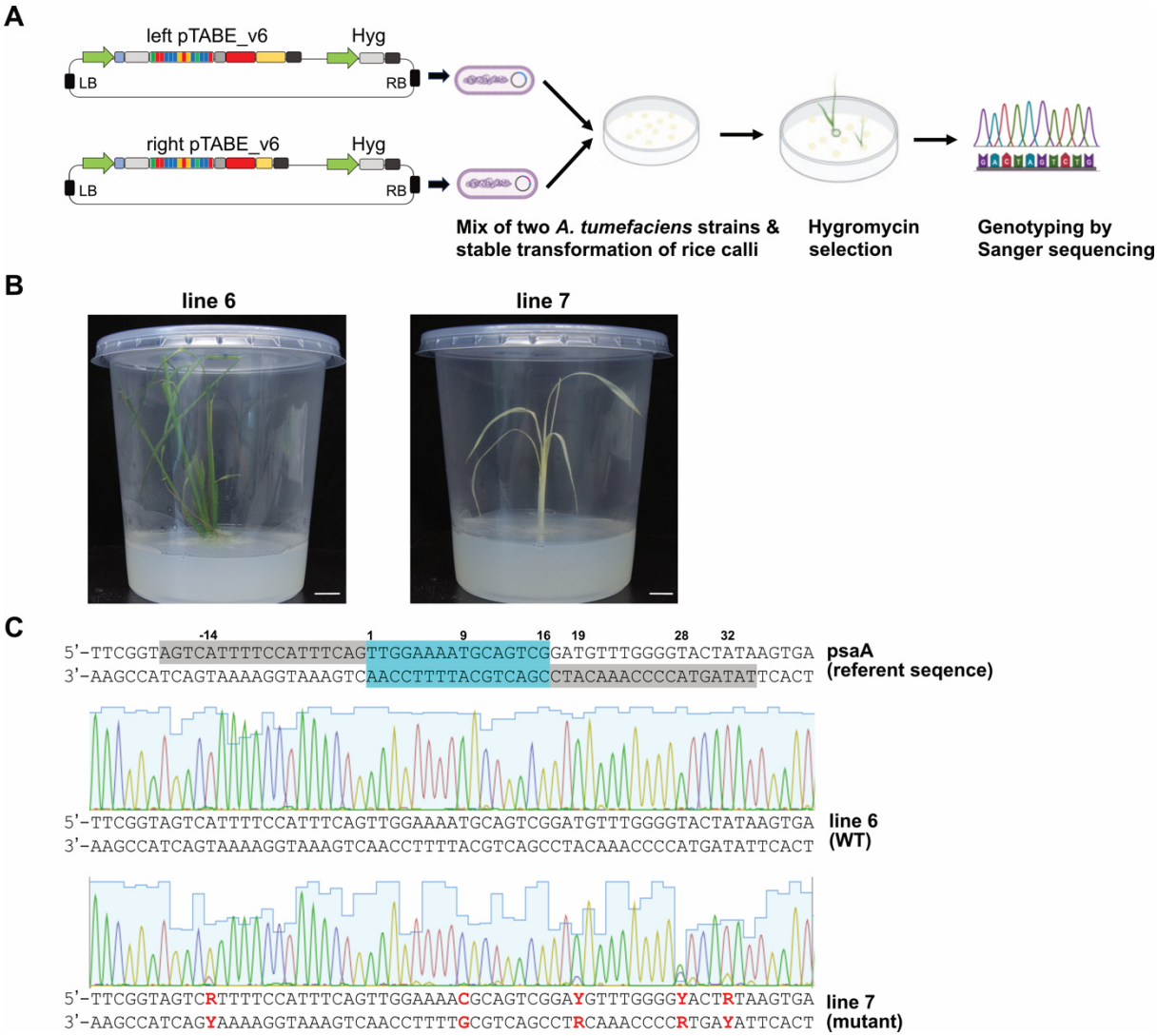
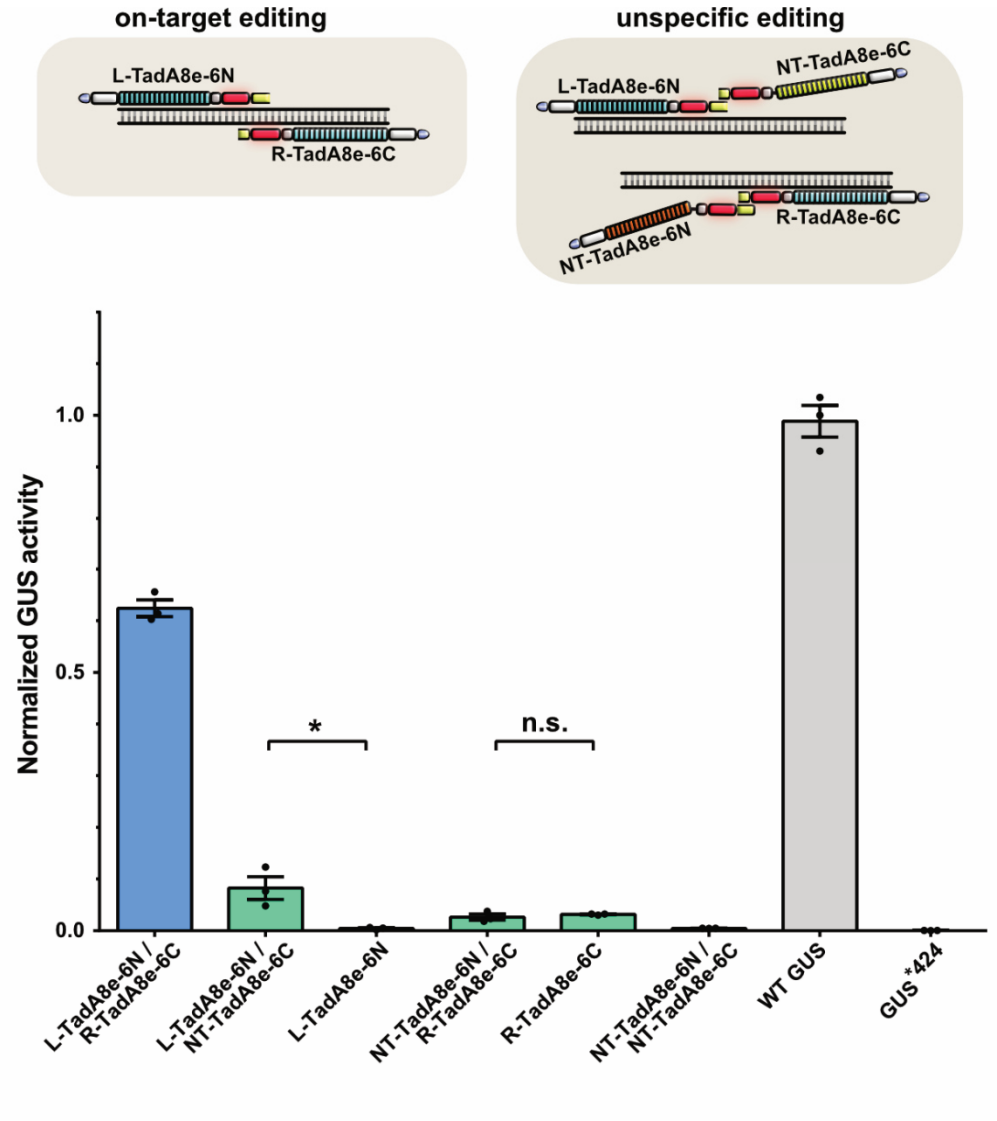


Figure 7. Off-target editing of pTABE_v6.

A•T-to-G•C editing efficiencies by pTABE_v6 with a pair of targeted pTABEs (L-TadA8e-6N / R-TadA8e-6C) or a combination of targeted and non-targeted TABE (L-TadA8e-6N / NT-TadA8e-6C or NT-TadA8e-6N / R-TadA8e-6C). GUS activities were determined from *A. tumefaciens*-infiltrated *N. benthamiana* leaves and normalized to 35S::GUS (WT GUS). Bottom: values and error bars indicate the mean ± SEM, n=3. *p < 0.05; n.s. (not significant) using Student's two-tailed unpaired t-test.



5 General discussion

5.1 Efficient gene targeting in allotetraploid brown mustard by CRISPR/Cas9.

The CRISPR/Cas system enables the induction of mutagenesis in a sequence-specific manner, thereby disrupting genes and facilitating their functional evaluation. This technology can be used for trait improvement in crops. During the course of plant evolution, certain plant species underwent polyploidization (Schaart *et al*., 2021). Polyploidy occurs when a diploid organism acquires one or more extra sets of chromosomes, and polyploids are classified into three types: autopolyploids, allopolyploids, and segmental allopolyploids (Levin, 1983). Genetic manipulation of polyploid plants via chemical and irradiation treatments commonly yields plants with single mutation in one allele only (Jacob *et al*., 2018). Different mutations can be combined in one genotype by crossing and backcrossing, however, this strategy is time consuming. Hence, the gene editing technology which can target multiple alleles simultaneously is particularly powerful to address traits in polyploid plants. CRISPR/Cas-mediated gene editing has been successfully applied in several polyploid crops, including wheat (Wang *et al*., 2014), potato (Andersson *et al*., 2017), rapeseed (Braatz *et al*., 2017), cotton (Gao *et al*., 2017a), Tragopogon (Shan *et al*., 2018), *Panicum virgatum* L. (Liu *et al*., 2018), and strawberry (Wilson *et al*., 2019).

Brassica juncea is a self-pollinated plant and an allotetraploid (AABB; 2n = 36) derived from interspecific hybridization between the diploid progenitors *Brassica rapa* (AA, 2n = 20) and *Brassica nigra* (BB, 2n = 16) (He *et al*., 2021). *Brassica juncea* (AABB) has a large genome (approximately 920 Mb) with sub-genomes A and B, and contains numerous structural variations between the two sub-genomes, making it challenging to generate mutations at multiple genomic sites simultaneously by conventional breeding.

The allergenic 2S albumin Bra j I, derived from the seeds of *Brassica juncea*, has been identified and characterized. It has been found to be reactive with human IgE antibodies in individuals with a sensitivity to mustard (Gonzalez de la Peña *et al*., 2009). In this thesis, two *Bra j I* homoeologs (named as *Bra j IA* and *Bra j IB*) were successfully cloned from two brown mustard lines Terratop (European line) and CR2664 (Indian line). Both *Bra j IA* and *Bra j IB*

possess a length of 537 base pairs encoding for 179 amino acids. Two CRISPR/Cas9 constructs with four independent sgRNA expression cassettes each were designed to target both *Bra j IA* and *Bra j IB* simultaneously. To modify both of the Bra j IA and Bra j IB homeologs simultaneously *Bra j I* homoeologs, two final binary constructs (pBraj1256 harbored four sgRNA sg1, sg2, sg5, and sg6; and pBraj3477 contained four sgRNAs sg3, sg4, sg7A, and sg7B) both carrying genes for SpCas9, hygromycin resistance, and sgRNA expression cassettes were generated. pBraj1256 had the potential to induce a complete deletion of both *Bra j IA* and *Bra j IB* coding sequences by targeting their 5′- and 3′-UTR regions while pBraj3477 was targeting the coding sequences of *Bra j IA* and *Bra j IB* themselves to induce frameshift mutations. For pBraj1256, the complete deletion of *Bra j I* may occur when Cas9 cuts the 5′- and 3′-UTR regions of *Bra j I* within a close time frame. If the cuts happen separately, there is a high possibility of small indels appearing after the DSBs repair. For pBraj3477, target cutting guided by one or more of the four sgRNAs is sufficient to introduce indels at the *Bra j I* coding region, and subsequence DSBs repair may result in frameshift mutations of *Bra j I*. A similar strategy has been applied to create low-gluten wheat by CRISPR/Cas9-mediated modification of immunoreactive α-gliadin genes (Sánchez-León *et al*., 2018). Sanger sequencing indicated that both CRISPR/Cas9 constructs showed a high editing efficiency in mustard. When utilizing pBraj3477, the mutagenesis frequency of *Bra j IA/Bra j IB* was observed to be approximately 33.3% (4/12) and 100% (6/6) in Terratop and CR2664, respectively. Similarly, pBraj1256 resulted in a *Bra j IA/Bra j IB* mutagenesis frequency of 100% (6/6) and 80% (8/10) in Terratop and CR2664, respectively. Those mutations induced by the CRISPR/Cas9 system were stably inherited to the T_1 progeny. Moreover, for the Bra j I deletion construct (pBraj1256) we found that two out of 16 regenerated T_0 mustard lines (T0-1 from CR2664 and T0-47 from Terratop) contain the whole deletion of *Bra j IB* but not *Bra j IA.* Unfortunately, the complete deletion of both *Bra j IA* and *Bra j IB* in the 16 regenerated T_0 mustard lines was not observed. This could be the occurrence of multiple double-strand breaks in genomic DNA generated by Cas9 resulting in a higher frequency of indels caused by error-prone non-homologous end joining repair, as opposed to fragment deletions that arise from the direct joining of two double-strand breaks. Currently, the precise molecular mechanism

underlying the generation of fragment deletion in vivo through the direct linkage of two double-strand breaks remains largely unclear. We speculated that the process leading to such a fragment deletion may require two double-strand breaks to occur within the same time frame, and is likely influenced by the inherent DNA repair mechanisms as well as the spatial separation between the DNA breaks. It is reported that the utilization of the endogenous tRNA processing system for the production of multiple sgRNA in multiplex genome editing has the potential to enhance the frequency of mutagenesis in chromosomal fragment deletion (Xie *et al*., 2015). The processing of pre-tRNAs, which act as spacers between multiple sgRNAs in a polycistronic tRNA-sgRNA transcript, can be accomplished by utilizing cellular RNase P and RNase Z under the control of a single Pol III promoter (Zhang *et al*., 2017; Ding *et al*., 2018). Compared to the conventional sgRNA processing system, the tRNA-sgRNA processing system revealed higher transcription levels of sgRNAs. This improvement can enhance simultaneous mutagenesis of multiple targets or deletion of short chromosomal fragments (Xie *et al*., 2015). This tRNA-sgRNA processing approach may be employed in the future to accomplish the deletion of target genes.

For efficient genome editing in mustard, high efficiency and specificity are necessary when designing sgRNAs. Target sites are easily identified by bioinformatic tools; however, it is much more challenging to predict the on-target and off-target score of each sgRNA *de novo*. The major concern with the CRISPR/Cas9 system is that Cas9 can have off-target effects. Off-target editing might occur at potential genomic sequences with homology to the protospacer which is known as sgRNA-dependent off-target editing. A potential approach to mitigate the sgRNA-dependent off-target editing is to substitute the conventional SpCas9 with high-fidelity SpCas9 variants, such as SpCas9-HF (Kleinstiver *et al*., 2016), eSpCas9(1.1) *(Slaymaker et al*., 2016), and HypaCas9 (Chen *et al*., 2017b). In this thesis, after analyzing 24 predicted sgRNA-dependent off-target sites in two edited T1 lines no off-target mutations were detected. Understanding the off-target mutations in genome edited crops is necessary. Off-target effects may cause unwanted phenotypes or cell toxicity, which could limit its commercial applications. In plants, the undesirable off-target mutations have been reported in *Arabidopsis thaliana* (Zhang *et al*., 2018a), Soybean (Jacobs *et al*., 2015), citrus (Peng *et al*., 2017), and rice (Xie

and Yang, 2013; Endo *et al*., 2015). Zhang and colleagues demonstrated that the utilization of high-specificity sgRNAs did not prevent a significant occurrence of unfavorable mutations induced by CRISPR/Cas9 in *A. thaliana*. And they also indicated that the off-target effects may be exacerbated in subsequent generations if the T-DNA containing the editing tools is still present (Zhang *et al*., 2018a).

Whole genome sequencing (WGS) has proven to be an effective method for detecting many kinds of mutations, including indels, SNPs, and structural variations such as large deletions, inversions, duplications, and rearrangements (Veres *et al*., 2014). Theoretically, the use of WGS has the potential to detect mutations that arise in an edited genome. However, this approach may be constrained by incomplete sequence coverage and alignments. Nevertheless, WGS has already been utilized to identify off-target mutations induced by Cas9 in several species of plants (Zhang *et al*., 2014; Feng *et al*., 2014; Peterson *et al*., 2016). To determine the specificity of SpCas9 in mustard, a more thorough analysis (such as WGS) is suggested.

5.2 The seeds derived from *bra j I* mustard exhibited a decreased in the accumulation of Bra j I protein and a modified phenotype.

The seed storage protein Bra j I is classified as a 2S albumin with an approximate molecular weight of 22 kDa, and is processed into two subunits of 9.5 kDa and 12 kDa. The protein electrophoretic profile of the protein from wild type mustard, as well as from a mustard line that overexpression GFP, and T_1 lines of *bra j I* mutants, were analyzed. The results indicated a deficiency or reduced accumulation of 2S albumin in different lines (T1-22, T1-32), these lines contained frameshift mutations in all *bra j I* alleles. Moreover, immunoblotting results suggested that the absence of Bra j I protein in the seeds of T_1 *bra j I* mutants was confirmed by the absence of any detected anti-Bra j I antibody binding in lines T1-22 and T1-32-1. These lines can serve as a valuable foundation for the development of low-allergenic mustard. An elevated abundance of certain proteins with an approximate molecular weight of 55 kDa was observed in the seed extracts of T1 lines, such as T1-23-16 and T1-22-5, which may suggest that the absence of Bra j I resulted in the accumulation of alternative seed storage proteins.

Researchers used CRISPR/Cas9 to mutate seed storage proteins in soybean and sesame also discovered the accumulation of unknown proteins in the seeds (Sugano *et al*., 2020; You *et al*., 2022). Further evaluation of seed storage proteins through high-performance liquid chromatography (HPLC) or trypsin inhibitor assay may be benificial in assessing the seed quality of these mustard mutants. Serum samples from individuals with mustard allergies could help to further clarify the allergenic properties of the mustard seeds developed in this thesis.

It was observed that the shape of the seeds of various *bra j* I mutants exhibited abnormalities. The observed phenotypes might indicate that Bra j I may have a significant impact on the process of seed development. One possible explanation for these seed phenotypes is that the multiple cutting of dsDNA in the genome could lead to an overabundance of DNA damage, resulting in abnormal cell development, particularly in multicopy genomic regions (Hart *et al*., 2015; Aguirre *et al*., 2016; Wang *et al*., 2017). It has been reported that the CRISPR-STOP method, which introduces early stop codons using CRISPR-cytosine base editors, is an efficient and less harmful alternative to wild-type SpCas9 for gene-knockout studies *(Kuscu et al*., 2017). One possible approach to prevent the occurrence of anomalous seeds is to direct the conversion of one DNA base to another at a programmable target locus without requiring DSBs. The conventional strategy for disrupting genes is through the creation of DSBs within DNA and the repair of DSBs by NHEJ or HDR. The process of gene disruption through NHEJ is highly effective, albeit leading to the formation of mosaic knockout alleles due to the variation of nucleotides inserted or deleted prior to DSB end joining (van Overbeek *et al*., 2016). Furthermore, these DSBs are harmful DNA lesions that can lead to genomic rearrangements and translocations, trigger DNA damage checkpoints, and result in cell death (Aguirre *et al*., 2016). As an alternative to NHEJ- and HDR-dependent genome editing, base editors (CBEs and ABEs) derived by Cas enzymes or TALE proteins can install C•G-to-T•A or A•T-to-G•C base changes in genome. Particularly for CBEs, it can efficiently inactivate genes by precisely converting four alternative codons (CAA, CAG, CGA, and TGG) into stop codons without the formation of DSBs.

Allergies to plant-based food products are a significant public health issue. Peanut allergy affects about 2% of people under the age of 18 in the United States (Bunyavanich *et al*., 2014).

Studies suggest that the prevalence of peanut allergy is rising (Sicherer *et al*., 2010). Ara h 2 has been identified as one of the major peanut allergens, Ara h 2 is recognized by IgE in over 90% of individuals who suffer from peanut allergies (Jong *et al*., 1998; Chatel *et al*., 2003). The successful application of RNAi to target the allergen Ara h 2 in peanuts has been established (Dodo *et al*., 2008). Furthermore, Dodo *et al.* discovered that transgenic peanut seeds containing the RNAi construct produced approximately 25% less Ara h 2 than the wild-type seeds. As a result, the IgE binding of peanut-allergic patient sera to transgenic peanut samples was dramatically reduced compared to wild-type peanut samples (Dodo *et al*., 2008). Soybeans are a major protein used for both human consumption and animal feed worldwide. However, it has been reported that several soy proteins have been identified as allergens (Heppell *et al*., 1987; Katz *et al*., 2014). Japanese scientists utilized CRISPR/Cas9 coupled with *Agrobacterium*-mediated transformation for simultaneous mutagenesis of two soybean allergens, Gly m Bd 28 K and Gly m Bd 30 K. Immunoblot analysis revealed that there was no accumulation of these two proteins in the seeds of the mutant soybeans (Sugano *et al*., 2020). The application of the CRISPR/Cas system for knocking out allergen genes demonstrates a proof-of-concept for developing hypoallergenic crops (Sugano *et al*., 2020; Assou *et al*., 2022). However, direct *in vivo* modification of allergen IgE-binding epitopes without gene dysfunction has not yet been reported. The IgE-binding epitopes present in allergens contribute to allergenicity (Renz *et al*., 2018). Several studies have demonstrated that peanut allergen epitope peptides can be mutated to non-IgE-binding epitopes by a single amino acid change (Burks *et al*., 1997; Stanley *et al*., 1997; Shinmoto *et al*., 2010). Therefore, modifying epitopes can be used as a novel strategy for producting hypoallergenic foods. Thereby, both CBEs and ABEs can be used to make amino acid codon changes, which enable the modification of the epitopes of Bra j I. Although there are multiple epitopes present in the allergens, modifying all of them can be achieved by using base editors for multiplex editing. Together, through the utilization of base editors, it is possible to generate a mustard variant with reduced allergenic properties while maintaining its original characteristics.

5.3 MoClo assembly system of TALE-base editors.

To facilitate the assembly of custom-designed TALE arrays and deaminases in TALE cytosine base editors (TALE-DdCBEs) or adenine base editors (TALE-ABEs), a modular cloning (MoClo) pipeline was applied in this thesis. The MoClo system used in this thesis is advantageous because it is simple, efficient, cost-effective, and standardized compared to other conventional cloning methods, such as restriction enzyme-ligase cloning, Gateway cloning, and Gibson assembly (Weber *et al*., 2011; Geiβler *et al*., 2011; Grützner and Marillonnet, 2020). The restriction enzyme-ligase cloning involves the utilization of type IIP restriction endonucleases (such as EcoRI, HindIII, BamHI, and others.) to cleave the vector plasmid and the inserted DNA fragment. Subsequently, DNA ligases are employed to reconstitute the fragments into a recombinant plasmid. The acceptor plasmids used in restriction enzyme-ligase cloning usually contain multiple cloning sites (MCS) to facilitate the integration of diverse DNA fragment, however, this cloning method is limited by the absence of suitable restriction sites. Gateway cloning technique, which involves donor vectors and destination vectors, is a restriction site-independent approach that utilizes the integration and excision mechanism of the lambda phage to achieve site-specific recombination cloning (Curtis and Grossniklaus, 2003). Nevertheless, Gateway cloning has its disadvantage, including the cost of enzymes used, the presence of undesired *attB* sequence in the final construct, and it is challenging to assemble multiple fragments at the same time. Gibson assembly allows the assembly of multiple overlapping DNA fragments in one reaction (Gibson, 2011). Gibson assembly is a restriction site-independent cloning method as well, but it requires the linearization of both the inserted fragments and acceptor vector containing the overlapping sequence through enzyme digestion or PCR before the final recombination. The MoClo method is a restriction enzyme-ligase cloning technique based on Golden Gate cloning (Weber *et al*., 2011). It involves the use of type IIS restriction endonucleases (such as BpiI and BsaI) to cut outside of their recognition site and generate a 4-base overhang. Those digested fragments and the acceptor vector can then be ligated through the recognition of the specific 4-base overhangs by DNA ligase. When these type IIS enzyme recognition sites are placed at the 5′- and 3′-end of any DNA fragment in inverse orientation, they are removed in the cleavage

process, allowing two DNA fragments flanked by compatible sequence overhangs to be ligated seamlessly. MoClo is convenient for assembling multiple fragments and the modular DNA parts in a defined order (Kang *et al*., 2021a; Li *et al*., 2021b).

There are two popular methods for assembling TALE base editors. One method involves synthesizing or amplifying modules (such as TALE repeats, deaminases, MTS/NLS, and UGI) as gene blocks, then using Gibson assembly to ligate these modules into a recipient acceptor vector (Mok *et al*., 2020; Lei *et al*., 2022). The other method involves using Golden Gate assembly to assemble the TALE repeats from a TALE arrays library in a subcloning step. Then the TALE repeat plasmids are assembled together with the recipient acceptor vectors (recipient acceptor vectors containing TALE N/C-terminal domains, transition peptide, and deaminase) using either Golden Gate assembly (Kang *et al*., 2021a; Li *et al*., 2021b) or Gateway cloning (Nakazato *et al*., 2021). In this thesis, we created a plasmid library to assemble TALE-DdCBEs and TALE-ABEs. All the components from TALE-DdCBEs and TALE-ABEs are modulating as single plasmid modules. By utilizing the Golden Gate assembly, these individual modules can be combined into a complete transcription unit for further experimentation. Our plasmid library is compatible with the established MoClo system (Weber *et al*., 2011; Engler *et al*., 2014), which enriches the genome editing toolbox in the MoClo system for plant research.

5.4 Limitations and improvement of TALE-DdCBEs and TALE-ABEs.

Although TALE-DdCBEs and TALE-ABEs exhibit effective C•G-to-T•A editing and A•T-to-G•C editing, respectively, within dsDNA, their editing capabilities are restricted to the editing window of the respective deaminase. In mammalian cells, the TALE-DdCBEs (G1397-split DddA) exhibit a preference for editing TCs that are situated approximately 4-7 (C4-C7) nucleotides upstream of the 3′-end of 14 to 18-nt spacer regions in both mitochondrial DNA strands (Mok *et al*., 2020) and nuclear targets (Boyne *et al*., 2022). The editing windows of TALE-DdCBEs containing canonical DddA, DddA6, or DddA11 are generally similar (Mok *et al*., 2022b). In this thesis, the TALE-DdCBEs prefer to edit TCs that are positioned at C5-C6 upstream of the 3′-end of the DNA strand in plants which is similar. Moreover, in this thesis, it has been shown that the spacer length can be shortened to 8-nt while maintaining high editing efficiency at C5

or C6 (Figure 9A). TALE-ABEs (pTABE_v2 and pTABE_v6) exhibit a preference for adenines in the target location, specifically at A4 within a spacer region consisting of 10-12 nucleotides (Figure 9B). Researchers have achieved various editing windows of SpCas9-base editors by using different deaminases (Anzalone *et al*., 2020). SpCas9-CBEs were first developed with cytosine deaminase APOBEC1, which exhibited an activity window ranging from 4-8 in the protospacer (where position 1 is distal to the PAM sequence) (Komor *et al*., 2016). Subsequent reported SpCas9-CBEs with different cytosine deaminases, such as *Petromyzon marinus* cytidine deaminase (PmCDA) exhibited an activity window ranging from 1-10 (Nishida *et al*., 2016), and the APOBEC family members APOBEC3A (A3A) exhibited an activity window ranging from 2-13 (Gehrke *et al*., 2018). (Figure 9C). For SpCas9-ABEs, Gaudelli *et al*. initially reported on the SpCas9-ABE variant ABE7.10, which uses a TadA-TadA* dimer as the adenine deaminase, showed an activity window ranging from 4-7 (Gaudelli *et al*., 2017), and the ABE7.10 was further evolved into ABE8e (which uses monomeric TadA8e as the adenine deaminase), which exhibited substantially increased deamination kinetics and an activity window ranging from 4-8 (Richter *et al*., 2020) (Figure 9D). Due to the PAM restriction of SpCas9, modifying certain cytosines or adenines outside the editing windows of SpCas9-base editors is challenging. To overcome the limitations of the editing window, using TALE repeat domains of varying lengths can enable paired TALE base editors to form an optimal editing spacer to introduce the desired modification.

Furthermore, the editing efficiencies of TALE-DdCBEs and TALE-ABEs may be influenced by the used of RVDs (Streubel *et al*., 2012; Becker and Boch, 2021). The reprogramming of the RVDs enables the generation of TALE proteins that binding to specific DNA sequences (Boch *et al*., 2009; Yang *et al*., 2014; Miller *et al*., 2015). The efficiency of RVDs represents a significant improvement to the fundamental DNA-specificity code of TALEs. RVDs with high affinity (strong RVDs), such as NN or HD, form hydrogen bonds with DNA bases, while certain RVDs with lower affinity (weak RVDs), such as NI or NG, only make van der Waals contacts (Deng *et al*., 2012; Mak *et al*., 2012). To design efficient TALEs, certain guidelines are recommended: TALEs should contain three to four strong RVDs. Strong RVDs should be positioned to avoid stretched of weak RVDs, especially at the ends. For high guanine

specificity, it is suggested to use NH to enhance the activity of TALE or use NK if enough strong RVDs are present (Streubel *et al*., 2012; Becker and Boch, 2021).

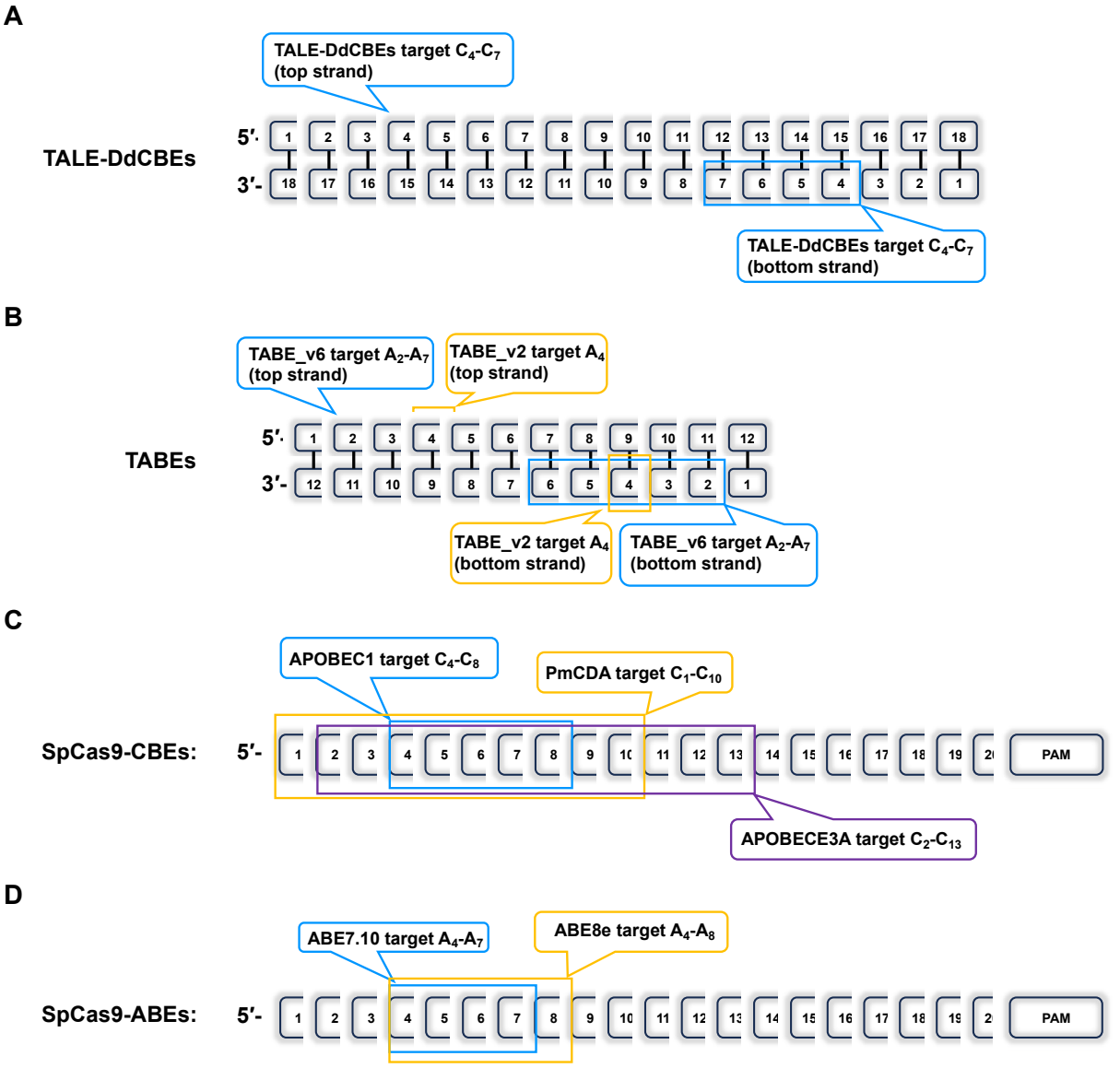


Figure 9. Activity windows of cytosine base editors and adenine base editors. (A) The TALE-DdCBEs' activity windows are shown in blue rectangles over the 18-nt dsDNA spacer region. **(B)** The TABEs' activity windows are shown in blue rectangles (TABE_v2) or yellow rectangles (TABE_v6) over the 12-nt dsDNA spacer region. **(C)** The SpCas9-CBEs' activity windows are shown in a blue rectangle (APOBEC1) or a yellow rectangle (PmCDA) or a purple rectangle (APOBEC3A) over the 20-nt protospacer sequence. **(D)** The SpCas9-ABEs' activity windows are shown in a blue rectangle (ABE7.10) or a yellow rectangle (ABE8e) over the 20-nt protospacer sequence. Numbers in the shaded rectangle represent the positions of individual nucleotides. PAM: protospacer adjacent motif.

In theory, base editors have the potential to avoid the unexpected side effects caused by Cas9, such as the activation of the p53-mediated DNA damage response and the induction of large genomic deletions. These effects are elicited by DSBs and may result in inhibition of cell growth (Haapaniemi *et al*., 2018; Kosicki *et al*., 2018). TALE-DdCBEs and TALE-ABEs have the ability to convert base pairs (C•G-to-T•A and A•T-to-G•C) by using double-stranded DNA as a substrate. However, these base editors are limited in their ability to induce transversion mutations. Researchers have initiated the process of overcoming this limitation by the manipulation of the DNA repair pathways in both cytosine base editing and adenine base editing. C•G-to-G•C base editors (CGBEs) have been developed by fusing the base excision repair enzyme uracil N-glycosylase (UNG) instead of UGI to the nSpCas9-CBE architecture (Zhao *et al*., 2021; Koblan *et al*., 2021). These CGBEs increase the frequency of C•G-to-G•C mutations by promoting the formation of abasic sites at the target cytidine which triggers a somewhat random repair process without template base (Kurt *et al*., 2021).

Modified bases are generally detected, excised, and replaced by the base excision repair (BER) pathway (Beard *et al*., 2019). BER is initiated by DNA glycosylases, which facilitate the cleavage of the glycosidic bond between the nitrogenous base and the deoxyribose sugar of particular damaged nucleotides. This enzymatic activity results in the formation of an abasic site (also known as an apurinic/apyrimidinic site or AP site) (Krokan and Bjørås, 2013). DNA glycosylases, such as UNG, generate abasic sites, which are subsequently processed by AP site DNA lyases (AP lyases) such as APEX1 or APEX2. Other factors involved in BER, such as PARP1 (poly-ADP-ribose polymerase-1) and XRCC1 (X-ray repair cross-complementing protein 1), assist in this process. XRCC1 acts as a scaffold for BER factors and also regulates PARP1, thereby preventing mis-regulation of BER and maintaining genomic stability (Lindahl, 2000; Demin *et al*., 2021). Upon the creation of an abasic site, XRCC1 recruits various BER factors to repair the abasic site, resulting in guanine as the major product (Gu *et al*., 2021). By fusing the XRCC1 protein with the nSpCas9-CBE architecture, researchers discovered an improvement in the editing efficiency of C•G-to-G•C in mammalian cells (Chen *et al*., 2021). A similar strategy has been implemented in nSpCas9-ABEs, an adenine transversion base editor, AYBE (A•T-to-T•A and A•T-to-C•G transversion editing in mammalian cells) was

developed by fusing hypoxanthine excision protein N-methylpurine DNA glycosylase (MPG) or mouse alkyladenine DNA glycosylase (mAAG) to the nSpCas9-ABE architecture *(Chen et al*., 2013; Tong *et al*., 2023). These findings emphasize the significance of comprehending the DNA repair mechanisms that generate diverse outcomes in base editing. It is expected that the creation of further novel base editors will necessitate the utilization of novel nucleic acid modification chemical processes and/or DNA repair manipulation tactics. Henceforth, CBE, ABE, CGBE, and AYBE, led by nSpCas9 or TALE proteins, will enable all types of base conversions, including transitions and transversions (Figure 10).

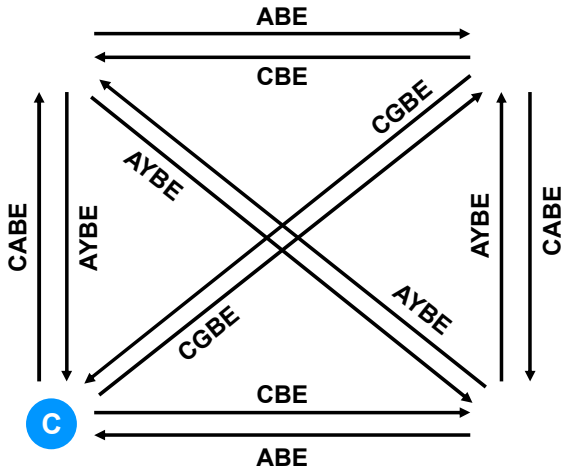


Figure 10. Overview of base editors. Diagram illustrates the several types of possible point mutations that can be carried out using the base editors that are now available. ABE: A•T-to-G•C base editors; CBE: C•G-to-T•A base editors; CGBE: C•G-to-G•C base editors; AYBE: A•T-to-T•A or A•T-to-C•G base editors; CABE: C•G-to-A•T base editors.

Base editing technologies hold great promise in diverse fields, such as gene therapy, plant breeding, and biofuel technology (Knott and Doudna, 2018). However, the issue of off-target mutations, which might lead to genetic instability and dysfunction, has been a major concern in the appliation of base editors (Jin *et al*., 2019; Zuo *et al*., 2019). In this thesis, off-target editing could be detected in both TALE-DdCBE and TALE-ABE. Furthermore, the low efficiency of regenerating plants edited by TALE-DdCBE or TALE-ABE could indicate that the off-target editing has a negative effect on viability of the edited cells.

Several techniques have been developed for detecting off-target activity of genome-wide base editing. Digenome-seq (Digested-genome sequencing) is a method that involves treating

extracted genomic DNA with a cytosine base editor and a mixture of DNA-modifying enzymes *in vitro* to generate DSBs at uracil-containing sites, allowing for the assessment of the specificity of cytosine base editors (Kim *et al*., 2017). GOTI (Genome-wide Off-target analysis by Two-cell embryo Injection) is a method used to identify the genome-wide off-target effects in edited mouse embryos generated by either CRISPR–Cas9 or base editors (Zuo *et al*., 2020). Through high-throughput whole genome sequencing, GOTI can directly compare edited and non-edited cells without the interference of genetic background, enabling it to detect potential off-target variants with high sensitivity. In addition, since the edited and non-edited cells are derived from a single ancestor cell, GOTI can effectively reduce the likehood of confounding genetic background and somatic mutations (Zuo *et al*., 2020). Detect-seq (dU-detection enabled by C-to-T transition during sequencing) has been developed for the genome-wide identification of off-target sites induced by cytosine base editors in a cellular context (Lei *et al*., 2021). Based on chemical labeling and enrichment of dU (deoxyuridine), a direct editing product of cytosine base editors, Detect-seq can trace the *in vivo* editing events in an unbiased manner.

While using GOTI, researchers found that the TALE-DdCBEs caused thousands of off-target single-nucleotide variants (SNVs) enriched for C•G-to-T•A editing in the nuclear genome even although the TALE-DdCBEs were designed to localize to mitochondria guided by a mitochondrial targeting signal (MTS) (Wei *et al*., 2022). Wei *et al*. reported that the TALE-DdCBEs induce a higher number of genome-wide SNVs in mouse embryos compared to SpCas9-APOBEC1 (also known as BE3). This effect is observed when TALE-DdCBEs are targeted towards mitochondrial DNA, whereas BE3 is targeted towards nuclear DNA. (Zuo *et al., 2019; Wei et al., 2022)*. The researchers also hypothesized that these unexpected off-target outcomes are likely due to the distinctive characteristics of the $DddA_{tox}$ cytosine deaminase employed in TALE-DdCBEs, because $DddA_{tox}$ exhibits a preference for dsDNA as its substrate (Mok *et al*., 2020), while the cytosine deaminase APOBEC1 in BE3 prefers ssDNA as its substate (Komor *et al*., 2016). Furthermore, the effectiveness of off-target editing using TALE-DdCBEs has been investigated by Detect-seq and targeted deep sequencing. These

studies have revealed two distinct types of off-targets editing exits: TALE-dependent off-targets and TALE-independent off-targets (Lei *et al*., 2022; Lee *et al*., 2022b).

Based on the specificity studies of TALE-DdCBEs mentioned above, we have concluded that the unexpected off-target outcomes in this thesis may be attributed to two main factors: nonspecific interactions between TALE proteins and DNA (TALE-dependent off-targets), and the spontaneous assembly of split DddAtox halves independent of TALE–DNA interactions (TALE-independent off-targets). Lei *et al*. found a frequent G-to-A mismatches between the TALE-dependent off-target sites and on-target sites (Lei *et al*., 2022). This observation can be attributed to the binding affinity of the RVD NN for both A and G nucleotides (Boch *et al*., 2009; Yang *et al*., 2014; Miller *et al*., 2015). Studies have altered the target nucleotides within the TALE binding site to investigate the impact of the number of mismatches on TALE activity. Their results reveal that TALE proteins can tolerate mismatches and the TALE activity decreases as the number of mismatches increases (Zhang *et al*., 2011). These research studies could collectively explain the TALE-dependent off-target effects. To limit this TALE-dependent off-target editing, it is important to select appropriate TALE pairs and choose specific RVDs for TALE-DdCBE. For TALE-independent off-target effects, Lee *et al*. discovered that these off-targets result from the spontaneous assembly of the split $DddA_{tox}$ cytosine deaminase enzyme (Lee *et al*., 2022b). Moreover, Lei *et al*. suggested that the presence of CTCF binding regions appears to be correlated with TALE-independent off-target sites *(Lei et al*., 2022). CTCF protein is a widely recognized factor that plays a crucial role in the organization of the three-dimensional genome architecture (Merkenschlager and Nora, 2016; Rowley and Corces, 2018).

In this thesis, it was shown that TALE-DdCBEs, and to a lesser extent TALE-ABEs (TABEs), are capable of editing target sites that are bound by one TALE monomer, only. Thus, it is possible that the DddA-halves can dimerize independently of TALE-binding. To avoid the spontaneous reassembly of split $DddA_{tox}$, a high-fidelity DdCBE was developed. This was achieved by mutating key residues at the split interface of $DddA_{tox}$. As a result, there was a decrease of off-target editing without sacrificing the on-target activity while using the high-fidelity DdCBEs (Lee *et al*., 2022b). Besides, nuclear off-target editing could also be

significantly decreased by incorporating nuclear export signals (NES) into the TALE-DdCBEs constructs, which hinder the localization of DdCBE into the nucleus (Lei *et al*., 2022; Lee *et al*., 2022a). Alternatively, by co-transfecting the DddA inhibitor ($DddI_A$) fused with two NLSs can block the activity of $DddA_{tox}$ present in the nuclei (Lei *et al*., 2022). Given that off-target effects are likely to be associated with the expression levels of the deaminase constructs, it is recommended to regulate the expression levels (Lei *et al*., 2022; Lee *et al*., 2022a; Silva-Pinheiro *et al*., 2023). The optimization of expression levels of TALE-DdCBEs has been achieved by incorporating a hammerhead ribozyme in the 3′-UTR and/or connecting the TALE-DdCBE monomers in tandem using a T2A element (Silva-Pinheiro *et al*., 2023). The incorporation of a hammerhead ribozyme in the 3′-UTR of TALE-DdCBE messenger RNA facilitates the processing of coding mRNA, resulting in a poly(A)-free 3′-end that is susceptible to degradation. This greatly reduces the expression of protein (Beilstein *et al*., 2015). Connecting the TALE-DdCBE monomers to the T2A element may lead to a decrease in the concentration of the downstream monomer (Liu *et al*., 2017). These improved architectures have significantly increased the precision of base editing of TALE-DdCBEs constructs by reducing off-target effects on mtDNA to background levels (Silva-Pinheiro *et al*., 2023).

This thesis has found that the unspecific editing effects occurred in both TALE-DdCBE and TALE-ABE. These effects were likely caused by the spontaneous assembly of DddA halves in the absence of TALE-DNA interaction. It is worthwhile to attempt the methods mentioned above to improve the base editing specificity of TALE-DdCBE and TALE-ABE in plants. Additionally, it is important to evaluate their genome-wide level of SNVs using an appropriate sequencing technique. Therefore, conducting more comprehensive investigations into the off-target editing of TALE-DdCBE and TALE-ABE is essential for both clinical applications and crop improvement.

6 Conclusion and outlook

Allergies to plant-based food products are a significant public health issue, and mustard is considered as one of the priority food allergens. This thesis aims to enhance the safety of food crops for human consumption by developing strategies to generate plant with low allergen content. This thesis presents the removal of the major allergen Bra J I, a seed storage protein, from brown mustard (*Brassica juncea*) using CRISPR/Cas9 genome editing technology. For this, a knock-out approach is needed to inactivate all four *Bra J I* alleles in the allotetraploid mustard genome. The seed production was generally decreased, however, there was variability observed among the T_0 and T_1 genome-edited plants. A specific antibody directed against an epitope of Bra J I was generated. Using this antibody, it was demonstrated that the Bra J I protein is reduced or completely absent in several of the genetically modified mustard lines. This research represents an initial step towards the development of a commercially viable, allergen-free mustard. This is one of the earliest examples of allergen reduction in plant species, and it represents a significant breakthrough in the field of genome editing in brown mustard. In addition to the elimination of allergens in mustard, the application of genome editing techniques to modify allergenic proteins (such as Ara h1, Ara h2, Ara h3, and others) in peanut has the potential to produce a hypoallergenic peanut. This high-lights the significant potential of these technologies in improving our understanding of allergenic proteins, and providing superior and alternative treatment options for allergic diseases.

Furthermore, this thesis successfully applied the protein-only base editors TALE-DdCBE and TALE-ABE in plants. TALE-DdCBEs exhibited a high efficiency of C•G-to-T•A editing in rice and tobacco protoplasts. Moreover, TALE-DdCBE that incorporate DddA11 exhibit a wider range of sequence compatibility for the purpose of editing non-T<u>C</u> targets. For TALE-ABE, the optimal architecture of TALE-ABE system has been refined in plant cells. Additional research is required in the future to explore the use of TALE-DdCBEs and TALE-ABEs for producing genetically modified plants and examining their specificity *in vivo*. The protein-only base editing tools TALE-DdCBEs and TALE-ABEs broaden the plant genome editing toolbox and provide a valuable resource for editing organelle or nucleus DNA.

Literature

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., *et al*. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353:aaf5573.

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., Verdine, V., Cox, D. B. T., Kellner, M. J., Regev, A., *et al*. (2017). RNA targeting with CRISPR–Cas13. *Nature* 550:280–284.

Aguirre, A. J., Meyers, R. M., Weir, B. A., Vazquez, F., Zhang, C.-Z., Ben-David, U., Cook, A., Ha, G., Harrington, W. F., Doshi, M. B., *et al*. (2016). Genomic copy number dictates a Gene-independent cell response to CRISPR/Cas9 targeting. *Cancer Discov.* 6:914–929.

Andersson, M., Turesson, H., Nicolia, A., Fält, A.-S., Samuelsson, M., and Hofvander, P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Rep.* 36:117–128.

Anzalone, A. V., Koblan, L. W., and Liu, D. R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol.* 38:824–844.

Assou, J., Zhang, D., Roth, K. D. R., Steinke, S., Hust, M., Reinard, T., Winkelmann, T., and Boch, J. (2022). Removing the major allergen Bra j I from brown mustard (*Brassica juncea*) by CRISPR/Cas9. *Plant J.* 109:649–663.

Bacman, S. R., Williams, S. L., Pinto, M., Peralta, S., and Moraes, C. T. (2013). Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat Med.* 19:1111–1113.

Beard, W. A., Horton, J. K., Prasad, R., and Wilson, S. H. (2019). Eukaryotic base excision repair: New approaches shine light on mechanism. *Annu Rev of Biochem.* 88:137–162.

Becker, S., and Boch, J. (2021). TALE and TALEN genome editing technologies. *Gene Genome Ed.* 2:100007.

Beilstein, K., Wittmann, A., Grez, M., and Suess, B. (2015). Conditional control of mammalian gene expression by tetracycline-dependent hammerhead ribozymes. *ACS Synth. Biol.* 4:526–534.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y.-G., and Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular Cell Biol.* 21:289–297.

Boch, J., and Bonas, U. (2010). Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annul Rev of Phytopathol.* 48:419–436.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., and Bonas, U. (2009). Breaking the code of dna binding specificity of TAL-type III effectors. *Science* 326:1509–1512.

Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors–pathogen strategies and plant resistance engineering. *New Phytol.* 204:823–832.

Bosshard, S., Duroy, P.-O., and Mermod, N. (2019). A role for alternative end-joining factors in homologous recombination and genome editing in Chinese hamster ovary cells. *DNA Repair* 82:102691.

Boyne, A., Yang, M., Pulicani, S., Feola, M., Tkach, D., Hong, R., Duclert, A., Duchateau, P., and Juillerat, A. (2022). Efficient multitool/multiplex gene engineering with TALE-BE. *Front Bioeng Biotechnol.* 10.

Braatz, J., Harloff, H.-J., Mascher, M., Stein, N., Himmelbach, A., and Jung, C. (2017). CRISPR-Cas9 targeted mutagenesis leads to simultaneous modification of different homoeologous gene copies in polyploid oilseed rape (*Brassica napus*). *Plant Physiol.* 174:935–942.

Breiteneder, H., and Ebner, C. (2000). Molecular and biochemical classification of plant-derived food allergens. *J Allergy Clin Immunol.* 106:27–36.

Breiteneder, H., and Radauer, C. (2004). A classification of plant food allergens. *J Allergy Clin Immunol.* 113:821–830.

Bunyavanich, S., Rifas-Shiman, S. L., Platts-Mills, T. A. E., Workman, L., Sordillo, J. E., Gillman, M. W., Gold, D. R., and Litonjua, A. A. (2014). Peanut allergy prevalence among school-age children in a US cohort not selected for any disease. *J Allergy Clin Immunol.* 134:753–755.

Burks, A. W., Shin, D., Cockrell, G., Stanley, J. S., Helm, R. M., and Bannon, G. A. (1997). Mapping and mutational analysis of the IgE-binding epitopes on Ara h 1, a legume vicilin protein and a major allergen in peanut hypersensitivity. *Eur J Biochem.* 245:334–339.

Caballero, T., Martin-Esteban, M., Garcia-Ara, C., Pascual, C., and Ojeda, A. (1994). Relationship between pollinosis and fruit or vegetable sensitization. *Pediatr Allergy Immunol.* 5:218–222.

Cade, L., Reyon, D., Hwang, W. Y., Tsai, S. Q., Patel, S., Khayter, C., Joung, J. K., Sander, J. D., Peterson, R. T., and Yeh, J.-R. J. (2012). Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res.* 40:8001–8010.

Cai, Y., Chen, L., Zhang, Y., Yuan, S., Su, Q., Sun, S., Wu, C., Yao, W., Han, T., and Hou, W. (2020). Target base editing in soybean using a modified CRISPR/Cas9 system. *Plant Biotechnol J.* 18:1996–1998.

Cameron, P., Coons, M. M., Klompe, S. E., Lied, A. M., Smith, S. C., Vidal, B., Donohoue, P. D., Rotstein, T., Kohrs, B. W., Nyer, D. B., *et al*. (2019). Harnessing type I CRISPR–Cas systems for genome engineering in human cells. *Nat Biotechnol.* 37:1471–1477.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. *Genome Biol.* 16:232.

Chatel, J.-M., Bernard, H., and Orson, F. M. (2003). Isolation and characterization of two complete Ara h 2 isoforms cDNA. *Int Arch Allergy Immunol.* 131:14–18.

Chen, S., Wan, Z., Nelson, M. N., Chauhan, J. S., Redden, R., Burton, W. A., Lin, P., Salisbury, P. A., Fu, T., and Cowling, W. A. (2013). Evidence from genome-wide simple sequence repeat markers for a polyphyletic origin and secondary centers of genetic diversity of *Brassica juncea* in China and India. *J Hered.* 104:416–427.

Chen, Y., Wang, Z., Ni, H., Xu, Y., Chen, Q., and Jiang, L. (2017a). CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. *Sci China Life Sci.* 60.

Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., Sternberg, S. H., Joung, J. K., Yildiz, A., and Doudna, J. A. (2017b). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* 550:407–410.

Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019a). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu Rev Plant Biol.* 70:667–697.

Chen, W., McKenna, A., Schreiber, J., Haeussler, M., Yin, Y., Agarwal, V., Noble, W. S., and Shendure, J. (2019b). Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* 47:7989–8003.

Chen, L., Park, J. E., Paa, P., Rajakumar, P. D., Prekop, H.-T., Chew, Y. T., Manivannan, S. N., and Chew, W. L. (2021). Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat Commun.* 12:1384.

Cho, S.-I., Lee, S., Mok, Y. G., Lim, K., Lee, J., Lee, J. M., Chung, E., and Kim, J.-S. (2022). Targeted A-to-G base editing in human mitochondrial DNA with programmable deaminases. *Cell* 185:1764-1776.e12.

Choo, Y., Sánchez-García, I., and Klug, A. (1994). In vivo repression by a site-specific DNA-binding protein designed against an oncogenic sequence. *Nature* 372:642–645.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., *et al*. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823.

Cuculis, L., Abil, Z., Zhao, H., and Schroeder, C. M. (2015). Direct observation of TALE protein dynamics reveals a two-state search mechanism. *Nat Commun.* 6:7277.

Cuculis, L., Abil, Z., Zhao, H., and Schroeder, C. M. (2016). TALE proteins search DNA using a rotationally decoupled mechanism. *Nat Chem Biol.* 12:831–837.

Curtis, M. D., and Grossniklaus, U. (2003). A Gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol.* 133:462–469.

De La Peña, M. A. G., Monsalve, R. I., Batanero, E., Villalba, M., and Rodríguez, R. (1996). Expression in Escherichia coli of Sin a 1, the major allergen from mustard. *Eurn J Biochem.* 237:827–832.

Demin, A. A., Hirota, K., Tsuda, M., Adamowicz, M., Hailstone, R., Brazina, J., Gittens, W., Kalasova, I., Shao, Z., Zha, S., *et al*. (2021). XRCC1 prevents toxic PARP1 trapping during DNA base excision repair. *Mol Cell* 81:3018-3030.e5.

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., and Yan, N. (2012). Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* 335:720–723.

Ding, D., Chen, K., Chen, Y., Li, H., and Xie, K. (2018). Engineering introns to express rna guides for Cas9- and Cpf1-mediated multiplex genome editing. *Mol Plant* 11:542–552.

Dodo, H. W., Konan, K. N., Chen, F. C., Egnin, M., and Viquez, O. M. (2008). Alleviating peanut allergy using genetic engineering: the silencing of the immunodominant allergen Ara h 2 leads to its significant reduction and a decrease in peanut allergenicity. *Plant Biotechnol J.* 6:135–145.

Dolan, A. E., Hou, Z., Xiao, Y., Gramelspacher, M. J., Heo, J., Howden, S. E., Freddolino, P. L., Ke, A., and Zhang, Y. (2019). Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using Type I CRISPR-Cas. *Mol Cell* 74:936-950.e5.

Endo, M., Mikami, M., and Toki, S. (2015). multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in Rice. *Plant Cell Physiol.* 56:41–47.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G., Patron, N. J., and Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synth Biol.* 3:839–843.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., *et al*. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Pro Natl Acad Sci U S A.* 111:4632–4637.

Forner, J., Kleinschmidt, D., Meyer, E. H., Fischer, A., Morbitzer, R., Lahaye, T., Schöttler, M. A., and Bock, R. (2022). Targeted introduction of heritable point mutations into the plant mitochondrial genome. *Nat Plants* 8:245–256.

Gammage, P. A., Rorbach, J., Vincent, A. I., Rebar, E. J., and Minczuk, M. (2014). Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. *EMBO Mol Med.* 6:458–466.

Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* 184:1621–1635.

Gao, H., Wu, X., Chai, J., and Han, Z. (2012). Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.* 22:1716–1720.

Gao, W., Long, L., Tian, X., Xu, F., Liu, J., Singh, P. K., Botella, J. R., and Song, C. (2017a). Genome editing in cotton with the CRISPR/Cas9 system. *Front Plant Sci.* 8.

Gao, L., Cox, D. B. T., Yan, W. X., Manteiga, J. C., Schneider, M. W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017b). Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol.* 35:789–792.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences* 109:E2579–E2586.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551:464–471.

Gehrke, J. M., Cervantes, O., Clement, M. K., Wu, Y., Zeng, J., Bauer, D. E., Pinello, L., and Joung, J. K. (2018). An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat Biotechnol.* 36:977–982.

Geiβler, R., Scholze, H., Hahn, S., Streubel, J., Bonas, U., Behrens, S.-E., and Boch, J. (2011). Transcriptional activators of human genes with programmable DNA-specificity. *PLoS One* 6:e19509.

Gibson, D. G. (2011). Chapter fifteen - Enzymatic assembly of overlapping DNA fragments. In *Methods in Enzymology* (ed. Voigt, C.), pp. 349–361. Academic Press.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C. V., Sánchez-León, S., Baltes, N. J., Starker, C., Barro, F., Gao, C., *et al*. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* 89:1251–1262.

Gonzalez de la Peña, M. A., Menéndez-Arias, L., Monsalve, R. I., and Rodríguez, R. (2009). Isolation and characterization of a major allergen from oriental mustard seeds, Bra j I. *Int Arch Allergy Appl Immunol.* 96:263–270.

Gould, H. J., and Wu, Y.-C. B. (2018). IgE repertoire and immunological memory: compartmental regulation and antibody function. *Int Immunol.* 30:403–412.

Grützner, R., and Marillonnet, S. (2020). Generation of MoClo standard parts using Golden Gate cloning. In *DNA Cloning and Assembly: Methods and Protocols* (ed. Chandran, S.) and George, K. W.), pp. 107–123. New York, NY: Springer US.

Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: Expanding the types of DNA damage products harnessed for genome editing. *Gene Genome Ed.* 1:100005.

Guyon-Debast, A., Alboresi, A., Terret, Z., Charlot, F., Berthier, F., Vendrell-Mir, P., Casacuberta, J. M., Veillet, F., Morosinotto, T., Gallois, J.-L., *et al*. (2021). A blueprint for gene function analysis through Base Editing in the model plant *Physcomitrium* (*Physcomitrella*) patens. *New Phytol.* 230:1258–1272.

Haapaniemi, E., Botla, S., Persson, J., Schmierer, B., and Taipale, J. (2018). CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat Med.* 24:927–930.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, K. R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., *et al*. (2015). High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163:1515–1526.

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A., Cedrone, F., *et al*. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol J.* 12:934–940.

He, Z., Ji, R., Havlickova, L., Wang, L., Li, Y., Lee, H. T., Song, J., Koh, C., Yang, J., Zhang, M., *et al*. (2021). Genome structural evolution in Brassica crops. *Nat Plants* 7:757–765.

Heppell, L. M. J., Sissons, J. W., and Pedersen, H. E. (1987). A comparison of the antigenicity of soya-bean-based infant formulas. *Br J Nutr.* 58:393–403.

Heyer, W.-D., Ehmsen, K. T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annu Rev Genet.* 44:113–139.

Hoffmann-Sommergruber, K. (2002). Pathogenesis-related (PR)-proteins identified as allergens. *Biocheml Soc Trans.* 30:930–935.

Hou, Z., Zhang, Y., Propson, N. E., Howden, S. E., Chu, L.-F., Sontheimer, E. J., and Thomson, J. A. (2013). Efficient genome engineering in human pluripotent stem cells using Cas9 from *Neisseria meningitidis*. *Proc the Natl Acad Sci U S A.* 110:15644–15649.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., *et al*. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556:57–63.

Huang, T.-K., and Puchta, H. (2019). CRISPR/Cas-mediated gene targeting in plants: finally a turn for the better for homologous recombination. *Plant Cell Rep.* 38:443–453.

Huang, P., Xiao, A., Zhou, M., Zhu, Z., Lin, S., and Zhang, B. (2011). Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol.* 29:699–700.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol.* 169:5429–5433.

Jacob, P., Avni, A., and Bendahmane, A. (2018). Translational research: Exploring and creating genetic diversity. *Trends Plant Sci.* 23:42–52.

Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15:16.

Jin, S., Zong, Y., Gao, Q., Zhu, Z., Wang, Y., Qin, P., Liang, C., Wang, D., Qiu, J.-L., Zhang, F., *et al*. (2019). Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 364:292–295.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821.

Jong, D. E., Zijverden, V., Spanhaak, Koppelman, Pellegrom, and Penninks (1998). Identification and partial characterization of multiple major allergens in peanut proteins. *Clin Exp Allergy* 28:743–751.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., *et al*. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589.

Jung, J. H., and Altpeter, F. (2016). TALEN mediated targeted mutagenesis of the caffeic acid O-methyltransferase in highly polyploid sugarcane improves cell wall composition for production of bioethanol. *Plant Mol Biol.* 92:131–142.

Kagan, R. S. (2003). Food allergy: an overview. *Environmental Health Perspectives* 111:223–225.

Kang, B.-C., Yun, J.-Y., Kim, S.-T., Shin, Y., Ryu, J., Choi, M., Woo, J. W., and Kim, J.-S. (2018). Precision genome engineering through adenine base editing in plants. *Nat Plants* 4:427–431.

Kang, B.-C., Bae, S.-J., Lee, S., Lee, J. S., Kim, A., Lee, H., Baek, G., Seo, H., Kim, J., and Kim, J.-S. (2021a). Chloroplast and mitochondrial DNA editing in plants. *Nat Plants* 7:899–905.

Kang, L., Qian, L., Zheng, M., Chen, L., Chen, H., Yang, L., You, L., Yang, B., Yan, M., Gu, Y., *et al*. (2021b). Genomic insights into the origin, domestication and diversification of Brassica juncea. *Nat Genet.* 53:1392–1402.

Katz, Y., Gutierrez-Castrellon, P., González, M. G., Rivas, R., Lee, B. W., and Alarcon, P. (2014). A comprehensive review of sensitization and allergy to Soy-based products. *Clinic Rev Allerg Immunol.* 46:272–281.

Kay, S., Hahn, S., Marois, E., Hause, G., and Bonas, U. (2007). a bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* 318:648–651.

Kazama, T., Okuno, M., Watari, Y., Yanase, S., Koizuka, C., Tsuruta, Y., Sugaya, H., Toyoda, A., Itoh, T., Tsutsumi, N., *et al*. (2019). Curing cytoplasmic male sterility via TALEN-mediated mitochondrial genome editing. *Nat Plants* 5:722–730.

Kim, D., Lim, K., Kim, S.-T., Yoon, S., Kim, K., Ryu, S.-M., and Kim, J.-S. (2017). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat Biotechnol.* 35:475–480.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., Gonzales, A. P. W., Li, Z., Peterson, R. T., Yeh, J.-R. J., *et al*. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523:481–485.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., and Joung, J. K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529:490–495.

Knott, G. J., and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science* 361:866–869.

Koblan, L. W., Arbab, M., Shen, M. W., Hussmann, J. A., Anzalone, A. V., Doman, J. L., Newby, G. A., Yang, D., Mok, B., Replogle, J. M., *et al*. (2021). Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol.* 39:1414–1425.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533:420–424.

Koonin, E. V., Makarova, K. S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol.* 37:67–78.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol.* 34:192–198.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol.* 36:765–771.

Krebitz, M., Wagner, B., Ferreira, F., Peterbauer, C., Campillo, N., Witty, M., Kolarich, D., Steinkellner, H., Scheiner, O., and Breiteneder, H. (2003). Plant-based heterologous expression of Mal d 2, a thaumatin-like protein and allergen of apple (*Malus domestica*), and its characterization as an antifungal protein. *J Molr Biology* 329:721–730.

Krokan, H. E., and Bjørås, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol.* 5:a012583.

Kunz, C., Saito, Y., and Schär, P. (2009). DNA repair in mammalian cells. *Cell Mol Life Sci.* 66:1021–1038.

Kurt, I. C., Zhou, R., Iyer, S., Garcia, S. P., Miller, B. R., Langner, L. M., Grünewald, J., and Joung, J. K. (2021). CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat Biotechnol.* 39:41–46.

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., and Adli, M. (2017). CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat Methods* 14:710–712.

Langner, T., Kamoun, S., and Belhaj, K. (2018). CRISPR Crops: Plant genome editing toward disease resistance. *Annu Rev Phytopathol.* 56:479–512.

Lee, L. A., and Burks, A. W. (2006). Food allergies: prevalence, molecular characterization, and treatment/prevention strategies. *Annu Rev Nutr.* 26:539–565.

Lee, S., Lee, H., Baek, G., Namgung, E., Park, J. M., Kim, S., Hong, S., and Kim, J.-S. (2022a). Enhanced mitochondrial DNA editing in mice using nuclear-exported TALE-linked deaminases and nucleases. *Genome Biol.* 23:211.

Lee, S., Lee, H., Baek, G., and Kim, J.-S. (2022b). Precision mitochondrial DNA editing with high-fidelity DddA-derived base editors. *Nat Biotechnol.* Advance Access published October 13, 2022, doi:10.1038/s41587-022-01486-w.

Lei, Z., Meng, H., Lv, Z., Liu, M., Zhao, H., Wu, H., Zhang, X., Liu, L., Zhuang, Y., Yin, K., *et al.* (2021). Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat Methods* 18:643–651.

Lei, Z., Meng, H., Liu, L., Zhao, H., Rao, X., Yan, Y., Wu, H., Liu, M., He, A., and Yi, C. (2022). Mitochondrial base editor induces substantial nuclear off-target mutations. *Nature* 606:804–811.

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. *Amer Natur.* 122:1–25.

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., and Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol.* 30:390–392.

Li, J.-F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G. M., and Sheen, J. (2013). Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nat Biotechnol.* 31:688–691.

Li, C., Zong, Y., Wang, Y., Jin, S., Zhang, D., Song, Q., Zhang, R., and Gao, C. (2018a). Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol.* 19:59.

Li, S., Zhang, X., Wang, W., Guo, X., Wu, Z., Du, W., Zhao, Y., and Xia, L. (2018b). Expanding the scope of CRISPR/Cpf1-mediated genome editing in rice. *Mol Plant* 11:995–998.

Li, G., Sretenovic, S., Eisenstein, E., Coleman, G., and Qi, Y. (2021a). Highly efficient C-to-T and A-to-G base editing in a Populus hybrid. *Plant Biotechnol J.* 19:1086–1088.

Li, R., Char, S. N., Liu, B., Liu, H., Li, X., and Yang, B. (2021b). High-efficiency plastome base editing in rice with TAL cytosine deaminase. *Mol Plant* 14:1412–1414.

Lieber, M. R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem.* 79:181–211.

Lin, S., Staahl, B. T., Alla, R. K., and Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* 3:e04766.

Lin, Q., Zhu, Z., Liu, G., Sun, C., Lin, D., Xue, C., Li, S., Zhang, D., Gao, C., Wang, Y., *et al*. (2021). Genome editing in plants with MAD7 nuclease. *J Genet Genomics* 48:444–451.

Lindahl, T. (2000). Suppression of spontaneous mutagenesis in human cells by DNA base excision–repair. *Mutat Res.* 462:129–135.

Liu, Z., Chen, O., Wall, J. B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep.* 7:2193.

Liu, Y., Merrick, P., Zhang, Z., Ji, C., Yang, B., and Fei, S. (2018). Targeted mutagenesis in tetraploid switchgrass (Panicum virgatum L.) using CRISPR/Cas9. *Plant Biotechnol J.*16:381–393.

Lloyd, A., Plaisier, C. L., Carroll, D., and Drews, G. N. (2005). Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Nal Acad Sci U S A.* 102:2232–2237.

Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J., and Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to Its DNA target. *Science* 335:716–719.

Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., *et al*. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol.* 18:67–83.

Malabarba, J., Chevreau, E., Dousset, N., Veillet, F., Moizan, J., and Vergne, E. (2021). New strategies to overcome present CRISPR/Cas9 limitations in apple and pear: efficient dechimerization and base editing. *Int J Mol Sci.* 22:319.

Matsuo, H., Yokooji, T., and Taogoshi, T. (2015). Common food allergens and their IgE-binding epitopes. *Allergol. Int.* 64:332–343.

Merkenschlager, M., and Nora, E. P. (2016). CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu Rev Genomics Hum Genet.* 17:17–43.

Miller, J., McLachlan, A. d., and Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* 4:1609–1614.

Miller, J. C., Zhang, L., Xia, D. F., Campo, J. J., Ankoudinova, I. V., Guschin, D. Y., Babiarz, J. E., Meng, X., Hinkley, S. J., Lam, S. C., *et al*. (2015). Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Methods* 12:465–471.

Mills, E. N. C., Jenkins, J. A., Alcocer, M. J. C., and Shewry, P. R. (2004). Structural, biological, and evolutionary relationships of plant food allergens sensitizing via the gastrointestinal tract. *Crit Rev Food Sci. Nutr.* 44:379–407.

Mok, B. Y., de Moraes, M. H., Zeng, J., Bosch, D. E., Kotrys, A. V., Raguram, A., Hsu, F., Radey, M. C., Peterson, S. B., Mootha, V. K., *et al*. (2020). A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* 583:631–637.

Mok, Y. G., Lee, J. M., Chung, E., Lee, J., Lim, K., Cho, S.-I., and Kim, J.-S. (2022a). Base editing in human cells with monomeric DddA-TALE fusion deaminases. *Nat Commun.* 13:4038.

Mok, B. Y., Kotrys, A. V., Raguram, A., Huang, T. P., Mootha, V. K., and Liu, D. R. (2022b). CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. *Nat Biotechnol.* 40:1378–1387.

Mok, Y. G., Hong, S., Bae, S.-J., Cho, S.-I., and Kim, J.-S. (2022c). Targeted A-to-G base editing of chloroplast DNA in plants. *Nat. Plants* Advance Access published December 1, 2022, doi:10.1038/s41477-022-01279-8.

Molla, K. A., Sretenovic, S., Bansal, K. C., and Qi, Y. (2021). Precise plant genome editing using base editors and prime editors. *Nat Plants* 7:1166–1187.

Monsalve, R. I., Gonzalez de la Peña, M. A., Menendez-Arias, L., Lopez-Otin, C., Villalba, M., and Rodriguez, R. (1993). Characterization of a new oriental-mustard (*Brassica juncea*) allergen, Braj IE: detection of an allergenic epitope. *Biochem J.* 293:625–632.

Monsalve, R. I., De La Peña, M. a. G., Lópezotíín, C., Fiandor, A., Fernández, C., Villalba, M., and Rodríguez, R. (1997). Detection, isolation and complete amino acid sequence of an aeroallergenic protein from rapeseed flour. *Clin Exp Allergy* 27:833–841.

Morisset, M., Moneret-Vautrin, D.-A., Maadi, F., Frémont, S., Guénard, L., Croizier, A., and Kanny, G. (2003). Prospective study of mustard allergy: first study with double-blind placebo-controlled food challenge trials (24 cases). *Allergy* 58:295–299.

Moscou, M. J., and Bogdanove, A. J. (2009). A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science* 326:1501–1501.

Nakazato, I., Okuno, M., Yamamoto, H., Tamura, Y., Itoh, T., Shikanai, T., Takanashi, H., Tsutsumi, N., and Arimura, S. (2021). Targeted base editing in the plastid genome of Arabidopsis thaliana. *Nat Plants* 7:906–913.

Nambiar, T. S., Billon, P., Diedenhofen, G., Hayward, S. B., Taglialatela, A., Cai, K., Huang, J.-W., Leuzzi, G., Cuella-Martin, R., Palacios, A., *et al*. (2019). Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat Commun.* 10:3395.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D. G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol.* 31:691–693.

Niinimäi, A., Björksten, F., Puukka, M., Tolonen, K., and Hannuksela, M. (1989). Spice allergy: results of skin prick tests and RAST with spice extracts. *Allergy* 44:60–65.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., *et al*. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353:aaf8729.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., Noda, T., Abudayyeh, O. O., Gootenberg, J. S., Mori, H., *et al*. (2018). Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361:1259–1262.

Nordlee, J. A., Taylor, S. L., Townsend, J. A., Thomas, L. A., and Bush, R. K. (1996). Identification of a brazil-nut allergen in transgenic soybeans. *N Engl J Med.* 334:688–692.

Pastorello, E. A., Farioli, L., Pravettoni, V., Ortolani, C., Ispano, M., Monza, M., Baroglio, C., Scibola, E., Ansaloni, R., Incorvaia, C., *et al*. (1999). The major allergen of peach (*Prunus persica*) is a lipid transfer protein. *J Allergy Clin Immunol.* 103:520–526.

Pavletich, N. P., and Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809–817.

Peng, A., Chen, S., Lei, T., Xu, L., He, Y., Wu, L., Yao, L., and Zou, X. (2017). Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol J.* 15:1509–1519.

Peterson, B. A., Haak, D. C., Nishimura, M. T., Teixeira, P. J. P. L., James, S. R., Dangl, J. L., and Nimchuk, Z. L. (2016). Genome-wide assessment of efficiency and specificity in crispr/cas9 mediated multiple site targeting in *Arabidopsis*. *PLoS One* 11:e0162169.

Porteus, M. (2016). Genome editing: a new approach to human therapeutics. *Annu Rev Pharmacol Toxicol.* 56:163–190.

Porteus, M. H., and Baltimore, D. (2003). chimeric nucleases stimulate gene targeting in human cells. *Science* 300:763–763.

Porto, E. M., and Komor, A. C. (2023). In the business of base editors: Evolution from bench to bedside. *PLoS Biology* 21:e3002071.

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. *Curr Opin Plant Biol.* 36:1–8.

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173–1183.

Qin, L., Li, J., Wang, Q., Xu, Z., Sun, L., Alariqi, M., Manghwar, H., Wang, G., Li, B., Ding, X., *et al*. (2020). High-efficient and precise base editing of C•G to T•A in the allotetraploid cotton (*Gossypium hirsutum*) genome using a modified CRISPR/Cas9 system. *Plant Biotechnol J.* 18:45–56.

Rahman, M., Khatun, A., Liu, L., and Barkla, B. J. (2018). Brassicaceae mustards: Traditional and agronomic uses in Australia and New Zealand. *Molecules* 23:231.

Rancé, F. (2003). Mustard allergy as a new food allergy. *Allergy* 58:287–288.

Rancé, F., Dutau, G., and Abbal M, M. (2000). Mustard allergy in children. *Allergy* 55:496–500.

Rees, H. A., Yeh, W.-H., and Liu, D. R. (2019). Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat Commun.* 10:2212.

Renz, H., Allen, K. J., Sicherer, S. H., Sampson, H. A., Lack, G., Beyer, K., and Oettgen, H. C. (2018). Food allergy. *Nat Rev Dis Primers* 4:1–20.

Richter, M. F., Zhao, K. T., Eton, E., Lapinaite, A., Newby, G. A., Thuronyi, B. W., Wilson, C., Koblan, L. W., Zeng, J., Bauer, D. E., *et al*. (2020). Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat Biotechnol.* 38:883–891.

Rönspies, M., Schmidt, C., Schindele, P., Lieberman-Lazarovich, M., Houben, A., and Puchta, H. (2022). Massive crossover suppression by CRISPR–Cas-mediated plant chromosome engineering. *Nat. Plants* 8:1153–1159.

Rowley, M. J., and Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nat Rev Genet.* 19:789–800.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., and Barro, F. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol J.* 16:902–910.

Schaart, J. G., van de Wiel, C. C. M., and Smulders, M. J. M. (2021). Genome editing of polyploid crops: prospects, achievements and bottlenecks. *Transgenic Res.* 30:337–351.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J. J., Qiu, J.-L., *et al*. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol.* 31:686–688.

Shan, S., Mavrodiev, E. V., Li, R., Zhang, Z., Hauser, B. A., Soltis, P. S., Soltis, D. E., and Yang, B. (2018). Application of CRISPR/Cas9 to Tragopogon (*Asteraceae*), an evolutionary model for the study of polyploidy. *Mol Ecol Resour* 18:1427–1443.

Shen, M. W., Arbab, M., Hsu, J. Y., Worstell, D., Culbertson, S. J., Krabbe, O., Cassa, C. A., Liu, D. R., Gifford, D. K., and Sherwood, R. I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* 563:646–651.

Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell* 7:945–956.

Shinmoto, H., Matsuo, Y., Naganawa, Y., Tomita, S., and Takano-Ishikawa, Y. (2010). Epitope analysis of peanut allergen Ara h1 with human monoclonal IgM antibody 92-2. *Cytotechnology* 62:307–311.

Shukla, V. K., Doyon, Y., Miller, J. C., DeKelver, R. C., Moehle, E. A., Worden, S. E., Mitchell, J. C., Arnold, N. L., Gopalan, S., Meng, X., *et al*. (2009). Precise genome modification in the crop species *Zea mays* using zinc-finger nucleases. *Nature* 459:437–441.

Sicherer, S. H., and Sampson, H. A. (2014). Food allergy: Epidemiology, pathogenesis, diagnosis, and treatment. *J Allergy Clin Immunol.* 133:291-307.e5.

Sicherer, S. H., Muñoz-Furlong, A., Godbold, J. H., and Sampson, H. A. (2010). US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. *J Allergy Clin Immunol.* 125:1322–1326.

Silva-Pinheiro, P., Mutti, C. D., Van Haute, L., Powell, C. A., Nash, P. A., Turner, K., and Minczuk, M. (2023). A library of base editors for the precise ablation of all protein-coding genes in the mouse mitochondrial genome. *Nat Biomed Eng.* 7:692–703.

Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* 351:84–88.

Sowka, S., Hsieh, L. S., Krebitz, M., Akasawa, A., Martin, B. M., Starrett, D., Peterbauer, C. K., Scheiner, O., and Breiteneder, H. (1998). Identification and cloning of prs a 1, a 32-kDa

endochitinase and major allergen of avocado, and its expression in the yeast Pichia pastoris. *J Biol Chem.* 273:28091–28097.

Stanley, J. S., King, N., Burks, A. W., Huang, S. K., Sampson, H., Cockrell, G., Helm, R. M., West, C. M., and Bannon, G. A. (1997). Identification and Mutational Analysis of the Immunodominant IgE Binding Epitopes of the Major Peanut AllergenAra h 2. *Arch Biochem Biophy.* 342:244–253.

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., and Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507:62–67.

Streubel, J., Blücher, C., Landgraf, A., and Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nat Biotechnol.* 30:593–595.

Sugano, S., Hirose, A., Kanazashi, Y., Adachi, K., Hibara, M., Itoh, T., Mikami, M., Endo, M., Hirose, S., Maruyama, N., *et al*. (2020). Simultaneous induction of mutant alleles of two allergenic genes in soybean by using site-directed mutagenesis. *BMC Plant Biol.* 20:513.

Szurek, B., Rossier, O., Hause, G., and Bonas, U. (2002). Type III-dependent translocation of the Xanthomonas AvrBs3 protein into the plant cell. *Mol Microbiol.* 46:13–23.

Tan, J., Zhao, Y., Wang, B., Hao, Y., Wang, Y., Li, Y., Luo, W., Zong, W., Li, G., Chen, S., *et al*. (2020). Efficient CRISPR/Cas9-based plant genomic fragment deletions by microhomology-mediated end joining. *Plant Biotechnol J.* 18:2161–2163.

Teuber, S. S., Jarvis, K. C., Dandekar, A. M., Peterson, W. R., and Ansari, A. A. (1999). Identification and cloning of a complementary DNA encoding a vicilin-like proprotein, Jug r 2, from English walnut kernel (*Juglans regia*), a major food allergen. *J Allergy Clin Immunol.* 104:1311–1320.

Tian, S., Jiang, L., Cui, X., Zhang, J., Guo, S., Li, M., Zhang, H., Ren, Y., Gong, G., Zong, M., *et al*. (2018). Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* 37:1353–1356.

Tong, H., Wang, X., Liu, Y., Liu, N., Li, Y., Luo, J., Ma, Q., Wu, D., Li, J., Xu, C., *et al*. (2023). Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat Biotechnol.* Advance Access published January 9, 2023, doi:10.1038/s41587-022-01595-6.

Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435:646–651.

Van den Ackerveken, G., Marois, E., and Bonas, U. (1996). Recognition of the bacterial avirulence protein AvrBs3 occurs inside the host plant cell. *Cell* 87:1307–1316.

van der Horst, S., Filipovska, T., Hanson, J., and Smeekens, S. (2020). Metabolite control of translation by conserved peptide uORFs: The ribosome as a metabolite multisensor. *Plant Physiol.* 182:110–122.

van Overbeek, M., Capurso, D., Carter, M. M., Thompson, M. S., Frias, E., Russ, C., Reece-Hoyes, J. S., Nye, C., Gradia, S., Vidal, B., *et al*. (2016). DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol Cell* 63:633–646.

Vaughan, J. G., Hemingway, J. S., and Schofield, H. J. (1963). Contributions to a study of variation in *Brassica juncea* Coss. & Czern. *Bot J Linn Soc.* 58:435–447.

Veillet, F., Perrot, L., Chauvin, L., Kermarrec, M.-P., Guyon-Debast, A., Chauvin, J.-E., Nogué, F., and Mazier, M. (2019a). Transgene-free genome editing in tomato and potato plants using Agrobacterium-mediated delivery of a CRISPR/Cas9 cytidine base editor. *Int J Mol Sci.* 20:402.

Veillet, F., Chauvin, L., Kermarrec, M.-P., Sevestre, F., Merrer, M., Terret, Z., Szydlowski, N., Devaux, P., Gallois, J.-L., and Chauvin, J.-E. (2019b). The *Solanum tuberosum* GBSSI gene: a target for assessing gene and base editing in tetraploid potato. *Plant Cell Rep.* 38:1065–1080.

Veres, A., Gosis, B. S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C. A., Talkowski, M. E., and Musunuru, K. (2014). Low Incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* 15:27–30.

Vieths, S. (1997). Allergenic cross-reactivity, food allergy and pollen. *Environ Toxicol Pharmacol.* 4:61–70.

Voytas, D. F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu Rev Plant Biol.* 64:327–350.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., and Qiu, J.-L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol.* 32:947–951.

Wang, T., Yu, H., Hughes, N. W., Liu, B., Kendirli, A., Klein, K., Chen, W. W., Lander, E. S., and Sabatini, D. M. (2017). Gene essentiality profiling reveals gene networks and synthetic lethal interactions with oncogenic ras. *Cell* 168:890-903.e15.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. *PLoS One* 6:e16765.

Wei, Y., Li, Z., Xu, K., Feng, H., Xie, L., Li, D., Zuo, Z., Zhang, M., Xu, C., Yang, H., *et al*. (2022). Mitochondrial base editor DdCBE causes substantial DNA off-target editing in nuclear genome of embryos. *Cell Discov.* 8:1–4.

Wilson, F. M., Harrison, K., Armitage, A. D., Simkin, A. J., and Harrison, R. J. (2019). CRISPR/Cas9-mediated mutagenesis of phytoene desaturase in diploid and octoploid strawberry. *Plant Met.* 15:45.

Wright, D. A., Townsend, J. A., Winfrey Jr, R. J., Irwin, P. A., Rajagopal, J., Lonosky, P. M., Hall, B. D., Jondle, M. D., and Voytas, D. F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J.* 44:693–705.

Wu, J., Chen, C., Xian, G., Liu, D., Lin, L., Yin, S., Sun, Q., Fang, Y., Zhang, H., and Wang, Y. (2020). Engineering herbicide-resistant oilseed rape by CRISPR/Cas9-mediated cytosine base-editing. *Plant Biotechnol J.* 18:1857–1859.

Xie, K., and Yang, Y. (2013). RNA-guided genome editing in plants using a CRISPR–Cas system. *Mol Plant* 6:1975–1983.

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci U S A.* 112:3570–3575.

Xing, S., Chen, K., Zhu, H., Zhang, R., Zhang, H., Li, B., and Gao, C. (2020). Fine-tuning sugar content in strawberry. *Genome Biol.* 21:230.

Yang, J., Zhang, Y., Yuan, P., Zhou, Y., Cai, C., Ren, Q., Wen, D., Chu, C., Qi, H., and Wei, W. (2014). Complete decoding of TAL effectors for DNA recognition. *Cell Res.* 24:628–631.

Yeh, C. D., Richardson, C. D., and Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat Cell Biol.* 21:1468–1478.

You, J., Li, D., Yang, L., Dossou, S. S. K., Zhou, R., Zhang, Y., and Wang, L. (2022). CRISPR/Cas9-Mediated Efficient Targeted Mutagenesis in Sesame (*Sesamum indicum* L.). *Front Plant Sci.* 13.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., *et al*. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–771.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E. M., Winblad, N., Choudhury, S. R., Abudayyeh, O. O., Gootenberg, J. S., *et al*. (2017). Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat Biotechnol.* 35:31–34.

Zhang, F., Cong, L., Lodato, S., Kosuri, S., Church, G. M., and Arlotta, P. (2011). Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol.* 29:149–153.

Zhang, H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H., Xu, N., *et al*. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol J.*12:797–807.

Zhang, D., Zhang, H., Li, T., Chen, K., Qiu, J.-L., and Gao, C. (2017). Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome Biol.* 18:191.

Zhang, Q., Xing, H.-L., Wang, Z.-P., Zhang, H.-Y., Yang, F., Wang, X.-C., and Chen, Q.-J. (2018a). Potential high-frequency off-target mutagenesis induced by CRISPR/Cas9 in Arabidopsis and its prevention. *Plant Mol Biol.* 96:445–456.

Zhang, H., Si, X., Ji, X., Fan, R., Liu, J., Chen, K., Wang, D., and Gao, C. (2018b). Genome editing of upstream open reading frames enables translational control in plants. *Nat Biotechnol.* 36:894–898.

Zhang, Y., Ren, Q., Tang, X., Liu, S., Malzahn, A. A., Zhou, J., Wang, J., Yin, D., Pan, C., Yuan, M., *et al*. (2021). Expanding the scope of plant genome engineering with Cas12a orthologs and highly multiplexable editing systems. *Nat Commun.* 12:1944.

Zhao, D., Li, J., Li, S., Xin, X., Hu, M., Price, M. A., Rosser, S. J., Bi, C., and Zhang, X. (2021). Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat Biotechnol.* 39:35–40.

Zhu, W., Yang, B., Chittoor, J. M., Johnson, L. B., and White, F. F. (1998). AvrXa10 Contains an Acidic Transcriptional Activation Domain in the Functionally Conserved C Terminus. *MPMI* 11:824–832.

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., *et al*. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat Biotechnol.* 34:1279–1286.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D., and Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat Biotechnol.* 35:438–440.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., and Yang, H. (2019). Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* 364:289–292.

Zuo, E., Sun, Y., Wei, W., Yuan, T., Ying, W., Sun, H., Yuan, L., Steinmetz, L. M., Li, Y., and Yang, H. (2020). GOTI, a method to identify genome-wide off-target effects of genome editing in mouse embryos. *Nat Protoc.* 15:3009–3029.

Abbreviations

ABEs	Adenine base editors
AID	Activation-induced cytidine deaminase
APOBEC	Apolipoprotein B mRNA editing enzyme, catalytic polypeptide
BER	Base excision repair
Cas	CRISPR-associated
CBEs	Cytosine base editors
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DddA	Double-strand DNA cytidine deaminase
DSBs	Double strand breaks
dCas9	Deactivated Cas9
GUS	β-glucuronidase
HDR	Homology-directed repair
IgE	Immune-globulin E
MoClo	Modular Cloning
MCS	Multiple cloning sites
NHEJ	Nonhomologous end-joining
PR	Pathogenesis-related
PACE	Phage-assisted continuous evolution
RAD51	Radiation Sensitive 51
PANCE	Phage-assisted non-continuous evolution
PAMs	Protospacer-adjacent motifs
QTLs	Quantitative trait loci
RVDs	Repeat variable di-residue
sgRNA	Single guide RNA
SNVs	Single nucleotide variants
ssDNA	Single-stranded DNA
SSNs	Sequence-specific nucleases
TadA	Transfer RNA adenosine deaminase
TALE	Transcription activator-like effector
TALENs	Transcription activator-like effector nucleases
TALEDs	TALE-based ABEs
TALE-DdCBEs	TALE-derived Ddda-based cytosine base editors
UGI	Uracil glycosylase inhibitor
UNG	Uracil N-glycosylase
WGS	Whole-genome sequencing
WT	Wild-type
ZFN	Zinc finger nuclease
ZFs	Zinc fingers

Acknowledgments

First and foremost, I am extremely grateful to my supervisor, Prof. Jens Boch for his insightful comments and suggestions throughout my PhD journey. Words cannot express my gratitude. Without your guidance, it would not have been possible for me to grow and improve research skills. The results presented in my thesis would not have been possible without your support.

From the bottom of my heart, I would like to express my gratitude to all the members of the Department of Plant Biotechnology, especially Dr. Jana Streubel, Dr. Sebastian Becker and M. Sc. Patrick Buchholz for their understanding, and invaluable assistance throughout my project.

I would like to acknowledge the financial and organizational support provided by the LACoP project, especially for the technical assistance of Prof. Traud Winkelmann and Dr. Juvenal Assou.

I would like to thanks Prof. Traud Winkelmann, and Prof. Claus-Peter Witte for proofreading this thesis.

I want to express my sincere gratitude to my dear mother Xiulei Su (苏秀磊). Although she is no longer with me, she consistently stood by my side. Your faith in me made it possible for me to complete this journey.

Last but not least, I would like to express my heartfelt gratitude to my girlfriend, Dr. Hong Zhao, who has generously given up her time and unwaveringly supported me throughout my journey.

Thank you.

Curriculum Vitae

Name	**Dingbo Zhang**
Date and place of birth	**Dec-30-1990, China**

Education

11/2018 — present	**Ph. D. student at the Gottfried Wilhelm Leibniz-Universität Hannover, Hannover, Germany** **Institute of Plant Genetics, Department of Plant Biotechnology, Prof. Dr. Jens Boch**
09/2013 — 07/2018	**Master of Engineering in Biotechnology Engineering** **Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, China** Master thesis: Enhanced on-target editing of the high-fidelity SpCas9 variants by tRNA-sgRNA architectures. Supervisor: Prof. Caixia Gao
09/2009 — 07/2013	**Bachelor of Biotechnology Engineering** **Northwest Minzu University, China**

List of publications

1. **Zhang, D.**, Zhang, H., Li, T., Chen, K., Qiu, J.L. and Gao, C., 2017. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. *Genome biology*, *18*(1), pp.1-7.
2. Zhang, Y., Li, D., **Zhang, D.**, Zhao, X., Cao, X., Dong, L., Liu, J., Chen, K., Zhang, H., Gao, C. and Wang, D., 2018. Analysis of the functions of *TaGW2* homoeologs in wheat grain weight and protein content traits. *The Plant Journal*, *94*(5), pp.857-866.
3. Li, C., Zong, Y., Wang, Y., Jin, S., **Zhang, D.**, Song, Q., Zhang, R. and Gao, C., 2018. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome biology*, *19*, pp.1-9.
4. Zong, Y., Song, Q., Li, C., Jin, S., **Zhang, D.**, Wang, Y., Qiu, J.L. and Gao, C., 2018. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nature Biotechnology*, *36*(10), pp.950-953.
5. He, X., Wang, Y., Yang, F., Wang, B., Xie, H., Gu, L., Zhao, T., Liu, X., **Zhang, D.**, Ren, Q. and Liu, X., 2019. Boosting activity of high-fidelity CRISPR/Cas9 variants using a $tRNA^{Gln}$-processing system in human cells. *Journal of Biological Chemistry*, *294*(23), pp.9308-9315.
6. Assou, J., **Zhang, D.**, Roth, K.D., Steinke, S., Hust, M., Reinard, T., Winkelmann, T. and Boch, J., 2022. Removing the major allergen Bra j I from brown mustard (*Brassica juncea*) by CRISPR/Cas9. *The Plant Journal*, *109*(3), pp.649-663.
7. **Zhang, D.**, Pries, V., Boch, J. Targeted C•G-to-T•A base editing with TALE cytosine deaminases in plants. (submitted)
8. **Zhang, D.**, Boch, J. Development of TALE adenine base editors in plants. (under review)