(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 January 2020 (16.01.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/011985 A1

(51) International Patent Classification:
*C12N 15/10* (2006.01) *C12N 9/22* (2006.01)
*C12N 15/82* (2006.01)

(21) International Application Number:
PCT/EP2019/068839

(22) International Filing Date:
12 July 2019 (12.07.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/697,021 12 July 2018 (12.07.2018) US
18194078.4 12 September 2018 (12.09.2018) EP

(71) Applicant: **KEYGENE N.V.** [NL/NL]; P.O. Box 216, 6700 AE Wageningen (NL).

(72) Inventor: **BUNDOCK, Paul**; c/o P.O. Box 216, 6700 AE Wageningen (NL).

(74) Agent: **NEDERLANDSCH OCTROOIBUREAU**; P.O.Box 29720, 2502 LS The Hague (NL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2020/011985 A1

(54) Title: TYPE V CRISPR/NUCLEASE-SYSTEM FOR GENOME EDITING IN PLANT CELLS

(57) Abstract: Provided is a method for targeted modification of DNA in a plant cell using crRNA-guided MAD7- nuclease and compositions and kits for doing the same.

# Type V CRISPR/nuclease-system for genome editing in plant cells

## Field of the invention

The present invention is in the field of molecular biology and plant biology. The invention concerns targeted DNA modifications, including methods and compositions for making such modifications.

## Background

The process of deliberately creating changes in the genetic material of living cells has the goal of modifying one or more genetically encoded biological properties of that cell, or of the organism of which the cell forms part or into which it can regenerate. These changes can e.g. take the form of deletion of parts of the genetic material, addition of exogenous genetic material, or changes in the existing nucleotide sequence of the genetic material. Methods of altering the genetic material of eukaryotic organisms have been known for over 20 years, and have found widespread application in plant, human and animal cells and micro-organisms for improvements in the fields of agriculture, human health, food quality and environmental protection. The most common methods consist of adding exogenous DNA fragments to the genome of a cell, which will then confer a new property to that cell or its organism over and above the properties encoded by already existing genes, including applications in which the expression of existing genes will thereby be suppressed. Although many such examples are effective in obtaining the desired properties, these methods have several drawbacks. For example, these conventional methods are not very precise, because there is not always control over the genomic positions in which the exogenous DNA fragments are inserted (and hence over the ultimate levels of expression), and the desired effect will have to manifest itself over the natural properties encoded by the original and well-balanced genome. On the contrary, methods of genome editing that will result in the addition, deletion or conversion of nucleotides in predefined genomic loci will allow the precise modification of existing genes.

By using site-specific nucleases, such as zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) nucleases, the field of targeted DNA alteration is rapidly developing.

CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) are loci containing multiple short direct repeats and are found in 40% of the sequenced bacteria and 90% of sequenced archaea. The CRISPR repeats form a system of acquired bacterial immunity against genetic pathogens such as bacteriophages and plasmids. When a bacterium is challenged with a pathogen, a small piece of the pathogen's genome is processed by CRISPR associated proteins (Cas) and incorporated into the bacterial genome between CRISPR repeats. The CRISPR loci are then transcribed and processed to form so called crRNAs which include approximately 30 bps of sequence identical to the pathogen's genome. These RNA molecules form the basis for the recognition of the pathogen upon a subsequent infection and lead to silencing of the pathogen genetic elements through direct digestion of the pathogen's genome. The Cas9 protein is an essential component of the type-II CRISPR/Cas system from *S. pyogenes* and forms an

endonuclease, when combined with the crRNA and a second RNA termed the trans-activating crRNA (tracrRNA), which targets the invading pathogenic DNA for degradation by the introduction of DNA double strand breaks (DSBs) at the position in the genome defined by the crRNA. This type-II CRISPR/Cas9 system has been proven to be a convenient and effective tool in biochemistry that, via the targeted introduction of double-strand breaks and the subsequent activation of endogenous repair mechanisms, is capable of introducing modification in eukaryotic genomes at sites of interest. Jinek et al. (2012, Science 337: 816-820) demonstrated that a single chain chimeric RNA (single guide RNA, sRNA, sgRNA), produced by combining the essential sequences of the crRNA and tracrRNA into a single RNA molecule, was able to form a functional endonuclease in combination with Cas9. Many different CRISPR/Cas systems have been identified from different bacterial species (Zetsche et al. 2015 Cell 163, 759-771; Kim et al. 2017, Nat. Commun. 8, 1-7 ; Ran et al. 2015. Nature 520, 186-191).

The CRISPR/Cas9 system can be used for genome editing in a wide range of different organisms and cell types. First a genomic sequence is identified at which the CRISPR/Cas endonuclease should induce a DSB and this is then screened for the presence of a protospacer adjacent motif (PAM). The PAM sequence is essential for the CRISPR/Cas endonuclease activity, is relatively short, and is therefore usually present multiple times in any given sequence of some length. For instance the PAM motif of the *S. pyogenes* Cas9 protein is NGG, which ensures that for any given genomic sequence multiple PAM motifs are present and so many different guide RNAs can be designed. In addition, guide RNAs can also be designed targeting the opposite strands of the same double strand sequence. The sequence immediately adjacent to the PAM is incorporated into the guide RNA. This can differ in length depending upon the CRISPR/Cas system being used. For instance, the optimal length for the targeting sequence in the Cas9 sgRNA is 20nt, and in most cases a sequence of this length is unique in a plant genome. For expression in plant cells a gene coding for a guide RNA can be linked to an RNA polymerase-III promoter, such as the U6 promoter from *Arabidopsis*, or the corresponding or functionally similar pol-III promoter from the cell type, organism, plant species or family in which the experiments are being performed.

The CRISPR/Cas endonuclease can be expressed in the cell from any form of constitutive or inducible promoter that is suitable for the organism or cell type in which the experiments are being performed. In some instances, the protein expression levels of the CRISPR/Cas endonuclease can be improved by optimization of its codon usage for the specific cell type or organism.

The two components of the CRISPR/Cas system, the endonuclease and the guide RNA(s) can be expressed in the cell from ectopic genomic elements such as (non-replicating) plasmid constructs, viral vectors or introduced directly in the cells or organism as protein (the CRISPR/Cas endonuclease) and RNA (guide RNA). In addition mRNA encoding the CRISPR/Cas endonuclease can be used. When the plasmid or viral vectors are unable to replicate in the transformed cells then the CRISPR/Cas and guide RNA(s) are expressed or present for a short period and then are eliminated from the cell. Stable expression of the CRISPR/Cas protein and guide RNA can be

achieved using a transgenic approach whereby the genes coding for them are integrated into the host genome.

Once the CRISPR/Cas endonuclease and the guide RNA is present/expressed in the cell then the complex of the two components scans the genomic DNA for the sequence complementary to the targeting sequence on the guide RNA and adjacent to a PAM sequence. Depending on the CRISPR/Cas endonuclease being used, the complex then induces nicks in both of the DNA strands at varying distances from the PAM. For instance the *S. pyogenes* Cas9 protein introduces nicks in the both DNA strands 3 bps upstream from the PAM sequence to create a blunt DNA DSB.

Once a DNA DSB has been produced the cellular DNA repair machinery, particularly proteins belonging to the non-homologous end joining (NHEJ) pathway, are involved in the re-ligation of the DNA ends. If this DSB is repaired accurately then the sequence again forms a target for cutting by the CRISPR/Cas-guide RNA complex. However, some re-ligation events are imprecise and can lead to the random loss or gain of a few nucleotides at the break, resulting in an indel mutation in the genomic DNA. This results in an alteration of the target sequence that prevents binding of the guide RNA and thus any further DSB induction.

Recently, a second CRISPR/Cas system capable of the programmed introductions of DSBs was characterized, i.e. type-V CRISPR/Cpf1 (also known as CRISPR/Cas12a). The main differences between Cpf1 and Cas9, is that Cpf1 only requires a crRNA instead of both a crRNA and tracrRNA, the optimal spacer of Cpf1 is at least 21 nucleotides and Cpf1 recognizes a T-rich PAM instead of the G-rich PAM of Cas9. Further, the PAM is located upstream of the guide sequence and Cpf1 generates a staggered DSBs distal from the PAM, instead of a blunt-ended DSB proximal to the PAM as generated by Cas9. In WO/2018/115390, which is incorporated herein by reference, effective genome editing in plant cell protoplasts has been demonstrated.

The present inventors now discovered a type-V CRISPR/nuclease system being surprisingly effective in plants cells. Till now, this system, called MAD7, has only been reported to be effective in a prokaryotic host cell (see US 9,982,279, which is incorporated herein by reference).

**Figure legend**

**Figure 1:** Comparison of MAD7 and *As*Cpf1. The percentage of reads showing indels at the target region is shown on the y axis. In each sample either the *As*Cpf1 (Cpf1) or MAD7 enzyme was expressed together with the appropriate crRNA for targeting the respective target sites as listed in Table 1.

**Definitions**

Various terms relating to the methods, compositions, uses and other aspects of the present invention are used throughout the specification and claims. Such terms are to be given their ordinary meaning in the art to which the invention pertains, unless otherwise indicated. Other specifically defined terms are to be construed in a manner consistent with the definition provided herein. Although any methods and materials similar or equivalent to those described herein can be used in

the practice for testing of the present invention, the preferred materials and methods are described herein.

It is clear for the skilled person that any methods and materials similar or equivalent to those described herein can be used for practicing the present invention.

Methods of carrying out the conventional techniques used in methods of the invention will be evident to the skilled worker. The practice of conventional techniques in molecular biology, biochemistry, computational chemistry, cell culture, recombinant DNA, bioinformatics, genomics, sequencing and related fields are well-known to those of skill in the art and are discussed, for example, in the following literature references: Sambrook et al.. Molecular Cloning. A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989; Ausubel et al.. Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1987 and periodic updates; and the series Methods in Enzymology, Academic Press, San Diego.

“A,” “an,” and “the”: these singular form terms include plural referents unless the content clearly dictates otherwise. The indefinite article "a" or "an" thus usually means "at least one". Thus, for example, reference to “a cell” includes a combination of two or more cells, and the like.

“About” and “approximately”: these terms, when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of ±20% or ±10%, more preferably ±5%, even more preferably ±1%, and still more preferably ±0.1% from the specified value, as such variations are appropriate to perform the disclosed methods. Additionally, amounts, ratios, and other numerical values are sometimes presented herein in a range format. It is to be understood that such range format is used for convenience and brevity and should be understood flexibly to include numerical values explicitly specified as limits of a range, but also to include all individual numerical values or sub-ranges encompassed within that range as if each numerical value and sub-range is explicitly specified. For example, a ratio in the range of about 1 to about 200 should be understood to include the explicitly recited limits of about 1 and about 200, but also to include individual ratios such as about 2, about 3, and about 4, and sub-ranges such as about 10 to about 50, about 20 to about 100, and so forth.

“And/or”: The term “and/or” refers to a situation wherein one or more of the stated cases may occur, alone or in combination with at least one of the stated cases, up to with all of the stated cases.

“Comprising”: this term is construed as being inclusive and open ended, and not exclusive. Specifically, the term and variations thereof mean the specified features, steps or components are included. These terms are not to be interpreted to exclude the presence of other features, steps or components.

“Exemplary": this terms means "serving as an example, instance, or illustration," and should not be construed as excluding other configurations disclosed herein.

“Plant" refers to either the whole plant or to parts of a plant, such as cells, tissue cultures or organs (*e.g.* pollen, seeds, ovules, gametes, roots, leaves, flowers, flower buds, branches, anthers, fruit, kernels, ears, cobs, husks, stalks, root tips, grains, embryos, etc.) obtainable from the plant, as well as derivatives of any of these and progeny derived from such a plant by selfing or crossing.

Non-limiting examples of plants include crop plants and cultivated plants, such as barley, cabbage, canola, cassava, cauliflower, chicory, cotton, cucumber, eggplant, grape, hot pepper, lettuce, maize, melon, oilseed rape, potato, pumpkin, rice, rye, sorghum, squash, sugar cane, sugar beet, sunflower, sweet pepper, tomato, water melon, wheat, and zucchini.

"Plant cell(s)" include protoplasts, gametes, suspension cultures, microspores, pollen grains, etc., either in isolation or within a tissue, organ or organism. The plant cell can *e.g.* be part of a multicellular structure, such as a callus, meristem plant organ or an explant.

The terms “construct”, “nucleic acid construct”, “vector”, and “expression vector” are used interchangeably herein and is herein defined as a man-made nucleic acid molecule resulting from the use of recombinant DNA technology. These constructs and vectors therefore do not include naturally occurring nucleic acid molecules although a nucleic acid construct may comprise (parts of) naturally occurring nucleic acid molecules.

The vector backbone may for example be a binary or superbinary vector (see e.g. U.S. Pat. No. 5,591,616, US 2002138879 and WO 95/06722), a co-integrate vector or a T-DNA vector, as known in the art and as described elsewhere herein, into which a chimeric gene is integrated or, if a suitable transcription regulatory sequence is already present, only a desired nucleotide sequence (e.g. a coding sequence, an antisense or an inverted repeat sequence) is integrated downstream of the transcription regulatory sequence. Vectors can comprise further genetic elements to facilitate their use in molecular cloning, such as e.g. selectable markers, multiple cloning sites and the like.

The term “gene” means a DNA fragment comprising a region (transcribed region), which is transcribed into an RNA molecule (e.g. a pre-mRNA or ncRNA) in a cell. The transcribed region can be operably linked to suitable regulatory regions (e.g. a promoter), which form part of the gene as defined herein. A gene can comprise several operably linked fragments, such as a 5’ leader sequence, a coding region and a 3’ non-translated sequence (3’ end) comprising a polyadenylation site.

“Expression of a gene” refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly a promoter, is transcribed into an RNA, and, in case the RNA encodes for a biologically active protein or peptide, subsequently translated into a biologically active protein or peptide.

The term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleotide sequence. For instance, a promoter, or rather a transcription regulatory sequence, is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked may mean that the DNA sequences being linked are contiguous.

“Promoter” refers to a nucleic acid fragment that functions to control the transcription of one or more nucleic acids. A promoter fragment is located upstream (5’) with respect to the direction of transcription of the transcription initiation site of the gene, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation site(s) and can further comprise any other DNA sequences, including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides

known to one of skill in the art to act directly or indirectly to regulate the amount of transcription from the promoter.

Optionally the term “promoter” may also include the 5’ UTR region (5’ Untranslated Region) (*e.g.* the promoter may herein include one or more parts upstream of the translation initiation codon of transcribed region, as this region may have a role in regulating transcription and/or translation). A “constitutive” promoter is a promoter that is active in most tissues under most physiological and developmental conditions. An “inducible” promoter is a promoter that is physiologically (*e.g.* by external application of certain compounds) or developmentally regulated. A “tissue specific” promoter is only active in specific types of tissues or cells.

The terms “protein” or “polypeptide” are used interchangeably and refer to molecules consisting of a chain of amino acids, without reference to a specific mode of action, size, 3 dimensional structure or origin. A “fragment” or “portion” of a protein may thus still be referred to as a “protein.” A protein as defined herein and as used in any method as defined herein may be an isolated protein. An “isolated protein” is used to refer to a protein which is no longer in its natural environment, for example *in vitro* or in a recombinant bacterial or plant host cell.

The term “regeneration” is herein defined as the formation of a new tissue and/or a new organ from a single plant cell, a callus, an explant, a tissue or an organ. Preferably, the regeneration is at least one of shoot regeneration, ectopic apical meristem formation, and root regeneration. Regeneration can occur through somatic embryogenesis or organogenesis. The regeneration may further include the formation of a new plant from a single plant cell or from *e.g.* a callus, an explant, a tissue or an organ. The plant cell for regeneration can be an undifferentiated plant cell. The regeneration process hence can occur directly from parental tissues or indirectly, *e.g.* via the formation of a callus.

“Conditions that allow for regeneration” is herein understood as an environment wherein a plant cell or a tissue can regenerate. Such conditions include at minimum a suitable temperature, nutrition, day/night rhythm and irrigation.

The term "deaminase" refers to an enzyme that catalyzes a deamination reaction. In some embodiments, the deaminase is a cytosine deaminase, catalyzing the hydrolytic deamination of cytosine to uracil. The deaminase may also be an adenine deaminase, catalyzing the deamination of adenine thereby converting it to inosine.

“Sequence” or “Nucleotide sequence”: This refers to the order of nucleotides of, or within a nucleic acid. In other words, any order of nucleotides in a nucleic acid may be referred to as a sequence or nucleotide sequence.

The terms “homology”, “sequence identity” and the like are used interchangeably herein. Sequence identity is herein defined as a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (polynucleotide) sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between amino acid or nucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Similarity" between two amino acid sequences is determined

by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

The term “complementarity” is herein defined as the sequence identity of a sequence to a fully complementary strand (defined herein below, e.g. the second strand). For example, a sequence that is 100% complementary (or fully complementary) is herein understood as having 100% sequence identity with the complementary strand and e.g. a sequence that is 80% complementary is herein understood as having 80% sequence identity to the (fully) complementary strand.

"Identity" and "similarity" can be readily calculated by known methods. “Sequence identity” and “sequence similarity” can be determined by alignment of two peptide or two nucleotide sequences using global or local alignment algorithms, depending on the length of the two sequences. Sequences of similar lengths are preferably aligned using a global alignment algorithm (e.g. Needleman Wunsch) which aligns the sequences optimally over the entire length, while sequences of substantially different lengths are preferably aligned using a local alignment algorithm (e.g. Smith Waterman). Sequences may then be referred to as "substantially identical” or “essentially similar” when they (when optimally aligned by for example the programs GAP or BESTFIT using default parameters) share at least a certain minimal percentage of sequence identity (as defined below). GAP uses the Needleman and Wunsch global alignment algorithm to align two sequences over their entire length (full length), maximizing the number of matches and minimizing the number of gaps. A global alignment is suitably used to determine sequence identity when the two sequences have similar lengths. Generally, the GAP default parameters are used, with a gap creation penalty = 50 (nucleotides) / 8 (proteins) and gap extension penalty = 3 (nucleotides) / 2 (proteins). For nucleotides the default scoring matrix used is nwsgapdna and for proteins the default scoring matrix is Blosum62 (Henikoff & Henikoff, 1992, PNAS 89, 915-919). Sequence alignments and scores for percentage sequence identity may be determined using computer programs, such as the GCG Wisconsin Package, Version 10.3, available from Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA, or using open source software, such as the program “needle” (using the global Needleman Wunsch algorithm) or “water” (using the local Smith Waterman algorithm) in EmbossWIN version 2.10.0, using the same parameters as for GAP above, or using the default settings (both for ‘needle’ and for ‘water’ and both for protein and for DNA alignments, the default Gap opening penalty is 10.0 and the default gap extension penalty is 0.5; default scoring matrices are Blosum62 for proteins and DNAFull for DNA). When sequences have a substantially different overall lengths, local alignments, such as those using the Smith Waterman algorithm, are preferred.

Alternatively percentage similarity or identity may be determined by searching against public databases, using algorithms such as FASTA, BLAST, etc. Thus, the nucleic acid and protein sequences of the present invention can further be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the BLASTn and BLASTx programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403—10. BLAST nucleotide searches can be performed with the

NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the BLASTx program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., BLASTx and BLASTn) can be used. See the homepage of the National Center for Biotechnology Information at http://www.ncbi.nlm.nih.gov/.

A “homolog” of a gene is a further gene by descent from a common ancestral DNA sequence. The term homolog may apply to the relationship between genes separated by the event of speciation (ortholog) or to the relationship between genes separated by the event of genetic duplication (paralog). An “ortholog” of a gene is a gene in a different species that evolved from a common ancestral gene by speciation, and is understood herein as having retained the same function in the course of evolution.

A “target sequence” is to denote an order of nucleotides within a nucleic acid that is to be targeted, e.g. wherein an alteration is to be introduced or to be detected. For example, the target sequence is an order of nucleotides comprised by a first strand of a DNA duplex.

An “endonuclease” is an enzyme that hydrolyses at least one strand of a duplex DNA upon binding to its recognition site. An endonuclease is to be understood herein as a site-specific endonuclease and the terms “endonuclease” and “nuclease” are used interchangeable herein. A restriction endonuclease is to be understood herein as an endonuclease that hydrolyses both strands of the duplex at the same time to introduce a double strand break in the DNA. A “nicking” endonuclease is an endonuclease that hydrolyses only one strand of the duplex to produce DNA molecules that are “nicked” rather than cleaved.

**Detailed description of the invention**

The inventors discovered a novel method for effectively modifying DNA in plant cells at a site-specific way, *i.e.* by using for the first time in plant cells a new type-V CRISPR/nuclease, known as MAD7 or MAD7-nuclease. Surprisingly, in plant cells, the crRNA-guided MAD7-nuclease appeared more effective as compared to a type-V CRISPR/nuclease counterpart, *i.e.* *Acidominococcus sp.* Cpf1 (*As*Cpf1), at 6 out of 7 independent plant target sites within in total 4 different plant genes, and being equally effective at the remainder target site. Until now, Cpf1 has been considered as one of the most, or even the most, effective CRISPR nuclease in plants. Hence, the inventors discovered a more effective new type-V CRISPR/nuclease for DNA modification in plant cells.

Therefore, provided is a method for targeted modification of DNA in a plant cell, comprising contacting the DNA with a crRNA-guided MAD7-nuclease. The DNA may be any type of DNA, endogenous or exogenous to the cell, for example genomic DNA, chromosomal DNA, artificial chromosomes, plasmid DNA, or episomal DNA. The DNA may be nuclear or organellar DNA. Preferably, the DNA is chromosomal DNA, preferably endogenous to the cell. The method of the

invention results in altered DNA at a site of interest, also indicated herein as the target sequence, which is preferably within a gene of interest, preferably within a gene at the plants endogenous chromosomal DNA. In other words, the method of the invention preferably results in a genomic modification, which may refer to an epigenetic modification and/or a genetic modification. The terms "alteration" and "modification" are used interchangeably herein. An epigenetic modification is a heritable modification that changes the gene function or activity, without changing the nucleotide sequence. In an embodiment, the modification is a genetic modification. A genetic modification is understood herein as the alteration of the nucleotide sequence of the DNA, such as a deletion, insertion, substitution or conversion of one or more nucleotides.

Preferably, the method for targeted modification of DNA as detailed herein, wherein the method comprises a step of contacting the DNA with a crRNA-guided MAD7-nuclease, is as least as efficient as compared to an identical or nearly identical method, with the exception that a Cpf1 nuclease is used instead of a crRNA-guided MAD7-nuclease. Preferably the Cpf1 nuclease has an amino acid sequence of SEQ ID NO: 6. Hence preferably, the efficiency to generate one or more INDEL mutations in a plant genome is comparable between MAD7 and Cpf1. MAD7 may be more efficient than Cpf1 in generating INDELs in a plant genome, preferably the MAD7 nuclease is at least 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 times more efficient than the Cpf1 nuclease, preferably when MAD7 and Cpf1 target the same sequence in a plant genome.

The efficiency of generating INDEL mutations in a plant genome may be expressed as percentage of the reads showing INDELs at the target region.

The method of the invention may further comprise a step of multiplication and/or genotyping using any conventional method known in the art, in order to screen or test for the nucleotide alteration. The method of the invention therefore comprises such step after contacting the DNA with the crRNA-guided MAD7 nuclease. In a preferred embodiment, the cells, preferably the plant cells, comprising the targeted modification as defined herein, may be genotyped using deep-sequencing technologies, such as Illumina or 454 sequencing. Preferably when using the method of the invention at least 0.5%, 0.8%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10% or more of the total sequencing reads have an INDEL at the target region.

The targeted DNA modification may be within a coding sequence in the plant genome, thereby resulting in a modified protein, *e.g.* a protein comprising one or more amino acid alterations or a protein comprising a truncation. The targeted DNA modification may be within a noncoding sequence in the plant genome, such as in an intronic sequence or in a sequence encoding a noncoding (nc)RNA. The targeted modification in a noncoding sequence may result in *e.g.* a modified splice site or an alteration in the regulatory function of a non-coding RNA. The targeted nucleotide alteration may also be in a regulatory sequence resulting in the down or upregulation of gene expression, optionally in knocking out gene expression. The method of the invention may comprise a step of screening or testing for protein modifications and/or protein expression levels. Such screening or testing may be directly on the protein itself or on altered functionality using any conventional means. In addition or alternatively, the DNA modification may result in a phenotypic alteration of the plant cell or plant. Therefore, the method may comprise a step of screening or

testing for a phenotypic alteration or characteristic in the plant cell or plant, preferably a step of screening or tested for a phenotypic characteristic as defined herein.

Preferably, the method of the invention results in a genetic modification of a gene of interest.

Gene of interest

Preferably, a gene of interest (GOI) is a gene that produces or alters a characteristic, preferably a phenotypic characteristic. A plant GOI thus preferably produces or alters a plant characteristic, preferably a phenotypic plant characteristic. The term “plant characteristic” means any characteristic of a plant, plant cell or plant tissue. Preferably, the plant characteristic is selected from the group consisting of plant development, plant growth, yield, biomass production, plant architecture, plant biochemistry, plant physiology, metabolism, survival capacity and stress tolerance. Alternatively or in addition, the plant characteristic is selected from the group consisting of DNA synthesis, DNA modification, endoreduplication, cell cycle, cell wall biogenesis, transcription regulation, signal transduction, storage lipid mobilization, and photosynthesis.

The term “altering a plant characteristic” as used herein encompasses any change in the plant characteristic such as increase, decrease or change in time or place. It is understood herein that the plant GOI can alter the plant characteristic by introducing, increasing, decreasing, or removing the expression of the GOI and/or by modifying the functionality of the encoded protein such as by altering the coding sequence thereby resulting in expression of a modified encoded protein. Whether the plant characteristic is altered due to an introduced expression of the GOI, increased expression of the GOI, decreased expression of the GOI, removed expression of the GOI and/or modified functionality of the encoded protein, is dependent on the type of GOI and/or the type of plant characteristic.

In an embodiment, the targeted modification is genetic modification that alters a plant characteristic. Such modification may be an early stop. Such modification may also be a single nucleotide modification (SNP) resulting in an amino acid change in the translated protein, which may result in a single amino acid change.

Detailed herein below are, non-limiting, examples of plant characteristics that can be modified by the method of the invention as described herein:

“Growth” refers to the capacity of the plant or of plant parts to expand and increase in biomass. Altered growth refers amongst others to altered growth rate, cycling time, the size, expansion or increase of the plant. Additionally and/or alternatively, growth characteristics may refer to cellular processes comprising, but not limited to, cell cycle (entry, progression, exit), cell division, cell wall biogenesis and/or DNA synthesis, DNA modification and/or endoreduplication.

“Yield” refers to the harvestable part of the plant. “Biomass” refers to any part of the plants. These terms also encompass an increase in seed yield, which includes an increase in the biomass of the seed (seed weight) and/or an increase in the number of (filled) seeds and/or in the size of the seeds and/or an increase in seed volume, each relative to corresponding wildtype plants. An increase in seed size and/or volume may also influence the composition of seeds. An increase in seed yield could be due to an increase in the number and/or size of flowers. An increase in yield

may also increase the harvest index, which is expressed as a ratio of the total biomass over the yield of harvestable parts, such as seeds.

"Plant development" means any cellular process of a plant that is involved in determining the developmental fate of a plant cell, in particular the specific tissue or organ type into which a progenitor cell will develop. Typical plant characteristics according to the present invention are therefore characteristics relating to cellular processes relevant to plant development such as for example, morphogenesis, photomorphogenesis, shoot development, root development, vegetative development, reproductive development, stem elongation, flowering, regulatory mechanisms involved in determining cell fate, pattern formation, differentiation, senescence, time of flowering and/or time to flower.

Plant architecture", as used herein refers to the external appearance of a plant, including any one or more structural features or a combination of structural features thereof. Such structural features include the shape, size, number, position, colour, texture, arrangement, and patternation of any cell, tissue or organ or groups of cells, tissues or organs of a plant, including the root, stem, leaf, shoot, petiole, trichome, flower, petal, stigma, style, stamen, pollen, ovule, seed, embryo, endosperm, seed coat, aleurone, fibre, fruit, cambium, wood, heartwood, parenchyma, aerenchyma, sieve element, phloem or vascular tissue, amongst others.

The term "stress tolerance" is understood as the capability of better survival and/or better performing in stress conditions such as environmental stress, which can be biotic or abiotic. Salinity, drought, heat, chilling and freezing are all described as examples of conditions which induce osmotic stress. The term "environmental stress" as used in the present invention refers to any adverse effect on metabolism, growth or viability of the cell, tissue, seed, organ or whole plant which is produced by a non-living or non-biological environmental stressor. More particularly, it can encompass environmental factors such as water stress (flooding, water logging, drought, dehydration), anaerobic (low level of oxygen, CO2 etc.), aerobic stress, osmotic stress, salt stress, temperature stress (hot/heat, cold, freezing, frost) or nutrients deprivation, pollutants stress (heavy metals, toxic chemicals), ozone, high light, pathogen (including viruses, bacteria, fungi, insects and nematodes) and combinations of these. Biotic stress is stress as a result of the impact of a living organism on the plant. Examples are stresses caused by pathogens (virus, bacteria, nematodes insects etc.). Another example is stress caused by an organism, which is not necessarily harmful to the plant, such as the stress caused by a symbiotic or an epiphyte. Accordingly, particular plant characteristics obtained by modification of the GOI can encompass early vigour, survival rate, stress tolerance.

Characteristics related to "plant physiology" can encompass characteristics of functional processes of a plant, including developmental processes such as growth, expansion and differentiation, sexual development, sexual reproduction, seed set, seed development, grain filling, asexual reproduction, cell division, dormancy, germination, light adaptation, photosynthesis, leaf expansion, fiber production, secondary growth or wood production, amongst others; responses of a plant to externally-applied factors such as metals, chemicals, hormones, growth factors, environment and environmental stress factors (e.g. anoxia, hypoxia, high temperature, low

temperature, dehydration, light, day length, flooding, salt, heavy metals, amongst others), including adaptive responses of plants to said externally-applied factors. Particular plant physiology characteristics which are altered by the GOI identified in the method of the invention can further encompass altered storage lipid mobilization, photosynthesis, transcription regulation and signal transduction.

Plant characteristics related to "plant biochemistry" are to be understood by those skilled in the art to preferably refer to the metabolic characteristics. “Metabolism” can be used interchangeable with biochemistry. Metabolism and/or biochemistry encompass catalytic or assimilation or other metabolic processes of a plant, including primary and secondary metabolism and the products thereof, including any element, small molecules, macromolecules or chemical compounds, such as but not limited to starches, sugars, proteins, peptides, enzymes, hormones, growth factors, nucleic acid molecules, celluloses, hemicelluloses, calloses, lectins, fibres, pigments such as anthocyanins, vitamins, minerals, micronutrients, or macronutrients, that are produced by plants.

The modification of the GOI can be identified by determining the altered plant characteristic, preferably by determining at least one or more altered plant characteristics as defined herein above. In an embodiment, the plant cell having the preferred altered plant characteristic is selected and isolated from plant cells not having the altered plant characteristic. The plant cell can first be generated into a plant prior to selecting the plant having the altered plant characteristic.

The plant cell or plant having an altered plant characteristic can be sequenced to identify and/or further analyse the gene of interest. The skilled person understands that the whole plant genome can be sequenced or a part of the plant genome. In an embodiment, at least the modified sequence is determined.

Preferably, in those cases wherein the modification of the GOI results in an altered plant characteristic, the modification identifies a gene of interest. Therefore the method of the invention can be used in order to screen for gene functionality.

crRNA-guided MAD7-nuclease

The MAD7-nuclease or the protein of the crRNA-guided MAD7-nuclease complex may be a MAD7-nuclease obtainable from *Eubacterium rectale*, or any homolog or ortholog thereof. Preferably the MAD7-nuclease has an amino acid sequence that has at least about 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 1. Preferably, the MAD7-nuclease is a protein that is encoded by a nucleotide sequence having at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 2. Preferably, the MAD7-nuclease is a protein that is encoded by a nucleotide sequence having at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 74.

The crRNA is capable of complexing with the MAD7-endonuclease and can hybridize with a target sequence, thereby directing the nuclease to the target sequence. A subject nuclease capable of complexing with a guide nucleic acid can be referred to as a nuclease that is compatible with

the guide nucleic acid. Likewise, a guide nucleic acid capable of complexing with a nuclease can be referred to as a guide nucleic acid that is compatible with the nucleases. Preferably the crRNA is RNA but may also comprise DNA. The MAD7-endonuclease thus may form a complex with the crRNA, resulting in a crRNA-guided MAD7-nuclease.

The crRNA can comprise or consist of non-modified or naturally occurring nucleotides. Alternatively, the crRNA can comprise or consist of modified or non-naturally occurring nucleotides, preferably such chemically modified nucleotides are for protecting the crRNA against degradation.

In an embodiment of the invention, the crRNA comprises ribonucleotides and non-ribonucleotides. The crRNA can comprise one or more ribonucleotides and one or more deoxyribonucleotides.

The crRNA may comprise one or more non-naturally occurring nucleotides or nucleotide analogues, such as a nucleotide with phosphorothioate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2' and 4' carbons of the ribose ring, bridged nucleic acids (BNA), 2'-O-methyl analogues, 2'-deoxy analogues, 2'-fluoro analogues or combinations thereof. The modified nucleotides may comprise modified bases selected from the group consisting of, but not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, and 7-methylguanosine.

The crRNA may be chemically modified by incorporation of 2'-O-methyl (M), 2'-O-methyl 3'phosphorothioate (MS), 2'-O-methyl 3'thioPACE (phosphonoacetate) (MSP), or a combination thereof, at one or more terminal nucleotides. Such chemically modified crRNAs can comprise increased stability and/or increased activity as compared to unmodified crRNAs. (Hendel et al, 2015, Nat Biotechnol. 33(9);985-989). In certain embodiments, a crRNA comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucleotides and/or nucleotide analogues in a region that binds to the MAD7-nuclease. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogues can be incorporated in engineered crRNA structures, such as, without limitation, in the guide sequence, the scaffolding sequence and/or in between the guide and scaffolding sequences. Alternatively or in addition, the chemically modified nucleotides can be located 5' of the scaffolding sequence and/or 3' or the guide sequence.

Preferably, the crRNA for guiding or targeting the MAD7-nuclease to the target sequence in the DNA comprises or consists of a scaffolding and guide sequence, preferably said scaffolding sequence is at, or close to, its 5' end and the guide sequence is at, or close to, its 3' end. Preferably, said scaffolding sequence has a sequence that is at least about 50%, 60%, 70%, 80%, 90%, 95%,96% ,97%, 98%, 99%, or 100% identical to any one of SEQ ID NO: 21-24, preferably SEQ ID NO: 24. Alternatively, the scaffolding sequence has a sequence that is at least about 50%, 60%, 70%, 80%, 90%, 95%,96% ,97%, 98%, 99%, or 100% identical to SEQ ID NO:73.

The guide sequence is designed to have sufficient complementarity with the target sequence to hybridize with the target sequence and direct sequence-specific binding of a complexed Mad-7 nuclease to the target sequence. Said target sequence preferably is within a GOI as defined herein. The target sequence is adjacent to a protospacer adjacent motif (PAM) sequence, which preferably is a T-rich motif, preferably TTTN, wherein N can be any one of T, G, A or C. The skilled person is

capable of engineering the crRNA to target any desired target sequence, preferably by engineering the guide sequence to be complementary to any desired target sequence, in order to hybridize thereto. Preferably, the complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is at least about 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 100%. The guide sequence may be at least about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20 nucleotides in length. Preferably the guide sequence is about 10-30 nucleotides in length, or about 15-20 nucleotides in length. Preferably the guide sequence is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length, preferably 21 nucleotides.

The MAD7-nuclease may comprise endogenous catalytical activity, thereby being capable of introducing a DSB, or may be engineered to comprise at least one catalytically inactive domain, *e.g.* one active and one inactive domain. Said MAD7-nuclease may be modified to have an inactive Nuc domain rendering a MAD7-nickase, *e.g.* an MAD7-R1173A, or an inactive RuvC domain rendering a catalytically inactive MAD7, *e.g.* MAD7-D1213A, or any homolog or ortholog of MAD7 having a mutation at a similar or equivalent position. Such catalytically inactive MAD7-nuclease may serve to guide a (fused) functional domain as detailed herein below to a specific site in the DNA as determined by the crRNA.

The MAD7-nuclease, or MAD7-nickase or catalytically inactive MAD7-nuclease may be fused to a functional domain. Optionally, such functional domain is for epigenetic modification, for example a histone modification domain. The domains for epigenetic modification can be selected from the group consisting of a methyltransferase, a demethylase, a deacetylase, a methylase, a deacetylase, a deoxygenase, a glycosylase and an acetylase (Cano-Rodriguez et al, Curr Genet Med Rep (2016) 4:170–179). The methyltransferase may be selected from the group consisting of G9a, Suv39h1, DNMT3, PRDM9 and Dot1L. The demethylase may be LSD1.The deacetylase may be SIRT6 or SIRT3. The methylase may be at least one of KYP, TgSET8 and NUE. The deacetylase may be selected from the group consisting of HDAC8, RPD3, Sir2a and Sin3a. The deoxygenase may be at least one of TET1, TET2 and TET3, preferably TET1cd (Gallego-Bartolomé J et al, Proc Natl Acad Sci U S A. (2018);115(9):E2125-E2134). The glycosylase may be TDG. The acetylase may be p300.

Optionally, the functional domain is a deaminase, or functional fragment thereof, selected from the group consisting of an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced cytosine deaminase (AID), an ACF1/ASE deaminase, an adenine deaminase, and an ADAT family deaminase. Alternatively or in addition, the deaminase or functional fragment thereof may be ADAR1 or ADAR2, or a variant thereof.

The apolipoprotein B mRNA-editing complex (APOBEC) family of cytosine deaminase enzymes encompasses eleven proteins that serve to initiate mutagenesis in a controlled and beneficial manner. Preferably, the APOBEC deaminase is selected from the group consisting of APOBEC1, APOBEC2, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, APOBEC4 and Activation-induced (cytidine) deaminase. Preferably, the

cytosine deaminase of the APOBEC family is activation-induced cytosine (or cytidine) deaminase (AID) or apolipoprotein B editing complex 3 (APOBEC3). These proteins all require a $Zn^{2+}$ -coordinating motif (His-X-Glu-X23-26-Pro-Cys-X2_4-Cys) and bound water molecule for catalytic activity. Preferably, in a method of the invention, the deaminase domain of the fused to the MAD7-nuclease is an APOBEC1 family deaminase. Preferably, the deaminase domain is rat deaminase (rAPOBEC1) encoded by a sequence having at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 25 or 37, preferably with SEQ ID NO: 37. In addition or alternatively, the amino acid sequence of the rat deaminase domain has at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 26. Preferably, the deaminase domain has deaminase activity.

Another exemplary suitable type of deaminase domain that may be fused to the MAD7-nuclease is an adenine (or adenosine) deaminase, for example an ADAT family of adenine deaminase. Further, the adenine deaminase may be TadA or a variant thereof, preferably as described in Gaudelli et al., 2017 (Gaudelli et al. 2017 Nature 551: 464-471). Preferably, the deaminase domain is TadA encoded by a sequence having at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 27. In addition or alternatively, the amino acid sequence of TadA has at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 28. Preferably, the deaminase domain has deaminase activity. Further, the MAD7-nuclease may be fused to an adenine deaminase domain, *e.g.* derived from ADAR1 or ADAR2. The deaminase domain of the present invention may comprise or consist of a whole deaminase protein or a fragment thereof which has catalytic activity.

The functional domain, *e.g.* the deaminase domain, may be fused to the N- or C-terminus of the MAD7-nuclease or catalytically inactive MAD7-nuclease. Preferably, the functional domain is fused to the N-terminus of the MAD7-nuclease. Optionally, the functional domain and the MAD7-nuclease used in the method of the invention are fused directly to each other or via a linker (also denominated herein as a spacer).

The linker may be any suitable linker in the art, e.g., ranging from very flexible linkers of the form (GGGGS)n, (GGS)n, and (G)n to more rigid linkers of the form (EAAAK)n (SEQ ID NO: 29), (SPKKKRKVEAS)n (SEQ ID NO: 30), or (SGSETPGTSESATPES)n (SEQ ID NO: 31), or (KSGSETPGTSESATPES)n (SEQ ID NO: 32), or any variant thereof, wherein n preferably is between 1 and 7, *i.e.* 1, 2, 3, 4, 5, 6, or 7.

The linker preferably has a length between 2 and 30 amino acids, or between 3 and 23 amino acids, or between 3 and 18 amino acids.

Optionally, the MAD7-nuclease is further fused to an UDG inhibitor (UGI) domain. The UGI domain may be fused to the N- or C-terminus of the MAD7-nuclease. Preferably, the deaminase domain is fused to the C-terminus of the MAD7-nuclease. The fusion may be direct or via a linker as indicated above. Preferably, the MAD7-nuclease is fused to a deaminase domain at the N-terminus of the MAD7-nuclease, and the MAD7-nuclease is fused to a UGI domain at the C-terminus of the MAD7-nuclease.

Uracil DNA glycosylases (UDGs) recognize uracil, inadvertently present in DNA and initiates the uracil excision repair pathway by cleaving the N-glycosidic bond between the uracil and the deoxyribose sugar, releasing uracil and leaving behind a basic site (AP-site). The AP-site is then processed and restored to a canonical base by the subsequent actions of AP-endonuclease, dRPase, DNA polymerase and DNA ligase enzymes. By fusing a UGI domain to the cytosine deaminase containing MAD7-nuclease fusion protein, the efficiency of base editing increases. Preferably, the UGI domain is or is a variant of UGI from B. subtilis bacteriophage PBS1 or PBS2 (UniProtKB - P14739). Preferably, the nucleotide sequence of the UGI domain may have at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 33 or 35, preferably with SEQ ID NO: 35. Preferably, the amino acid sequence of the UGI domain may have at least about 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity with SEQ ID NO: 34 or 36, preferably with SEQ ID NO: 36. Preferably, the UGI domain inhibits UDG.

In an embodiment, the UDG inhibitor is not fused to the MAD7-nuclease protein as defined herein, but is contacted to the DNA to be edited as a further functional protein, preferably together with the crRNA-guided MAD7-nuclease. In this embodiment, the cell, preferably the plant cell, may be transfected using the UDG inhibitor or a construct encoding the UDG inhibitor. In the latter case, said construct may further comprise a sequence encoding the MAD7-nuclease (fusion) protein as defined herein, or alternatively, the UDG inhibitor and MAD7-nuclease (fusion) protein may be encoded on separate constructs.

The invention also pertains to the use of a MAD7-nuclease or nucleic acid encoding the same, and/or a crRNA for guiding said MAD7-nuclease as defined herein or nucleic acid encoding the same, in a method of the invention, *i.e.* for use in modifying a plant cell DNA and/or for stably expressing a plant cell.

Oligonucleotide mediated targeted nucleotide exchange (ODTNE)

In an embodiment, the crRNA-guided MAD7-nuclease as defined herein is capable of introducing a single-stranded or double stranded break. In this embodiment, the method of the invention may further comprise a step of introducing into the plant cell single-stranded oligonucleotide having a sequence that is at least partly complementary to a target sequence. Such single-stranded oligonucleotide can also be annotated as a template oligonucleotide. Within the context of the current invention “single-stranded” refers to a linear stretch of nucleotides, with a 5’end and a 3’end, without the presence of its fully complementary strand.

Preferably, the single-stranded oligonucleotide comprises the target sequence except for at least one mismatch with respect to the target sequence. In other words, the single-stranded oligonucleotide comprises the information with respect to the alteration that is to be introduced in the target sequence of the DNA to be modified in the method of the invention. In addition to the target sequence with the one or more mismatches, the single-stranded oligonucleotide, may, in certain embodiment also comprise additional stretches of nucleotides adjacent to the target sequence with the one or more mismatches. However, preferably the oligonucleotide substantially

consists of the target sequence with the one or more mismatches relative to the target sequence comprised in the first strand of the DNA duplex. Preferably, the single-stranded oligonucleotide comprises at least one mismatch with respect to the PAM sequence preferably present in the target sequence. Preferably, the single-stranded oligonucleotide comprises at least one mismatch with respect to the PAM sequence preferably present in the target sequence, and in addition at least one mismatch outside said PAM sequence. Preferably said at least one mismatch outside said PAM sequence, is a mismatch resulting in a codon alteration, which preferably results in an amino acid change, or early stop, in the encoded protein.

The use of ODTNE and the structure and design of the oligonucleotides that are functional in this technology are well described, inter alia in WO98/54330, W099/25853, WO01/24615, W001/25460, W02007/084294, W02007/073149, W02007/073166, W02007/073170 W02009/002150, WO2012/074385, WO2012/074386, WO2018/115389 and WO2015/139008. The skilled person thus straightforwardly understands how to design the first and second oligonucleotide for use in the current invention.

The mutagenic oligonucleotides used in the present invention preferably have a length that is in line with other mutagenic oligonucleotides used in the art, *i.e.* typically between 10-60 nucleotides, preferably 20-55 nucleotides, more preferably 25-50 nucleotides. The single-stranded oligonucleotide can comprise a chemical modification, *e.g.* to make the oligonucleotides at least partly resistant to nucleases.

In an embodiment, the oligonucleotide comprises at least about 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, or at least about 50 contiguous nucleotides that have 100% sequence identity with respectively at least about 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, or at least about 50 contiguous nucleotides of the DNA comprising the target sequence, preferably have 100% sequence identity with respectively at least about 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, or at least about 50 contiguous nucleotides of the DNA comprising the target sequence.

The single-stranded oligonucleotide may comprise both RNA and DNA nucleotides, preferably the single-stranded oligonucleotide does not comprise RNA nucleotides. Preferably, the single-stranded oligonucleotide consists of DNA nucleotides. In some embodiments, as will be detailed herein, one or more of the nucleotides of the single-stranded oligonucleotide comprise chemical modifications. Preferably, the single-stranded oligonucleotide is a chemically protected oligonucleotide, preferably comprising at least one, two, three or four chemically protected nucleotides. Such a chemically protected oligonucleotide may be more resistant to nucleases and may in addition provide for higher binding affinity.

The type of modifications to provide a chemically protected oligonucleotide is not in particular limited. Examples of suitable modifications include the introduction of a reverse base (idC) at the 3′ end of the oligonucleotide to create a 3′ blocked end on the repair oligonucleotide; introduction of one or more 2′O-methyl nucleotides or bases which increase hybridization energy (see WO2007/073149) at the 5′ and/or 3′ of the repair oligonucleotide; conjugated (5′ or 3′) intercalating dyes such as acridine, psoralen, and ethidium bromide; introduction of a 5′ terminus

cap such as a T/A clamp, a cholesterol moiety, SIMA (HEX), and riboC; backbone modifications such as phosphothioate, methyl phosphonates, MOE (methoxyethyl), di PS and peptide nucleic acid (PNA); or ribose modifications such as 2' O methyl and locked nucleic acids (LNA). Preferred chemical modifications are either phosphorothioates (PS) that help to protect the single-stranded oligonucleotide from degradation (PS are normally placed at the ends of the single-stranded oligonucleotide) or locked nucleic acids (LNAs) that give both protection against nucleases and also a higher binding affinity (LNAs can be placed either at the ends of the single-stranded oligonucleotide or internally).

Preferably the at least one mismatch is not a chemically protected nucleotide. According to another preference a chemically protected nucleotide is at least one nucleotide from the at least one mismatch. In other words, preferably a mismatch in the single-stranded oligonucleotide is not a chemically protected nucleotide and the nucleotide adjacent to (either side) the mismatch is also not a chemically protected (or modified) nucleotide.

Homologous Recombination

In an embodiment, the crRNA-guided MAD7-nuclease as defined herein is capable of introducing a single-stranded or double stranded break, e.g. because the nuclease comprises, preferably endogenous, nuclease activity. In this embodiment, the method of the invention may further comprise a step of introducing into the plant cell a double-stranded oligonucleotide, a double stranded DNA, or a double-stranded DNA fragment (donor fragment) having a sequence that is at least partly complementary to a sequence of the DNA to be modified by the method of the invention and comprising the target sequence. Such double-stranded oligonucleotide or DNA is annotated herein as a "donor construct" or a "donor nucleotide", which are used interchangeably herein.

The double-stranded oligonucleotide can be used to modify the sequence of the DNA as defined herein, preferably by means of homologous recombination. Hence, the double-stranded oligonucleotide preferably comprises a sequence at its 3'-end and at its 5'-end that has 100% sequence identity or 100% sequence complementarity with a sequence in the DNA to be modified. Preferably, the sequence in the DNA that is identical to the double-stranded oligonucleotide is located immediately adjacent to the cleavage site generated by the crRNA-guided MAD7-nuclease.

In the presence of the homologous sequence, which can either be the sister chromatid, the donor fragment or the double-stranded oligonucleotide, the DSB can be repaired by homologous recombination (HR). This is the basis for the process of gene targeting whereby, rather than the sister chromatid being used for repair, information is copied from a double-stranded oligonucleotide or donor fragment that is introduced into the cell. The double-stranded oligonucleotide or donor fragment contains alterations or an exogenous sequence compared with the original chromosomal locus, and thus the process of HR incorporates these alterations or exogenous sequence in the genome. Preferably, the exogenous sequence to be introduced by HR is located within the double-stranded oligonucleotide or donor fragment in between the sequences at its 3'-end and at its 5'-end that has 100% sequence identity or 100% sequence complementarity with a sequence in the DNA.

The double-stranded oligonucleotide or double-stranded DNA, or DNA fragment, can be linearized or circular. Preferably, the double-stranded oligonucleotide is a linearized molecule. The first and/or second double-stranded oligonucleotide can comprise a chemical modification, *e.g.* to make the oligonucleotides at least partly resistant to nucleases. Preferably, the double-stranded oligonucleotide comprises RNA, DNA or is an RNA-DNA hybrid. Preferably, the double-stranded oligonucleotide comprises DNA.

Introducing the crRNA-guided MAD7-nuclease

The method of the invention comprises the step of contacting the DNA of the plant cell with crRNA-guided MAD7-nuclease. This may be accomplished by introducing into the plant cell a MAD7-nuclease and crRNA for guiding the MAD7-nuclease to the DNA. The method of the invention may therefore also be defined as a method for targeted modification of DNA in a plant cell, comprising the step of introducing into the plant cell a MAD7-nuclease and crRNA for guiding the MAD7-nuclease to the DNA. The MAD7-nuclease and crRNA may be delivered in the plant cell directly in the form of the MAD7-nuclease protein and/or crRNA (preferably as pre-assembled ribonucleo-protein complexes), as mRNA encoding the MAD7-nuclease and/or crRNA and/or precursor crRNA into a cell, or as one or more nucleic acids encoding the MAD7-nuclease protein and/or crRNA. The latter may be performed by the transfection of one or more vectors comprising a gene encoding the MAD7-nuclease and/or crRNA, wherein the vectors are for the transient expression of said gene and/or crRNA. The introduction in the plant cell or transfection may be performed by any conventional method known in the art. Optionally, sequences encoding MAD7 and/or crRNA(s) are stably introduced in the genome of the plant cell. In case the MAD7-nuclease and crRNA are delivered in the cell as a ribonucleo-protein complex, the method of the invention further comprises the step of forming said complex prior to the step of introducing said complex in the plant cell.

Optionally, the method further comprises a step of introducing a single-stranded oligonucleotide for ODTNE as defined herein or a step of introducing a double-stranded oligonucleotide or donor fragment for HR as defined herein. Optionally, the MAD7-nuclease protein or construct encoding the protein, crRNA or construct encoding the crRNA, and single-stranded oligonucleotide for ODTNE or double-stranded oligonucleotide or donor fragment for HR, are introduced in the plant cell in a single step, i.e. at substantially the same time. In other words, preferably, the MAD7-nuclease protein or construct encoding the protein, crRNA or construct encoding the crRNA, and single-stranded oligonucleotide for ODTNE or double-stranded oligonucleotide or donor fragment for HR are introduced in the plant cell in a single transfection step. Optionally, these components are transfected in two or more transfections steps.

The MAD7-nuclease protein may contain one or more nuclear localization signal sequences (NLS), mutations, deletions, alterations or truncations. Preferably, the NLS-linked MAD7-nuclease protein has an amino acid sequence that has at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95% , 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 3. Preferably, the NLS-linked MAD7-nuclease is a protein that is encoded by a nucleotide sequence having at least

about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95% , 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 4. In addition, the MAD7-nuclease encoding genes may be codon optimized, *e.g.* for expression in plants, comprise transcription regulatory sequences and may be driven by either a constitutive, inducible, tissue-specific or species-specific promoter when applicable. Preferably, by a constitutive, inducible, tissue-specific or species-specific promoter that is suitable for expression in plant cells. Preferably, the codon optimized NLS-linked MAD7-nuclease encoding sequence has at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95% , 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 4.

Exemplary transcript termination and polyadenylation signals are either NosT, RBCT, HSP18.2T or other gene specific or species-specific terminators. The MAD7-nuclease gene cassettes or mRNA may contain introns, either native or in combination with gene-specific promoters and or synthetic promoters.

In a preferred embodiment, the cell is transformed with at least one MAD7-nuclease, *i.e.* the MAD7-nuclease protein is delivered directly into the cell. In a further embodiment, the cell is transformed with at least one crRNA. Optionally, the cell is transformed with one MAD7-nuclease and two or more crRNAs, wherein each crRNA comprises a different guide sequence, *i.e.* targets a different sequence within the plant genome. Preferably, the method of the invention comprises a step of complexing the MAD7-nuclease protein with at least one crRNA before introducing the MAD7-nuclease and crRNA in the cell. In such embodiment, the crRNA-guided MAD7-nuclease is introduced in the plant cell as a complex. Optionally, said introduction step also comprise the introduction of a single-stranded oligonucleotide for ODTNE as defined herein, or a double-stranded oligonucleotide or donor fragment for HR as defined herein.

In another preferred embodiment, the cell is transfected with a nucleic acid construct encoding at least one MAD7-nuclease. The cell is may further be transfected with an additional nucleic acid construct encoding at least one crRNA, wherein optionally, said nucleic acid constructs encodes for two or more crRNAs, wherein each crRNA comprises a different guide sequence, *i.e.* targets a different sequence within the plant genome. Preferably, within this embodiment, the cell is transfected with a single construct encoding at least one MAD7-nuclease and one or more crRNAs as defined herein. Preferably, the nucleotide sequence encoding the MAD7-nuclease protein and the nucleotide sequence encoding the crRNA are under control of different promoters. For example, the MAD7-nuclease protein may, preferably, be under control of a constitutive promoter, preferably suitable for expression in plant, such as the 35 S promoter (e.g. the 35 S promoted from cauliflower mosaic virus (CaMV; Odell et al. Nature 313:810-812; 1985). Other suitable constitutive promoters include, but are not limited to, the cassava vein mosaic virus (CsVMV) promoter, and the sugarcane bacilliform badnavirus (ScBV) promoter (see e.g. Samac et al. Transgenic Res. 2004 Aug;13(4):349-61.) Other constitutive promoters include, for example, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43 838 and US 6072050; ubiquitin (Christensen et al., Plant Mol. Biol. 12:619-632, 1989 and Christensen et al., Plant Mol. Biol. 18:675- 689, 1992); pEMU (Last et al., Theor. Appl. Genet. 81:581-588, 1991); AA6 promoter (WO2007/069894); and the like. The nucleic acid constructs may also include

transcription termination regions. Where transcription termination regions are used, any termination region may be used in the preparation of the nucleic acid constructs. In a preferred embodiment, the nucleic acid construct is for transient expression. In other words, the expression in the plant material is temporary as a consequence of the non-permanent presence of the nucleic acid construct. Expression may, for instance, be transient when the construct is not integrated into the host genome. For example, MAD7-nuclease protein and crRNA may be transiently provided to a plant cell, followed by a decline in the amount of either or both of the components. Subsequently, the plant cell, progeny of the plant cell, and plants which comprise the plant cell or have been derived from the plant protoplast wherein the duplex DNA has been altered, comprise a reduced amount of either or both of the components used in the method of the invention, or no longer contain one or more of the components. Preferably, said plant cell, progeny of the plant cell, and plants which comprise the plant cell or have been derived from the plant protoplast wherein the duplex DNA has been altered, still comprise the DNA modification.

The nucleic acid construct encoding the MAD7-nuclease may be optimized for increased expression in the transformed plant, i.e. codon-optimized for expression in the plant cell. For instance, the nucleotide sequence encoding the MAD7-nuclease may be codon-optimized for expression in tomato, wherein said tomato preferably is *Solanum lycopersicum*. That is, the nucleic acid construct encoding the MAD7-nuclease protein can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri, (Plant Physiol. 92: 1-11, 1990) for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes (see, for example, Murray et al., Nucleic Acids Res. (1989) 17:477-498, or Lanza et al. (2014) BMC Systems Biology 8:33-43).

The invention also pertains to nucleic acids and/or constructs encoding MAD7-endonuclease as defined herein, wherein said coding sequence is preferably characterized in that it is codon optimized for expression plant cells, such as but not limited to expression in *Solanum Lycopersicum*. Preferably, the coding sequence has at least about at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 2 and/or the coding sequence has at least about at least about 50%, 60%, 70%, 80%, 85% 90%, 91%, 92%, 93%, 94%, 95%, 96% ,97%, 98%, 99%, or 100% identity to SEQ ID NO: 74.

The invention also pertains to nucleic acids and/or constructs encoding one or more crRNAs as defined herein, which are characterized in that the guide sequence is for targeting a sequence in the plant genome, preferably a sequence within the GOI as defined herein. The crRNAs may comprise a scaffolding sequence as defined herein.

Such nucleic acids may be single-stranded, double stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. Expression vectors according to the invention is suitable for introducing gene expression in a cell, preferably a plant cell. A preferred expression vector is a naked DNA, a DNA complex or a viral vector, wherein the DNA molecule can be a plasmid. A plasmid refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard

molecular cloning techniques. A DNA complex can be a DNA molecule coupled to any carrier suitable for delivery of the DNA into the cell. A preferred carrier is selected from the group consisting of a lipoplex, a liposome, a polymersome, a polyplex, a dendrimer, an inorganic nanoparticle, a virosome and cell-penetrating peptides. In a preferred embodiment the expression vector is a viral vector, preferably a Tobacco Rattle Virus (TRV), a Bean yellow dwarf virus (BeYDV), a Cabbage leaf curl virus (CaLCuV), a tobravirus and a Wheat dwarf virus (WDV). Preferably, the viral vector is a Tobacco Rattle Virus as defined herein above.

The invention further pertains to a composition comprising one or more, preferably two or more crRNAs as defined herein. Preferably, the crRNA, or crRNAs, is/are characterized in that the guide sequence is for targeting a sequence in the plant genome, preferably a sequence within the GOI as defined herein. The crRNA, or crRNAs, may comprise a scaffolding sequence as defined herein.

There are many suitable approaches known in the art for delivering the nucleic acids (encoding the MAD7-nuclease and/or (encoding) the crRNAs) or the proteins or ribonucleo-protein complexes into the cell. The delivery system may for example constitute a viral-based delivery system or a non-viral delivery system.

Non-limiting examples of non-viral delivery systems include chemical-based transfection (e.g. using calcium phosphate, dendrimers, cyclodextrin, polymers, liposomes, or nanoparticles), non-chemical-based methods (e.g. electroporation, cell squeezing, sonoporation, optical transfection, protoplast fusion, impalefection, heat shock and hydrodynamic delivery), particle-based methods (e.g. a gene gun or magnet-assisted transfection) and bacterial-based delivery systems (e.g. *agrobacterium*-mediated delivery). Non-limiting examples of a viral delivery system includes lentivirus and adenovirus.

In a preferred embodiment, the nucleic acids and/or proteins are introduced into the cell using an aqueous medium, wherein the aqueous medium comprises PEG. Any suitable method can be used, preferably the medium has a pH value of between 5 – 8, preferably between 6 – 7.5. Next to the presence in the aqueous medium of the MAD7-nuclease and optionally the crRNA, the medium comprises polyethylene glycol. Polyethylene glycol (PEG) is a polyether compound with many applications from industrial manufacturing to medicine. PEG is also known as polyethylene oxide (PEO) or polyoxyethylene (POE). The structure of PEG is commonly expressed as H−(O−CH2−CH2)n−OH. Preferably, the PEG used is an oligomer and/or polymers, or mixtures thereof with a molecular mass below 20,000 g/mol.

The aqueous medium comprising the population of e.g. plant cells preferably comprises 100 – 400 mg/ml PEG. So the final concentration of PEG is preferably between 100 – 400 mg/ml, for example, between 150 and 300 mg/ml, for example between 180 and 250 mg/ml. A preferred PEG is PEG 4000 Sigma-Aldrich no. 81240. (i.e. having an average Mn 4000 (Mn, the number average molecular weight is the total weight of all the polymer molecules in a sample, divided by the total number of polymer molecules in a sample.). Preferably the PEG used as a Mn of about 1000 – 10 000, for example between 2000 – 6000).

In a further preferred embodiment, the aqueous medium comprising PEG does not comprise more than about 0.001%, 0.01%, 0.05%, 0.1%, 1%, 2%, 5%, 10% or 20% (v/v) glycerol. Preferably, the medium comprises less than about 0.001%, 0.01%, 0.05%, 0.1%, 1%, 2%, 5%, 10% or 20% (v/v) glycerol. In particular for the introduction of a MAD7-nuclease protein, the aqueous medium comprises less than about 0.1 % (for example, less than 0.09%, 0.08%, 0.07%, 0.06%, 0.05%, 0.04%, 0.03%, 0.02%, 0.01%, 0.009%, 0.008%, 0.007%, 0.006%, 0.005%, 0.004%, 0.003%, 0.002%, 0.001% , 0.0009%, 0.0008%, 0.0007%, 0.0006%, 0.0005%, 0.0004%, 0.0003%, 0.0002% or 0.0001% (v/v) glycerol. Optionally, the aqueous medium comprising the population of plant cells is completely free of glycerol.

Preferably, the cell cycle of e.g. plant cells is synchronized when exposing the duplex DNA to the crRNA-guided MAD7-nuclease. The synchronization preferably takes places when the crRNA-guided MAD7-nuclease or nucleic acid(s) encoding the same is introduced into the cell as detailed herein. Synchronization is preferably performed by contacting the (plant) cell with a synchronizing agent.

Such method of synchronizing the cell cycle of the (plant) cell has been described in detail in European patent EP2516652, incorporated herein by reference. More particular, synchronizing the (plant) cells, for example, the plant protoplasts may be advantageous in certain embodiments of the invention to further enhance efficacy of the introduction of the alteration in the duplex DNA. Thus, in certain embodiments, the method comprises a step of synchronizing the cell cycle of the cell, preferably a plant cell.

The synchronization preferably takes places when the crRNA-guided MAD7-nuclease or nucleic acid(s) encoding the same is introduced into the cell as detailed herein, such that most of the (plant) cells will be in the same phase of the cell cycle when the duplex DNA is exposed to the site-specific nucleases as defined herein. This may be advantageous and increase the rate of introduction of the alteration in the duplex DNA.

Synchronizing the (plant) cell may be accomplished by any suitable means. For example, synchronization of the cell cycle may be achieved by nutrient deprivation such as phosphate starvation, nitrate starvation, ion starvation, serum starvation, sucrose starvation, auxin starvation.

Synchronization can also be achieved by adding a synchronizing agent to the (plant) cell. Preferably, the synchronizing agent is selected from the group consisting of aphidocolin, hydroxyurea, thymidine, colchicine, cobtorin, dinitroaniline, benefin, butralin, dinitramine, ethalfluralin, oryzalin, pendimethalin, trifluralin, amiprophos-methyl, butamiphos dithiopyr, thiazopyr propyzamide, tebutam DCPA (chlorthal-dimethyl), mimosine, anisomycin, alpha amanitin, lovastatin, jasmonic acid, abscisic acid, menadione, cryptogeine, hydrogenperoxide, sodiumpermanganate, indomethacin, epoxomycin, lactacystein, icrf 193, olomoucine, roscovitine, bohemine, staurosporine, K252a, okadaic acid, endothal, caffeine, MG 132, cycline dependent kinases and cycline dependent kinase inhibitors, as well as their target mechanism. The amounts and concentrations and their associated cell cycle phase are described for instance in "Flow Cytometry with plant cells", J. Dolezel c.s. Eds. Wiley-VCH Verlag 2007 pp 327 ff. Preferably, the synchronizing agent is aphidicolin and/or hydroxyurea.

Preferably, in the method of the invention, synchronizing the cell cycle synchronizes the (plant) cell in the S-phase, the M-phase, the G1 and/or G2 phase of the cell cycle.

In a preferred embodiment, the MAD7-nuclease comprises two catalytically active endonuclease domains. Within this embodiment, the crRNA-guided MAD7-nuclease will introduce a double-strand break in the target sequence. Subsequent activation of the repair mechanism results in alteration of the target sequence of the plant genome. The targeted alteration may comprise the insertion, deletion or modification of at least one base pair. For example, the targeted alteration may comprise the deletion of at least one base pair and the insertion of at least one base pair. For example, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more base pairs may be altered with the method of the invention. More than one modification may be introduced in a single experiment, and/or the experiment may be repeated to introduce subsequent alterations in the genome of the plant cell, optionally at other or at the same gene as targeted in the first event.

The invention further pertains to a method for targeted alteration of a coding sequence (CDS) in duplex DNA, preferably as described in PCT/EP2018/074150 which is incorporated herein by reference, wherein the method comprises a step of exposing the duplex DNA to at least two crRNA-guided MAD7 nucleases, wherein a first site-specific nuclease cleaves the DNA generating a first indel at a first location within the ORF and wherein a second site-specific nuclease cleaves the DNA generating a second indel at a second location within the same CDS, wherein the CDS before the first indel and after the second indel remain in the same reading frame, and wherein the altered CDS does not comprise a stop codon.

Plant cell

The method of the invention may further comprise the step of providing a plant cell, preceding the step of introducing into said plant cell a MAD7-nuclease and crRNA for guiding the MAD7-nuclease as defined herein. The skilled person understands that the method of the invention is not limited to a certain plant cell type. In particular, the method of the invention as disclosed herein can be applied to dividing as well as non-dividing cells. The cell may be transgenic or non-transgenic. The plant cell can for example be obtainable from plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, grains and the like.

In a preferred embodiment, the plant cell is a plant protoplast. The skilled person is aware of methods and protocols for preparing and propagating plant protoplasts, see for example Plant Tissue Culture (ISBN: 978-0-12-415920-4, Roberta H. Smith). The plant protoplasts for use in the method of the current invention can be provided using common procedures used for the generation of plant cell protoplasts (e.g. the cell wall may be degraded using cellulose, pectinase and/or xylanase).

Plant cell protoplasts systems have for example been described for tomato, tobacco and many more (Brassica napus, Daucus carota, Lactucca sativa, Zea mays, Nicotiana benthamiana, Petunia hybrida, Solanum tuberosum, Oryza sativa). The present invention is generally applicable

to any protoplast system, including those, but not limited to, the systems described in any one of the following references: Barsby et al. 1986, Plant Cell Reports 5(2): 101–103; Fischer et al. 1992, Plant Cell Rep. 11(12): 632-636; Hu et al. 1999, Plant Cell, Tissue and Organ Culture 59: 189–196; Niedz et al. 1985, Plant Science 39: 199-204; Prioli and Söndahl, 1989, Nature Biotechnology 7: 589 - 594; S. Roest and Gilissen 1989, Acta Bot. Neerl. 38(1): 1-23; Shepard and Totten, 1975, Plant Physiol.55: 689-694; Shepard and Totten, 1977, Plant Physiol. 60: 313-316, which are incorporated herein by reference.

The plant cell is preferably obtainable from a crop plant such as a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. A crop plant is plant species which is cultivated and bred by humans. A crop plant may be cultivated for food purposes (e.g. field crops), or for ornamental purposes (e.g. production of flowers for cutting, grasses for lawns, etc.). A crop plant as defined herein also includes plants from which non-food products are harvested, such as oil for fuel, plastic polymers, pharmaceutical products, cork and the like.

The plant cell may also be of an alga, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus *Brassica*; plants of the genus *Lactuca*; plants of the genus *Spinacia*; plants of the genus *Capsicum*; plants of the genus *Solanum,* preferably *Solanum lycopersicum*).

In another preferred embodiment, the cell is obtainable from a plant selected from the group consisting of asparagus, barley, blackberry, blueberry, broccoli, cabbage, canola, carrot, cassava, cauliflower, chicory, cocoa, coffee, cotton, cucumber, eggplant, grape, hot pepper, lettuce, maize, melon, oilseed rape, pepper, potato, pumpkin, raspberry, rice, rye, sorghum, spinach, squash, strawberry, sugar cane, sugar beet, sunflower, sweet pepper, tobacco, tomato, water melon, wheat, and zucchini.

Preferably, the obtained plant cell comprising the targeted alteration is regenerated into a plant or descendent therefore. Therefore a preferred embodiment of the invention, the method further comprises a step of regenerating a plant or descendent thereof comprising the targeted alteration.

Plant cell and plant or plant products

The method may further comprise the step of regenerating a plant or descendent thereof comprising the targeted modification. Preferably, such regeneration is performed using conditions suitable for regeneration. The skilled person is well aware of methods and protocols of regenerating a plant from a plant protoplast. Progeny, descendant's, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the targeted alteration introduced with the method taught herein.

In addition to the plant cell or plant comprising the targeted modification, the invention also pertains to a plant cell transiently or stably expressing MAD7-endonuclease. Preferably, the plant cell is a transgenic plant modified to comprise a sequence encoding MAD7-endonuclease in its

genome as defined herein. Included within the scope of the invention is such transgenic plant cell and plant regenerated thereof, any progeny, descendant's, variants, and mutants comprising the sequence encoding MAD7-endonuclease, preferably under the control of an inducible promoter and/or a meristem promoter which may be constitutive active meristem promoter. The invention therefore also pertains to a method for producing such transgenic plant cell and/or plant derived therefrom comprising the step of integrating into its genome the MAD7-endonuclease encoding sequence as defined herein.

The cell or organism obtainable by a method of the invention may subsequently be propagated to *e.g.* obtain a culture of cells, (part of) an organism or any descendants thereof.

Preferably, the cell is a plant cell obtainable from a crop plant such as a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. The plant cell may also be of an alga, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus *Brassica*; plants of the genus *Lactuca*; plants of the genus *Spinacia*; plants of the genus *Capsicum*; plants of the genus *Solanum,* preferably *Solanum lycopersicum*).

In another preferred embodiment, the cell is obtainable from a plant selected from the group consisting of Arabidopsis, asparagus, barley, blackberry, blueberry, broccoli, cabbage, canola, carrot, cassava, cauliflower, chicory, cocoa, coffee, cotton, cucumber, eggplant, grape, hot pepper, lettuce, maize, melon, oilseed rape, pepper, potato, pumpkin, raspberry, rice, rye, sorghum, spinach, squash, strawberry, sugar cane, sugar beet, sunflower, sweet pepper, tobacco, tomato, water melon, wheat, and zucchini.

The invention also pertains to the progeny of a plant cell or plant obtainable by a method of the invention. Further, the invention pertains to a plant product obtainable from the plant cell or plant as defined herein, *e.g.* fruits, leaves, plant organs, plant fats, plant oils, plant starch, and plant protein fractions, either crushed, milled or still intact, mixed with other materials, dried, frozen, and so on. These products may be non-propagating. Preferably, said plant product comprises at least or at least part of one of:

i) the modified DNA, preferably the modified GOI as defined herein, or encoded products thereof,

ii) the MAD7-nuclease as define herein, and

iii) the MAD7-nuclease encoding sequence.

Preferably, these products comprise at least fractions of modified DNA the MAD7-nuclease or encoding sequence, which allows to assess that the plant product is derived from a plant obtained by a method as defined herein.

Kit of parts

The invention also concerns a kit of parts, preferably a kit of parts for use in the method as described herein. Preferably, the kit of parts comprises at least one of:

- A container comprising a MAD7-nuclease as defined herein;

- A container comprising a crRNA for guiding the MAD7-nuclease as defined herein, a single-stranded oligonucleotide for ODTNE as defined herein, a double-stranded oligonucleotide for HR as defined herein, or any combination thereof. Preferably, the container comprises two or more crRNAs each comprising different guide sequences for targeting different target sequences;
- A container comprising one or more nucleic acid constructs or vectors encoding a MAD7-nuclease as defined herein. Optionally, said construct or vector further encodes one or more crRNAs for guiding said MAD7-nuclease as defined herein; and
- A container comprising one or more nucleic acid constructs or vectors encoding one or more crRNAs for guiding said MAD7-nuclease as defined herein.

The kit of parts may further comprise a container comprising one or more substances for transfection as defined herein. Also included within the kit may be a manual for performing the method of the invention of modifying a DNA within the plant cell as specified herein.

The reagents may be present in lyophilized form, or in an appropriate buffer. The kit may also contain any other component necessary for carrying out the present invention, such as buffers, pipettes, microtiter plates and written instructions. Such other components for the kits of the invention are known to the skilled person.

Having now generally described the invention, the same will be more readily understood through reference to the following example which is provided by way of illustration and is not intended to be limiting of the present invention.

**Example 1**

**Use of MAD7 to generate targeted mutations in tomato**

Constructs

The sequences of the type-V CRISPR nucleases MAD7 and *As*Cpf1, which have been optimized for translation in tomato, are presented by SEQ ID NO: 3 and SEQ ID NO: 5 for the protein sequence of MAD7 and *As*Cpf1, respectively, and by SEQ ID NO: 4 and SEQ ID NO: 6 for the nucleotide sequence encoding the Mad7 and *As*Cpf1 protein, respectively. These ORFs were synthesized and then cloned in a vector behind the constitutive 35S promoter for expression in plant cells. The sequences of the crRNA cassettes used (U6 promoter + crRNA) are shown in table 1. These were also synthesized and cloned into a vector. All of the plasmid constructs were introduced into *E. coli* and then transformed colonies were used to inoculate 50ml cultures. After overnight growth plasmid DNA was isolated from these cultures using standard methods.

Table 1: Sequences of the crRNA cassettes. *Sequences of the crRNA cassettes are represented herein by SEQ ID NO: 7-20. The Arabidopsis U6 promoter sequence is underlined, the non-variant crRNA sequence is shown in bold and the targeting sequence is shown in italics. Each crRNA is followed by seven thymine bases that act as a polIII terminator. Target site 1 and 2 are in PDS1, target site 3 and 4 are in LIN5, target site 5 and 6 are in eIF4e and target site 7 is in ETR1.*

| **SEQ ID NO:** | **Description** | **Sequence** |
|---|---|---|
| SEQ ID NO: 7 | KG10731 (crRNA 1 for guiding MAD7 to target site 1) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***AC TTCTGAGGTTTGTGGATCT* TTTTTTT |
| SEQ ID NO: 8 | KG10732 (crRNA 2 for guiding MAD7 to target site 2) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***AG TTCCCAAAGAAGACGACCT* TTTTTTT |
| SEQ ID NO: 9 | KG10733 (crRNA 3 for guiding MAD7 to target site 3) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***AT CGTCAGGTAATACATCGGA* TTTTTTT |
| SEQ ID NO: 10 | KG10734 (crRNA 4 for guiding MAD7 to target site 4) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***GC CATACTTGTCGCGGAATAC* TTTTTTT |

| | | |
|---|---|---|
| SEQ ID NO: 11 | KG10735 (crRNA 5 for guiding MAD7 to target site 5) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***TG GACCAAGAATGCTGCAAAT* TTTTTTT |
| SEQ ID NO: 12 | KG10736 (crRNA 6 for guiding MAD7 to target site 6) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***AT TTGCAGCATTCTTGGTCCA* TTTTTTT |
| SEQ ID NO: 13 | KG10737 (crRNA 7 for guiding MAD7 to target site 7) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**GTCAAAAGACCTTTTTAATTTCTACTCTTGTAGAT***TC CATTCCAGTGGAGTTGATA* TTTTTTT |
| SEQ ID NO: 14 | KG9943 (crRNA 1.1 for guiding *As*Cpf1 to target site 1) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***ACTTCTGAGGTTTGTGG ATCT*TTTTTTTT |
| SEQ ID NO: 15 | KG9942 (crRNA 2.1 for guiding *As*Cpf1 to target site 2) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***AGTTCCCAAAGAAGACG ACCT*TTTTTTTT |
| SEQ ID NO: 16 | KG10533 (crRNA 3.1 for guiding *As*Cpf1 to target site 3) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***ATCGTCAGGTAATACAT CGGA*TTTTTTT |
| | | |
| SEQ ID NO: 17 | KG10106 (crRNA 4.1 for guiding *As*Cpf1 to target site 4) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***GCCATACTTGTCGCGGA ATAC*TTTTTTT |

| | | |
|---|---|---|
| SEQ ID NO: 18 | KG10101 (crRNA 5.1 for guiding *As*Cpf1 to target site 5) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***TGGACCAAGAATGCTGC AAAT*TTTTTTTT |
| SEQ ID NO: 19 | KG10102 (crRNA 6.1 for guiding *As*Cpf1 to target site 6) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***ATTTGCAGCATTCTTGG TCCA*TTTTTTTT |
| SEQ ID NO: 20 | KG10104 (crRNA 7.1 for guiding *As*Cpf1 to target site 7) | GGAGTGATCAAAAGTCCCACATCGATCAGGTGATATAT AGCAGCTTAGTTTATATAATGATAGAGTCGACATAGCGA TT**TAATTTCTACTCTTGTAGAT***TCCATTCCAGTGGAGTT GATA*TTTTTTTT |

Tomato protoplast isolation and transfection

*In vitro* shoot cultures of *Solanum lycopersicon* var Moneyberg were maintained on MS20 medium with 0.8% agar in high plastic jars at 16/8 h photoperiod of 2000 lux at 25°C and 60-70% RH. Young leaves (1 g) were gently sliced perpendicularly to the mid nerve to ease the penetration of the enzyme mixture. Sliced leaves were transferred to the enzyme mixture (2% Cellulase Onozuka RS, 0.4% Macerozyme Onozuka R10 in CPW9M) and cell wall digestion was allowed to proceed overnight in the dark at 25ºC. The protoplasts were filtered through a 50 µm nylon sieve and were harvested by centrifugation for 5 minutes at 800 rpm. Protoplasts were resuspended in CPW9M (Frearson, 1973) medium and 3 mL CPW18S (Frearson *et al.* 1973. Developmental Biology, 33:130-137) was added at the bottom of each tube using a long-neck glass Pasteur pipette. Live protoplasts were harvested by centrifugation for 10 minutes at 800 rpm as the cell fraction at the interface between the sucrose and CPW9M medium. Protoplasts were counted and resuspended in MaMg (Negrutiu *et al.* 1987. Plant Molecular Biology, 8: 363-373) medium at a final density of $10^6$ per mL.

For the protoplast transfections 10µg of a MAD7 or *As*Cpf1 expression plasmid and 20µg of one of the sgRNA expressing plasmids were mixed with 500 µL (500000 protoplasts) of the protoplast suspension and 500 µL of PEG solution (400g/l poly(ethylene glycol) 4000, Sigma-Aldrich #81240; 0.1M $Ca(NO_3)_2$) was then added and the transfection was allowed to take place for 20 minutes at room temperature. Control samples were also produced by omitting one or both of the plasmids from the transfection. Then, 10 mL of 0.275 M $Ca(NO_3)_2$ solution was added and thoroughly, but gently mixed in. The protoplasts were harvested by centrifugation for 5 minutes at 800 rpm and

resuspended in 9M culture medium at a density of 0.5 x $10^6$ per ml and transferred to a 4cm diameter petri dish and an equal volume of 2% alginate solution (20g/l Alginate-Na (Sigma-Aldrich #A0682), 0.14g/l $CaCl_2.2H_2O$, 90g/l mannitol) was added. Then 1 ml aliquots (125000 transfected protoplasts) were spread over Ca-Agar plates (72.5g/l mannitol, 7.35g/l $CaCl_2.2H_2O$, 8g/l agar, pH5.8) and allowed to polymerize for 1 hour. The alginate disc containing the embedded protoplasts was then transferred to a 4cm tissue culture dish containing 4ml of K8p (Kao *et al.* 1975. Planta, 126: 105-110) culture medium. To determine the frequency of indel formation at the target sequences the disc of transfected protoplasts was removed from the dish after 48 hours, the alginate was dissolved, and the protoplasts were isolated by centrifugation. For the regeneration of calli, the protoplasts are incubated in the K8p medium for 21 days at 28°C in the dark. After this period the discs of transfected protoplasts are transferred to solid GM medium (Tan *et al.* 1987. Plant Cell Reports,6: 172-175) supplemented with 1 $mg.l^{-1}$ zeatin , 0.2 $mg.l^{-1}$ GA3 and 20nM chlorsulfuron. The discs are then transferred to fresh plates of the same GM medium every 3 weeks until the surviving calli are large enough to be picked with tweezers and are subsequently grown for genotyping on GM medium. Once calli containing mutations at the target site are identified then they are regenerated into fertile plants. Calli are maintained on GM medium without the herbicide until the first shoots develop. The shooting calli are then placed on MS medium supplemented with 2 $mg.l^{-1}$ zeatin and 0.1 $mg.l^{-1}$ IAA media. After some time the regenerated tomato plantlets can be excised and rooted on MS medium supplemented with 0.5 $mg.l^{-1}$ IBA before transfer to the greenhouse.

Genotyping protoplasts and calli

Total genomic DNA was isolated from tomato protoplasts (48 hrs post transfection) using the DNeasy Plant Mini Kit (Qiagen). This gDNA was then used in a PCR reaction to amplify either the target regions. The PCR products were then used as templates to generate a library from each sample which were then pooled and sequenced using a 126 nt paired run on the MiSeq platform (Illumina) . Each sample was identified using a unique 5 bp tag. After sequencing the reads derived from each sample were processed to identify the number and types of sequence changes present at the target site. To identify calli with mutations at the target site the transfected protoplasts are regenerated to calli approximately 3mm in diameter. A direct PCR kit (Phire Plant Direct PCR kit, Thermo Scientific) is then used together with gene specific primers to amplify the target sequence from each callus. The resulting PCR products are then sequenced or genotyped in a different way identify the calli that have mutations at the target site. These calli are then selected for regeneration.

Results

We tested the ability of the class V-CRISPR nuclease MAD7 to generate INDEL mutations in the genome plant cells by expressing the MAD7 protein and crRNAs targeting seven different sequences ectopically in tomato protoplasts. First a MAD7 ORF optimized for codon usage in tomato was constructed together with a nuclear localization signal (NLS) fused at the C terminus. This was then cloned behind the constitutive CaMV 35S promoter for expression in plant cells. The

MAD7 protein requires the presence of a PAM sequence adjacent to the targeted sequence, in this case TTTN, which is identical to the PAM of another class V-CRISPR nuclease *As*Cpf1 that has been reported to be effective in plant cells as well as other eukaryotic cell types. This common PAM sequence allows the design of crRNAs for both MAD7 and *As*Cpf1 containing exactly the same target sequence so that a direct comparison of the mutagenesis efficiencies produced by each nuclease can be performed. Vectors for the expression of either MAD7 or *As*Cpf1 were introduced into tomato mesophyll protoplasts together with a second vector expressing a crRNA driven by the *Arabidopsis thaliana* U6 promoter. In such a system the MAD7 or *As*Cpf1 mRNA and the crRNA will be expressed at high levels for a short period, 24-48 hours, at which point the introduced plasmids will become degraded by cellular nucleases and CRISPR reagents will disappear from the cell. While they are present they are able to find the specific target site in the genome and create INDEL mutations. The introduced plasmids rarely integrate into the genome and so this approach does not result in transgenic lines. The protoplasts were then cultured for 48 hours and then analyzed by sequencing for the presence of an INDEL at the target site.

In total, seven different target sites were investigated, located in the exons of 4 different genes in the tomato genome (*PDS, LIN5, eIF4e and ETR1*). As shown in Figure 1, the MAD7 class V-CRISPR nuclease was able to introduce INDEL mutations at all seven of the target sites at varying efficiencies. Surprisingly, the efficiency of INDEL formation when using MAD7 was in nearly all cases higher than when the same sequence was targeted using the *As*Cpf1 class V-CRISPR nuclease. Several publications have reported the use of *As*Cpf1 in plants and other eukaryotic cell types, but our data shows that MAD7 is unexpectedly superior and is therefore more suitable for mutagenesis experiments in plants. Table 2 summarizes the types of INDEL mutations that are introduced at the target sites by both MAD7 and *As*Cpf1. The size and position of the INDELs created by either nuclease are similar and in most cases would lead to the elimination of gene activity (null allele) due to the introduction of a frameshift in the coding sequence. In order to generate plants carrying MAD7 INDEL mutations, tomato mesophyll protoplasts are transfected with the MAD7 expression vector and a crRNA expressing vector for the appropriate target site and then grown to calli. These calli are then genotyped and the ones containing an INDEL mutation regenerated into intact plants.

Table 2. INDEL mutations introduced at the target sites by both MAD7 and *As*Cpf1. *For each sample the enzyme used (MAD7 or AsCpf1) and the crRNA is indicated together with the percentage of total reads that contain an indel. The upper sequence in each sample shows the unmodified target (underlined) with the PAM sequence is shown in bold. The top two most common indels produced in each sample are shown along with the percentage of reads in which they occur.*

| | MAD7 + KG10731 (crRNA 1), 2.85% indels | |
|---|---|---|
| SEQ ID NO: 38 | CG**TTTA**ACTTCTGAGGTTTGTGGATCTTTTA | |
| SEQ ID NO: 45 | CG**TTTA**ACTTCTGAGGTTT----------TA | 0.24% |
| SEQ ID NO: 46 | CG**TTTA**ACTTCTGAGGTTTGT---------A | 0.11% |

| | |
|---|---|
| SEQ ID NO: 38<br>SEQ ID NO: 47<br>SEQ ID NO: 48 | AsCpf1 + KG9943 (crRNA 1.1), 2.7% indels<br>CG**TTTA**ACTTCTGAGGTTTGTGGATCTTTTA<br>CG**TTTA**ACTTCTGAGGTTT----------TA 0.42%<br>CG**TTTA**ACTTCTGAGGTTT---------TTA 0.28% |
| SEQ ID NO: 39<br>SEQ ID NO: 49<br>SEQ ID NO: 50 | MAD7 + KG10732 (crRNA 2), 1.07% indels<br>AC**TTTC**AGTTCCCAAAGAAGACGACCTCGAGCTCC<br>AC**TTTC**AGTTCCCAAAGAAGACGA------GCTCC 0.34%<br>AC**TTTC**AGTTCCCAAAGAAGAC---CTCGAGCTCC 0.14% |
| SEQ ID NO: 39<br>SEQ ID NO: 51<br>SEQ ID NO: 52 | AsCpf1 + KG9942 (crRNA 2.2), 0.32% indels<br>AC**TTTC**AGTTCCCAAAGAAGACGACCTCGAGCTCC<br>AC**TTTC**AGTTCCCAAAGAAGACGA------GCTCC 0.06%<br>AC**TTTC**AGTTCCCAAAGAAGA---------GCTCC 0.03% |
| SEQ ID NO: 40<br>SEQ ID NO: 53<br>SEQ ID NO: 54 | MAD7 + KG10733 (crRNA 3), 1.9% indels<br>AA**TTTC**ATCGTCAGGTAATACATCGGATTCA<br>AA**TTTC**ATCGTCAGGTAATA--------TCA 0.18%<br>AA**TTTC**ATCGTCAGGTAATAC-----ATTCA 0.16% |
| SEQ ID NO: 40<br>SEQ ID NO: 55<br>SEQ ID NO: 56 | AsCpf1 + KG10533 (crRNA 3.1), 0.73% indels<br>AA**TTTC**ATCGTCAGGTAATACATCGGATTCA<br>AA**TTTC**ATCGTCAGGTAATAC--------CA 0.09%<br>AA**TTTC**ATCGTCAGGTAATA-----GATTCA 0.15% |
| SEQ ID NO: 41<br>SEQ ID NO: 57<br>SEQ ID NO: 58 | MAD7 + KG10734 (crRNA 4), 6.07% indels<br>GG**TTTA**GCCATACTTGTCGCGGAATACCTTG<br>GG**TTTA**GCCATACTTGTCG--------CTTG 0.97%<br>GG**TTTA**GCCATACTTGTCGC---------TG 0.51% |
| SEQ ID NO: 41<br>SEQ ID NO: 59<br>SEQ ID NO: 60 | AsCpf1 + KG10106 (crRNA 4.1), 0.94% indels<br>GG**TTTA**GCCATACTTGTCGCGGAATACCTTG<br>GG**TTTA**GCCATACTTGTCGC---------TG 0.18%<br>GG**TTTA**GCCATACTTGTCGC----------G 0.09% |
| SEQ ID NO: 42<br>SEQ ID NO: 61<br>SEQ ID NO: 62 | MAD7 + KG10735 (crRNA 5), 0.93% indels<br>GC**TTTG**TGGACCAAGAATGCTGCAAATGAA<br>GC**TTTG**TGGACCAAGAATG---------AA 0.24%<br>GC**TTTG**TGGACCAAGAATGCTG------AA 0.16% |
| SEQ ID NO: 42 | AsCpf1 + KG10101 (crRNA 5.1), 0.05% indels<br>GC**TTTG**TGGACCAAGAATGCTGCAAATGAA |

| | | |
|---|---|---|
| SEQ ID NO: 63 | GC**TTTG**TGGACCAAGA----------AA | 0.01% |
| SEQ ID NO: 64 | GC**TTTG**TGGACCAAGAATGCTG------AA | 0.03% |
| | MAD7 + KG10736 (crRNA 6), 0.88% indels | |
| SEQ ID NO: 43 | TG**TTTC**ATTTGCAGCATTCTTGGTCCACA | |
| SEQ ID NO: 65 | TG**TTTC**ATTTGCAGCATT---------CA | 0.15% |
| SEQ ID NO: 66 | TG**TTTC**ATTTGCAGCATTC---------- | 0.07% |
| | AsCpf1 + KG10102 (crRNA 6.1), 0.08% indels | |
| SEQ ID NO: 43 | TG**TTTC**ATTTGCAGCATTCTTGGTCCACA | |
| SEQ ID NO: 67 | TG**TTTC**ATTTGCAGCATTCTT-------A | 0.04% |
| SEQ ID NO: 68 | TG**TTTC**ATTTGCAGCATTCT--------- | 0.02% |
| | MAD7 + KG10737 (crRNA 7), 0.4% indels | |
| SEQ ID NO: 44 | TA**TTTC**TCCATTCCAGTGGAGTTGATATACT | |
| SEQ ID NO: 69 | TA**TTTC**TCCATTCCAGTGGAGTT-----ACT | 0.03% |
| SEQ ID NO: 70 | TA**TTTC**TCCATTCCAGTGGA-------TACT | 0.03% |
| | AsCpf1 + KG10104 (crRNA 7.1), 0.07% indels | |
| SEQ ID NO: 44 | TA**TTTC**TCCATTCCAGTGGAGTTGATATACT | |
| SEQ ID NO: 71 | TA**TTTC**TCCATTCCAGTGGA-------TACT | 0.02% |
| SEQ ID NO: 72 | TA**TTTC**TCCATTCCAGTGGA---------CT | 0.02% |

**Claims**

1. Method for targeted modification of DNA in a plant cell, comprising contacting the DNA with a crRNA-guided MAD7-nuclease.

2. Method according to claim 1, wherein the MAD7-nuclease comprises two catalytically active endonuclease domains.

3. Method according to claim 1, wherein the MAD7-nuclease comprises at least one catalytically inactive endonuclease domain.

4. Method according to any one of claims 1-3, wherein the MAD7-nuclease is fused to a functional domain, preferably, a deaminase domain.

5. Method according to any one of claims 1-4, wherein the MAD7-nuclease is introduced in the cell by transfecting the cell with a nucleic acid construct encoding said MAD7-nuclease.

6. Method according to any one of claims 1-4, wherein the MAD7-nuclease is introduced in the cell by transfecting the cell with the MAD7-nuclease.

7. Method according to any one of claims 1-6, wherein the crRNA is introduced in the cell by transfecting the cell with a nucleic acid construct encoding said crRNA.

8. Method according to any one of claims 1-6, wherein a crRNA is introduced in the cell by transfecting the cell with the crRNA, and wherein preferably the crRNA is chemically modified.

9. Method according to any one of claims 1-8, wherein the cell is further transfected with a template oligonucleotide, wherein preferably the template oligonucleotide is chemically modified.

10. Method according to any one of claims 1-8, wherein the cell is further transfected with a donor construct, wherein preferably the donor construct is chemically modified.

11. The method according to any one of the preceding claims, wherein the MAD7-nuclease, crRNA and/or optionally the template oligonucleotide or donor construct, are introduced into the plant cell using polyethylene glycol mediated transfection, preferably using an aqueous medium comprising PEG.

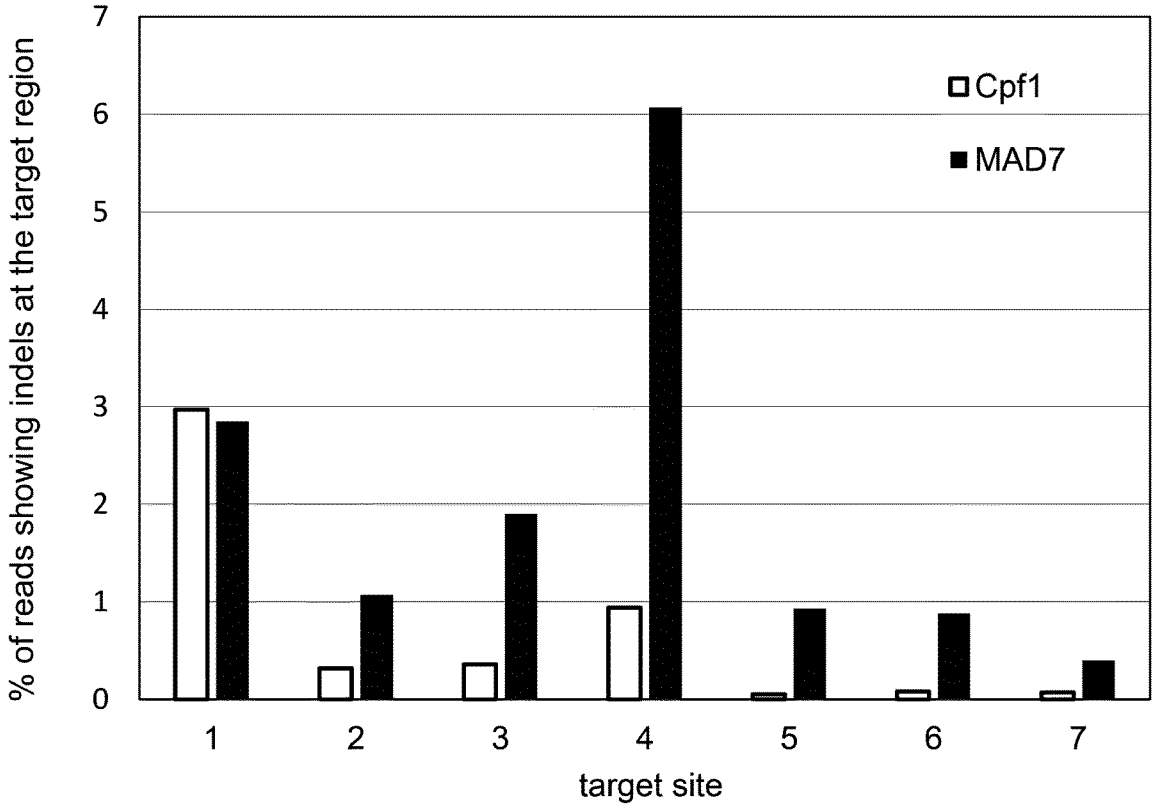
12. The method according to any one of the preceding claims, wherein the method further comprises the step of regenerating a plant or descendent thereof comprising the targeted modification.

13. A MAD7-endonuclease as defined in any one of claims 1-12, or one or more constructs encoding the same, for targeted modification of DNA in a plant cell, wherein preferably the MAD7-nuclease is complexed with a crRNA.

14. Kit for targeted modification of DNA in a plant cell comprising at least one of i) a container comprising the MAD7-endonuclease of claim 13; and
ii) a container comprising one or more constructs of claim 13,
and optionally a container comprising one or more crRNAs and/or constructs encoding the same.

15. Use of a crRNA guided MAD7-endonuclease as defined in claim 13, or one or more constructs encoding the same, or a kit as defined in claim 14, for targeted modification of DNA in a plant cell.

Fig. 1


% of reads showing indels at the target region
0
1
2
3
4
5
6
7
1
2
3
4
5
6
7
target site
Cpf1
MAD7

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/068839

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10 C12N15/82 C12N9/22
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C40B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | WO 2017/223538 A1 (THE REGENTS OF THE UNIV OF COLORADO A BODY CORPORATE [US]; MUSE BIOTEC) 28 December 2017 (2017-12-28) paragraphs [0052], [0120], [0195]; figure 33; example 32; table 4; sequence 7 ----- | 1-15 |
| X | US 9 982 279 B1 (GILL RYAN T [US] ET AL) 29 May 2018 (2018-05-29) cited in the application column 7, line 58; claims 1, 7,10, 15; example 6; tables 1,5; sequence 7 column 38, line 55 ----- | 1-15 |
| X | WO 2018/071672 A1 (UNIV COLORADO REGENTS [US]; INSCRIPTA INC [US]) 19 April 2018 (2018-04-19) paragraphs [0167], [0187], [0190]; table 1; sequence 2 ----- -/-- | 1-15 |

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 9 August 2019 | 09/09/2019 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Krüger, Julia |

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | ZAIDI SYED SHAN-E-ALI ET AL: "CRISPR-Cpf1: A New Tool for Plant Genome Editing", TRENDS IN PLANT SCIENCE, vol. 22, no. 7, 1 July 2017 (2017-07-01), pages 550-553, XP085093481, ISSN: 1360-1385, DOI: 10.1016/J.TPLANTS.2017.05.001 the whole document<br>----- | 1-15 |
| A | MALZAHN AIMEE ET AL: "Plant genome editing with TALEN and CRISPR", CELL & BIOSCIENCE, BIOMED CENTRAL LTD, LONDON, UK, vol. 7, 24 April 2017 (2017-04-24), pages 1-18, XP002785201, ISSN: 2045-3701, DOI: 10.1186/S13578-017-0148-4 abstract page 12, column 1, paragraph 3 - column 2, paragraph 1<br>----- | 1-15 |

1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2019/068839

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| WO 2017223538 A1 | 28-12-2017 | AU 2017280353 A1 | 24-01-2019 |
| | | CA 3029254 A1 | 28-12-2017 |
| | | CN 109688820 A | 26-04-2019 |
| | | EP 3474669 A1 | 01-05-2019 |
| | | JP 2019518478 A | 04-07-2019 |
| | | US 2017369870 A1 | 28-12-2017 |
| | | US 2018230460 A1 | 16-08-2018 |
| | | US 2018230461 A1 | 16-08-2018 |
| | | US 2019194650 A1 | 27-06-2019 |
| | | WO 2017223538 A1 | 28-12-2017 |
| US 9982279 B1 | 29-05-2018 | US 9982279 B1 | 29-05-2018 |
| | | US 2018371498 A1 | 27-12-2018 |
| WO 2018071672 A1 | 19-04-2018 | EP 3526326 A1 | 21-08-2019 |
| | | WO 2018071672 A1 | 19-04-2018 |