



(86) Date de dépôt PCT/PCT Filing Date: 2015/10/16

(87) Date publication PCT/PCT Publication Date: 2016/04/21

(45) Date de délivrance/Issue Date: 2022/01/11

(85) Entrée phase nationale/National Entry: 2017/04/13

(86) N° demande PCT/PCT Application No.: US 2015/055980

(87) N° publication PCT/PCT Publication No.: 2016/061481

(30) Priorité/Priority: 2014/10/17 (US62/065,093)

(51) Cl.Int./Int.Cl. *C12N 15/63* (2006.01),
C07H 21/02 (2006.01), *C12N 15/00* (2006.01),
C12N 15/11 (2006.01), *C12N 15/113* (2010.01),
C12N 15/82 (2006.01), *C12N 15/85* (2006.01),
C12N 5/10 (2006.01)

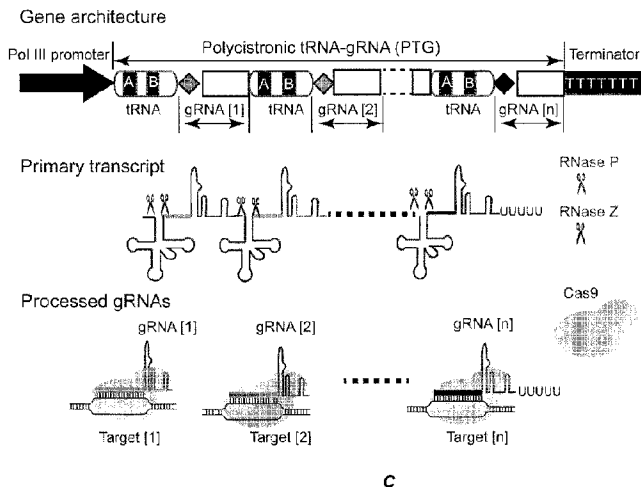
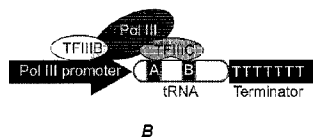
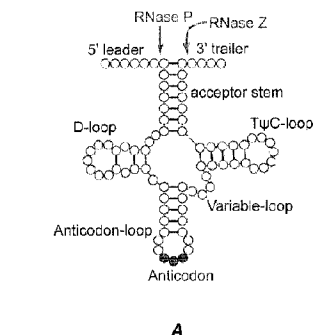
(72) Inventeurs/Inventors:
YANG, YINONG, US;
XIE, KABIN, US

(73) Propriétaire/Owner:
THE PENN STATE RESEARCH FOUNDATION, US

(74) Agent: BERESKIN & PARR LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : PROCÉDES ET COMPOSITIONS POUR L'ÉDITION GÉNOMIQUE MULTIPLEX À GUIDAGE ARN ET AUTRES TECHNOLOGIES D'ARN

(54) Title: METHODS AND COMPOSITIONS FOR MULTIPLEX RNA GUIDED GENOME EDITING AND OTHER RNA TECHNOLOGIES



(57) **Abrégé/Abstract:**

The invention includes materials and methods to generate numerous small RNAs from one polynucleotide construct (synthetic gene) to facilitate RNA-guided multiplex genome editing, modification, inhibition of expression and other RNA-based technologies. The synthetic gene/polynucleotide construct encodes polycistronic RNA components separated by tRNAs, and preferably also includes regulatory components such as a promoter or terminator to form an expression cassette. Once transcribed in a cell, the transcript is processed by the cell to multiple RNA molecules by the endogenous tRNA processing system. The system can be used for any RNA based gene manipulation method including RNA-mediated genome editing, artificial microRNA mediated gene silencing, small RNA mediated genetic manipulation, double-stranded RNA mediated gene silencing, antisense mechanisms and the like.

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
21 April 2016 (21.04.2016)

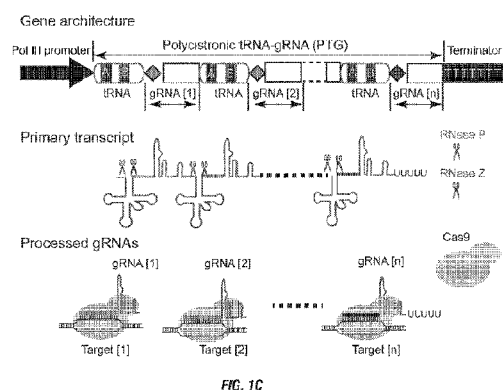
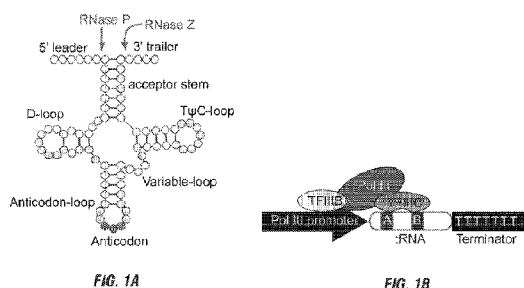


(10) International Publication Number
WO 2016/061481 A1

- (51) **International Patent Classification:**
C12N 15/113 (2010.01) *C12N 15/82* (2006.01)
C12N 15/11 (2006.01)
- (21) **International Application Number:**
PCT/US2015/055980
- (22) **International Filing Date:**
16 October 2015 (16.10.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/065,093 17 October 2014 (17.10.2014) US
- (71) **Applicant:** THE PENN STATE RESEARCH FOUNDATION [US/US]; 304 Old Main, University Park, Pennsylvania 16802 (US).
- (72) **Inventors:** YANG, Yinong; 1315 N. Foxpointe Drive, State College, Pennsylvania 16803 (US). XIE, Kabin; 10 Vairo Blvd., Apt. 214B, State College, Pennsylvania 16803 (US).
- (74) **Agents:** NEBEL, Heidi S. et al.; McKee, Voorhees & Sease, P.L.C., 801 Grand Avenue, Suite 3200, Des Moines, Iowa 50309-2721 (US).
- (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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,

[Continued on next page]

(54) **Title:** METHODS AND COMPOSITIONS FOR MULTIPLEX RNA GUIDED GENOME EDITING AND OTHER RNA TECHNOLOGIES



(57) **Abstract:** The invention includes materials and methods to generate numerous small RNAs from one polynucleotide construct (synthetic gene) to facilitate RNA-guided multiplex genome editing, modification, inhibition of expression and other RNA-based technologies. The synthetic gene/polynucleotide construct encodes polycistronic RNA components separated by tRNAs, and preferably also includes regulatory components such as a promoter or terminator to form an expression cassette. Once transcribed in a cell, the transcript is processed by the cell to multiple RNA molecules by the endogenous tRNA processing system. The system can be used for any RNA based gene manipulation method including RNA-mediated genome editing, artificial microRNA mediated gene silencing, small RNA mediated genetic manipulation, double-stranded RNA mediated gene silencing, antisense mechanisms and the like.

WO 2016/061481 A1



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, — *with sequence listing part of description (Rule 5.2(a))*
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

**TITLE: METHODS AND COMPOSITIONS FOR MULTIPLEX RNA GUIDED
GENOME EDITING AND OTHER RNA TECHNOLOGIES**

FIELD OF THE INVENTION

5 This invention relates to methods for gene targeting and genome editing in the field of molecular biology and genetic engineering. More specifically, the invention describes the use of the tRNA processing system to enable multiplex RNA based or guided genome engineering, gene inhibition etc.

10 **BACKGROUND OF THE INVENTION**

Methodologies for specific gene targeting or precise genome editing are of great importance to functional characterization of plant or animal genes. In recent years, sequence-specific nucleases have been developed to increase the efficiency of gene targeting or genome editing in plant, animal, or microbial systems.

15 Recently, a new genome editing tool has been developed in microbial and mammalian systems based on the cluster regularly interspaced short palindromic repeats (CRISPR)-associated nuclease system. The CRISPR-associated nuclease is part of adaptive immunity in bacteria and archaea. The Cas9 endonuclease, a component of *Streptococcus pyogenes* type II CRISPR/Cas system, forms a complex with two short RNA molecules
20 called CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA), which guide the nuclease to cleave non-self DNA on both strands at a specific site. The crRNA-tracrRNA heteroduplex could be replaced by one chimeric RNA (so-called guide RNA (gRNA)), which can then be programmed to targeted specific sites. The minimal constrains to program gRNA-Cas9 is at least 15-base-pairing between engineered 5'-RNA and targeted
25 DNA without mismatch, and an NGG motif (so-called protospacer adjacent motif or PAM) follows the base-pairing region in the targeted DNA sequence. Generally, 15-22 nt in the 5'-end of the gRNA region is used to direct Cas9 nuclease to generate DSBs at the specific site. The CRISPR/Cas system has been demonstrated for genome editing in human, mice,

zebrafish, yeast and bacteria. Distinct from animal, yeast, or bacterial cells to which recombinant molecules (DNA, RNA or protein) could be directly transformed for Cas9-mediated genome editing, recombinant plasmid DNA is typically delivered into plant cells via the *Agrobacterium*-mediate transformation, biolistic bombardment, or protoplast transformation due to the presence of cell wall. Thus, specialized molecular tools and methods need to be created to facilitate the construction and delivery of plasmid DNAs as well as efficient expression of Cas9 and gRNAs for genome editing in, plant or animal cells. Compositions and methods for making and using CRISPR-Cas systems are described in U.S. Patent No. 8,697,359, entitled "CRISPR-CAS SYSTEMS AND METHODS FOR ALTERING EXPRESSION OF GENE PRODUCTS."

In principle, multiplex genetic manipulation could be achieved by expressing multiple gRNAs with Cas9 (or Cas9 derived effectors) for corresponding target sites. Due to limitations in delivery method and capacity of current gRNA expressing devices, however, simultaneously producing numerous gRNAs in plant, animal, or microbial organisms is still a challenge.

It is an objective of the present invention to provide a strategy to use a cell's endogenous tRNA processing system as an efficient and precise approach to produce numerous gRNAs from a single polynucleotide construct/synthetic gene, thereby boosting the multiplex editing capability of CRISPR/Cas9 or other RNA mediated genetic manipulation tools for genome engineering.

In another aspect, the present invention relates to a polynucleotide construct, expression cassette, vector or recipient modified cell comprising a nucleic acid as described, *supra*.

It is a further objective, feature or advantage of the present invention to provide compositions and methods for manipulating multiple target genes, or multiple positions within a single target gene with introduction of a single expression cassette that includes sequences to utilize the tRNA system in the recipient cell.

Additional objectives, features and advantages may become obvious based on the

disclosure contained herein.

SUMMARY OF THE INVENTION

Applicants have produced materials and methods to generate numerous small
5 RNAs from one polynucleotide construct (synthetic gene) to facilitate RNA-guided
multiplex genome editing, inhibition of expression, genetic and epigenetic modification,
and other RNA-based technologies. The synthetic gene/polynucleotide construct encodes
polycistronic RNA components separated by tRNAs, and preferably also includes
regulatory components such as a promoter or terminator to form an expression cassette.
10 Once transcribed in a cell, the transcript is processed by the cell to multiple RNA
molecules by the endogenous tRNA processing system.

Methods of the invention can be used to improve any RNA-based genome editing
and modulation technology, allowing for multiplex RNA generation and separation via a
single polynucleotide construct and the endogenous tRNA processing system. Such
15 methods can include but are not limited to Cas9-mediated genome editing, artificial
microRNA mediated gene silencing, and other RNA-mediated mechanisms and the like.

According to the invention a method for producing multiplex RNA genetic
manipulation in a recipient cell includes, providing a heterologous polynucleotide
sequence with two or more RNA mediated genetic manipulation sequences in tandem
20 with one or more tRNA cleavage sequences (pre-tRNA). The polynucleotide is
introduced to the recipient cell by any of a number of standard techniques, and upon
expression, said recipient cells tRNA processing system cleaves the heterologous
polynucleotide at the tRNA sequences. In particular the invention provides tRNA
cleavage sequences for RNA-nucleolytic activity such as in pre-tRNA splicing, 3' end
25 pre-mRNA endonuclease activity, pre-tRNA cleavage activity, and/or the pre-ribosomal
RNA cleavage activity. The RNA genetic manipulation sequences and the RNA cleavage
sequence are operably linked to regulatory sequences for expression in a recipient cell,
such as promoter and terminator sequences.

A tRNA cleavage sequence includes any sequence and/or structural motif that actively interacts with and is cleaved by a cell's endogenous tRNA system such as RNase P, RNase Z and RNase E (bacteria). This can include structural recognition elements such as the acceptor stem, D-loop arm, T Psi C loop as well as specific sequence motifs. There are numerous tRNA active sequences and motifs known and available to those of skill in the art through sources such as the tRNA-SE program available at world wide web lowelab.ucsc.edu/tRNAscan-SE/ or world wide web trna.bioinf.uni-leipzig.de/DataOutput/Organisms (for all organisms), or world wide web at plantrna.ibmp.cnrs.fr/ (for plants). Numerous articles and Genbank resources are also available and are recited herein.

The invention also includes polynucleotide constructs for multiplex RNA genetic manipulation, expression constructs, vectors and recipient cells that have been manipulated. The constructs comprise a polynucleotide sequence which encodes two or more RNA mediated genetic manipulation sequences in tandem with a sequence which encodes a tRNA cleavage sequence. In one embodiment the RNA mediated genetic manipulation sequence includes a guide RNA for CAS9-mediated genome editing. The guide RNA, or gRNA is introduced along with an endonuclease, in this case the CRISPR associated (Cas) nuclease, Cas9. The guide RNA provides the scaffold and a spacer sequence complementary to the target site. In another embodiment the RNA genetic manipulation sequence is an interfering RNA such as an siRNA, or a microRNA sequence designed for gene silencing according to standard methods in the art. The construct also includes one or more tRNA cleavage sequences which may be actively processed by the endogenous tRNA cleavage system within the cell. The construct is preferably an expression construct or synthetic gene which includes promoter and terminator sequences. Surprisingly, the applicant have found that when used with a CRISPR system, the RNA cleavage components provide for up to 30 times higher expression likely due to internal tRNA promoter elements.

The invention also includes cells and progeny thereof which have been genetically

manipulated, such cells may include genomic insertions or deletions or manipulation of gene expression by gene silencing. Progeny of such cells may be used to develop lines, and plant and animal breeding materials. In some embodiments the cells are human cells.

While multiple embodiments are disclosed, still other embodiments of the present invention will become apparent to those skilled in the art from the following detailed description, which shows and describes illustrative embodiments of the invention. Accordingly, the drawings and detailed description are to be regarded as illustrative in nature and not restrictive.

10 DETAILED DESCRIPTION OF THE FIGURES

Figure 1. Engineering the endogenous tRNA system for multiplex genome editing with CRISPR/Cas9. (A) The eukaryotic pre-tRNA with 5' leader and 3' trailer is cleaved by RNase P and RNase Z at specific sites. (B) Transcription of tRNA gene with RNA polymerase III (Pol III). The box A and box B elements in the tRNA gene function as internal transcriptional elements and are bound by transcription factor IIIC (TFIIIC), which recruits TFIIIB and Pol III to start transcription. (C) Schematic depiction of PTG/Cas9 method for simultaneously targeting multiple sites. The synthetic *PTG* consists of tandemly arrayed tRNA-gRNA units with each gRNA containing target-specific spacer (labeled as diamonds) and conserved gRNA scaffold (rectangle). The tRNA containing box A and B elements is shown as round rectangles. The primary transcript of *PTG* is cleaved by endogenous RNase P and RNase Z (labeled as scissors) to release mature gRNAs and tRNA. The excised mature gRNAs direct Cas9 to multiple targets.

Figure 2. Precise excision of functional gRNAs *in vivo* from synthetic *PTG* genes. (A) The architecture of two *sgRNA* genes and four *PTGs* to produce one gRNA. (B) Sequence and predicted 2nd RNA structure of tRNA-gRNA-tRNA fusion of *PTG* gene (SEQ ID NO:87). The bases of tRNA region are shown and the tRNA 5'-leader is indicated in lowercase. The gRNA is shown and the gRNA spacer sequence is indicated as *N*. (C-F) Examination of mature gRNAs produced from *sgRNA* or *PTGs* with cRT-PCR. Total RNAs

from the protoplasts expressing empty vector were used as control (CK). Arrows indicate mature gRNAs amplified by cRT-PCR, and asterisks indicate the nonspecifically amplified rRNA. (G) Summary of excision sites in *PTG* (SEQ ID NO:88) according to mapped gRNA ends from cRT-PCR. Arrows indicate the cleavage sites in *PTG* to release gRNA.

5 The mature gRNA 5'-ends were excised from *PTG* exactly at the tRNA-gRNA fusion site in all cRT-PCR results whereas its 3'-ends shifted 1-4 nt within the tRNA 5' leader (lowercase). (H) gRNA produced from *U3p:sgRNA* (SEQ ID NO:89). All detected *U3p:sgRNA* produced gRNA started with ribonucleotide A and terminated with multiple Us.

(I) Introduction of indels at the desired sites by *PTG1:Cas9* or *PTG2:Cas9* in rice

10 protoplasts as shown by PCR/RE. Arrows indicate mutated fragments resistant to RE digestion. The indel frequency is indicated at the bottom. (J) Relative expression of *sgRNA1/2* and *PTG1/2* in rice protoplasts. Data represent mean $\pm$ SD. ND, not detected. CK, empty vector control.

Figure 3. Simultaneous editing of multiple genomic sites in rice protoplasts with the

15 *PTG:Cas9* system. (A) Architecture, gRNA components and targets of *PTGs*. (B) PCR detection of chromosomal fragment deletion at targeted loci in rice protoplasts expressing respective *PTGs* with Cas9. Successful deletion is shown as truncated PCR product (indicated with arrows). The chromosomal fragment deletion frequency (del %) is indicated at the bottom of each lane. The protoplast samples expressing an empty vector

20 were used as control (CK). (C) Representative sequences of chromosomal fragment deletion, including SEQ ID NOs 90-99, aligned with that of wild type (WT) (MPK1 WT, SEQ ID NO:79; MPK2 WT, SEQ ID NO:81; MPK5 WT, SEQ ID NO:82; MPK6 WT, SEQ ID NO:84). The gRNA paired region is highlighted. Dots in WT sequence indicate bases not shown and the total length between two Cas9 cut sites (labeled with scissor) is

25 indicated on top. Short lines indicate gaps in alignment.

Figure 4. Highly efficient targeted mutagenesis in transgenic rice expressing *PTG:Cas9*. (A) Chromosomal fragment deletion in *PTG7:Cas9* plant at T₀ generation (SEQ ID NOs 100-101). Of note, only 5 *mpk1* with 358 bp deletion (Δ 358) was detected in

genomic DNA. (B) Sequence analysis of targeted deletions at MPK1 and MPK5 loci in PTG7:Cas9 plants. All sequenced PCR products (the number in parentheses) show identical deletion pattern. (C) Albino seedlings were regenerated from calli transformed with *PTG10:Cas9*. All T₀ seedlings (n=12) exhibited a similar photo-bleach phenotype, suggesting a very high efficiency of knocking out PDS with *PTG10:Cas9*. Vec, control plants transformed with empty vector; Bar= 5cm.

Figure 5. Schematic diagram of rice MAPK genes (SEQ ID NOs 102-109) targeted by gRNA1-gRNA8 (SEQ ID NOs 3-10). (A) MPK5 target gene. (B) MPK1 target gene. (C) MPK2 target gene. (D) MPK6 target gene. The rectangles indicate exons while filled rectangles indicate protein coding region. The relative location of gRNA targeting site is shown as a vertical line and PCR primers for genotyping are indicated with an arrow. The restriction enzyme sites for PCR/RE assay are also shown for each gene. The shaded bar in (A) indicates the qPCR amplicon for deletion efficiency estimation. Only a partial region of MPK1 locus is shown here.

Figure 6. Illustration of pRGE32 and pRGE32 plasmids used in this study. (A and B) Schematic depiction of pRGE32 and pRGE32 plasmid vectors. The pRGE32 vector was used for transient expression of sgRNAs and PTGs along with Cas9 in plant protoplasts, and pRGE32 is a binary vector for the Agrobacterium-mediated transformation. (C) The cloning site (SEQ ID NOs 110-111) for insertion of gRNA spacer sequence or PTG genes into both vectors. Italic letters indicate overhangs in linearized vectors after BsaI digestion. The appropriate overhangs of oligo-duplex or synthetic PTG genes are shown at the bottom. Amp, ampicillin resistance gene; Kan, kanamycin resistance gene; 35S, cauliflower mosaic virus 35S promoter; UB1p, rice ubiquitin promoter; U3p, rice U3 snoRNA promoter; HPT II, hygromycin phosphotransferase II; Ter, nopaline synthase terminator; LB, T-DNA left border; RB, T-DNA right border.

Figure 7. Schematic depiction of cRT-PCR to map gRNA 5' and 3' ends. The primary PTG transcript or processed mature gRNA could be circularized by self-ligation with T4 RNA ligase. The circular RNA was reverse-transcribed to cDNA with gRNA

spacer-specific primers (Rev). Then the fragment containing 5'- and 3'-end junction could be amplified with a pair of specific primers (Fwd and Rev) to obtain sequences with mature gRNA extremities. In the cRT-PCR, the linear RNA fraction could not be amplified (right panel) because the primers were not paired.

5 Figure 8. Sequences of cRT-PCR products from *sgRNA1* (SEQ ID NOs 112-114), *sgRNA2* (SEQ ID NO:116), *PTG1* (SEQ ID NO:115) and *PTG2* (SEQ ID NO:117). An example of sequence chromatographs is shown for each cRT-PCR product. The number of individual colonies with the identical sequences is shown in parentheses. The sequencing results indicate that variable length of poly(U) tails were added to the 3'-end of gRNAs
10 produced by these genes.

 Figure 9. Sequences of cRT-PCR products from *PTG1.1* (SEQ ID NO:118) and *PTG2.1* (SEQ ID NO:119). (A and B) Representative sequencing chromatographs of cRT-PCR products. (C) Summary of the mapped 3'-end of mature gRNAs from *PTG1.1* and *PTG2.1*. All sequences show identical 1st nt of gRNA 5'-end. The number of
15 individually cloned PCR products with the identical sequences is shown in parentheses. Dots indicate bases not shown in the alignment (nucleotides 107-189 of SEQ ID NO:18 or SEQ ID NO:19).

 Figure 10. Strategy to synthesize PTGs with one step (level 1) Golden Gate (GG) assembly. (A) Schematic guide to design gRNA spacer specific primers with 4 bp
20 overlapping for Golden Gate assembly. The primers could be overlapped on any 4 consecutive nucleotides within the spacer. (B) Mechanism to generate a complete gRNA spacer during GG assembly. After PCR amplification and BsaI digestion, the 4 bp overlapped sequence in the gRNA spacer was generated as overhangs to ligate two parts and the resulting ligation product would produce a complete gRNA without extra
25 nucleotides. The DNA sequences in the box indicate BsaI cut site. (C) Schematic diagrams for level 1 GG assembly to synthesize PTGs from PCR parts and clone them into plasmid vectors (pRGE32 or pRGE32). A PTG with $n-1$ gRNAs are divided into n parts (Part[1] — Part[n], see the bottom). Each part was amplified with spacer-specific primers

containing BsaI adaptor, except two terminal parts using gRNA spacer primer and terminal specific primers containing Fok I site (L5AD5-F and L3AD5-R). These PCR parts were ligated together using GG assembly to produce the PTG with complete gRNA spacers. The assembled product was amplified with short terminal specific primers (S5AD5-F and S3AD5-R). After Fok I digestion, the PTG fragment was inserted into the BsaI digested pRGE32 and pRGE32.

Figure 11. Strategy to synthesize a large PTG from two PTG parts using level 2 Golden Gate (GG) assembly. Two PTG genes (PTG[1] and PTG[2]) containing different numbers (2–6) of gRNAs were synthesized by level 1 GG assembly. To concatenate PTG[1] with PTG[2], the last gRNA (bridge gRNA) in PTG[1] should contain the same spacers as the first gRNA in PTG[2]. Then level 2 GG parts were amplified from PTG[1] and PTG[2] with bridge gRNA spacer specific primers (Ln-gR-F and Ln-gR-R) containing BsaI site and terminal specific primers (S5AD5-F and S3AD5-R). After Fok I digestion, these two PTG parts were ligated together with GG assembly and inserted into BsaI digested pRGE32 or pRGE32 vector.

Figure 12. Targeted mutagenesis at MPK5 locus in transgenic plants expressing sgRNA:Cas9. Targeted DNA fragment was amplified from independent transgenic plants and analyzed with PCR/RE assays, and a wild type plant (WT) was included as control. Kpn I and SacI were used to examine mutation at sgRNA1/gRNA1 (A) and sgRNA2/gRNA2 (B) targets, respectively. On the agarose gels, mutated DNA fragments resistant to KpnI or SacI digestion are indicated by lighter arrow whereas digested WT DNA fragments are indicated by darker arrow. The indel frequency, which was estimated based on the intensity of digested and undigested bands, is indicated at the bottom of each lane. The putative biallelic mutants are indicated by asterisk.

Figure 13. Targeted mutagenesis at MPK5 locus in transgenic plants expressing PTG6:Cas9. Mutations at MPK5 locus of PTG6:Cas9 plants were examined with PCR/RE assay. The putative biallelic mutants at gRNA1 or gRNA2 site are indicated with asterisk, and biallelic mutations at both sites are indicated with vertical rectangle. In the gel,

mutated DNA fragments resistant to restriction enzyme (KpnI or SacI) digestion are indicated by lighter arrow whereas digested WT DNA fragments are indicated by darker arrow. The indel frequency, which was estimated based on the intensity of digested and undigested bands, is indicated at the bottom of each lane.

5 Figure 14. Targeted mutagenesis at MPK5 locus in transgenic plants expressing PTG7:Cas9. The mutations at gRNA1 and gRNA2 targets within MPK5 were examined by PCR/RE assays. The putative biallelic mutations at gRNA1 or gRNA2 site are indicated with asterisk, and biallelic mutations at both sites are marked with vertical rectangle. The deletion of a chromosomal fragment between gRNA1 and gRNA2 is indicated by the
10 arrows on the gels. Mutated DNA fragments resistant to restriction enzyme digestion are indicated by lighter arrow whereas digested WT DNA fragments are indicated by darker arrow. The indel frequency is indicated at the bottom of each lane.

Figure 15. Targeted mutagenesis at five genomic sites in transgenic plants expressing PTG9:Cas9. The mutations at the targeting sites of gRNA1/2/3/5/7 were examined by
15 PCR/RE assays. The transgenic lines carrying potential biallelic mutations at each target are indicated with asterisk, and biallelic mutations at all five sites are marked with vertical rectangles. The deletion of chromosomal fragments between paired gRNAs is indicated by the arrows on the gels. Mutated DNA fragments resistant to restriction enzyme digestion are indicated with lighter arrows whereas digested WT DNA fragments are indicated with
20 darker arrow. The indel frequency and fragment deletion frequency in each line are indicated at the bottom. Of note, there are two EcoN I sites within the *MPK2* amplicon and therefore the digestion of WT DNA fragment yielded three bands.

Figure 16. Targeted mutation at eight genomic sites in four MPK loci in PTG9:Cas9 lines. The mutated sequences from three transgenic lines (#1, #10 and #11) expressing
25 PTG9:Cas9 were aligned with wild type sequences. Insertion or substitution is shown in lowercase. Short line (-) indicates base pair deletion and dot (.) indicates sequences not shown in the alignments. The Cas9/gRNA cut sites are indicated with scissor. MPK1 WT (SEQ ID NO:79); MPK1 #1 (SEQ ID NO:120); MPK1 #10 (SEQ ID NO:121); MPK1 #11

(SEQ ID NO:122); MPK2 WT (SEQ ID NO:81); MPK2 #1 (SEQ ID NO:123); MPK2 #10 (SEQ ID NO:124); MPK2 #11 (SEQ ID NO:125); MPK5 WT (SEQ ID NO:82); MPK5 #1 (SEQ ID NO:126); MPK5 #10 (SEQ ID NO:127); MPK5 #11 (SEQ ID NO:128); MPK6 WT (SEQ ID NO:84); MPK6 #1 (SEQ ID NO:129); MPK6 #10 (SEQ ID NO:130); and
5 MPK6 #11 (SEQ ID NO:131).

Figure 17. Targeted mutation of rice *PDS* gene in transgenic plants with PTG10:Cas9. (A) Schematic depiction of rice *PDS* locus. Rectangles represent coding region exons. The targeting sites of gRNA9 and gRNA10 and the location of primers (F and R) for genotyping are indicated with vertical line and arrow, respectively. (B) Targeting sequences of gRNA9 and gRNA10 are highlighted. (C) Gene architecture of *PTG10*. (D) PCR
10 amplification of targeted DNA sites from transgenic plants expressing *PTG10:Cas9*. The deletion of chromosomal fragment between gRNA1 and gRNA2 is indicated with an arrow. (E) Examples of mutation at the *PDS* target sites (SEQ ID NO: 133 & 134).

Figure 18. Schematic map of the pSico-sgRNAs-EF1a-mCherry-T2A-Cas9 vector. The guide sequence or PTG genes could be expressed after inserting into the *Bbs*I sites of this
15 vector.

Figure 19. Mature gRNAs with desired guide sequences were precisely produced from hPTG genes. (A) Electrophoresis of cRT-PCR products in an acrylamide gel. Arrows indicate the mature gRNAs produced from hPTG1 and hPTG2. Vec, empty vector control. (B) Chromatography of mature gRNA sequences as revealed by cRT-PCR and DNA
20 sequencing. The arrow indicates the first nucleotide at the 5' -end of mature gRNAs. (C) The mapped cleavage site in hPTG1 and hPTG2 according to cRT-PCR results. The scissor indicates the cleaved site of the tRNA processing system. Lowercase letter, gRNA scaffold sequence; underlined letter, Pol III terminator.

Figure 20. Targeted deletion of chromosomal fragment at HDAC loci. (A and B) chromosomal fragment deletion were revealed by truncated PCR products (indicated by arrows) at respective HDAC loci in human cells expressing the Cas9-hPTG constructs. (C) Deletion efficiency of hPTG constructs expressing different number of gRNAs.
25

Figure 21. DNA sequences of the amplicons with the chromosomal deletions at HDAC
30 gene loci.

Figure 22. Schematic diagram of a polycistronic tRNA-gRNA gene (PTG) processing

system and design of PTGs to target four rice *MPKs*. A: A polycistronic tRNA-gRNA gene (PTG) consists of tandemly arrayed tRNA-gRNA repeats with a Pol III promoter to initiate transcription that is stopped by a Pol III terminator. The cell's endogenous RNaseP and RNaseZ can recognize the secondary structure of the tRNAs within the primary transcript and splice multiple functional gRNAs from the transcript. The processed gRNAs can complex with the Cas9 endonuclease to simultaneously target multiple independent genomic sites. B: Eight gRNAs were designed to target four closely-related rice *MPK* genes for gene knock-out. Chromosomal deletion could occur on each gene in addition to small insertion and deletion of nucleotides by non-homologous end-joining because each gene is targeted by two gRNAs. C: A total of eight PTG co-expressing one to eight gRNAs were used in this study. PTG2 and PTG6 target *MPK5*. PTG3, PTG4, and PTG5 target *MPK1*, *MPK2*, and *MPK6*, respectively. PTG7 and PTG8 target *MPK1/MPK5* and *MPK2/MPK6* to create double mutants. PTG9 co-expressed eight gRNAs to simultaneously target all four *MPK* genes.

Figure 23. T₁ generation rice plants of four PTGb9 lines carry eight biallelic mutations in four genes. A: Detection of mutations on five sites via PCR-RE assay. The PCR product of mutated DNA cannot be digested by the appropriate restriction enzyme (RE; underlined in B). PCR products of the corresponding genes in all four T₁ plants are resistant to digestion (arrow) and scissors indicate indel mutations at these sites. Note that *MPK2* possesses two EcoNI sites and digestion of PCR product from mutated DNA yields two bands instead of three bands in wildtype DNA control. CK: wildtype DNA control. B: Exact indel mutations at all eight targeted sites in the T₁ plants of four PTGb9 lines. Sequences were obtained by direct sequencing of the PCR product or TA clones in case of *mpk6* sequences to plant 1-2. Smaller case letters a and b behind the plant number indicate different alleles. The sequences show that the RE sites (underline) and all remaining target sites are mutated. The scissors indicate the predicted DSB site 3 bp downstream the PAM. The protospacer sequences used for targeting are highlighted. Dashes indicate deleted nucleotides when compared to wildtype and lower case letters indicate inserted or

substituted nucleotides. MPK1 WT (SEQ ID NO:79); MPK1 1-2a (SEQ ID NO:135);
 MPK1 1-2b
 (SEQ ID NO:136); MPK1 2-1 (SEQ ID NO:137); MPK1 3-2a (SEQ ID NO:138); MPK1
 3-2b (SEQ ID NO:139); MPK1 4-2 (SEQ ID NO:140); MPK2 WT (SEQ ID NO:81);
 5 MPK2 1-2 (SEQ ID NO:141); MPK2 2-1 (SEQ ID NO:142); MPK2 3-1 (SEQ ID
 NO:143); MPK2 4-2 (SEQ ID NO:144); MPK2 4-2b (SEQ ID NO:145); MPK5 WT (SEQ
 ID NO: 82); MPK5 1-2 (SEQ ID NO:146); MPK5 2-1 (SEQ ID NO:147); MPK5 3-2 (SEQ
 ID NO:148); MPK5 4-2 (SEQ ID NO:149); MPK6 WT (SEQ ID NO:84); MPK6 1-2a
 (SEQ ID NO:150); MPK6 1-2b (SEQ ID NO:151); MPK6 2-1 (SEQ ID NO:152); MPK6
 10 3-2 (SEQ ID NO:153); MPK6 4-2a (SEQ ID NO:154); MPK6 4-2b (SEQ ID NO:155).

Figure 24. PTG/Cas9 induced insertions and deletions and their frequency. Shown is a
 summary of the detected indel mutations and how frequent they were observed in a sample
 of 54 sequences from independent mutational events. The inserted box shows the
 calculated frequencies of all insertions versus all deletions.

15 Figure 25. Inheritance of chromosomal deletion in T₁ and T₂ generation plants. A:
 Detection of full length (*MPK5*) and chromosomal deletion (Δ) alleles on T₀ generation
 mutant plants. A chromosomal deletion at the targeted *MPK5* locus is shown by a truncated
 PCR product compared to a full length PCR product. Lines 5, 4, and 6 are heterozygous for
 the deletion as indicated by amplification of full and truncated PCR product. B: Inheritance
 20 of chromosomal deletion into the T₁ generation. Note that plant 4-2 and 6-1 each only
 inherited one of the both alleles and became homozygous. C: Inheritance of the fixed
 deletion into the T₂ generation. Only the truncated PCR product indicating deletion was
 detected in the progeny of plant 6-1. CK: wildtype DNA control.

Figure 26. Mutations in the *MPK1* and *MPK6* locus tend to preserve the open reading
 25 frame (ORF) of the sequence. A: Frequency of ORF-preserving mutations versus
 non-preserving mutations on eight targeted sites in four genes of PTGb9 T₁ plants. B:
 Frequency of ORF-preserving mutations when only considering target sites in exons. C:
 Predicted MPK1 protein sequence of the mutant alleles (lower case) (SEQ ID NOs

156-158) compared to the wildtype sequence (upper case) (SEQ ID NO:80). The mutations in *MPK1* delete one to five amino acids in the MPK1 protein but keep the predicted protein kinase domains intact, which start at amino acid 67 in the wildtype protein. D: Predicted MPK5 protein sequence of the mutant alleles (SEQ ID NOs 159-162) compared to the wildtype sequence (SEQ ID NO:83). Mutations in all alleles produce a premature stop of the protein sequence before the predicted protein kinase domains of the wildtype protein starting from amino acid 35. E: PTGb3 T₀ generation plants from the callus homozygous for chromosomal deletion on *MPK1* compared to an empty vector control. Mature plants 4-A and 4-B stayed severely dwarfed and sterile compared to the control. Scale bar is 5 cm in the main photo and 1 cm in magnified insert.

Figure 27. Schematic diagram showing that targeted sites of four *MPK* genes. The blank rectangles indicate the 5' untranslated regions of the first exon and the filled rectangles indicate the protein coding exons. The protospacer regions used as target sites for the gRNAs and the restriction enzyme sites used for detection of mutation events in PCR-RE assays are marked with vertical bars. Note that *MPK2* possesses two EcoNI sites and digestion of PCR product from mutated DNA yields two bands instead of three bands. The primers used for amplification of the target regions are marked with arrows.

Figure 28. Detection of mutagenesis in PTGb3 callus and PTGb5 plant lines. A: Successful and complete editing of *MPK1* in four independent transgenic callus pieces was confirmed by PCR-RE assay (arrow). Line 4 callus carries a chromosomal deletion on *MPK1* as shown by the truncated PCR product (Δ). B: The PCR-RE assay confirmed successful and complete editing of *MPK6* in PTGb5 T₀ generation plants as the PCR product is indigestible (arrow). CK: wildtype DNA control.

Figure 29. Detection of mutagenesis PTGb2 and PTGb4 plant lines. A: The PCR-RE assay confirmed successful and complete editing on the *MPK5* locus of PTGb2 transformed plants as PCR product from all nine independent lines was resistant to digestion with the appropriate enzyme (arrow). B: Similarly, the PCR-RE assays confirmed complete editing of *MPK2* of all thirteen examined lines of PTGb4. Note that *MPK2*

possesses two EcoNI sites and digestion of PCR product from mutated DNA yields two bands and that the unusual band pattern of lines 9,10, and 13 are result of longer deletions on the *MPK2* locus. C: Undigested PCR product from lines 9 to 13. The truncated PCR product (Δ) detected in lines 9, 10, and 13 indicate chromosomal deletions. CK: wildtype DNA control.

Figure 30. Detection of mutagenesis in PTGb8 lines. The digestion resistant band (arrows) on the *MPK6* and *MPK2* locus indicate lines with successful genome-editing. Note that the total number of lines with successful editing could theoretically be higher as only two out of four gRNA target sites could be examined by PCR-RE assay, and that *MPK2* possesses two EcoNI sites and digestion of PCR product from mutated DNA yields two bands. The boxes indicate transgenic lines that possess simultaneous mutations at both sites and are putative *MPK6/MPK2* double knock-out lines. CK: wildtype DNA control.

Figure 31. Transgene-free T₁ plants of PTGb9 lines. The presence of the genome-editing device (*U3:PTG* and a 1 kb fragment of Cas9) was screened in T₀ and T₁ plants of PTGb9 lines. While T₀ lines contain a copy of genome-editing device in their genome, self-pollination of T₀ lines allows the removal of the genome-editing device from the T₁ generation plants. Genomic DNAs from plants 2-1, 3-2, and 4-2 (arrows) do not contain any detectable fragment of *U3:PTG* or Cas9. The control PCR product was readily amplified from the endogenous rice gene LOC_Os05g49730, indicating that the DNAs used have sufficient quality for PCR. WT: wildtype DNA control.

Figure 32. Sequence of chromosomal deletion on *MPK5*. Shown is the sequencing result of the truncated fragment (Δ) from PTGb9 line 4 (SEQ ID NO:163). The DNA was precisely joined at the two predicted double strand break sites of the protospacers target sites (highlight) 3 bp downstream of the PAM producing a chromosomal deletion of 727 bp on the *MPK5* locus.

Figure 33. Inheritance pattern of the full-length and truncated alleles of *MPK5*. Eight T₁ plants from line 4, ten T₁ plants from line 5, and nine T₁ plants from lines 6 were screened for inheritance of the full-length (*MPK5*) or truncated *mpk5* (Δ) allele. The plants

were categorized into heterozygous (H) when containing both alleles, homozygous for full-length (F) if only the longer PCR product was detected, or homozygous for deletion (D) if only the truncated PCR product was detected. CK: wildtype DNA control.

5

DETAILED DESCRIPTION OF THE INVENTION

The technology provides a method to simultaneously manipulate multiple genes or DNA sequences based on RNA-guided genome editing or gene silencing through a single polynucleotide expression cassette. Generally, conventional genome editing or gene
10 silencing technology is capable of manipulating only one gene target. Current methods for multiplex genome editing involve multiple promoters/terminators and cumbersome cloning steps.

Applicant's invention provides a simple, precise and effective approach to produce multiple small RNAs (e.g., gRNA of Cas9, siRNA or miRNA) in vivo from one
15 polynucleotide construct (synthetic gene), and thus equips the current tools (like CRISPR/Cas9 system, RNAi) with the ability to manipulate multiple targets. As an general platform, this method also simplifies and improves the capability and fidelity of gRNA:Cas9/Cas9-nickase mediated genome editing, gRNA:deactive-Cas9 mediated gene activation/repression/labeling/epigenome editing, dsRNA-mediated gene silencing,
20 artificial microRNA and other small RNA mediated gene silencing.

In plants this technology can be used to improve crop yield, quality, stress tolerance, insect and disease resistance, and/or herbicide resistance by simultaneously manipulating multiple essential genes or quantitative trait loci (QTLs). In animals this technology can be used to increase fertility, health and well-being of animals and help in the development
25 of founder lines with improved traits such as weight gain, feed efficiency etc. In humans this technology can facilitate the virus-mediated multiple gRNA delivery because of its compact size and can be used to treat or ameliorate various disease states. Because this technology is based on the tRNA processing system that exists in all organisms, its

applications in genome engineering extends to microbes, animals and humans for industrial and medical biotechnology.

Definitions

5 Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. See, e.g., Sambrook *et al.* MOLECULAR
10 CLONING: A LABORATORY MANUAL, 2d ed., Cold Spring Harbor Laboratory Press, 1989; 3d ed., 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998;
15 METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P. M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P. B. Becker, ed.) Humana Press, Totowa, 1999.

The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used
20 interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (e.g.,
25 phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in

which one or more amino acids are chemical analogues or modified derivatives of a corresponding naturally-occurring amino acids.

"Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. Such interactions are generally characterized by a dissociation constant (K_d) of 10^{-6} M⁻¹ or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d .

10 A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

20 The term "sequence" refers to a nucleotide sequence of any length, which can be DNA or RNA; can be linear, circular or branched and can be either single-stranded or double stranded. The term "donor sequence" refers to a nucleotide sequence that is inserted into a genome. A donor sequence can be of any length, for example between 2 and 10,000 nucleotides in length (or any integer value there between or thereabove), preferably between about 100 and 1,000 nucleotides in length (or any integer there between), more preferably between about 200 and 500 nucleotides in length.

25 Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic

sequences can also be determined and compared in this fashion. In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more sequences (polynucleotide or amino acid) can be compared by
5 determining their percent identity. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981).
10 This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by
15 the Genetics Computer Group (Madison, Wis.) in the "BestFit" utility application. The default parameters for this method are described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, Wis.). A preferred method of establishing percent identity in the context of the present disclosure is to use the MPSRCH package of programs copyrighted by the
20 University of Edinburgh, developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, Calif.). From this suite of packages the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap extension penalty of one, and a gap of six). From the data generated the "Match" value reflects sequence identity. Other
25 suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10;

Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE;
Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS
translations+Swiss protein+Spuupdate+PIR. With respect to sequences described herein,
the range of desired degrees of sequence identity is approximately 80% to 100% and any
5 integer value therebetween. Typically the percent identities between sequences are at least
70-75%, preferably 80-82%, more preferably 85-90%, even more preferably 92%, still more
preferably 95%, and most preferably 98% sequence identity.

Alternatively, the degree of sequence similarity between polynucleotides can be
determined by hybridization of polynucleotides under conditions that allow formation of
10 stable duplexes between homologous regions, followed by digestion with
single-stranded-specific nuclease(s), and size determination of the digested fragments.
Two nucleic acid, or two polypeptide sequences are substantially homologous to each other
when the sequences exhibit at least about 70%-75%, preferably 80%-82%, more preferably
85%-90%, even more preferably 92%, still more preferably 95%, and most preferably 98%
15 sequence identity over a defined length of the molecules, as determined using the methods
above. As used herein, substantially homologous also refers to sequences showing
complete identity to a specified DNA or polypeptide sequence. DNA sequences that are
substantially homologous can be identified in a Southern hybridization experiment under,
for example, stringent conditions, as defined for that particular system. Defining
20 appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook et al.,
supra; Nucleic Acid Hybridization: A Practical Approach, editors B. D. Hames and S. J.
Higgins, (1985) Oxford; Washington, D.C.; IRL Press).

Selective hybridization of two nucleic acid fragments can be determined as follows.
The degree of sequence identity between two nucleic acid molecules affects the efficiency
25 and strength of hybridization events between such molecules. A partially identical nucleic
acid sequence will at least partially inhibit the hybridization of a completely identical
sequence to a target molecule. Inhibition of hybridization of the completely identical
sequence can be assessed using hybridization assays that are well known in the art (e.g.,

Southern (DNA) blot, Northern (RNA) blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low

5 stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is

10 chosen that is complementary to a reference nucleic acid sequence, and then by selection of appropriate conditions the probe and the reference sequence selectively hybridize, or bind, to each other to form a duplex molecule. A nucleic acid molecule that is capable of hybridizing selectively to a reference sequence under moderately stringent hybridization conditions typically hybridizes under conditions that allow detection of a target nucleic acid

15 sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe. Hybridization conditions useful for

20 probe/reference sequence hybridization, where the probe and reference sequence have a specific degree of sequence identity, can be determined as is known in the art (see, for example, *Nucleic Acid Hybridization: A Practical Approach*, editors B. D. Hames and S. J. Higgins, (1985) Oxford; Washington, D.C.; IRL Press).

Conditions for hybridization are well-known to those of skill in the art.

25 Hybridization stringency refers to the degree to which hybridization conditions disfavor the formation of hybrids containing mismatched nucleotides, with higher stringency correlated with a lower tolerance for mismatched hybrids. Factors that affect the stringency of hybridization are well-known to those of skill in the art and include, but are not limited to,

temperature, pH, ionic strength, and concentration of organic solvents such as, for example, formamide and dimethylsulfoxide. As is known to those of skill in the art, hybridization stringency is increased by higher temperatures, lower ionic strength and lower solvent concentrations.

5 With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of the sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization
10 solutions (e.g., dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

15 "Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of
20 either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

 A "chromosome," is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a
25 cell can comprise one or more chromosomes.

 A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist. For example, the sequence 5'-GAATTC-3' is a target site for the Eco RI

restriction endonuclease.

An "exogenous" or "heterologous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell

By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

"Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of a mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

"Modulation" of gene expression refers to a change in the activity of a gene. Modulation of expression can include, but is not limited to, gene activation and gene repression.

The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a

function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in cis with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

As used herein, the terms "coding region" and "coding sequence" are used interchangeably and refer to a nucleotide sequence that encodes a polypeptide and, when placed under the control of appropriate regulatory sequences expresses the encoded polypeptide. The boundaries of a coding region are generally determined by a translation start codon at its 5' end and a translation stop codon at its 3' end. A "regulatory sequence" is a nucleotide sequence that regulates expression of a coding sequence to which it is operably linked. Non-limiting examples of regulatory sequences include promoters, enhancers, transcription initiation sites, translation start sites, translation stop sites, and transcription terminators.

A polynucleotide that includes a coding region may include heterologous nucleotides that flank one or both sides of the coding region. As used herein, "heterologous nucleotides" refer to nucleotides that are not normally present flanking a coding region that is present in a wild-type cell. For instance, a coding region present in a wild-type microbe and encoding a Cas9 polypeptide is flanked by homologous sequences, and any other nucleotide sequence flanking the coding region is considered to be heterologous. Examples of heterologous nucleotides include, but are not limited to regulatory sequences. Typically, heterologous nucleotides are present in a polynucleotide disclosed herein through the use of standard genetic and/or recombinant methodologies well known to one skilled in the art. A polynucleotide disclosed herein may be included in a suitable vector.

As used herein, "genetically modified cell" refers to a cell, which has been altered

"by the hand of man." A genetically modified cell, includes a cell, callus, tissue, plant, or animal into which has been introduced an exogenous polynucleotide. Genetically modified cell, also refers to a cell that has been genetically manipulated such that endogenous nucleotides have been altered to include a mutation, such as a deletion, an insertion, a transition, a transversion, or a combination thereof. For instance, an endogenous coding region could be deleted. Such mutations may result in a polypeptide having a different amino acid sequence than was encoded by the endogenous polynucleotide. Another example of a genetically modified cell, callus, tissue, plant, or animal is one having an altered regulatory sequence, such as a promoter, to result in increased or decreased expression of an operably linked endogenous coding region.

Unless otherwise specified, "a," "an," "the," and "at least one" are used interchangeably and mean one or more than one.

Also herein, the recitations of numerical ranges by endpoints include all numbers subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.80, 4, 5, etc.).

As used herein, the term "tRNA cleavage sequence" refers to any nucleotide sequence recognized and excised by a tRNA splicing endonuclease, such as RNase P, RNase Z or RNase E.

As used herein, the term "tRNA splicing endonuclease" refers to the enzyme that is responsible for the recognition of the splice sites contained in precursor tRNA and the cleavage of the introns present in precursor tRNA. The archaeal tRNA splicing endonuclease recognizes the bulge-helix-bulge motif in archaeal precursor tRNA. The eukaryotic tRNA splicing endonuclease recognizes the splice sites contained in precursor tRNA by measuring the distance from the mature domain to the splice sites.

The expression "antisense RNA" denotes a RNA that is capable of binding to a target nucleic sequence (DNA or RNA) so as to limit or prevent its functioning; in particular, antisense RNAs are able to bind to a target messenger RNA in order to prevent its translation (see, for example, Tafech et al. (2006) Curr. Med. Chem. 13: 863-881).

The expression "interfering RNA" denotes an RNA capable of preventing or limiting

the expression of a target gene by the phenomenon of interference (see, for example, Tafech et al. (2006) Curr. Med. Chem. 13: 863-881).

5 Methods and Compositions of the Invention

tRNA

The general characteristics of a tRNA are well-known to the person skilled in the art. Preferably, a tRNA is formed of a single ribonucleotide chain which is capable of folding to adopt a characteristic, so-called cloverleaf secondary structure. This characteristic

10 secondary structure comprises:

(i) an acceptor stem composed of the first 7 ribonucleotides of the 5' end of the ribonucleotide chain and the 7 ribonucleotides that precede the last 4 ribonucleotides of the 3' end of the ribonucleotide chain, thus forming a double-stranded structure comprising 6 or 7 pairs of ribonucleotides, it being possible for the ribonucleotides constituted by the
15 first ribonucleotide of the 5' end of the ribonucleotide chain and the ribonucleotide that precedes the last 4 ribonucleotides of the 3' end of the ribonucleotide chain not to be paired;

ii) a D arm constituted by 4 pairs of ribonucleotides and a D loop constituted by 8 to 10 ribonucleotides, formed by the folding of a part of the ribonucleotide chain that follows
20 the first 7 ribonucleotides of the 5' end of the ribonucleotide chain;

(iii) a stem of the anticodon constituted by 5 pairs of ribonucleotides, and a loop of the anticodon constituted by 7 ribonucleotides (stem-loop of the anticodon), formed by the folding of a part of the ribonucleotide chain that follows the D arm and the D loop;

(iv) a variable loop constituted by from 4 to 21 ribonucleotides and formed by a part
25 of the ribonucleotide chain that follows the stem of the anticodon and the loop of the anticodon;

(v) a T arm constituted by 5 pairs of ribonucleotides, and a T loop constituted by 8 ribonucleotides, formed by the folding of a part of the ribonucleotide chain that follows the

variable loop and precedes the ribonucleotides of the 3' end of the ribonucleotide chain which are involved in the constitution of the acceptor stem.

Likewise preferably, from the 5' end in the direction towards the 3' end, 2
ribonucleotides are present between the first 7 ribonucleotides of the 5' end of the
5 ribonucleotide chain and the D arm and loop, 1 ribonucleotide is present between the D
arm and loop, on the one hand, and the stem and the loop of the anticodon, on the other
hand, and 1 ribonucleotide is present between the stem and the loop of the anticodon, on
the one hand, and the variable loop, on the other hand.

Still preferably, and according to the numbering well-known to the person skilled in
10 the art and defined by Sprinzl et al. (1998) "Compilation of tRNA sequences and sequences
of tRNA genes". Nucleic Acids Res. 26: 148-153, the tRNA comprises 17 ribonucleotides,
ensuring the three-dimensional structure of the tRNA and recognition by the cell enzymes,
namely: U.sub.8, A.sub.14, (A or G).sub.15, G.sub.18, G.sub.19, A.sub.21, G.sub.53,
U.sub.54, U.sub.55, C.sub.56, (A or G).sub.57, A.sub.58, (C or U).sub.60, C.sub.61,
15 C.sub.74, C.sub.75, A.sub.76. The indicated ribonucleotides correspond to the sequence of
the tRNA as transcribed before any post-transcriptional modifications of certain
ribonucleotides by the cell machinery.

In particular, the tRNA defined above may be selected from the group constituted by
Archean, bacterial, viral, protozoan, fungal, algal, plant or animal tRNAs.

20 The tRNAs which can be used according to the invention also include all the tRNAs
described by Sprinzl et al. (1998) "Compilation of tRNA sequences and sequences of tRNA
genes". Nucleic Acids Res. 26: 148-153.

In the context of the invention, the term "tRNA" also includes structures obtained by
modifying a tRNA as defined above or natural variants of a tRNA as defined above,
25 provided that those modified structures or those variants retain the functionalities of the
unmodified tRNA, namely especially the interaction with proteins such as EF-Tu factor
(see, for example, Rodnina et al. (2005) FEBS. Lett. 579: 938-942) or CCAse (see, for
example, Augustin et al. (2003) J. Mol. Biol. 328: 985-994). There are numerous

tRNA active sequences and motifs known and available to those of skill in the art through sources such as the tRNA-SE program. Numerous articles and Genbank resources are also available and are recited herein.

A substrate for a tRNA endonuclease of RNase may be produced by any method well-known to one of skill in the art. For example, the substrate may be chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (see, e.g., Caruthers et al., U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers et al., 1982, Genetic Engineering 4:1-17; User's Manual Model 392 and 394 Polynucleotide Synthesizers, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, et al., 1997, Biochemistry, 36:6033-6045). After synthesis, the substrate can be purified using standard techniques known to those skilled in the art (see Hwang et al., 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the substrate and the method of its synthesis, such purification techniques include, but are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification.

There are numerous RNase resources available to those of skill in the art including the following:

20 RNase E

1. Soderbom F, Svard SG, Kirsebom LA. 2005. RNase E cleavage in the 5' leader of a tRNA precursor. J Mol Biol 352(1): 22-27.
2. Nakajima N, Ozeki H, Shimura Y. 1981. Organization and structure of an E. coli tRNA operon containing seven tRNA genes. Cell 23(1): 239-249.
- 25 3. Li Z, Deutscher MP. 2002. RNase E plays an essential role in the maturation of Escherichia coli tRNA precursors. RNA 8(1): 97-109.

RNase P

1. Kirsebom LA. 2007. RNase P RNA mediated cleavage: substrate recognition and catalysis. *Biochimie* 89(10): 1183-1194.
2. Gutmann B, Gobert A, Giege P. 2012. PRORP proteins support RNase P activity in both organelles and the nucleus in Arabidopsis. *Genes Dev.* 26(10): 1022-1027.
- 5 3. Forster AC, Altman S. 1990. External guide sequences for an RNA enzyme. *Science* 249(4970): 783-786.
4. Yuan Y, Altman S. 1995. Substrate recognition by human RNase P: identification of small, model substrates for the enzyme. *EMBO J* 14(1): 159-168.
- 10 RNase Z
1. Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, Marchfelder A, Echeverria M. 2003. Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J* 22(3): 621-632.
2. Schiffer S, Rosch S, Marchfelder A. 2002. Assigning a function to a conserved group of
- 15 proteins: the tRNA 3'-processing enzymes. *EMBO J* 21(11): 2769-2777.
3. Barbezier N, Canino G, Rodor J, Jobet E, Saez-Vasquez J, Marchfelder A, Echeverria M. 2009. Processing of a dicistronic tRNA-snoRNA precursor: combined analysis in vitro and in vivo reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol* 150(3): 1598-1610.
- 20 4. Canino G, Bocian E, Barbezier N, Echeverria M, Forner J, Binder S, Marchfelder A. 2009. Arabidopsis encodes four tRNase Z enzymes. *Plant Physiol* 150(3): 1494-1502.

As used herein, the phrase "a substrate for a 3' end pre-mRNA endonuclease" refers to any nucleotide sequence recognized and excised by a 3' end pre-mRNA endonuclease. For example, a nucleotide sequence comprising a hexanucleotide with the sequence AAUAAA

25 upstream and a G/U-rich sequence element downstream of the cleavage site may be utilized as a substrate for 3' end pre-mRNA endonuclease in an assay described herein. A nucleotide sequence recognized and excised by a 3' end pre-mRNA endonuclease may comprise 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 25 nucleotides, 30

nucleotides, 40 nucleotides, 45 nucleotides, 50 nucleotides, 55 nucleotides, 60 nucleotides, 65 nucleotides, 75 nucleotides, 100 nucleotides, 125 nucleotides, 150 nucleotides, or more. In a specific embodiment, the substrates for 3' end pre-mRNA endonuclease utilized in the assays described herein comprise a cleavage and polyadenylation site.

5

Vectors and Nucleic Acids

A variety of nucleic acids may be introduced into cells for purposes of the invention. As used herein, the term nucleic acid includes DNA, RNA, and nucleic acid analogs, and nucleic acids that are double-stranded or single-stranded (i.e., a sense or an antisense single strand). Nucleic acid analogs can be modified at the base moiety, sugar moiety, or phosphate backbone to improve, for example, stability, hybridization, or solubility of the nucleic acid. Modifications at the base moiety include deoxyuridine for deoxythymidine, and 5-methyl-2'-deoxycytidine and 5-bromo-2'-doxycytidine for deoxycytidine. Modifications of the sugar moiety include modification of the 2' hydroxyl of the ribose sugar to form 2'-O-methyl or 2'-O-allyl sugars. The deoxyribose phosphate backbone can be modified to produce morpholino nucleic acids, in which each base moiety is linked to a six membered, morpholino ring, or peptide nucleic acids, in which the deoxyphosphate backbone is replaced by a pseudopeptide backbone and the four bases are retained. See, Summerton and Weller (1997) *Antisense Nucleic Acid Drug Dev.* 7(3):187; and Hyrup et al. (1996) *Bioorgan. Med. Chem.* 4:5. In addition, the deoxyphosphate backbone can be replaced with, for example, a phosphorothioate or phosphorodithioate backbone, a phosphoroamidite, or an alkyl phosphotriester backbone.

The nucleic acid sequence can be operably linked to a regulatory region such as a promoter. Regulatory regions can be porcine regulatory regions or can be from other species. As used herein, operably linked refers to positioning of a regulatory region relative to a nucleic acid sequence in such a way as to permit or facilitate transcription of the target nucleic acid.

Any type of promoter can be operably linked to a target nucleic acid sequence.

Examples of promoters include, without limitation, tissue-specific promoters, constitutive promoters, inducible promoters, and promoters responsive or unresponsive to a particular stimulus. Suitable tissue specific promoters can result in preferential expression of a nucleic acid transcript in beta cells and include, for example, the human insulin promoter. Other
5 tissue specific promoters can result in preferential expression in, for example, hepatocytes or heart tissue and can include the albumin or alpha-myosin heavy chain promoters, respectively. In other embodiments, a promoter that facilitates the expression of a nucleic acid molecule without significant tissue or temporal-specificity can be used (i.e., a constitutive promoter). For example, a beta-actin promoter such as the chicken beta-actin
10 gene promoter, ubiquitin promoter, miniCAGs promoter, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter, or 3-phosphoglycerate kinase (PGK) promoter can be used, as well as viral promoters such as the herpes simplex virus thymidine kinase (HSV-TK) promoter, the SV40 promoter, or a cytomegalovirus (CMV) promoter. In some embodiments, a fusion of the chicken beta actin gene promoter and the CMV enhancer is used as a promoter.
15 See, for example, Xu et al. (2001) Hum. Gene Ther. 12:563; and Kiwaki et al. (1996) Hum. Gene Ther. 7:821.

Additional regulatory regions that may be useful in nucleic acid constructs, include, but are not limited to, polyadenylation sequences, translation control sequences (e.g., an internal ribosome entry segment, IRES), enhancers, inducible elements, or introns. Such regulatory
20 regions may not be necessary, although they may increase expression by affecting transcription, stability of the mRNA, translational efficiency, or the like. Such regulatory regions can be included in a nucleic acid construct as desired to obtain optimal expression of the nucleic acids in the cell(s). Sufficient expression, however, can sometimes be obtained without such additional elements.

25 A nucleic acid construct may be used that encodes signal peptides or selectable markers. Signal peptides can be used such that an encoded polypeptide is directed to a particular cellular location (e.g., the cell surface). Non-limiting examples of selectable markers include puromycin, ganciclovir, adenosine deaminase (ADA), aminoglycoside phosphotransferase

(neo, G418, APH), dihydrofolate reductase (DHFR), hygromycin-B-phosphotransferase, thymidine kinase (TK), and xanthin-guanine phosphoribosyltransferase (XGPRT). Such markers are useful for selecting stable transformants in culture. Other selectable markers include fluorescent polypeptides, such as green fluorescent protein or yellow fluorescent
5 protein.

Nucleic acids can be incorporated into vectors. A vector is a broad term that includes any specific DNA segment that is designed to move from a carrier into a target DNA. A vector may be referred to as an expression vector, or a vector system, which is a set of components needed to bring about DNA insertion into a genome or other targeted DNA
10 sequence such as an episome, plasmid, or even virus/phage DNA segment. Vector systems such as viral vectors (e.g., retroviruses, adeno-associated virus and integrating phage viruses), and non-viral vectors (e.g., transposons) used for gene delivery in animals have two basic components: 1) a vector comprised of DNA (or RNA that is reverse transcribed into a cDNA) and 2) a transposase, recombinase, or other integrase enzyme that recognizes both
15 the vector and a DNA target sequence and inserts the vector into the target DNA sequence. Vectors most often contain one or more expression cassettes that comprise one or more expression control sequences, wherein an expression control sequence is a DNA sequence that controls and regulates the transcription and/or translation of another DNA sequence or mRNA, respectively.

20 Many different types of vectors are known. For example, plasmids and viral vectors, e.g., retroviral vectors, are known.

Polynucleotide constructs may be introduced into the genome of a desired host or recipient cell by a variety of conventional techniques. For reviews of such techniques see, for example, Weissbach & Weissbach *Methods for Cell, Molecular Biology* (1988,
25 Academic Press, N.Y.) Section VIII, pp. 421-463; and Grierson & Corey, *Cell Molecular Biology* (1988, 2d Ed.), Blackie, London, Ch. 7-9. For example, the DNA construct may be introduced directly into the genomic DNA of the cell using techniques such as electroporation and microinjection, or the DNA constructs can be introduced directly to

plant or animal tissue using biolistic methods, such as DNA particle bombardment (see, e.g., Klein et al (1987) Nature 327:70-73).

Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector.

5 *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch et al (1984) Science 233:496-498, and Fraley et al (1983) Proc. Nat'l. Acad. Sci. USA 80:4803. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the cellular DNA when the cell is
10 infected by the bacteria using binary T DNA vector (Bevan (1984) Nuc. Acid Res. 12:8711-8721) or the co-cultivation procedure (Horsch et al (1985) Science 227:1229-1231). See Hernalsteen et al (1984) EMBO J 3:3039-3041; Hooykass-Van Slogteren et al (1984) Nature 311:763-764; Grimsley et al (1987) Nature 325:1677-179; Boulton et al (1989) Cell Mol. Biol. 12:31-40; and Gould et al (1991) Cell Physiol.
15 95:426-434.

Alternative gene transfer and transformation methods include, but are not limited to, protoplast transformation through calcium-, polyethylene glycol (PEG)- or electroporation-mediated uptake of naked DNA (see Paszkowski et al. (1984) EMBO J3:2717-2722, Potrykus et al. (1985) Molec. Gen. Genet. 199:169-177; Fromm et al.
20 (1985) Proc. Nat. Acad. Sci. USA 82:5824-5828; and Shimamoto (1989) Nature 338:274-276) and electroporation of cell, plant or animal tissues (D'Halluin et al. (1992) Cell, 4:1495-1505). Additional methods for cell transformation include microinjection, silicon carbide mediated DNA uptake (Kaeppeler et al. (1990) Cell Reporter 9:415-418), and microprojectile bombardment (see Klein et al. (1988) Proc. Nat. Acad. Sci. USA
25 85:4305-4309; and Gordon-Kamm et al. (1990) Cell 2:603-618).

A transformed cell, callus, tissue plant or animal may be identified and isolated by selecting or screening the engineered cell, plant or animal material for traits encoded by the marker genes present on the transforming DNA. For instance, selection may be

performed by growing the engineered cell, on media containing an inhibitory amount of the antibiotic or herbicide to which the transforming gene construct confers resistance. Further, transformed cells may also be identified by screening for the activities of any visible marker genes (e.g., the β -glucuronidase, luciferase, B or C1 genes) that may be
5 present on the recombinant nucleic acid constructs. Such selection and screening methodologies are well known to those skilled in the art.

Physical and biochemical methods also may be used to identify transformants containing inserted gene constructs. These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the
10 recombinant DNA insert; 2) Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct; 4) protein gel electrophoresis, Western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the
15 gene construct products are proteins. Additional techniques, such as in situ hybridization, enzyme staining, and immunostaining, also may be used to detect the presence or expression of the recombinant construct in specific cell, plant or animal organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

Genomic changes including exogenous polynucleotides that are stably incorporated
20 into cells, or indels created by gene editing can be introduced into other cells, tissues, plants or animals using, for example, standard breeding techniques.

Genetically modified cells which are produced by any of the above transformation techniques can be cultured to regenerate a whole plant or animal tissues which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques
25 rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. plant or animal regeneration from cultured protoplasts is described in Evans, et al., "Protoplasts Isolation and Culture" in Handbook of

Cell, plant or animal Cell Culture, pp. 124-176, Macmillian Publishing Company, New York, 1983; and Binding, Regeneration of Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from callus, excell, organs, pollens, embryos or parts thereof. Such regeneration techniques are described generally in Klee et al (1987)
5 Ann. Rev. of Cell, plant or animal Phys. 38:467-486.

Nucleic acids introduced into a cell, can be used to confer desired traits on essentially any cell, plant or animal. A wide variety of cell, plant or animal systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation
10 methods mentioned above. In preferred embodiments, target cells, for engineering in plants can include, but are not limited to, those monocotyledonous and dicotyledonous cell, plant or animals, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops
15 (e.g., lettuce, spinach); flowering cell, plant or animals (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); cell, plant or animals used in phytoremediation (e.g., heavy metal accumulating cell, plant or animals); oil crops (e.g., sunflower, rape seed) and cell, plant or animals used for experimental purposes (e.g., Arabidopsis). Thus, the disclosed methods and compositions have use over a broad range
20 of cell, plant or animals, including, but not limited to, species from the genera Asparagus, Avena, Brassica, Citrus, Citrullus, Capsicum, Cucurbita, Daucus, Glycine, Hordeum, Lactuca, Lycopersicon, Malus, Manihot, Nicotiana, Oryza, Persea, Pisum, Pyrus, Prunus, Raphanus, Secale, Solanum, Sorghum, Triticum, Vitis, Vigna, and Zea. One of skill in the art will recognize that after the expression cassette is stably incorporated in transgenic
25 cell, plant or animals and confirmed to be operable, it can be introduced into other cell, plant or animals by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

Effects of gene manipulation using the methods disclosed herein can be observed by, for example, northern blots of the RNA (e.g., mRNA) isolated from the tissues of interest. Typically, if the amount of mRNA has increased, it can be assumed that the corresponding endogenous gene is being expressed at a greater rate than before. Other
5 methods of measuring gene activity can be used. Different types of enzymatic assays can be used, depending on the substrate used and the method of detecting the increase or decrease of a reaction product or by-product. In addition, the levels of and/or protein expressed can be measured immunochemically, i.e., ELISA, RIA, EIA and other antibody based assays well known to those of skill in the art, such as by electrophoretic detection
10 assays (either with staining or western blotting). The transgene may be selectively expressed in some tissues or at some developmental stages, or the transgene may be expressed in substantially all cells, substantially along its entire life cycle. However, any combinatorial expression mode is also applicable.

The present disclosure further encompasses the progeny, clones, cell lines or cells
15 of the transgenic cell described above wherein said progeny, clone, cell line or cell has the transgene or gene construct or has been edited.

Plasmid vectors for gene targeting and genome editing

According to one aspect of the invention, compositions are provided that allow
20 gene targeting and genome editing in cells. In one aspect, specific RNA-guided Genome Eediting vectors are provided. In a preferred embodiment, the vectors include a first regulatory element operable in a plant or animal cell operably linked to at least one nucleotide sequence encoding a CRISPR-Cas system guide RNA that hybridizes with the target sequence; and a second regulatory element operable in a plant or animal cell
25 operably linked to a nucleotide sequence encoding a Type-II CRISPR-associated nuclease. The nucleotide sequence encoding a CRISPR-Cas system guide RNA and the nucleotide sequence encoding a Type-II CRISPR-associated nuclease may be on the same or different vectors of the system. The guide RNA targets the target sequence, and the CRISPR-associated nuclease cleaves the DNA molecule, whereby expression of at least

one gene product is altered.

In a preferred embodiment, the vectors include a nucleotide sequence comprising a DNA-dependent RNA polymerase III promoter, wherein said promoter operably linked to a gRNA molecule and a Pol III terminator sequence, wherein said gRNA molecule includes a DNA target sequence; and a nucleotide sequence comprising a DNA-dependent RNA polymerase II promoter operably linked to a nucleic acid sequence encoding a type II CRISPR-associated nuclease. The CRISPR-associated nuclease is preferably a Cas9 protein.

In another embodiment, vectors are provided for the *Agrobacterium*-mediated transient expression or stable transformation in plant or animal tissues. In particular the plasmid vectors for transient expression in protoplasts, tissue cultures contain: (1) a DNA-dependent RNA polymerase III (Pol III) promoter (for example, rice snoRNA U3 or U6 promoter) to control the expression of engineered gRNA molecules in the cell, plant or animal cell, where the transcription was terminated by a Pol III terminator (Pol III Term), (2) a DNA-dependent RNA polymerase II (Pol II) promoter (e. g., 35S promoter) to control the expression of Cas9 protein; (3) a multiple cloning site (MCS) located between the Pol III promoter and gRNA scaffold, which is used to insert a 15-30 bp DNA sequence for producing an engineered gRNA. To facilitate the *Agrobacterium*-mediated transformation, binary vectors are provided (e.g., the pCAMBIA 1300 vector), wherein gRNA scaffold/Cas9 cassettes are inserted into the plant, animal and fungal cells via *Agrobacterium* transformation,. To program gRNA, a 15-30 bp long synthetic DNA sequence complementary to the targeted genome sequence can be inserted into the MCS site of the vector.

Methods to introduce engineered gRNA-Cas9 constructs into cells for genome editing and genetic modification.

According to another aspect of the invention, gene constructs carrying gRNA-Cas9 nuclease can be introduced into cells by various methods, which include but are not limited

to PEG- or electroporation-mediated protoplast transformation, tissue culture or cell, plant or animal tissue transformation by biolistic bombardment, or the *Agrobacterium*-mediated transient and stable transformation. The transformation can be transient or stable transformation.

5 Target gene sequences for genome editing and genetic modification can be selected using methods known in the art, and as described elsewhere in this application. In a preferred embodiment, target sequences are identified that include or are proximal to protospacer adjacent motif (PAM). Once identified, the specific sequence can be targeted by synthesizing a pair of target-specific DNA oligonucleotides with appropriate cloning
10 linkers, and phosphorylating, annealing, and ligating the oligonucleotides into a digested plasmid vector, as described herein. The plasmid vector comprising the target-specific oligonucleotides can then be used for transformation of a cell, plant or animal.

Methods of designing specific gRNAs with minimal off-target risk

15 According to one aspect, the invention provides methods to design DNA/RNA sequences that guide Cas9 nuclease to target a desired site at a high specificity. The specificity of engineered gRNA could be calculated by sequence alignment of its spacer sequence with genomic sequence of targeting organism.

Approaches to produce non-transgenic, genetically modified plant or animals

Using the aforementioned plasmid vectors and delivery methods, genetically engineered cells, plants or animals can be produced through specific gene targeting and genome editing. In many cases, the resulting genetically modified crops contain no foreign genes and basically are non-transgenic. A DNA sequence encoding gRNA can be designed
25 to specifically target any genes or DNA sequences for targeted mutation via insertion or deletion or other modifications (e.g., transcriptional activation and repression, methylation and demethylation, labeling) through this technology. The ability to efficiently and specifically create targeted mutations or modifications in the cell, plant or animal genome

greatly facilitates the development of many new progeny with improved or novel traits. Because the CRISPR/Cas gene constructs are only transiently expressed in protoplasts and are not integrated into the genome, genetically modified plants or animals regenerated from protoplasts contain no foreign DNAs and are basically non-transgenic. In addition,
5 transgene-free plants or animals can be obtained through genetic segregation after the backcrossing or selfing of genome edited transgenic lines.

Other RNA-Based Genetic Manipulation Techniques

The invention is equally applicable to other RNA mediated genetic manipulation
10 techniques such as RNA based inhibition of gene expression. Typically, an expression cassette capable of producing an RNA molecule that inhibits the transcription and/or translation of at least one polypeptide is included in the invention.

In some embodiments of the present invention, a cell is transformed with an expression cassette that is capable of expressing a polynucleotide that inhibits the expression
15 of polypeptide. The term "expression" as used herein refers to the biosynthesis of a gene product, including the transcription and/or translation of said gene product. The "expression" or "production" of a protein or polypeptide from a DNA molecule refers to the transcription and translation of the coding sequence to produce the protein or polypeptide, while the "expression" or "production" of a protein or polypeptide from an RNA molecule
20 refers to the translation of the RNA coding sequence to produce the protein or polypeptide.

Examples of polynucleotides that inhibit the expression of a polypeptide include sense Suppression/Cosuppression, where an expression cassette is designed to express an RNA molecule corresponding to all or part of a messenger RNA encoding the polypeptide in the "sense" orientation and over expression of the RNA molecule can result in reduced
25 expression of the native gene; *Antisense Suppression* where the expression cassette is designed to express an RNA molecule complementary to all or part of a messenger RNA encoding the ACC synthase polypeptide and over expression of the antisense RNA molecule can result in reduced expression of the native gene; *Double-Stranded RNA Interference*,

where a sense RNA molecule like that described above for cosuppression and an antisense RNA molecule that is fully or partially complementary to the sense RNA molecule are expressed in the same cell, resulting in inhibition of the expression of the corresponding endogenous messenger RNA, *Hairpin RNA Interference and Intron-Containing Hairpin RNA Interference*, where the expression cassette is designed to express an RNA molecule that hybridizes with itself to form a hairpin structure that comprises a single-stranded loop region and a base-paired stem, *Small Interfering RNA or Micro RNA*, where the expression cassette is designed to express an RNA molecule that is modeled on an endogenous miRNA gene.

In addition, the multiplex gRNA-mediated genome editing described in this invention can be used together with a deactive Cas protein for transcriptional activation and repression of multiple genes, epigenome editing (e.g., methylation and demethylation) of multiple sites or genes, labeling of genomic elements, and other multiplex genetic manipulations in the living organisms.

EXAMPLE 1

The CRISPR/Cas9 system is being harnessed as a powerful tool for genome engineering in basic research, molecular therapy and crop improvement. This system utilizes a small guide RNA (gRNA) to direct Cas9 endonuclease to a specific DNA site, thus its targeting capability is largely constrained by the gRNA expressing device. Applicants developed a general strategy to produce numerous gRNAs from a single polycistronic gene. The endogenous tRNA processing system, which precisely cleaves both ends of the tRNA precursor, was engineered as a simple and robust platform to boost the targeting and multiplex editing capability of the CRISPR/Cas9 system. Applicants herein demonstrate that synthetic genes with tandemly arrayed tRNA-gRNA architecture were efficiently and precisely processed into gRNAs with desired 5' targeting sequences *in vivo*, which directed Cas9 to edit multiple chromosomal targets. Using this strategy, multiplex genome editing and chromosomal fragment deletion were readily achieved in stable transgenic rice plants with a high efficiency (up to 100%). Because tRNA and its processing system are virtually conserved in all living organisms, this method could be

broadly utilized to boost targeting capability and editing efficiency of CRISPR/Cas9 toolkits.

Higher organisms often employ complicate genetic networks with functionally redundant genes to fine tune cellular processes. Therefore, molecular tools with the capability to simultaneously manipulate multiple genes are of great value in both basic research and practical applications of genetic engineering.

5 Recently, the bacterial type II clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein (Cas) system is emerging as a promising tool for this purpose. The Cas9 endonuclease from *Streptococcus pyogenes*, coupled with an artificial guide RNA (gRNA), is able to target DNA sequence of 5'-N20-NGG-3' (N indicates any base) in which N20 is the same as 20 bases of gRNA 5' sequence (referred as gRNA spacer hereafter) and NGG is the protospacer-adjacent-motif
10 (PAM) (1-3). The simple RNA-guided programmable rule and high occurrence of PAM in genomes allowed Cas9-gRNA to readily target almost all genetic elements for genome editing. Owing to its simplicity and high efficiency, Cas9-based tools have been rapidly developed for genome/epigenome editing, transcriptional regulation and other applications in genetic engineering (4-6).

In principle, multiplex genome editing could be achieved by expressing Cas9 (or Cas9 derived effectors)
15 along with multiple gRNAs for respective target sites. Conventional delivery methods involve either microinjection or expression of gene constructs containing multiple single gRNA (sgRNA) expressing cassettes for multiplex genome editing (2, 3, 7-11). Direct microinjection of *in vitro* synthesized gRNAs and Cas9 protein (or Cas9 mRNA) into the cell or embryo is only suitable for very few systems. As a result, the most common strategy is to stack multiple sgRNA expressing cassettes in one plasmid
20 construct or to use multiple constructs. The typical size of one sgRNA expressing cassette is about 400-500 bp and consists of RNA Polymerase III (Pol III) promoter, gRNA and Pol III terminator. Due to the limitation of delivery method and plasmid vector capacity, simultaneous production of numerous gRNAs would be a challenge with such a gRNA expressing strategy for most organisms. Moreover, eukaryotic Pol III transcribed RNA is obligated to start with a specific ribonucleotide which may reduce
25 the Cas9/gRNA targetable sites. An advanced strategy is to compact multiple gRNAs in one synthetic gene and engineered an RNA processing system to excise individual gRNAs from the primary transcript. The only proved method based on such a strategy is the expression of Csy4 endoribonuclease (endo-RNase) with Cas9 to processed a transcript containing gRNAs fused with Csy4-cleavable RNA

(12, 13). Nevertheless, more robust and precise methods for simultaneous production of multiple gRNAs are needed to improve multiplex editing capability and facilitate more sophisticated Cas9 applications.

RNA is a fundamental cellular component and its production is guaranteed by the conserved and precise RNA processing systems in different organisms. This inspired us to engineer an endogenous RNA processing system to produce multiple gRNAs from a single transcript, rather than introducing any additional RNases along with Cas9/gRNA components. Applicants herein demonstrate that multiple gRNAs could be efficiently produced from a single synthetic gene with the tRNA-gRNA architecture after precise excision of transcripts *in vivo* by the endogenous RNases. This gRNA expressing strategy was shown to not only allow multiplex targeting but also improve editing efficiency of the CRISPR/Cas9 system in plants. Because the tRNA processing system exists virtually in all organisms, this strategy could be broadly utilized to boost the multiplex editing capability of CRISPR/Cas9 tools for genome engineering.

Results

Strategy to engineer tRNA processing system for producing numerous gRNAs

In order to simultaneously produce multiple gRNAs from one primary transcript, we aimed to compact a cluster of gRNAs with different spacers in one polycistronic gene and hijack an endogenous RNA processing system for cleavage of this transcript into individual gRNAs in nucleus. In search of RNA and endo-RNases that potentially meet this requirement, tRNA attracted our attentions for its characteristics as follows. (I) In nucleoplasm, the tRNA precursor (pre-tRNA) are cleaved at specific sites in eukaryotes by RNase P and RNase Z (or RNase E in bacterium) to remove 5' and 3' extra sequences (Fig. 1A), respectively (14-18). (II) RNase P and RNase Z recognize the tRNA structure, regardless of the pre-tRNA sequence (17, 18). Previous studies revealed that only the acceptor stem, D-loop arm, and T ψ C-loop arm of tRNA are necessary for RNase P and RNase Z cleavage of pre-tRNA (16-19). (III) tRNA was found in polycistronic transcription unit in bacteria (20) and occasionally in eukaryotes (19, 21), suggesting that the tRNA processing system is likely utilized as an intrinsic mechanism to produce different small RNAs (*e.g.*, snoRNA) with tRNA from a single polycistronic gene (19). (IV) Since tRNA

is one of the most abundant cellular components, the endogenous tRNA processing system should be robust to process a large amount of substrates. (V) The tRNA genes contain internal promoter elements (box A and B) to recruit RNA polymerase III (Pol III) complex (Fig. 1B) (22, 23). Hence the tRNA sequence may also work as a potential transcriptional enhancer for Pol III. Based on these features, we hypothesized that the endogenous tRNA system could be engineered as a general platform for precise processing of gRNAs.

To utilize the endogenous tRNA system for multiplex genome editing with Cas9/gRNA, we designed polycistronic tRNA-gRNA (*PTG*) gene for simultaneous production of numerous gRNAs. As shown in Fig. 1C, this *PTG* gene consists of tandem repeats of tRNA-gRNA and would be transcribed as a normal snoRNA or sgRNA gene under the control of Pol III promoter. The endogenous tRNA processing RNases (*e.g.*, RNase P and RNase Z in plants) would recognize the tRNA components and excise individual gRNAs from the *PTG* transcript. The resulting gRNAs would then direct Cas9 to multiple target sites for genome editing (Fig. 1C).

15 **Precise processing of *PTG* to produce functional gRNAs with desired targeting sequences**

To explore whether the synthetic *PTG* gene would be transcribed, processed and function as we predicted (Fig. 1C), we synthesized four *PTG* genes with a structure of tRNA-gRNA (*PTG1* and *PTG2*) or tRNA-gRNA-tRNA (*PTG1.1* and *PTG2.1*, Fig. 2A). These four *PTG* genes were designed to produce gRNA1 (*PTG1* and *PTG1.1*) and gRNA2 (*PTG2* and *PTG2.1*) which have been tested previously (24) with sgRNA genes (*sgRNA1* and *sgRNA2*) in the Cas9-mediated genome editing (*Example 2, Figure 5A* and see gene sequence in Table 2.1). Both gRNAs targeted rice *MPK5* gene which encodes a mitogen-activated protein kinase (MAPK) involved in biotic and abiotic signaling pathways. These *PTGs* were constructed with a 77-bp long pre-tRNA^{Gly} gene which recognizes the *GGC* codon and is widely present among various genomes (25). The chosen pre-tRNA^{Gly} sequence consists of 5' leader (5'-*AACAAA*-3', 6 bp) and mature tRNA (71 bp) (Fig. 2B). Such tRNA-gRNA fusion in *PTGs* mimics the native tRNA-snoRNA43 dicistron in plants (19).

In this study, we used a plasmid vector (*Example 2, Fig. 6*) in which *sgRNA* or *PTG* is expressed with rice U3 snoRNA promoter (*U3p*) and Cas9 is expressed with a rice ubiquitin promoter plus the complete

5' untranslated region (*UBI*p). After transfecting rice protoplasts with plasmids containing *U3p:sgRNA* or *U3p:PTG*, circularized reverse transcription PCR (cRT-PCR)(26, 27) was performed to map both 5' - and 3'-ends of mature gRNAs (See *SI Appendix*, Fig. S7 and Table S2 for the principle of cRT-PCR). Mature gRNAs with the expected size were detected from the protoplasts expressing *PTGs* or *sgRNAs* (Fig. 2C-2F). However, the tRNA-gRNA fused transcript from *PTGs* was not detectable by cRT-PCR, probably due to the highly robust tRNA processing system that cleaved most (if not all) primary transcripts of *PTGs*. Sequence analysis of the cRT-PCR products revealed that all four *PTG* transcripts were precisely cleaved at the tRNA-gRNA junction and produced mature gRNAs carrying the desired 5' spacer sequences without extra nucleotides (Fig. 2G and *Example 2*, Fig. 8 and 9). On the other hand, the 3'-end of mature gRNAs carried a 1-7 nt long poly (U) tail if it preceded the Pol III terminator (*sgRNA*, *PTG1* and *PTG2*, see *Example 2*, Fig. 8), or 1-4 nt ribonucleotides from the tRNA-leader when it preceded the tRNA (*PTG1.1* and *PTG2.1*, see *Example 2*, Fig. 9). The cRT-PCR data also confirmed that gRNAs transcribed from *U3p:sgRNA* were obligated to start with nucleotide *A*, whereas *U3p:PTGs* produced gRNAs without constraint of the first nucleotide (*Example 2*, Fig. 8 and 9). In summary, the gRNAs produced from *PTG* are the same as the *sgRNAs* but are not obligated to start with a specific nucleotide (Fig. 2G and 2H).

The functionality of *PTG1* and *PTG2* was confirmed by examining the insertion/deletion (indel) mutations introduced by non-homologous end joining (NHEJ) repairing at the predicted Cas9:gRNA cleavage site. Because gRNA1 and gRNA2 targets contain the KpnI and SacI restriction enzyme (RE) sites (*Example 2*, Fig. 5A), respectively, the mutations induced by *PTG1/Cas9* and *PTG2/Cas9* could be readily analyzed by the digestion of PCR products encompassing the targeted sites with corresponding RE (PCR/RE assay). In rice protoplasts transfected with *PTG1/Cas9* and *PTG2/Cas9*, 15% and 9% of the target sites were found to carry indels, respectively (Fig. 2I), which are slightly higher than the mutation rate of *sgRNA:Cas9* constructs we used previously (24). In consistence with our hypothesis that tRNA may function as a transcriptional enhancer for Pol III, the quantitative RT-PCR with gRNA specific primers revealed that the transcript levels of *PTG1* and *PTG2* were about 3 and 31 times higher than those of *sgRNA1* and *sgRNA2* in protoplasts (Fig. 2J), respectively. Taken together, our results demonstrated that the endogenous tRNA system could be utilized as a precise and robust tool to produce gRNAs from *PTG* for Cas9-mediated genome editing.

Efficient multiplex genome editing in rice protoplasts via PTG/Cas9

To test the targeting capacity and efficiency of the PTG system, up to eight gRNAs (gRNA1-gRNA8, *Example 2*, Table 2.3 and Fig. 5) have been tandemly arrayed to construct PTGs for targeting four homologous rice MAPKs (*MPK1/2/5/6*) which may function redundantly in diverse cellular signaling pathways. These gRNAs were divided into four pairs and each pair targeted two genomic sites within a gene locus with a distance of 350-750 bp between them (*Example 2*, Fig. 5). By combining different gRNA pairs, we designed *PTG* genes encoding 2 (*PTG3/4/5/6*), 4 (*PTG7/8*), and 8 (*PTG9*) gRNAs to simultaneously target 1, 2, and 4 *MPK* loci, correspondingly (Fig. 3A and *Example 2*, Table 2.1). We expected that such a design might result in a deletion of short chromosomal fragment between two Cas9 cut sites and allowed us to readily examine the efficacy of *PTGs*.

To synthesize *PTG* with repetitive tRNA-gRNA architecture, we designed a scalable and flexible approach to assemble *PTGs* from PCR components based on the principle of Golden Gate assembly strategy (28) (See *Example 2*, Fig. 10 and 11 for details). Our gene assembly approach allowed fast synthesis of *PTGs* with different combination of gRNAs from common oligonucleotide primers. Using one- or two-steps assembly, we synthesized *PTG3-PTG9* genes and cloned them into the CRISPR/Cas9 expressing vectors (*Example 2*, Fig. 6 and Table 2.1-2.3).

We then transfected rice protoplasts with these plasmids containing both *U3p:PTG3-PTG9* and *UBIp:Cas9*. Chromosomal fragment deletions at four MAPK loci, which were revealed by amplification of truncated PCR products with target specific primers, were detected in protoplasts expressing respective *PTGs* with 4% – 45% efficiency (Fig. 3B). Even though 8 gRNAs were simultaneously produced from *PTG9*, they still guided Cas9 to efficiently excise chromosomal fragments at all four targeted loci with 4% – 20% frequency (Fig. 3B). The fragment deletion efficiency at *MPK5* loci was further confirmed by quantitative PCR using primers encompassing the gRNA2 cut site (*SI Appendix*, Table S4). Sequencing of these truncated PCR products confirmed that fragments between two gRNA/Cas9 cut sites were excised from targeted loci with or without additional indels (Fig. 3C). In general, the fragment deletion efficiency was negatively correlated to the distance between two paired cut sites (correlated efficiency $r = -0.95$) despite different gRNAs may have variable efficiencies. We noticed that the total number of gRNAs in one PTG affected the deletion efficiency with variable extent at different targets in protoplasts (Fig. 3B).

At three targeting loci (*MPK2/5/6*), the PTGs (*PTG4/5/6*) containing 2 gRNAs exhibited only slightly higher efficiency than the PTGs with 4 (*PTG7/8*) and 8 (*PTG9*) gRNAs. But at the *MPK1* locus, *PTG3* (2 gRNAs) showed ~2 times higher deletion frequency than *PTG7* (4 gRNA) or *PTG9* (8 gRNAs). Such a reduction of efficiency of PTGs with a higher number of gRNAs was likely due to the competition for
 5 Cas9 among gRNAs. Nevertheless, *PTG* with tandemly arrayed tRNA-gRNA architecture is capable of simultaneously producing numerous gRNAs and guiding Cas9 to multiple chromosomal targets with a high efficiency.

Improving multiplex genome editing in stable transgenic plants with PTG/Cas9

10 Because many plants, including important crops like rice, could not be readily regenerated from protoplasts, we used the conventional *Agrobacterium*-mediated transformation to produce stable transgenic lines and evaluated the efficacy of PTG/Cas9 system for multiplex genome editing in intact rice plants. Mutagenesis frequency at gRNA1 and gRNA2 targets in transgenic plants expressing sgRNA1:Cas9, sgRNA2:Cas9, PTG6:Cas9 or PTG7:Cas9 were examined at T₀ generation. Among T₀
 15 generation of sgRNA1:Cas9 (*n* = 32) and sgRNA2:Cas9 plants (*n* = 20), 44% and 60% of them carried indels while 13% and 20% of them had biallelic mutations, respectively (Table 1 and *Example 2*, Fig. 12). By contrast, indels at both targets were detected in all PTG6:Cas9 plants (100%, *n* = 17) including 35% (gRNA1) and 76% (gRNA2) biallelic mutations (Table 1 and *Example 2*, Fig. 13). But the chromosomal fragment deletion between gRNA1 and gRNA2 target sites was not detected in
 20 PTG6:Cas9 plants with PCR, which may occur at a lower frequency in regenerated calli/plants than in protoplasts. Nevertheless, the results showed that PTG6 not only expressed two gRNAs simultaneously, but also enhanced the mutagenesis efficiency at individual targets than sgRNA did (student's t-test *P* < 0.01, Table 1 and *Example 2*, Fig. 12 and 13).

When total number of targets increased to four in PTG7:Cas9 plants, a comparable indel frequency
 25 to sgRNA1:Cas9 plants was observed at gRNA1 site, but a significantly higher frequency of indels than sgRNA2:Cas9 was found at gRNA2 site (Table 1 and *Example 2*, Fig. 14). Interestingly, we obtained one PTG7:Cas9 line (6%) carrying biallelic deletion of ~350 bp in *MPK1* and monoallelic deletion of ~750 bp in *MPK5* (Fig. 4A and 4B). This prompted us to further examine the efficiency of all eight

gRNAs in PTG9:Cas9 lines. At five gRNA targets (gRNA1/2/3/5/7) whose mutation would destroy RE sites, 50% ($n = 14$) of PTG9:Cas9 T₀ lines carried biallelic mutations at all five targets (*Example 2*, Fig. 15). Interestingly, PCR/RE assays suggested that these five gRNAs exhibited comparable efficiency and all of them showed significantly higher mutagenesis activities than what we observed in

5 sgRNA1/2:Cas9 lines (Table 1). Sanger sequencing of the fragments from four *MPK* loci of PTG9:Cas9 plants confirmed that mutations were introduced at all eight sites (*Example 2*, Fig.16). However, the fragment deletion was only detected at *MPK2* and *MPK5* loci in two PTG9:Cas9 lines (*Example 2*, Fig. 15). In comparison with protoplasts, the efficiency of targeted chromosomal fragment deletion between paired gRNA/Cas9 cut sites is lower in transgenic plants, which might be due to a relatively lower

10 expression of gRNAs and Cas9 in rice calli/regenerated plants (typically only a single copy of transgene is integrated into rice genome after the *Agrobacterium*-mediated transformation). Our data demonstrate that PTG method not only increase the gRNA number and targeting sites, but also likely enhance mutagenesis efficiency at individual sites, especially when multiple gRNAs are expressed using PTG. To further substantiate the high efficiency of mutagenesis with PTG/Cas9, we synthesized *PTG10* with

15 two gRNAs (gRNA9 and gRNA10) to target the rice phytoene desaturase (*PDS*) gene (*Example 2*, Fig. 13A). *PDS* is frequently used as a convenient gene target to examine knock-out efficiency because plants with null functional *PDS* would lead to visible photo-bleached phenotype. Strikingly, all PTG10:Cas9 transgenic seedlings (T₀ generation, $n = 15$) regenerated from calli showed

20 photo-bleaching phenotype, and 13 of them were completely albino, indicating these plants carried null functional *PDS* (Fig. 4C). Though we only identified one line carrying chromosomal deletions between two gRNA targeted sites, sequencing of the *PDS* targets from all the lines confirmed that indels were introduced at the target sites (*Example 2*, Fig. 13). By comparing these data with the previously reported efficiency of sgRNA/Cas9 system (9, 11, 29-32), to our knowledge, the PTG/Cas9 approach yielded the highest efficiency of targeted mutagenesis (up to 100%, Fig. 4 and Table 1) in plants. Our results also

25 demonstrate that targeting one gene with two gRNAs using PTG will greatly increase the efficiency of complete gene knock-out.

Discussion

We developed a general strategy and platform to produce multiple gRNAs from a single synthetic *PTG* gene by hijacking the endogenous tRNA processing system (Fig. 1). We also provided a framework to design, synthesize and utilize *PTG* for multiplex genome editing with Cas9. These *PTGs* were expressed with Pol III promoters (e.g. *U3p*) in the same manner as *sgRNA* genes but not obligated to start with a specific nucleotide (Fig. 2). As a result, current CRISPR/Cas9 vectors for expressing *sgRNA* could be directly used to express *PTG* for multiplex genome editing as we demonstrated in this study.

By producing multiple gRNAs from a single polycistronic gene, the PTG technology could be employed to improve simultaneous mutagenesis of multiple genomic loci or deletion of short chromosomal fragments (Fig.3 and 4). Such a genome engineering approach may lead to simultaneous knock-out of multiple protein coding genes or deletion of non-coding RNA regions and other genetic elements. In addition to targeted mutagenesis/deletion, the PTG approach could facilitate other Cas9-based applications in which multiple gRNAs are required. For example, PTG could be utilized with Cas9 nickase to improve targeting fidelity (13, 33, 34), or with deactivated Cas9 transcriptional-activator or -repressor to manipulate multiple gene expression (35, 36). Given the high processing accuracy and capability of RNase P and RNase Z we observed (Fig. 2), the tRNA processing system also could be utilized as a general platform to produce other regulatory RNAs (e.g. short hairpin RNA or artificial microRNA) from a single synthetic gene. These different classes of regulatory RNAs, like gRNA and short hairpin RNA, also could be compacted into a single polycistronic gene to develop more sophisticated device for genetic engineering.

Recently, polycistronic transcripts that fused gRNA with a 28 nt RNA (referred as gRNA-28nt) was successfully utilized to guide Cas9 to target up to four targets in human cells (12, 13). The synthetic gene with gRNA-28nt architecture produced mature gRNAs with 28 nt extra 3' sequence and also required co-expressing the endonuclease Csy4 from *Pseudomonas aeruginosa* to cleave the transcript.

In comparison with the gRNA-28nt strategy, our approach utilizes the robust endogenous tRNA processing system that enables precise production of gRNAs with only 1-4 nt extra sequence at the gRNA 3' end (Fig. 1 and 2) and carries no additional risk of the endonuclease Csy4 toxicity to recipients. Given the extremely large number of tRNA genes with variable sequences and the fact that RNase P and

RNase Z precisely recognize RNA substrates with tRNA-like structures (18, 37), there are many choices of tRNA sequence to be embedded in *PTG*. Furthermore, the tRNA processing system is universal in all living organisms, thus the PTG technology could be directly adapted to other organisms for Cas9-mediated genome engineering.

5 When multiple DSBs were generated by PTG/Cas9 in rice plants, indels resulting from the error-prone NHEJ repairing occurred more frequently than fragment deletions generated by directly joining two DSBs (*Example 2*, Fig. 14 and 15). To date, the molecular mechanism by which two DSBs directly link together to generate chromosomal translocation or fragment deletion *in vivo* is largely unclear. We speculate that the process leading to such a chromosomal disorder may require two DSBs at
10 the same time interval and is likely determined by the highly dynamic interaction between gRNA/Cas9 cutting and endogenous DNA repairing and also by the distance between DSBs. Due to the differences in the delivery, expression and activity of gRNAs and Cas9, it is not surprising to see some discrepancies in fragment deletion frequency between protoplasts (Fig. 3B) and stable transgenic plants and among different PTG transgenic lines (Fig. 4A and *Example 2*, Fig. 13-15). Because the PTG
15 technology enables to generate many DSBs in genomic DNAs, it may provide an efficient tool to help dissect the molecular process of chromosomal deletion. More importantly, the PTG technology significantly improves multiplex editing capability and efficiency and is expected to facilitate more sophisticated Cas9 applications such as targeted mutagenesis and deletion of redundant genes or genetic elements, transcriptional modulation of multiple genes and pathways, modification and labeling of
20 numerous genomic sites, site-specific integration and gene replacement.

Materials and Methods

Plant materials

Rice (*Oryza sativa* L. ssp) cultivars Kitaake and Nipponbare were used in this study. Rice plants were
25 grown in a green house or growth chamber at 28°C day/23°C night with 12 h of light.

Plasmid vector construction

The plasmid vectors pRGE32 was used to transiently express U3p:sgRNA or U3p:PTG along with

UBIp:Cas9 in rice protoplasts and pRGEB32 was used for the *Agrobacterium*-mediated rice transformation (Example 2, Fig. 6). See Example 2, SI Methods for the details about plasmid vector construction.

5 **sgRNA:Cas9 and PTG:Cas9 expression constructs**

The U3p:sgRNA1 and U3p:sgRNA2 constructs were made as described previously (24). The specific spacer sequences for gRNA3-gRNA8 (Example 2, Table 2.3) were selected using the CRISPR-PLANT database at world wide web at genome.arizona.edu/crispr/ (38). The PTG genes and U3p:PTG constructs were generated as described in Example 2, SI Methods, Fig. 10-11 and Table 2.3.

- 10 The synthesized PTGs were inserted into the Bsa I digested pRGE32 or pRGEB32 for transient protoplast expression or stable rice transformation, respectively. The sequences of sgRNA1, sgRNA2 and PTG genes used in this study were listed in Example 2, Table 2.1.

Rice protoplast transfection

- 15 The rice protoplast preparation and transfection were performed as we described previously(24). Briefly, 20 µg of plasmid DNA was used to transfected 2×10^5 protoplasts with a transfection efficiency of ~40%-50%. Total rice genomic DNAs were extracted from the protoplast samples at 36h after transfection and then used for PCR and sequence analysis.

20 ***Agrobacterium*-mediated rice transformation**

Transgenic rice plants were generated by the *Agrobacterium tumefaciens*-mediated transformation using rice mature seed-derived calli according to a conventional protocol (39).

Genomic DNA extraction and PCR/RE assay

- 25 Rice genomic DNA was extracted as described previously with the CTAB method (24). To detect mutagenesis at desire sites, target regions were amplified with specific primers (See *Example 2*, Table 2.2 for primer sequences) using GoTaq DNA polymerase (Promega). The PCR product was separated in 1% agarose gel and stained with ethidium bromide to detect chromosomal fragment deletion. To detect

indel at specific sites with PCR/RE, the PCR products encompassing the target sites were digested with appropriate restriction enzymes (RE) for 5 hours, and then were analyzed with 1% agarose gel electrophoresis and ethidium bromide staining. The stained gels were imaged using Gel Doc XRS system (Bio-Rad) and quantified with the Quantity One 4.6 program (Bio-Rad). Selected PCR products
5 were cloned into pGEM T-easy vector (Promega) for DNA sequencing.

RNA extraction, cRT-PCR and quantitative PCR

See *Example 2*, SI Methods for details.

10 Gene accession number in Genebank

Genes and their Genebank RefSeq accession numbers are as follow: *MPK1*, Os06g0154500; *MPK2*, Os08g0157000; *MPK5*, Os03g0285800; *MPK6*, Os10g0533600; *UBI*, Os02g0161900; and *PDS*, Os03g0184000.

15 References

1. Jinek M, et al. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816-821.
2. Cong L, et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339(6121):819-823.
- 20 3. Mali P, et al. (2013) RNA-guided human genome engineering via Cas9. *Science* 339(6121):823-826.
4. Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods* 10(10):957-963.
5. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and
25 targeting genomes. *Nat Biotechnol* 32(4):347-355.
6. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157(6):1262-1278.
7. Wang H, et al. (2013) One-Step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153(4):910-918.
- 30 8. Jinek M, et al. (2013) RNA-programmed genome editing in human cells. *eLife* 2:e00471.
9. Li JF, et al. (2013) Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol* 31(8):688-691.
- 35 10. Zhou H, Liu B, Weeks DP, Spalding MH, Yang B (2014) Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice.

Nucleic acids research 42(17):10903-10914.

11. Shan Q, et al. (2013) Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol* 31(8):686-688.
12. Nissim L, Perli SD, Fridkin A, Perez-Pinera P, Lu TK (2014) Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell* 54(4):698-710.
13. Tsai SQ, et al. (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* 32(6):569-576.
14. Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. *Genes Dev* 24(17):1832-1860.
15. Schiffer S, Rosch S, Marchfelder A (2002) Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J* 21(11):2769-2777.
16. Gutmann B, Gobert A, Giege P (2012) PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*. *Genes Dev* 26(10):1022-1027.
17. Barbezier N, et al. (2009) Processing of a dicistronic tRNA-snoRNA precursor: combined analysis *in vitro* and *in vivo* reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol* 150(3):1598-1610.
18. Canino G, et al. (2009) *Arabidopsis* encodes four tRNase Z enzymes. *Plant Physiol* 150(3):1494-1502.
19. Kruszka K, et al. (2003) Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J* 22(3):621-632.
20. Nakajima N, Ozeki H, Shimura Y (1981) Organization and structure of an E. coli tRNA operon containing seven tRNA genes. *Cell* 23(1):239-249.
21. Nakaar V, Dare AO, Hong D, Ullu E, Tschudi C (1994) Upstream tRNA genes are essential for expression of small nuclear and cytoplasmic RNA genes in trypanosomes. *Mol Cell Biol* 14(10):6736-6742.
22. White RJ (2011) Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet* 12(7):459-463.
23. Dieci G, Fiorino G, Castelnovo M, Teichmann M, Pagano A (2007) The expanding RNA polymerase III transcriptome. *Trends Genet* 23(12):614-622.
24. Xie K, Yang Y (2013) RNA-guided genome editing in plants using a CRISPR-Cas system. *Mol Plant* 6(6):1975-1983.
25. Chan PP, Lowe TM (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res* 37(Database issue):D93-97.
26. Kuhn J, Binder S (2002) RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. *Nucleic Acids Res* 30(2):439-446.
27. Yokobori S, Paabo S (1995) Transfer RNA editing in land snail mitochondria. *Proc Natl Acad Sci USA* 92(22):10432-10435.
28. Engler C, Kandzia R, Marillonnet S (2008) A one pot, one step, precision cloning method with high throughput capability. *PloS one* 3(11):e3647.
29. Nekrasov V, Staskawicz B, Weigel D, Jones JD, Kamoun S (2013) Targeted

- mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol* 31(8):691-693.
30. Feng Z, et al. (2014) Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc Natl Acad Sci USA* 111(12):4632-4637.
- 5 31. Wang Y, et al. (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat Biotechnol* 32(9):947-951.
32. Zhang H, et al. (2014) The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol J* 12(6):797-807.
- 10 33. Ran FA, et al. (2013) Double nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154(6):1380-1389.
34. Guilinger JP, Thompson DB, Liu DR (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* 32(6):577-582.
- 15 35. Gilbert LA, et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154(2):442-451.
36. Mali P, et al. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31(9):833-838.
- 20 37. Forster AC, Altman S (1990) External guide sequences for an RNA enzyme. *Science* 249(4970):783-786.
38. Xie K, Zhang J, Yang Y (2014) Genome-wide prediction of highly specific guide RNA spacers for CRISPR-Cas9-mediated genome editing in model plants and major crops. *Mol Plant* 7(5):923-926.
- 25 39. Hiei Y, Komari T (2008) Agrobacterium-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat Protoc* 3(5):824-834.

30

Table 1. Targeted mutation efficiency in PTG:Cas9 vs. sgRNA:Cas9 plants.

Gene ID	# of T ₀ Lines	Editing at gRNA1		Editing at gRNA2		Editing at gRNA3/5/7	
		target		target		targets	
		Mut.	Bi-Mut	Mut.	Bi-mut.	Mut.	Bi-mut.
<i>sgRNA1</i>	32	44%	13%	n.a.	n.a.	n.a.	n.a.
<i>sgRNA2</i>	20	n.a.	n.a.	60%	20%	n.a.	n.a.
<i>PTG6</i>	17	100%**	35%**	100%**	76%**	n.a.	n.a.
<i>PTG7</i>	17	47%	6%	100%*	24%	n.a.	n.a.
<i>PTG9</i>	14	86%**	78%**	86%**	57%**	86%	86%

The frequency of mutation (Mut.) and biallelic mutation (Bi-Mut.) at the targeting sites in transgenic lines was examined and calculated based on PCR/RE assays (*Example 2*, Fig. S8-S11). The editing

frequencies at gRNA1 and gRNA2 targets were compared between sgRNA and PTG plants. **, student's t-test $P < 0.01$; *, student's t-test $P < 0.05$; n.a., not applied.

5

EXAMPLE 2

Supporting Information

Plasmid vector construction

10 To construct pRGE32, the rice *UBI*p fragment was amplified from genomic DNA of Nipponbare cultivar with a pair of specific primers. After amplifying the *U3p-gRNA* fragment from pRGE31 (Addgene plasmid 50929) with primers UGW-U3-F and UGW-gRNA-R, the *U3p-gRNA* and *UBI*p fragments were linked together by overlapping extension PCR with primers UGW-U3-F and UBI-R. Then pRGE32 was constructed by Gibson Cloning (New England Biolabs) to replace the
15 *U3p-gRNA-35S* fragment in pRGE31 with *U3p-gRNA-UBI*p. The pRGE32 binary vector (see Fig. 6 for vector map) used for the *Agrobacterium*-mediated rice transformation was created by inserting the *U3p-gRNA-UBI*p-*Cas9* fragment from pRGE32 into pCAMBIA1301-BsaI with Gibson Cloning (New England Biolabs). The pCAMBIA1300-BsaI is derived from pCAMBIA1300 after removing all of BsaI sites through site-directed mutagenesis. The primers used in plasmid construction are listed in Table 2.2.

20 The pGTR plasmid, which contains a gRNA-tRNA fused fragment, was used as a template to synthesize PTGs in this study. To construct pGTR, the gRNA scaffold fragment was amplified by PCR using a pair of specific primers (Bsa-gRNA-F and gRNA-R) whereas the tRNA^{Gly} fragment was amplified as a primer dimer of g-tRNA-F and tRNA-R. Then these two fragments were fused as gRNA-tRNA by overlapping extension PCR using primers Bsa-gRNA-F and tRNA-R. The overlapping
25 PCR product was separated and purified from an agarose gel, and then inserted into pGEM-T easy (Promega) to generate pGTR plasmid. The sequence of gRNA-tRNA fusion in pGTR is shown in Table S1. The primers used in plasmid construction are listed in Table 2.2.

RNA extraction, cRT-PCR and quantitative PCR

30 Total RNAs were extracted from protoplasts using the TRIzol Reagent (Life Technologies) according to the manufacturer's instruction. The cRT-PCR was schematically depicted in Fig. 7 and

performed as follow. To circularize RNA, 1 µg of total RNA was incubated in a 20 µl reaction containing 1 × T4 RNA ligase buffer (New England Biolabs), 50 µM of ATP, 10% PEG8000, 20 U of RNase inhibitor (New England Biolabs) and 10 U of T4 RNA ligase (New England Biolabs). The ligation was carried out at 25 °C for 4 h. Then 10 U of DNase I (New England Biolabs) was added into
5 ligation reaction to remove genomic DNA contamination at 25 °C for 20 min. Then circularized RNA was purified with TRIzol Reagent (Life Technologies) and dissolved in nuclease free water. A total of 200 ng circularized RNA was mixed with 0.5 mM dNTP and 1 µM of oligos specific to gRNA spacer (gRNA1-R and gRNA2-R, see Table S2) and denatured at 70 °C for 5 min. After chilling on ice, 1× MuMLV Reverse Transcriptase buffer, 20 U of RNase inhibitor (New England Biolabs), and 10 U of
10 MuMLV reverse transcriptase (New England Biolabs) were added to synthesize 1st cDNA at 42 °C for 1 hour. The negative controls without adding MuMLV reverse transcriptase (-RT) were also performed for all samples. After reverse transcription, MuMLV was inactivated at 70 °C for 10 min. Then PCR was performed with the following 50 µl reaction: one twentieth of 1st cDNA, 1× Phusion HF buffer, 0.2 mM dNTPs, 0.5 µM of forward primers, 0.5 µM reverse primers, and 1 U of Phusion DNA polymerase
15 (Thermo Scientific). The resulting PCR products were analyzed with 2% agarose gel electrophoresis. The cRT-PCR products were then sequenced after cloning into pGEM-T easy vector (Promega). See Table 2.2 for primer sequences used in cRT-PCR.

For quantitative RT-PCR, the DNase I treated total RNAs were reverse-transcribed to produce cDNAs using a specific primer gRNA-R along with MuMLV (New England Biolabs) according to the
20 manufacturer's instructions. The real-time PCR was performed using GoTaq qPCR Master Mix (Promega) in the StepOnePlus Realtime PCR system (Life technologies). The gRNA1-F and gRNA-R primers were used for *sgRNA1* and *PTG1*, and gRNA2-F and gRNA-R primer were used for *sgRNA2* and *PTG2*. The rice *UBI* gene was used as the internal reference for relative quantification. See Table S2 for primer sequences.

25 To quantify deletion efficiency, quantitative PCR was performed using 10 ng of the genomic DNA as a template to amplify the genomic fragments within *MPK5* or *UBI* loci. A pair of specific primers (MPK5-qF and MPK5-R611, Table S2) which encompass the gRNA2 cut site were used to amplify *MPK5* copies containing no fragment deletion (Fig. S1A). The relative quantity (RQ) of *MPK5*

fragment without deletion was calculated using *UBI* as a reference gene. Because only *MPK5* fragments without deletion were amplified, the fragment deletion efficiency could be estimated as 100%-RQ (Table 2.4).

5 **Synthesis of Polycistronic tRNA-gRNA (*PTG*) Genes by Golden Gate Assembly**

The PTG genes were synthesized based on the principle of Golden Gate (GG) assembly which is broadly used to assemble DNA parts like customized transcription activator-like effector (TALE). Our assembly approach allows synthesizing PTGs with different combinations of gRNAs using the same
10 components. For example, we made PTG3-PTG9 with the same set of oligo primers. By hierarchical GG assembly reaction, two PTGs could be assembled together to create a longer PTG (like *PTG9*). PTGs with no more than 6 gRNAs (e.g. *PTG1-PTG8*) could be synthesized by one step GG assembly (Level 1, Fig. 10), whereas PTGs with more than 6 gRNAs (e.g., *PTG9*) require two or more steps of GG assembly (Level 2, Fig. 11). The schematic diagrams of PTG synthesis approach are shown in Fig.
15 10 and 11, and details of primer design, GG assembly and plasmid construction are described as follow.

Step 1. Design primers to amplify gRNA-tRNA parts

In order to ligate multiple DNA parts in a desired order, GG assembly requires distinct 4-bp overhangs to ligate two DNA parts after digestion with *BsaI* (or other type II endonucleases such as *AarI*,
20 *BbsI*, *BsmAI*, *BsmBI*). The gRNA spacer is the only unique sequence in *PTG* (Fig. 1C), thus PTGs should be divided into DNA parts within the gRNA spacer region. As shown in Fig. 10A and 10B, a gRNA spacer was split into two parts with 4 bp overlap and each half of the spacer was synthesized within oligo primers with a *BsaI* site. Details to design gRNA specific primers are described below:

25 1.1. Select a 4-bp long sequence within each gRNA spacer as *BsaI* overhangs in GG assembly. The overhang could be any 4 consecutive nucleotides of the gRNA spacer. Of note, DNA parts that assembled in the same GG reaction should have a distinct 4-bp overhang, and could not be 5'-*GGCA*-3' or 5'-*AAAC*-3' which are used in terminal adaptors for cloning to pRGE32 and pRGE32 (Fig. 6).

1.2. Design the primer sequences as follow (also see Fig. 6):

In this example, the 9th to 12th nucleotides of a 20 nt long spacer of gRNA[x] is chosen as the BsaI overhang for GG cloning:

5'-N₁- N₂- N₃- N₄- N₅- N₆- N₇- N₈-N₉-N₁₀-N₁₁- N₁₂- N₁₃- N₁₄- N₁₅- N₁₆- N₁₇- N₁₈-N₁₉-N₂₀-3'

The primer should be:

gRNA[x]-F (Forward primer, anneal to 5'-end of gRNA scaffold):

5'-ta-*GGTCTC*-N-N₉N₁₀N₁₁N₁₂N₁₃N₁₄N₁₅N₁₆N₁₇N₁₈N₁₉N₂₀-gttttagagctagaa-3'

gRNA[x]-R (Reverse primer, anneal to 3'-end of pre-tRNA):

5'-cg-*GGTCTC*-N-N₁₂N₁₁ N₁₀ N₉N₈ N₇ N₆ N₅N₄N₃N₂ N₁- tgcaccagccggg-3'

Note: Any 4 consecutive nucleotides in the spacer could be selected as overhangs for GG assembly. This allows the selection of a specific overhang for each DNA part in GG assembly. The two 5'-terminal bases (shown in lowercase) are randomly added nucleotides to enhance BsaI digestion of PCR products. The red color indicates the reverse complementary sequence to the spacer. The lowercase letters at the 3'-end indicate bases that anneal to gRNA (Forward primer) or tRNA (Reverse primer).

Step 2. Level 1 GG assembly (Construction of *PTG1-PTG8* genes)

5 See Fig. S6C for the overall strategy of level 1 GG assembly of PTGs.

2.1. Set up 50 µL PCR reactions to amplify DNA parts for PTG construction.

pGTR plasmid	0.1 ng
5X Phusion HF buffer	10 µl
dNTPs (10mM)	1 µl
Forward primer (10 µM)	2.5 µl
Reverse Primer (10 µM)	2.5 µl
Phusion (2U/µl, NEB)	0.5 µl
H ₂ O	x µl
Total	50 µl

To construct PTGs used in this study, the forward and reverse primers to amplify level 1 parts were added as follow:

PCR ID	Forward primer	Reverse primer	Level 1 parts symbol
P1	L5AD5-F	gR1-R	L5AD-gR1
P2	gR1-F	L3AD5-R	gR1-L3AD
P3	L5AD5-F	gR2-R	L5AD-gR2
P4	gR2-F	L3AD5-R	gR2-L3AD
P5	gR1-F	gR2-R	gR1-gR2

P6	L5AD5-F	gR3-R	L5AD-gR3
P7	gR3-F	gR4-R	gR3-gR4
P8	gR4-F	L3AD5-R	gR4-L3AD
P9	L5AD5-F	gR5-R	L5AD-gR5
P10	gR5-F	gR6-R	gR5-gR6
P11	gR6-F	L3AD5-R	gR6-L3AD
P12	L5AD5-F	gR7-R	L5AD-gR7
P13	gR7-F	gR8-R	gR7-gR8
P14	gR8-F	L3AD5-R	gR8-L3AD
P15	gR2-F	gR3-R	gR2-gR3
P16	gR8-F	gR5-R	gR8-gR5
P17	gR4-F	gR7-R	gR4-gR7

PCRs were run with the following program:

Temperature	Time	Cycles
98 °C	2 min	1
98 °C	10 sec	35
50 °C	20 sec	
72 °C	20 sec	
72 °C	2.5 min	1
4 °C	Hold	1

2.2. The PCR products were purified with Spin Column PCR Products Purification kit (BioBasic).

2.3. Individual parts were ligated together by GG assembly with the following reaction:

Level 1 parts	25-50 ng
(add equal amount for each parts)	
2 x T7 DNA ligase Buffer (NEB)	10 µl
Bovine Serum Albumin (1mg/ml)	2 µl
Bsa I (10 U/µl, NEB)	0.5 µl
T7 DNA Ligase (3000 U/µl, NEB)	0.5 µl
Total Volume	20 µl

5

For GG assembly reactions to construct *PTG1-PTG8*, the level 1 parts were added as follow:

Gene ID	Encoding gRNAs	Level 1 parts used
<i>PTG1</i>	gRNA1	L5AD-gR1 and gR1-L3AD
<i>PTG2</i>	gRNA2	L5AD-gR2 and gR2-L3AD
<i>PTG3</i>	gRNA3-gRNA4	L5AD-gR3, gR3-R4, and gR4-L3AD
<i>PTG4</i>	gRNA5-gRNA6	L5AD-gR5, gR5-gR6, and gR6-L3AD
<i>PTG5</i>	gRNA7-gRNA8	L5AD-gR7, gR7-gR8, and gR8-L3AD
<i>PTG6</i>	gRNA1-gRNA2	L5AD-gR1, gR1-gR2, and gR2-L3AD

PTG7	gRNA1-gRNA2-gRNA3-gRNA4	L5AD-gR1, gR1-gR2, gR2-gR3, gR3-gR4, and gR4-L3AD
PTG8	gRNA7-gRNA8-gRNA5-gRNA6	L5AD-gR7, gR7-gR8, gR8-gR5, gR5-gR6, and gR6-L3AD

2.4. GG reactions were performed in a thermal cycler (Bio-Rad) by incubation at 37 °C, 5 min and 20 °C, 10 min for 30-50 cycles; and then held at 20 °C for 1 hour.

2.5. The GG reaction product was diluted with 180 µl of H₂O.

- 5 2.6. The level 1 GG assembly products were amplified in the 50 µl PCR reaction:

Ligation product (1:10 dilution)	1 µl
5 X Go Green buffer (Promega)	10 µl
dNTPs (10mM)	1 µl
S5AD5-F (10 µM)	1 µl
S3AD5-R (10 µM)	1 µl
GoTaq DNA polymerase (2U/µl, Promega)	1 µl
H ₂ O	35 µl
Total	50 µl

PCR was run in a thermal cycler (Bio-Rad) with the following program:

Temperature	Time	Cycles
95 °C	2 min	1
95 °C	10 sec	35
60 °C	20 sec	
72 °C	1 min/kb	
72 °C	5 min	1
4 °C	Hold	1

2.7. Purify the PCR product with Spin Column PCR Products Purification kit (Bio Basic).

- 10 2.8. Digest the purified PCR product with Fok I (NEB).

2.9. Separate the Fok I digested products in 1% agarose gel; Excise the DNA bands with the expected size from the gel and then purify them with Spin Column DNA Gel Extraction Kit (Bio Basic).

2.10. Ligate the Fok I digested GG fragment into the BsaI digested pRGE32 or pRGE32 vectors with T4 DNA ligase (NEB).

- 15 2.11. Transform the ligation product to *E. coli* DH5α, purify the recombinant plasmids and confirm the constructs by Sanger sequencing.

Step 3. Level 2 GG assembly (Construct *PTG9* by two-step GG assembly)

The schematic diagram of level 2 GG assembly is shown in Fig.11. Level 2 GG assembly is used to synthesize PTGs with more than six gRNAs. Such a large PTG is constructed by ligating two smaller PTGs (level 2 parts) together. These two small PTGs are synthesized by level 1 GG assembly and contain one overlapped gRNA as a bridge (bridge gRNA, Fig. 11) to ligate level 2 parts in a next GG assembly reaction. To amplify level 2 parts from level 1 assembled PTGs, a pair of specific primers annealing only to the bridge gRNA spacer are required for PCR with the terminal adaptor primers (S5AD5-F and S3AD5-R). In this study, the *PTG9* was synthesized with this approach and gRNA7 was used as the bridge gRNA.

3.1. Design bridge gRNA spacer-specific primers to amplify level 2 parts.

Bridge gRNA spacer:

5'-N₁- N₂- N₃- N₄- N₅- N₆- N₇- N₈-[N₉-N₁₀-N₁₁- N₁₂]- N₁₃- N₁₄- N₁₅- N₁₆- N₁₇- N₁₈-N₁₉-N₂₀-3'

Any consecutive 4-bp could be selected as an overhang for GG assembly. In this example,

N₉-N₁₀-N₁₁-N₁₂ was selected as an overhang.

Ln-gR[x]-F (Forward primer, only annealing to bridge gRNA spacer):

5'-ta-*GGTCTC*-N-[N₉N₁₀N₁₁N₁₂]N₁₃N₁₄N₁₅N₁₆N₁₇N₁₈N₁₉N₂₀-3'

Ln-gR[x]-R (Reverse primer, only annealing to bridge gRNA spacer):

5'-cg-*GGTCTC*-N-[N₁₂N₁₁ N₁₀ N₉] N₈ N₇ N₆ N₅N₄N₃N₂ N₁-3'

Note: The two 5'-terminal bases (shown in lowercase) are randomly added nucleotides to enhance Bsa I digestion of PCR products. The red color indicates the reverse complementary sequence to the spacer.

3.2. Set up GG reactions with the following level 1 parts generated from 2.1-2.2.

PTG ID	Encoding gRNAs	Level 1 Parts used in GG assembly reaction
PTG7-Ln	gRNA1-gRNA2-gRNA3-gRNA4-gRNA7	L5AD-gR1, gR1-gR2, gR2-gR3, gR3-R4, gR4-gR7, and gR7-L3AD

The level 1 GG assembly was carried out in the same manner as steps 2.4-2.5.

3.3. Amplify level 2 DNA parts with the following primers and templates:

PCR ID	Forward primer	Reverse primer	Template (1:10 diluted Level 1 GG reaction)
L2-P1	S5AD5-F	Ln-gR7-R	PTG7-Ln (step 3.2)
L2-P2	Ln-gR7-F	S3AD5-R	PTG8 (step 2.5)

The PCR condition is the same as step 2.6.

3.4. Separate the PCR products in 1% agarose gel. Excise the PCR bands with the expected size from the gel and purify them with Spin Column DNA Gel Extraction kit (Bio Basic).

3.5. Set up GG assembly reactions to ligate two DNA parts (L2-P1 and L2-P2) together.

L2-P1 PCR product	50 ng
L2-P2 PCR product	50 ng
2 x T7 DNA ligase Buffer (NEB)	10 µl
Bovine Serum Albumin (1mg/ml)	2 µl
Bsa I (10 U/µl, NEB) TABEL OF SEQUENCES	0.5 µl
T7 DNA Ligase (3000 U/µl, NEB)	0.5 µl
Total Volume	20 µl

3.6 Perform GG assembly reactions in a thermal cycler (Bio-Rad) using the following program: 37

5 °C, 5 min; 20 °C, 10 min for 25 cycles; and 20 °C for 1 hour.

3.7. Amplify Level 2 GG assembled product with S5AD5-F and S3AD5-R, then inserted the amplified product into pRGE32 or pRGE32 using the same procedure as steps 2.6-2.11.

10

Example 2 Tables

Table 2.1. Sequence of synthetic genes used in this study

Gene (Architecture)	Sequence (5'→3')
Sequence of gRNA-tRNA fusion in pGTR plasmid SEQ ID NO:15	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC AACAAAGC ACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACC CGGGTTCGATTCCCGGCTGGTGCA
<i>sgRNA1</i> SEQ ID NO:1	GATCCGTGGCAAGATGTCTAGAGCAGGTACGTTTTAGAGCTAGA AATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGCACCGAGTCGGTGCTTTTTT
<i>sgRNA2</i> SEQ ID NO:2	GATCCGTGGCAGTCTACATCGCCACGGAGCTCAGTTTTAGAGCT AGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA AAAAGTGGCACCGAGTCGGTGCTTTTTT
<i>PTG1</i>	GATCCGTGGC AACAAAGCACCGTGGTCTAGTGGTAGAATAGTAC

(tRNA-gRNA1) SEQ ID NO:16	<u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>AGA</u> TGTCGTAGAGCAGGTACGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTTTTTTTTTT
PTG2 (tRNA-gRNA2) SEQ ID NO:17	GATCCGTGGC <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC</u> <u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>ATCT</u> ACATCGCCACGGAGCTCAGTTTTAGAGCTAGAAATAGCAAGTTA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCTTTTTTTTTT
PTG1.1 (tRNA-gRNA1-tRNA) SEQ ID NO:18	GATCCGTGGC <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC</u> <u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>AGA</u> TGTCGTAGAGCAGGTACGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCA <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCC</u> <u>GCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>TTTTTT</u> TTTT
PTG2.1 (tRNA-gRNA2-tRNA) SEQ ID NO:19	GATCCGTGGC <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC</u> <u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>ATCT</u> ACATCGCCACGGAGCTCAGTTTTAGAGCTAGAAATAGCAAGTTA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCA <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACC</u> <u>CTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>TTTT</u> TTTTT
PTG3 (tRNA-gRNA3 -tRNA-gRNA4) SEQ ID NO:20	GATCCGTGGC <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC</u> <u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>ATC</u> CAGGCGACGCTGAGCCAGTTTTAGAGCTAGAAATAGCAAGTTA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCA <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACC</u> <u>CTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>TGGC</u> CCACCGGGGTATAAAAGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTTTTTTTTTT
PTG4 (tRNA-gRNA5-tRNA A-gRNA6) SEQ ID NO:21	GATCCGTGGC <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC</u> <u>CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA</u> <u>GAA</u> CCCGGTCGCCTCAAGGAGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCA <u>AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCC</u>

	GCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA GAATGC GCAGACTCGTCAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAAT AAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG GTGCTTTTTTTTTT
PTG5 (tRNA-gRNA7-tRN A-gRNA8) SEQ ID NO:22	GATGATCCGTGGC AACAAAGCACCAAGTGGTCTAGTGGTAGAATAG TACCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA GTGTCCGCTTGGCATCGATAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGCA AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTAC CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA AGC GGGTGCGGGTCAATCAAAGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTTTTTTT
PTG6 (tRNA-gRNA1-tRN A-gRNA2) SEQ ID NO:23	GATCCGTGGC AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTAC CCTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA AGA TGTCGTAGAGCAGGTACGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCA AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACCCT GCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA TCTACA TCGCCACGGAGCTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCTTTTTTTTTT
PTG7 (tRNA-gRNA1-tRN A-gRNA2-tRNA-gR NA3-tRNA-gRNA4) SEQ ID NO:24	GATCCGTGGC AATAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACC CTGCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA AGAT GTCGTAGAGCAGGTACGTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGT CGGTGCA AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACCCT GCCACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA TCTACA TCGCCACGGAGCTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGCA AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA ATCCAGGC GACGCTGAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGC AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACCCTGCCA CGGTACAGACCCGGGTTTCGATTCCCGGCTGGTGCA TGGCCACCG GGGTATAAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTG TTTTTTTTT

PTG8 (tRNA-gRNA7-tRN A-gRNA8-tRNA-gR NA5-tRNA-gRNA6) SEQ ID NO:25	GATCCGTGGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC CCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA AGTG TCCGCTTGGCATCGATAGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCT GCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA AGCGGG TGCGGGTCAATCAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA GAACCCGG TCGCCTCAAGGAGTTTTAGAGCTAGAAATAGCAAGTTAAAAATA GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGT GC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCAC GGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA GAATGCGCAGA CTCGTCAGGGTTTTAGAGCTAGAAATAGCAAGTTAAAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC TTTTTTTT
PTG9 (tRNA-gRNA1-tRN A-gRNA2-tRNA-gR NA3-tRNA-gRNA4- tRNA-gRNA7-tRN A-gRNA8-tRNA-gR NA5-tRNA-gRNA6) SEQ ID NO:26	GATCCGTGGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTAC CCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA AGA TGTCGTAGAGCAGGTACGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCT GCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA ATCTACA TCGCCACGGAGCTCAGTTTTAGAGCTAGAAATAGCAAGTTAAAA TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC GGTGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGC CACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA ATCCAGGC GACGCTGAGCCAGTTTTAGAGCTAGAAATAGCAAGTTAAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCA CGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA TGGCCACCG GGGTATAAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAAATAAGG CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC AACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGG TACAGACCCGGGTTCGATTCCCGGCTGGTGCA GTGTCCGCTTGGC ATCGATAGTTTTAGAGCTAGAAATAGCAAGTTAAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC AAC AAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTAC AGACCCGGGTTCGATTCCCGGCTGGTGCA AGCGGGTGCGGGTCAAT

	CAAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGCGGCACCGAGTCGGTGCAACAA AGCACCAAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAG ACCCGGGTTCGATTCCCGGCTGGTGCA<i>GAACCCGGTCGCCTCAAG</i> GAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCC GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAACAAAG CACCAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGAC CCGGGTTCGATTCCCGGCTGGTGCA<i>GAATGCGCAGACTCGTCAGG</i> GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTTT
PTG10 (tRNA-gRNA9-tRN A-gRNA10) SEQ ID NO:27	GATCCGTGGCAACAAAGCACCAGTGGTCTAGTGGTGGGAATAGTAC CCTGCCACGGTACAGACCCGGGTTCGATTCCCGGCTGGTGCA<i>AACA</i> AGCCAGGAGAATTCAGCGTTTTAGAGCTAGAAATAGCAAGTTAA AATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCAACAAAGCACCAGTGGTCTAGTGGTAGAATAGTACCCT GCCACGGTACAGACCCGGGTCCGATTCCCGGCTGGTGCA<i>CACTGC</i> ATGGATAACTCATCGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGG TGCTTTTTTTTTT

The sequences are annotated as follows:

Last 10 bp of *U3p* (SEQ ID NO:14: underlined

gRNA scaffold: bold

pre-tRNA (SEQ ID NO:13): boxed, bold italics

5 gRNA spacer: italics

Pol III Terminator (TTTT..T): grey background

10

Table 2.2. Primers used for plasmid construction, cRT-PCR, and genotyping

Oligo Name	Sequence (5'→3')	Purpose
gRNA1-R (SEQ ID NO:28)	AAACGTACCTGCTCTACGAC	cRT-PCR of gRNA1
gRNA2-R (SEQ ID NO:29)	AAACTGAGCTCCGTGGCGAT	cRT-PCR of gRNA2
Bsa-gRNA-F (SEQ ID NO:30)	GGAGACCGAGGTCTCGTTTGTAGAGCTAGAAAT A	cRT-PCR and amplify gRNA
gRNA1-F (SEQ ID NO:31)	GGC AAGATGTCGTAGAGCAGGTAC	Quantitative RT-PCR
gRNA2-F (SEQ ID NO:32)	GTCTACATCGCCACGGAGCTCA	Quantitative RT-PCR
gRNA-R (SEQ ID NO:33)	GCACCGACTCGGTGCCAC	Quantitative RT-PCR
UBI-qF (SEQ ID NO:34)	TGGTCAGTAATCAGCCAGTTTG	Quantitative RT-PCR
UBI-qR (SEQ ID NO:35)	CAAATACTTGACGAACAGAGGC	Quantitative RT-PCR
MPK5-qF (SEQ ID NO:36)	GATCCCGCCGCCGATCCCTC	Quantitative PCR
g-tRNA-F (SEQ ID NO:37)	GCACCGAGTCGGTGC AACAAAGCACCAGTGGTCTAGTGGT <u>AGTAATAGTA</u> <u>CCCTG</u>	Construct pGTR. Overlapped region is underlined.
tRNA-R (SEQ ID NO:38)	CTGCCATGCACCAGCCGGAATCGAACCCGGG TCTGTACCGTGGCAGGGTACTATTCTAC	
UBI-F (SEQ ID NO:39)	TGCATGCCTGCAGGTCCACAAATTCGGGTCAA GGCGG	Amplify <i>UBI</i> to construct pRGE32 and pRGE32
UBI-R: (SEQ ID NO:40)	CAAAC TTGTGATACTATCTGCAAGAAATAATC ACCAAAC	
UGW-U3-F (SEQ ID NO:41)	GACCATGATTACGCCAAGCTTAAGGAATCTTTA AACATACG	Construct pRGE32
UGW-gRNA-R (SEQ ID NO:42)	GGACCTGCAGGCATGCACGCGCTAAAAACGGA CTAGC	
MPK1-F (SEQ ID NO:43)	GGGTCGGGCACAGCATCTC	Genotyping of MPK1
MPK1-R2 (SEQ ID NO:44)	TGCGCCTAAAAATCGAGGGT	
MPK2-F (SEQ ID NO:45)	TTTGGGAAGCATGTATGAAGC	Genotyping of MPK2
MPK2-R2 (SEQ ID NO:46)	TATGCCAGCCAATGAGCCAA	
MPK5-F256 (SEQ ID NO:47)	GCCACCTTCCTTCCTCATCCG	Genotyping of MPK5
MPK5-R611 (SEQ ID NO:48)	GTTGCTCGGCTTCAGGTGCG	
MPK6-F (SEQ ID NO:49)	TTGACGCCCCAACATAAATAA	Genotyping of MPK6
MPK6-R (SEQ ID NO:50)	TGTTGCTGCCGCTTTTCT	

PDS-F (SEQ ID NO:51)	GGTAGAAATGCCATGCGGGA	Genotyping of PDS
PDS-R (SEQ ID NO:52)	ATTCAGCCGAACCTCACCAC	
L5AD5-F (SEQ ID NO:53)	CG GGTCTC A GGCA <u>GGATG GGCAGTCTG</u> <u>GGCA</u> ACAAAGCACCACTGG	PTG synthesis and cloning A Fok I site (underlined) was used to generate compatible overhangs (labeled with red color) for cloning into pRGE32 and pRGE32.
L3AD5-R (SEQ ID NO:54)	TA GGTCTC C AAAC <u>GGATG AGCGACAGC AAAC</u> AAAAAAAAA GCACCGACTCG	
S5AD5-F (SEQ ID NO:55)	CG GGTCTC A GGCA <u>GGATG GGCAGTCTG</u> <u>GGCA</u>	
S3AD5-R (SEQ ID NO:56)	TA GGTCTC C AAAC <u>GGATG AGCGACAGC AAAC</u>	

Table 2.3. Oligo nucleotides used to synthesize *PTG* genes

gRNA ID (Target)	Spacer/Protospacer (5'→3') ^a	Oligo ID	Sequence (5'→3') ^b
gRNA1 (MPK5)	AGATGTCG <u>TAGAG</u> CAGGTAC	gR1-F (SEQ ID NO:57)	TA <i>GGTCTCC</i> <u>TAGAG</u> CAGGTAC gtttagagctagaa
		gR1-R (SEQ ID NO:58)	AT <i>GGTCTCA</i> <u>TCTACGACATCT</u> tgcaccagccgggaa
gRNA2 (MPK5)	TCTACATCG <u>CCAC</u> GGAGCTC A	gR2-F (SEQ ID NO:59)	TA <i>GGTCTCC</i> <u>CCACGGAGCTCA</u> gtttagagctagaa
		gR2-R (SEQ ID NO:60)	AT <i>GGTCTCA</i> <u>GTGGCGATGTAGA</u> tgcaccagccgggaa
gRNA3 (MPK1)	ATCCAGGC <u>GACG</u> CTGAGCC A	gR3-F (SEQ ID NO:70)	TA <i>GGTCTCC</i> <u>GACGCTGAGCCA</u> gtttagagctagaa
		gR3-R (SEQ ID NO:71)	AT <i>GGTCTCA</i> <u>CGTCGCCTGGAT</u> tgcaccagccgggaa
gRNA4 (MPK1)	TGGCC <u>ACCG</u> GGGTATAAAA	gR4-F (SEQ ID NO:72)	TA <i>GGTCTCC</i> <u>ACCGGGGTATAAAA</u> gtttagagctagaa
		gR4-R (SEQ ID NO:73)	CG <i>GGTCTCA</i> <u>CGGTGGGCCA</u> tgcaccagccggg
gRNA5 (MPK2)	GAACCCG <u>GTCTG</u> CCTCAAGG A	gR5-F (SEQ ID NO:74)	TA <i>GGTCTCC</i> <u>GTCTGCCTCAAGGA</u> gtttagagctagaa
		gR5-R (SEQ ID NO:75)	CG <i>GGTCTCA</i> <u>CGACCGGGTTC</u> tgcaccagccggg
gRNA6 (MPK2)	GAATGCG <u>CAGA</u> CTCGTCAG G	gR6-F (SEQ ID NO:76)	TA <i>GGTCTCC</i> <u>CAGACTCGTCAGG</u> gtttagagctagaa
		gR6-R (SEQ ID NO:77)	CG <i>GGTCTCA</i> <u>TCTGCGCATTC</u> tgcaccagccggg
gRNA7 (MPK6)	GTGTCCCG <u>TTGG</u> CATCGATA	gR7-F (SEQ ID NO:78)	TA <i>GGTCTC</i> <u>C TTGGCATCGATA</u> gtttagagctagaa
		gR7-R (SEQ ID NO:79)	AT <i>GGTCTCA</i> <u>CCAAGCGGACAC</u> tgcaccagccgggaa
gRNA8 (MPK6)	GCGGGTG <u>CGGG</u> TCAATCAA A	gR8-F (SEQ ID NO:80)	TA <i>GGTCTCC</i> <u>CGGGTCAATCAA</u> gtttagagctagaa
		gR8-R (SEQ ID NO:81)	CG <i>GGTCTCA</i> <u>CCCGCACCCGC</u> tgcaccagccggg
gRNA7	Linkers for level 2 GG assembly	Ln-gR7-F (SEQ ID NO:82)	TA <i>GGTCTC</i> <u>C TTGGCATCGATA</u>
		Ln-gR7-R (SEQ ID NO:83)	AT <i>GGTCTC</i> <u>A CCAAGCGGACAC</u>
gRNA9 (PDS)	ACAAGC <u>CAGG</u> AGAATTCAGC	gR9-F (SEQ ID NO:84)	TA <i>GGTCTC</i> <u>C CAGGAGAATTCAGC</u> gtttagagctagaa
		gR9-R (SEQ ID NO:85)	CG <i>GGTCTC</i> <u>A CCTGGCTTGT</u> tgcaccagccggg
gRNA10 (PDS)	CACTGC <u>ATGG</u> ATAACTCATC	gR10-F (SEQ ID NO:86)	TA <i>GGTCTC</i> <u>C ATGGATAACTCATC</u> gtttagagctagaa
		gR10-R (SEQ ID NO:87)	CG <i>GGTCTC</i> <u>A CCATGCAGTG</u> tgcaccagccggg

^a The boxed letters indicate the overhang sequences in Golden Gate assembly.

5 ^b The first two letters are randomly added nucleotides. Italic bold sequences indicate the BsaI sites (5'-GGTCTCN-3', N indicates any nucleotide), underlined sequences are part of gRNA spacer whereas bold-underlined sequences are overhangs after BsaI digestion. See SI Methods for details about primer design and PTG assembly. Sequences in lower case are specific for gRNA scaffold (5'-gttttagagctagaa-3', in forward primers) or tRNA (5'-tgcaccagccggg-3', in reverse primers).

10

Table 2.4. Determination of chromosomal fragment deletion frequency at the *MPK5* locus in rice protoplasts by qPCR

Sample	Gene	Ct mean	Ct SD	RQ	RQ-Min	RQ-Max	Del efficiency (100%-RQ)
PTG6	<i>MPK5</i>	19.70	0.12	74%	61%	89%	26%
PTG7	<i>MPK5</i>	20.51	0.06	77%	65%	92%	23%
PTG9	<i>MPK5</i>	19.70	0.05	89%	82%	97%	11%
CK	<i>MPK5</i>	19.61	0.18	100%			
PTG6	<i>UBI</i>	20.02	0.12				
PTG7	<i>UBI</i>	20.84	0.15				
PTG9	<i>UBI</i>	20.30	0.04				
CK	<i>UBI</i>	20.38	0.14				

5 The fragment deletion efficiency at the *MPK5* locus was estimated with qPCR using genomic DNA as a template and a pair of specific primers that encompass the gRNA2 cut site within *MPK5* (Fig. S1A). Because the *MPK5* locus with fragment deletion would not be amplified, the deletion (Del.) efficiency could be estimated as 100%-RQ. The same genomic DNAs in Fig. 3B were used and the *UBI* gene serves as the reference for relative quantification. Ct, threshold cycle; SD, standard deviation; RQ, relative quantity; RQ-Min and RQ-Max indicate the 95% confidential interval of RQ.

EXAMPLE 3

Polycistronic tRNA and CRISPR guide-RNA enables highly efficient multiplexed genome engineering in human cells

15 During the past three years, the clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein nuclease has emerged as the most powerful tool for genome editing in many organisms (1-4). The most commonly used Cas endonuclease for genome editing is *Streptococcus pyogenes* Cas9 (referred as Cas9 hereafter). The Cas9 protein is guided by an artificial short guide RNA (gRNA) to cleave the DNA whose sequence is complement to the 5'-end of gRNA and preceded with the protospacer-adjacent motif (PAM, 5'-NGG-3') (5-7). This simple RNA guided DNA targeting

system fundamentally enhanced our ability to access specific genomic sites for genetic manipulation. The CRISPR-Cas9 system has been engineered for targeted mutagenesis, site-specific integration of DNA fragment, and precise manipulation of chromosomes such as large segment deletion and translocations, etc. The nuclease defective Cas9 (dCas9) and gRNA were also engineered to control
5 expression of targeted genes (8-11) and labeling specific loci of the chromosomes (12) in vivo. These Cas9 based tools greatly facilitate basic research and the practice of biotechnology in various medical and agricultural fields.

The robust and efficient Cas9-mediated genome engineering requires control of the expression of Cas9 and gRNA in vivo. The expression of Cas9 could be readily and precisely controlled using DNA
10 Polymerase II promoter, while the efficient expression of gRNA, which is a small non-coding RNA, remains a bottleneck. Particularly, many Cas9 based applications require simultaneous expression of multiple gRNAs. For example, a pair of gRNAs are needed to edit one site in the Cas9 nickase (13,14) and dCas9-FokI (15,16) mediated genome editing which helps reduce off-targeting risk associated with the original CRISPR-Cas9 system. In dCas9 mediated transcriptional regulation, multiple gRNAs are
15 required for robust gene activation or suppression, or generating sophisticated device of transcriptional circuit (10,17,18). Generally, the gRNA expression is driven by polymerase III promoters such as the widely used snoRNA U3 and U6 gene promoters. However, these Pol III promoters transcribed gRNA from specific nucleotides. For example, U3 promoter transcribed from “A” and U6 promoter starts a transcript with “G”. This restrained the targeting spacer of Cas9-gRNA. Therefore, a more robust gRNA
20 expression strategy is needed to boost the capability of CRISPR-Cas9 tools.

Recently, we demonstrated that the endogenous transfer RNA (tRNA) processing system could be engineered to boost Cas9 multiplex targeting capability (19). Multiple gRNAs were expressed in tandem from an artificial polycistronic-tRNA-gRNA (PTG) gene. The endogenous RNases P and Z recognize tRNA and precisely cleave the PTG to release tRNAs and gRNAs. This system not only
25 enables simultaneous expression of many gRNAs but also enhances the Pol III transcription since the tRNA also acts as an internal enhancer or promoter. The PTG system was implemented in plants but we speculated that it would be functional in all organisms because the tRNA processing system is highly conserved in all living organisms. In this study, we adapted the Cas9-PTG technology for human

genome editing. The PTG genes was made to simultaneously express 2 to 6 gRNAs for precise genome editing of one to four histone deacetylase (HDAC) genes. Our study demonstrates that the tRNA system could be engineered as a versatile tool for efficient and multiplex genome editing in human and animal systems.

5 MATERIALS AND METHODS

Plasmid vector construction

To construct pSicoR-sgRNAs-mCherry-Cas9 vector, the puromycin fragment in pSico-EF1-mCh-Puro (Addgene Plasmid 31845) (20) was fused with the NLS-Cas9-NLS fragment from pX260 (Addgene Plasmid 42229). Then the gRNA scaffold with the cloning site containing two
10 *BbsI* sites was inserted downstream of the U6 promoter. V The schematic of the pSico-sgRNA-mCherry-Cas9 vector is shown in Figure 18.

Assembly of PTG genes

To edit HDAC genes (*HDAC1/2/3/4/6*) in human cells, a pair of target sites separated by 250-500 bp was selected for each gene based on the presence of PAM and protospacer sequence specificity
15 (Table 3.1). A total of seven PTG genes (hPTG1 to hPTG7, Table 3.2) were assembled in vitro using synthesized oligonucleotides (Table 3.3). The in vitro synthesis of hPTG genes was performed as we described previously (19).

Cell cultures and transfection

Human Embryonic Kidney 293 (HEK293) cells were cultured in DMEM (Sigma) with 10% fetal
20 bovine serum (Atlanta Biologicals), 10 unit/mL penicillin and 10 µg/mL streptomycin (Gibco). The cultural medium was replaced every 2 days and was sub-cultured when cells reach to 80% to 90% confluency.

For transfection, HEK293 cells were grown in six-well plates at 3×10^5 cells per well. After reaching 70% confluency, cells were transfected with plasmids and polyethylenimine (PEI). Briefly, cells were
25 fed with fresh culture medium one hour prior to transfection. And then, 3 µg of plasmid DNA and 12

µg of PEI were added to each well. Transfected cells were incubated for 5 hours in a CO₂ incubator and cultured for additional 48 hours with fresh culture medium.

cRT-PCR and quantitative RT-PCR

5 The total RNAs were extracted from HEK293 cells with Trizol Reagent (Life Technologies) and treated with DNase I (New England Labs). To amplify the gRNA, cRT-PCR was performed as we described previously. Briefly, the total RNA was self-ligated with T4 RNA ligase (New England Labs) and then purified with Trizol Reagent. The ligated total RNA was reverse transcribed using gRNA specific primers and MMLV reverse transcriptase (New England Labs). The cDNA was amplified using
10 specific primers (See Table 3.3 for sequences) and cloned into pGEM-T easy (Promega) for Sanger sequencing.

DNA extraction and PCR

Transfected cells were washed with phosphate buffered saline and lysed in Lysis buffer (10mM Tris pH 8.0, 100 mM NaCl and 10mM EDTA) and genomic DNA was then precipitated with
15 isopropanol. To detect fragment deletion between two target sites within each gene, PCR primers encompass the targeting sites were designed for each HDAC gene (Table S2). PCR amplifications were performed in reactions containing 800 ng genomic DNA, 0.2 mM dNTP, 0.4 uM primers, 1x PCR buffer and 1 unit DreamTaq DNA polymerase (Thermo Fisher Scientific). PCR product was separated in 1.5% agarose gel containing ethidium bromide to detect chromosomal fragment deletion. The fragment
20 deletion efficiency was estimate with Image J. In addition, the PCR fragment was cloned into pGEM-T easy (Promega) for Sanger sequencing.

RESULTS AND DISCUSSION

PTG enables efficient production of multiple gRNAs in human cells

25 We demonstrated previously that a PTG with tandem-arrayed tRNA-gRNAs not only allowed precise processing and efficient production of multiple gRNAs, but also enabled Cas9 to simultaneously

target multiple genomic sites in plants. We speculated that the *PTG* genes should also boost CRISPR/Cas9 targeting efficiency in animals because tRNA and its processing system is highly conserved in all living organism. To demonstrate the utility of PTG for genome editing in the animal system, we first examined the gRNA excision accuracy from *hPTG1* and *hPTG2* transgenes in human
5 cells. Each of the artificial PTG genes encodes two gRNAs. Whereas *hPTG1* expresses HDAC1-sg1 and HDAC2-sg2, *hPTG2* expresses HDAC2-sg1 and HDAC2-sg2 (Table 3.2).

After transfection of *hPTG1* and *hPTG2* constructs into HEK293 cells, the cRT-PCR was performed to map the 5' and 3' end of mature gRNAs. Predicted cDNA products with expected size of single gRNA (~96nt) were detected in cRT-PCR, despite the presence of some nonspecifically amplified
10 products. (Figure 19A). As shown in Figures 19A and 19C DNA sequencing of these cDNA products indicated that mature gRNAs were precisely processed from *hPTG* genes with desired 5'-end which contains targeting guide sequences. As we observed previously in plants, the gRNA derived from PTG has two additional nucleotides (5'-AA-3') at 3'-ends if it proceeds tRNA or have two additional T if it precedes a Pol III terminator. We also detected a putative polyadenylation at the 3'-end of mature
15 HDAC2-sg2). These results suggest that PTG genes could be processed by the human tRNA processing system to produce multiple gRNAs.

Cas9-PTG enables multiplex genome editing in human cells

To examine the efficiency of PTG method for multiplex genome editing, we transfected human cells with seven plasmid constructs expressing different *hPTG* genes and Cas9. These *hPTG* genes encode
20 multiple gRNAs targeting five HDAC genes located in different chromosomes. *hPTG1* to *hPTG5* was used to target two genomic sites for each gene within HDAC1, 2, 3, 4, and 6; and the *hPTG6* and *hPTG7* were designed to simultaneously target HDAC3 and 2, respectively (Table 3.2). Because we used two gRNAs to edit one gene, the efficiency of Cas9-PTG could be estimated by measuring the chromosomal fragment deletion frequency within each targeting gene. As predicted, truncated PCR
25 amplicons were detected in all samples (Figure 20). DNA sequence analysis further confirmed these chromosomal deletions were introduced by Cas9-PTGs in transfected cells (Figure 21). The deletion frequencies between two gRNA targeting sites were 40-50% based on the calculation of DNA band

intensity. Interestingly, the efficiency of different hPTG genes with variable number of targets were comparable. For example, hPTG2 (2 gRNAs) and hPTG6 (6 gRNAs) resulted a fragment deletion from HDAC2 locus at ~40% frequency despite their difference in gRNA numbers. Our results demonstrate that the PTG method can be used to efficiently express multiple gRNAs and simultaneous editing of multiple sites not only in plants, but also in human cells. Because the PTG strategy does not use multiple promoters and terminators, it drastically reduces the size of gRNA construct. As a result, it is more suitable and effective for various genome editing purposes such as the virus-mediated delivery of multiple gRNAs for human gene therapy.

10 REFERENCES

1. Doudna, J.A. and Charpentier, E. (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science*, **346**, 1258096.
2. Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262-1278.
- 15 3. Sander, J.D. and Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.*, **32**, 347-355.
4. Mali, P., Esvelt, K.M. and Church, G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat. Methods*, **10**, 957-963.
5. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
- 20 6. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823-826.
- 25 7. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819-823.

8. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A. *et al.* (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, **154**, 442-451.
- 5 9. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P. and Lim, W.A. (2013) Repurposing CRISPR as an RNA-Guided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173-1183.
- 10 10. Cheng, A.W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T.W., Rangarajan, S., Shivalila, C.S., Dadon, D.B. and Jaenisch, R. (2013) Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.*, **23**, 1163-1171.
- 11 11. Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L. and Church, G.M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.*,
15 **31**, 833-838.
12. Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S. *et al.* (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, **155**, 1479-1491.
- 20 13. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. *et al.* (2013) Double nicking by RNA-Guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380-1389.
14. Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A.,
25 Iyer, V., Huang, X. *et al.* (2014) Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods*, **11**, 399-402.
15. Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J. and Joung, J.K. (2014) Dimeric CRISPR RNA-guided FokI

- nucleases for highly specific genome editing. *Nat. Biotechnol.*, **32**, 569-576.
16. Guilinger, J.P., Thompson, D.B. and Liu, D.R. (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.*, **32**, 577-582.
 - 5 17. Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H. and Joung, J.K. (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods*, **10**, 977-979.
 18. Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W. *et al.* (2013)
 - 10 RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods*, **10**, 973-976.
 19. Xie, K., Minkenberg, B. and Yang, Y. (2015) Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *PNAS*, **112**, 3570-3575.
 - 15 20. Salomonis, N., Schlieve, C.R., Pereira, L., Wahlquist, C., Colas, A., Zambon, A.C., Vranizan, K., Spindler, M.J., Pico, A.R., Cline, M.S. *et al.* (2010) Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *PNAS*, **107**, 10514-10519.

20

TABLES AND FIGURE LEGENDS

Table 3.1. Target Sequences for genome editing

sgRNA	Guide sequence (5'→3')	Targets	Deletion
HDAC1-sg1 (SEQ ID NO:164)	GTAAGACCACCGCACTAGGC	<i>HDAC1</i>	298bp
HDAC1-sg2 (SEQ ID NO:165)	GCCCTGCAGCTATTACCATT		
HDAC2-sg1 (SEQ ID NO:166)	GTATTTTAGGATATTGGTGC	<i>HDAC2</i>	269bp
HDAC2-sg2 (SEQ ID NO:167)	TGTTTCAATCTAACAGTCAA		
HDAC3-sg1 (SEQ ID NO:168)	GAGCAGAACTCAAAGAGCCC	<i>HDAC3</i>	437bp
HDAC3-sg2 (SEQ ID NO: 169)	GCTCAAGTAAGTAGCCCAGG		
HDAC4-sg1 (SEQ ID NO:170)	GAGCTCCTGAATACGTCGCA	<i>HDAC4</i>	275bp
HDAC4-sg2 (SEQ ID NO:171)	GCTCCTTTGCCGTCCCCGAG		
HDAC6-sg1 (SEQ ID NO:172)	CGACTGGCAGCAGGACGTGC	<i>HDAC6</i>	331bp
HDAC6-sg2 (SEQ ID NO:173)	CAAGCTGATCCTGTCTCTGG		

5

Table 3.2. Structure of hPTG genes

Gene Name	Target Genes	# of gRNAs	Encoding gRNAs
<i>hPTG1</i>	<i>HDAC1</i>	2	HDAC1-sg1, HDAC1-sg2
<i>hPTG2</i>	<i>HDAC2</i>	2	HDAC2-sg1, HDAC2-sg2
<i>hPTG3</i>	<i>HDAC3</i>	2	HDAC2-sg1, HDAC2-sg2
<i>hPTG4</i>	<i>HDAC5</i>	2	HDAC4-sg1, HDAC4-sg2
<i>hPTG5</i>	<i>HDAC6</i>	2	HDAC6-sg1, HDAC6-sg2
<i>hPTG6</i>	<i>HDAC1, HDAC2,</i> <i>HDAC3</i>	6	HDAC1-sg1, HDAC1-sg2, HDAC2-sg1, HDAC2-sg2, HDAC2-sg1, HDAC2-sg2
<i>hPTG7</i>	<i>HDAC4, HDAC6</i>	4	HDAC1-sg1, HDAC1-sg2, HDAC6-sg1, HDAC6-sg2

5

10

15

Table 3.3A Primers used for the synthesis of hPTG

Oligo Name	Sequence (5' -> 3')
HDAC1-sg1-tR	CG GGTCTC A cggtggtcttac tgcaccagccggg
HDAC1-sg1-gF	TA GGTCTC C accgcactaggc gtttagagctagaa
HDAC1-sg2-tR	CG GGTCTC A tagctgcagggc tgcaccagccggg
HDAC1-sg2-gF	TA GGTCTC C gctattaccatt gtttagagctagaa
HDAC2-sg1-tR	CG GGTCTC A alcctaaaatac tgcaccagccggg
HDAC2-sg1-gF	TA GGTCTC C ggatattggtgc gtttagagctagaa
HDAC2-sg2-tR	CG GGTCTC A tagattgaaaca tgcaccagccggg
HDAC2-sg2-gF	TA GGTCTC C tctaacagicaa gtttagagctagaa
HDAC3-sg1-tR	CG GGTCTC A tgagtctgtctc tgcaccagccggg
HDAC3-sg1-gF	TA GGTCTC C ctcaaagagccc gtttagagctagaa
HDAC3-sg2-tR	CG GGTCTC A acttactgagc tgcaccagccggg
HDAC3-sg2-gF	TA GGTCTC C aagtagcccagg gtttagagctagaa
HDAC4-sg1-tR	CG GGTCTC A attcaggagctc tgcaccagccggg
HDAC4-sg1-gF	TA GGTCTC C gaatacgtcgca gtttagagctagaa
HDAC4-sg2-tR	CG GGTCTC A cggcaaaggagc tgcaccagccggg
HDAC4-sg2-gF	TA GGTCTC C gccgtccccgag gtttagagctagaa
HDAC6-sg1-tR	CG GGTCTC A tgctgccagtcg tgcaccagccggg
HDAC6-sg1-gF	TA GGTCTC C agcaggacgtgc gtttagagctag
HDAC6-sg2-tR	CG GGTCTC A aggatcagcttg tgcaccagccggg
HDAC6-sg2-gF	TA GGTCTC C tcctgtctctgg gtttagagctagaa
HDAC3-sg2-InF	TA GGTCTC C ctcaaagagccc
HDAC3-sg2-InR	CG GGTCTC A acttactgagc
hL5AD6-F	CG GGTCTC A TGTT GGATG GGCAGTCTG TGTT acaagcaccagtg
hS5AD6-F	CG GGTCTC A TGTT GGATG GGCAGTCTG TGTT

Table 3.3B Primer sequences for genotyping

Primer Name	Sequence (5'→3')	Gene	Amplicon size
hHDAC1-del-gPCR-F	gagagggaggccattctagg	<i>HDAC1</i>	505bp
hHDAC1-del-gPCR-R	gtccagacaaaaagcaggag		
hHDAC2-del-gPCR-F	tttgattgtccggtctcaac	<i>HDAC2</i>	531bp
hHDAC2-del-gPCR-R	agtatgttggggcaaaatgg		
hHDAC3-del-gPCR-F	tagcccacgtgtgtaggagag	<i>HDAC3</i>	559bp
hHDAC3-del-gPCR-R	catggtggatgtcaatgtcaa		
hHDAC4-del-gPCR-F	gtggggtgtacgtgtgtctct	<i>HDAC4</i>	547bp
hHDAC4-del-gPCR-R	cacacgcaaagtagatgtgga		
hHDAC6-del-gPCR-F	ctctaactggtccaccacagg	<i>HDAC6</i>	493bp
hHDAC6-del-gPCR-R	tcaggaaaggaccagacatga		

EXAMPLE 4

Characterization of heritable mutations on multiple closely related rice genes created by the PTG/Cas9 mediated genome editing

5 In this study we showed that PTG/Cas9 is a powerful tool to mutate closely related genes by creating one to eight mutations at different combinations of four rice *MITOGEN-ASSOCIATED PROTEIN KINASE* (MPK) genes with efficiencies of 66 to 100%. PTG/Cas9 was removed by self-pollination and the transgene-free T₁ and T₂ plants carried all eight mutations or chromosomal deletions. The induced mutations resulted from
10 multiple base pairs of deletion or insertion and sometimes chromosomal deletions. We Demonstrated that PTG/Cas9 reliably produces multiple mutations which could be stably transmitted to the subsequent generations.

 New genome editing tools recently promise to mutate almost any gene in a given genome. Mutation of genes is achieved by inducing a double strand break (DSB) at the
15 target site that is be repaired by the cell. The imperfect non-homologous end-joining (NHEJ) pathway is the prevalent DSB repair pathway and its erroneous nature leads to insertions or deletions (indels) of nucleotides during the repair. These indels can cause a gene knock-out if the mutations prevent or alter the gene's transcription or translation. A more precise repair via homologous recombination is theoretically possible if a donor
20 template with sufficient homology to the DSB site is present, but higher plants possess a low intrinsic homology recombination rate (Voytas, 2013). The primary tools to achieve site-directed DSBs in a genome are Zinc Finger Nucleases (ZFNs), Transcription Activator-Like Nucleases (TALENs), and most recently the bacterial type II clustered regularly interspaced short palindromic repeat (CRISPR) and CRISPR-associated protein 9
25 nuclease (Cas9) system. The targeting ability of ZFNs and TALENs relies on protein:DNA interaction and therefore new binding domains need to be designed to change the target site of the nucleases (Voytas, 2013). The CRISPR/Cas9 system is much easier to program because the Cas9 endonuclease can be recycled to cut a different target site based on

RNA:DNA interactions of the so-called single guide RNA (gRNA; Jinek et al., 2012). The 20 bases at the 5'-end of the gRNA (gRNA spacer) are complementary to the genomic target DNA sequence (protospacer) next to the protospacer adjacent motif (PAM, 5'-NGG-3' in case of the *Streptococcus pyogenes* Cas9; Jinek et al., 2012; Cong et al., 2013; Mali et al., 2013). The CRISPR/Cas9 system has been successfully used in genome editing of *Arabidopsis*, tobacco, *N. benthamiana*, potato, tomato, soybean, sweet orange, liverwort, maize, sorghum, wheat, and rice to knock-out genes (Brooks et al., 2014; Cai et al., 2015; Jia and Wang, 2014; Jiang et al., 2013; Li et al., 2013; Miao et al., 2013; Shan et al., 2013; Sugano et al., 2014; Xie and Yang, 2013). In addition to its easy design, the CRISPR/Cas9 system can be used to target almost any gene. When using *SpCas9*, 97.1% of all transcription units in *Arabidopsis* and 89.6% of the transcription units in rice can be targeted with highly specific gRNA spacers (Xie et al., 2014a). The remaining genes could still be targeted by Cas9 variants with different PAM-requirements for specific gRNAs.

One of the CRISPR/Cas9 system's biggest advantages might be its ability for multiplexed genome editing. The Cas9 endonuclease can theoretically target multiple targets simultaneously if several gRNAs are co-expressed. This enables researchers to target multiple genes for knock-out and to create double, triple, or even decuple mutants in a single step. However, most multiplex editing vectors use several gRNA expression cassettes that drive each gRNA with its own Pol III promoter and terminator (Li et al., 2013; Lowder et al., 2015; Ma et al., 2015a). Each single gRNA expression cassette has a typical length of 500 to 820 bp and leave multiplex genome editing in plants still a challenge because the number of simultaneous expressible gRNAs is limited by vector capacity and cloning efficiency. Even if more than six of these expression cassettes can be stacked, the assembly is inefficient and frequently fails (Lowder et al., 2015). Common plant transformation techniques typically deliver the genome-editing device as plasmids via protoplast transfection or particle bombardment, or as T-DNA via *Agrobacterium*. Expressing multiple gRNAs from a single transcript instead of multiple individual cassettes is a promising alternative to overcome this limitation.

We previously enhanced the multiplexing capability of CRISPR/Cas9 genome-editing by introducing a polycistronic tRNA-gRNA gene (PTG) that produces multiple gRNAs from a single artificial gene (Figure 1; Xie et al., 2015). This gene consists of tandemly arrayed tRNA-gRNA sequences preceded by a Pol III promoter and followed by a Pol III terminator. The tRNA sequence used, rice glycine pre-tRNA as disclosed in Examples 1 and 2 has been used successfully in various organisms, including *C. elegans*, fruit fly and mammalian cells for multiplex genome editing. Every tandem repeat is less than 180 bp long, compared to at least 500 bp needed for an individual cassette. Each gRNA can contain a unique spacer that recognizes a different target, enabling simultaneous targeting of multiple genomic sites in one transformation (Figure 22). The endogenous RNase P and RNase Z recognize the tRNA's secondary structure (Barbezier et al., 2009; Canino et al., 2009; Gutmann et al., 2012; Phizicky and Hopper, 2010; Schiffer et al., 2002) and cut the tRNAs at specific sites to release the mature gRNAs that direct Cas9 endonuclease to multiple specific targets (Figure 22). In addition, expression of gRNAs with a PTG is up to 30 times higher than expression with a simple Pol III promoter, probably because the A- and B-box of the tRNA can enhance transcription of the PTG primary transcript (Xie et al., 2015). The ability to simultaneously mutate several genes is especially useful to analyze closely related genes, which tend to have similar or overlapping functions. Researchers currently need to find reliable mutant lines and cross them to obtain double or higher order mutants. The PTG/Cas9 technology may significantly improve analysis of closely related genes because a high number of multiple genes can be efficiently targeted in a single transformation.

In this study, we tested the ability of PTG/Cas9 to create mutant resources for studying closely related genes by successfully mutating four members of the rice TEY-type *MITOGEN-ASSOCIATED PROTEIN KINASE (MPK)* gene family in single, double, and quadruple mutant combinations with up to eight co-expressed gRNAs. MPKs are important signal transducers that are known to have partly overlapping functions in *Arabidopsis* (Beckers et al., 2009; Pitzschke et al., 2009). We furthermore show that the induced

mutations have a high variety and that mutations and chromosomal deletions are reliable inherited into transgene-free future generations. We discovered on the example of *MPK1* that essential genes are harder to knock out as the mutations tend to preserve the existing open reading frame. This study demonstrates that closely related genes could be simultaneously edited and mutated with the multiplexed PTG/Cas9 technology and transgene-free, multiple gene mutants could be obtained via stable inheritance for studying complex gene families and gene networks.

RESULTS AND DISCUSSION

Highly efficient targeting of four closely related rice *MPK* genes with PTG/Cas9

We designed gene constructs for *Agrobacterium*-mediated transformation to mutate single and combinations of stress-responsive *MPKs* in rice and to explore if PTGs can reliably produce stable mutations to dissect the function of closely related genes and gene families (Table 1). The orthologous *MPK* genes in *Arabidopsis* overlap in their function and partly compensate for each other in single mutants (Asai et al., 2002; Beckers et al., 2009). The PTG/Cas9 carrying gene constructs were designed to target single *MPK* genes or combinations of them in double and quadruple mutants (Table 4.1). PTGb3, PTGb4, PTGb5, and PTGb6 carried Cas9 driven by a rice ubiquitin promoter (*UBIp:Cas9*) and PTGs driven by a rice U3 promoter (*U3p:PTG*) to target *MPK1*, *MPK2*, *MPK6*, and *MPK5*, respectively, for creation of single mutants (Table 4.1). PTGs for single mutants encoded two gRNAs targeting one gene. An additional construct, PTGb2, targeted *MPK5* with only one gRNA at protospacer 2 (PS2; Figure 17). PTGb7 and PTGb8 contained PTGs with each four gRNAs targeting *MPK5/MPK1* and *MPK6/MPK2* to create double mutants (Table 4.1). We chose these pairs of *MPK* genes based on their close phylogenetic relationship in the rice *MPK* family (Reyna and Yang, 2006). Creating double mutants of rice *MPK* genes might help to uncover similar redundant functionality as previously observed in *Arabidopsis*. PTGb9 included a PTG of eight gRNAs to mutate all four closely

related rice *MPK* genes in a single transformation event.

Our previous study confirmed the functionality of all gRNAs used in rice protoplasts and determined the mutation efficiency (86 – 100%) of PTGb6 (*MPK5*), PTGb7 (*MPK5/1*), and PTGb9 (*MPK5/1/6/2*) in transgenic rice plants produced via the *Agrobacterium*-mediated transformation (Xie et al., 2015). In this study, we targeted *MPK1*, *MPK2*, *MPK6* and *MPK6/2* to investigate if we could mutate these single and combination of *MPK* genes with a similar high percentage of genome-edited plants. Transformation with PTGb3 (*MPK1*), PTGb4 (*MPK2*), and PTGb5 (*MPK6*) each yielded a 100% genome-editing efficiency for the tested lines (Table 4.1; Figures 28 and 29). However, the efficiencies were lower when targeting *MPK6* and *MPK2* simultaneously (PTGb8) with 83% for the *MPK6* and 66% for the *MPK2* locus (Figure 30). We tested the transformation efficiency at only one of the two target sites per gene because only one protospacer provided a convenient restriction enzyme (RE) site for PCR-RE assay (Figure 27). Therefore, the actual editing efficiency for PTGb8 could be higher than the measured 83% and 66% if all four target sites would be considered. The main objective of the PTGb8 transformation was to obtain *mpk6/mpk2* double mutants. A high percentage of 66% showed genome editing at both genes, which yielded putative double mutants (Figure 30). Out of these putative double mutant lines, 75% had biallelic mutations on *MPK6* and *MPK1* simultaneously.

While transformation with PTGb4, PTGb6, PTGb2, PTGb7, PTGb8 and PTGb9 produced an abundance of hygromycin-resistant and putative edited callus cells, efforts to mutate only *MPK1* with PTGb3 or *MPK6* with PTGb5 were less successful. We could only recover four hygromycin-resistant calli for PTGb3 and two resistant calli for PTGb5 out of a total of 600 transformed calli for each construct (three independent repeats of 200 calli for each vector). Nevertheless, we showed that the small percentage of recovered lines was mutated with an efficiency of 100% (Table 4.1; figure 28). Therefore, PTG/Cas9 can be designed to induce gene specific mutations at closely related genes and combinations of those with efficiencies of 66 – 100%. One critical objective to study phenotypes that might

be masked by redundant gene is to successfully obtain multiple mutated genes in a single line. The analysis of PTGb8 showed that 50% of all obtained lines carried biallelic mutations at both targeted genes. Our previous study achieved an even higher efficiency when editing four genes simultaneously with 86% of all lines carrying biallelic mutations at all four targeted genes (Xie et al., 2015). This high frequency of biallelic mutations on all targets should minimize the time and effort to screen multiple knock-out events. Our results indicate that PTG/Cas9 technology is a feasible tool to investigate the function of gene families or redundant genes by multiple gene knock-out.

10 **Transgene-free T₁ rice plants inherited all eight simultaneously induced mutation from their parents**

After we previously showed that PTG/Cas9 can simultaneously induce mutations at eight genomic sites using only a single transformation event (Xie et al., 2015), we now investigated if these mutations could be inherited into the next generation when the genome-editing device was removed. We chose to analyze the seeds of self-pollinated T₀ PTGb9 plants that previously showed biallelic mutations. The T₁ generation of these biallelic mutants is easy to genotype because five of the eight targeted genomic sites encompass RE sites that were destroyed at both alleles by the introduced mutations (Figure 23; Figure 27). Self-pollination could also remove the genome-editing device through the genetic segregation of the T-DNA fragment. Any mutations detected in these transgene-free plants should be the result of inheritance because the genome-editing device is no longer present to induce new mutations. We analyzed one T₁ generation plant of four different PTGb9 lines with three plants being transgene-free. The functional PTG/Cas9 system consists of a multiple gRNA encoding PTG and the Cas9 protein. A transgene-free plant should not contain any DNA encoding either the PTG or the Cas9 protein. PCRs with primers amplifying the *U3:PTG* cassette and a 1 kb fragment of the Cas9 gene confirmed that plants 2-1, 3-2 and 4-2 lack the genome-editing device (Figure 31, blue arrows). In contrast, the transgenic parent plants (T₀) of all four T₁ lines showed the presence of

U3:PTG and Cas9 DNA in their genomic DNAs (figure 31). Primers amplifying an endogenous rice gene confirmed that the DNA in all samples was of good quality because the PCR product could be readily amplified from genomic DNAs extracted from transgenic and transgene-free plants (figure 30, *control*).

5 To genotype the mutation of the T₁ generation lines, we then amplified pieces of the four genes that encompass the targeted eight protospacers for a PCR-RE assay. The T₁ progeny of biallelic T₀ PTGb9 lines should produce completely indigestible PCR product during the PCR-RE assay if both mutated alleles are inherited as expected. Indeed, the PCR-RE assay found that the mutation carried at the protospacers PS3, PS5, PS2, PS1, and
10 PS7 (figure 27) were normally transmitted to the T₁ generation (Figure 23A, 23B). While the appropriate RE digested the PCR product from wildtype DNA, the treatment did not affect the PCR product from any of the tested progeny of biallelic T₀ plants, indicating mutations that destroy the RE site (Figure 23A, red arrows). We further tested if mutations could also be found at the remaining three targeted sites. Because it is not possible to detect
15 mutation at PS4, PS6, and PS8 (Figure 22) with a PCR-RE assay and a T7 Endonuclease I assay is not suitable to detect homozygous mutations (Xie et al., 2014b), we decided to directly sequence the PCR products. The results from direct sequencing of PCR products can also inform about the zygosity of the T₁ mutant plants. The sequencing result consists of distinct single peaks if both alleles carry the same mutation and are homozygous. If the
20 gene carries a different mutation on each allele and is heterozygous, the result will consist of ambiguous double peaks, usually starting from Cas9 targeting site. Double peaks from heterozygous mutations could be deciphered with degenerate sequence decoding (Ma et al., 2015b). Sequencing the PCR products revealed mutations at all eight genomic target sites of the four *MPK* genes in T₁ generation plants of PTGb9 and confirmed that eight
25 mutations are faithfully inherited into future generations (Figure 23B). However, each plant had a different degree of heterozygosity. The plant 2-1 carried eight homozygous mutations at all four genes. Plant 3-2 was homozygous for *mpk2*, *mpk5*, and *mpk6* but heterozygous for PS4 on *mpk1*. Plant 1-2 was homozygous for *mpk2* and *mpk5*, but heterozygous on all

four PS regions of *mpk1* and *mpk6*. The plant 4-2 carried homozygous mutations for *mpk1* and *mpk5*, but is heterozygous for PS6 of *mpk2* and PS7 of *mpk6* (Figure 18b). In addition to carrying eight mutations, three of the examined T₁ generation plants were also transgene-free as seen by the lack of detectable *U3:PTG* and Cas9 fragments from genomic DNA (figure 31).

We showed that PTG/Cas9 technology is a highly efficient tool to mutate several members of a gene family simultaneously and with efficiencies of up to 100% (Table 4.1) and that the transgene-free T₁ generation carries mutations at all targeted sites (Figure 23; figure 30). Recently, Lowder et al. (2015) showed that their module system based on individual expression cassettes for only one gRNA (each cassette is up to 820 bp long) can be stacked to eight cassettes, but the assembly of six or more cassettes often failed and was inefficient. In addition, the work only analyzed constructs expressing up to three gRNAs simultaneously and it is unknown if expression of more gRNAs with individual cassettes results in satisfying genome-editing efficacy (Lowder et al., 2015). Our PTGs, on the contrary, express multiple gRNAs from a single expression cassette and hijack the robust endogenous tRNA processing machinery to convert the primary transcript into multiple functional gRNAs (Figure 22). Even when eight gRNAs were simultaneously expressed from one PTG, an editing efficiency of 86% was achieved in all eight target sites in the stable transgenic lines (Table 1; Xie et al., 2015). Therefore, PTG/Cas9 is a highly efficient genome-editing device that enables rapid and reliable mutagenesis of multiple genes.

PTG/Cas9 produces a variety of mutations

Most of the previous studies using the CRISPR/Cas9 system in *Arabidopsis* and rice reported a strong tendency to 1 bp indel mutations in the targeted genes of stably transformed T₀ plants (Endo et al., 2014; Feng et al., 2014; Hyun et al., 2014; Mikami et al., 2015; Zhang et al., 2014). In contrast, other studies identified longer deletions (≥ 3 bp) as main mutational type (Xu et al., 2015; Zhou et al., 2014). We analyzed all available sequencing results (n=54) from PTG/Cas9-mutated T₀ lines to investigate if our multiplex

genome-editing system produces mainly 1 bp indels or a higher variety of mutations. PTG/Cas9-induced mutations in a total of 54 independent sites were slightly biased towards 1 bp insertions (29.6%; Figure 24), but to a much lower degree than previous reports (37 - 54%; Feng et al., 2014; Zhang et al., 2014). A recent study found TALENs (transcription activator-like effector nucleases) to produce 69.9% deletions with 81.1% affecting multiple basepairs (Zhang et al., 2015). However, the overall mutation efficiency TALENs reached only 25% even though the scaffolds were optimized for targeting rice (Zhang et al., 2015). In our PTG/Cas9 system deletions accounted for 64.8% of the detected mutations (Figure 24) with 59.3% of all detected mutations affecting multiple base pairs. This mutation rate shows that PTG/Cas9 enriches the variety of mutations similar to TALENs (Figure 24), but advantageously exhibits a much higher targeting efficiency on even multiple genomic sites with 66-100% (Table 4.1). Deletions displayed the highest variety in mutation types with up to 74 bp affected (Table 4.2). In contrast, insertions and conversions were only observed for up to 4 or 3 bp, respectively. The variety in mutations was greater for the *MPK5* and *MPK6* locus (74 to 1 and 48 to 1 bp) compared to *MPK1* and *MPK2* (15 to 1 bp and 11 to 1 bp), suggesting that the target region at least partially influences the mutational variety (Table 4.2).

Zhang et al. (2015) claimed that the difference in mutation types between TALENs and CRISPR/Cas9 is due to the nature of the DSB. TALENs produce two spaced-out single strand breaks to induce the DSB, while the Cas9 protein cuts both strands at the same position. However, it seems possible to shift the mutational pattern of CRISPR/Cas9 by boosting its editing efficiency (Table 4.2; Figure 24). Early results of CRISPR/Cas9 in protoplasts, where expression of the components is assumed to be high, showed a great variety of mutational patterns (Li et al., 2013; Shan et al., 2013; Xie and Yang, 2013). The PTG system drastically increases the amount of gRNAs inside the cell compared to the conventional method (Xie et al., 2015). We hypothesize that the high titer of gRNAs causes the observed high efficiency and enrichment for multiple base pair mutations. The PTG/Cas9 system is a preferable tool to TALENs, even when it is necessary to mutate

multiple base pairs, because it provides an easy-to-use multiplex genome-editing device with high efficiencies and a variety of mutations. Furthermore, it is possible to program the PTG/Cas9 system to delete several smaller or larger chromosomal fragments. This enables researchers to remove whole genes or regulatory elements from the genome.

5

PTG-induced chromosomal deletions of rice *MPK* genes are faithfully inherited into T1 and T2 generations

PTGs provide a convenient and efficient way to express multiple gRNAs simultaneously. Targeting a chromosomal region with two gRNAs can result in the deletion of the fragment between both sites. Our previous research showed that PTG/Cas9 can
10 induce chromosomal deletion in transiently transformed protoplasts and in stably transformed rice callus (Xie et al., 2015). But so far it was unknown if next generations would inherit these PTG-induced deletions, and if the inheritance follows mendelian genetics. We analyzed the T₁ progeny of PTGb9 lines that were heterozygous for a
15 fragment deletion in *MPK5* to address this question. We assumed a full length allele if the PCR product of primers flanking both target sites was the same size as predicted from the wildtype sequence. On the other hand, a 727 bp smaller PCR product would indicate a chromosomal deletion. Simultaneous occurrence of both bands was interpreted as heterozygosity. The T₀ PTGb9 lines 4, 5, and 6 carried a full sized *MPK5* gene on one
20 allele and a copy with a 727 bp deletion on the other allele, as confirmed by sequencing (Figure 25A; figure 32). The deletion occurred as exact joining of the predicated PS1 and PS2 breaking sites three bp upstream of the protospacer-adjacent motifs (figure 32). Line 3 was selected as control and is homozygous for the full length gene (Figure 25A).
Genotyping of two T₁ plants from each line already revealed that the chromosomal deletion
25 were faithfully inherited (Figure 25B). T₁ plants 4-2 and 6-1 were homozygous for either the full length allele or the allele with deletion, respectively, showing that each allele can be fixed in the genome by selfpollination.

To investigate if the full length and deletion allele would be inherited by T₁ plants in a

Mendelian fashion, we further analyzed the progeny of the lines 4, 5, and 6. The expected ratio for heterozygous T_0 plants in T_1 progeny would be 1:2:1 for D:H:F (D: deletion, homozygous; H: heterozygous; F: full length, homozygous). We tested a subset of eight T_1 plants from line 4, ten T_1 plants from line 5, and nine T_1 plants from line 6. Line 4 and line 5 displayed ratios approximately fitting the expectation with 0.8:2:0.4 (D:H:F) and 1.5:2:1 (D:H:F), respectively, while progeny of line 5 was slightly enriched in heterozygous plants with a ratio of 0.25:2:0.25 (D:H:F; figure 33). The slight off-ratio in line 5 could be a result of the small number of tested progeny. But even though the subset was small, we detected from all tested lines at least one T_1 plant that was either homozygous for the deletion or homozygous for the full length allele (figure 33). Additionally, the T_1 plant 6-1, which was homozygous for the deletion, produced a T_2 progeny only showing the smaller size fragment with the deletion (Figure 25C). This result further demonstrates that 6-1 is indeed a plant carrying a homozygous deletion in the *MPK5* gene, and that these deletions can be inherited to the T_2 generation.

We hereby showed that targeting a single gene with two PTG encoded gRNAs can cause chromosomal deletions in the T_0 generation that can be inherited into future generations in a Mendelian mannner. If a T_0 plant is heterozygous for the deletion, its self-pollination can produce progeny with homozygous deletion and subsequent generations will stay homozygous, as shown by the progeny of plant 6-1. One previous report on similar chromosomal deletions in tobacco only showed deletions in protoplast transfections (Gao et al., 2015). Another report showed larger deletions in T_0 generation rice plants but did not demonstrate inheritance of the deletion to future generations (Zhou et al., 2014). In addition, both studies used systems that express each gRNA with their own promoter and terminator, which limited the number of gRNAs that could be coexpressed simultaneously to two gRNAs in the tobacco and four gRNAs in the rice study (Gao et al., 2015; Zhou et al., 2014). The new PTG/Cas9 system used in this study enables expression of a much higher number of gRNAs and also increases the titer of gRNAs inside the cells (Xie et al., 2015). Theoretically, our system could allow four

chromosomal deletions when eight gRNAs are simultaneously co-expressed and two deletions when four gRNAs are co-expressed. In PTGb7 T₀ lines (four gRNAs in one PTG), however, we previously only detected one line that simultaneously carried a monoallelic deletion in *MPK1* and biallelic deletion in *MPK5* (Xie et al., 2015). PTGb9 T₀ lines (eight gRNAs in one PTG) showed a similar trend with a fragment deletion in *MPK5* but not the other three targeted *MPK* genes (Xie et al., 2015). It is likely that a large number of genome-edited plants may be needed to identify a line with simultaneous deletion of two or four genes. In addition, factors other than gene-targeting efficiency may also affect the formation of chromosomal deletion. That deletions were mainly detected in *MPK5* suggests a potential positional effect on fragment deletion efficiency. It is also possible that chromosomal deletions in two or more closely related *MPK* genes might lead to drastic or detrimental phenotypes considering the importance of MPKs in plant development and stress response. Such lethal phenotypes might prevent the formation of multiple deletions in a singleregenerated plant.

Mutations that might preserve protein function are favored for *MPK1* and *MPK6*

In conformity with the previous speculation that mutations or deletions in some rice *MPK* genes are unfavorable, we detected an enrichment of mutations in *MPK1* and *MPK6* that can be divided by three without remainder in the four analyzed T₁ plants of PTGb9 lines (Figure 26). Such mutations are likely to preserve the open reading frame of the coding sequence because three bases constitute the coding unit for one amino acid. In contrast, all detected mutations in *MPK2* and *MPK5* were indels that are likely to interrupt the open reading frame of the sequence. 67% of the *MPK1* mutations affected three or a multiples of three base pairs (Figure 26A), and mutations in *MPK6* had an even higher tendency of this mutation type with 75% (Figure 25A). However, when only considering mutations in exons, 100% of the mutations in *MPK1* preserve the open reading frame but only 66.7% in *MPK6* (Figure 26B). This result suggests that keeping a functional protein might be more important for the *MPK1* gene than the *MPK6* gene in rice plants

To predict the effect of the mutations in *MPK1*, we translated the coding sequence of the mutants into a protein sequence and compared the first 60 amino acids with the wildtype sequence (Figure 26C). Plant 1-2 possessed two different mutant alleles of which one translated into a protein shortened by five amino acids (QATLS; 1-2a) and another into a protein shortened by one amino acid (S; 1-2b). Plants 2-1, 3-2, and 4-2 carry a different nucleotide mutation from 1-2b, but translate into the same protein sequence that is shortened by a serine on position 45 (Figure 26C). An analysis with InterProScan revealed that the mutations of all three protein sequences leave the predicted protein kinase domains intact, which start at amino acid 67 in the wildtype protein. In contrast to the results from *mpk1* alleles, all *mpk5* alleles detected in the PTGb9 lines resulted in premature stops of the *mpk5* proteins before the predicted protein kinase domains of the wildtype protein starting from amino acid 35 (Figure 26D).

The results indicate that *MPK1* plays an important role in rice and mutations in the *MPK1* gene produce a detrimental or lethal phenotype. Indeed, we encountered problems when we tried to mutate callus with PTGb3 (two gRNAs targeting *MPK1*). While PTGb3 line 1 (PTGb3-1) and PTGb3-2 produced normal plantlets on regeneration medium, callus of lines PTGb3-3 and PTGb3-4 turned black on the medium and mostly died. PTGb3-3 was not able to regenerate plantlets. Interestingly, even though callus of line PTGb3-4 carried a homozygous deletion on the *MPK1* gene (Figure 28A), we were able to recover two plantlets from the regeneration medium (Figure 26E). However, the recovered plantlets stayed severely dwarfed and only produced one sterile panicle, supporting the hypothesis that knock-out of *MPK1* causes detrimental or lethal effects (Figure 26E). We did not observe any negative effects for the lines of PTGb7 and PTGb9, which also target *MPK1*. But as shown for PTGb9, all analyzed T₁ mutant plants did carry *mpk1* alleles with mutations that preserve the existing open reading frame (Figure 26C). These other lines might carry a mutated but still functional *mpk1* protein that protects the plantlets from the observed detrimental phenotypes of lines PTGb3-3 and PTGb3-4. Previous reports of other groups used an *mpk1* knock-out mutant cell line in their research that was induced by

insertion of the retroposon *Tos17* (Kishi-Kaboshi et al., 2010; Kurusu et al., 2005). It is likely that MPK1 may play a so far undiscovered and essential role in tissue differentiation or plant development.

One difference of nuclease induced mutations compared to insertional mutation via T-DNA or *Tos17* is that nucleases like CRISPR/Cas9 merely cut the DNA but the mutation that causes the knock-out is dependent on the cell's repair mechanism. We found that in case of *MPK1* the mutations on viable T₁ generations always preserve the existing open reading frame (Figure 26B) and that a chromosomal deletion of *MPK1* caused callus to die early or yield severely dwarfed and sterile plants (Figure 26E). While knock-out of a truly essential gene is by definition impossible because of lethality, it seems that CRISPR/Cas9 also fails in reliably producing knock-outs for important but not truly essential genes. This might cause problems for analyzing gene families or redundant genes when knocking out several members may cause a drastic phenotype but the transformation and mutation system selects for open reading frame preserving mutations in one or more of the targeted genes. Even though the researcher wishes to discover this drastic phenotype, it might stay undetected because the mutations are unable to knock-out all genes. Shi et al. (2015) made similar observations in an attempt to identify essential genes in cancer cells by negative selection. When they targeted previously known essential genes in the 5' coding exons with CRISPR/Cas9, the induced mutations often preserved the existing open reading frame and the variants remained functional (Shi et al., 2015). They overcame this limitation by directly targeting protein domains instead of the 5' exon and achieved more severe negative selection phenotypes with this strategy. Therefore, it is suggested to target key protein domains with PTG/Cas9 to increase the chance of generating true knock-out mutants when studying gene families or redundant genes. Alternatively, mono-allelic mutation or heterozygous mutant lines may be generated and used to study essential or lethal genes.

MATERIAL AND METHODS

Plant materials and growth conditions

Rice cultivar Kitaake (*Oryza sativa* spp. japonica) was used in this study. Seeds were dried
5 for 36-48 h at 45 °C in a food dehydrator to break dormancy before germination in 37 °C
warm water for two days. The germinated seeds were planted into METROMIX 360 Soil
(SUNGRO HORTICULTURE, Agawam, MA) and grown in a greenhouse with 12 h
supplemental light per day at 28 °C day / 23 °C night temperature. The plants were
fertilized with 0.25% urea and 0.1% Sprint iron solution after the first week and fertilized
10 with 0.25% urea in subsequent weeks until flowering.

PTG/Cas9 gene constructs

The PTG constructs PTGb6, PTGb7, and PTGb9 were previously described (Xie et al.,
2015). PTGb3, PTGb4, PTGb5, PTGb2, and PTGb8 were constructed by inserting the
15 previously described and assembled PTGs PTG3, PTG4, PTG5, PTG2, and PTG8 (Xie et
al., 2015) into the BsaI-digested pRGE32 binary vector (Addgene Plasmid #63142).

Agrobacterium-mediated rice transformation

Binary vectors were transformed via electroporation into the *Agrobacterium tumefaciens*
20 strain EHA105.. Rice calli derived from mature seeds of the cultivar Kitaake were
transformed with the *Agrobacterium*-mediated method according to the previously
described protocol (Hiei and Komari, 2008).

Genomic DNA extraction

25 Genomic DNA was extracted from 100-200 µl volume of N₂-ground leaf material by
adding 0.9 ml of prewarmed CTAB buffer (140mM sorbitol, 220mM Tris-HCl pH 8.0,
22mM EDTA, 800mM NaCl, 34mM sarkosyl, and 22mM CTAB) and incubation at 65 °C
for 1 h in a 1.5 ml reaction tube. The sample was vortexed after adding 400 µl of

chloroform:isoamylalcohol (24:1) and mixed on a rotator for 20 min at room temperature before centrifugation at 12,000 rpm for 15 min. 2/3 volume of isopropanol was added to the upper aqueous phase before incubation at -20 °C for 30 min to precipitate DNA. The DNA was pelleted by centrifugation at 12,000 rpm for 30 sec to 5 min, and the pellet
5 washed with 70% ethanol before incubation with TE buffer containing 0.1 mg/ml RNase A at 37 °C for 30 min. The DNA was again precipitated by adding 1/10 volume of 3M sodium acetate pH 5.2 and 2-2.5 volumes of absolute ethanol and incubation at -20 °C overnight. The DNA was pelleted and washed with 70% ethanol before dissolving the dried pellet in an appropriate amount of TE buffer. The concentration was measured by a
10 spectrophotometer.

Genotyping of genome-edited progeny

All PCRs were performed with GoTaq DNA Polymerase in GoTaq Reaction Buffer (PROMEGA, Madison, WI). PCR products and digestions were analyzed by electrophoresis
15 with a 1% agarose gel and stained with ethidium bromide. The primers used for genotyping can be found in supplemental Table 4.1. The presence or absence of the genome-editing device in the progeny was detected by PCR with primers specific for *U3p:PTG*. Chromosomal deletions were detected by PCR with primers flanking the two target sites of each gene. The product size decreased by a specific number of base pairs if a chromosomal
20 deletion occurred. Indels resulting from genome-editing events on target sites with RE sites were detected by PCR-RE assay. PCR products encompassing the target sites were digested with the appropriate RE for 2-3 h. Mutated DNA could not be digested by the REs. Selected PCR products were sequenced to determine the specific mutation. Double peaks produced by different mutations in each allele were dissolved using degenerate sequence
25 decoding (Ma et al., 2015b). If the double peaks could not be decoded, the PCR product was cloned into pGEM-T EASY vector (PROMEGA, Madison, WI) by TA cloning for sequencing of single alleles.

Alignment of protein sequences and prediction of functional domains

The predicted protein sequences of the *mpk1* and *mpk5* mutant alleles were aligned with their corresponding wildtype sequence (MSU RGAP Release 7) by the Clustal Omega multiple sequence alignment tool (Li et al., 2015). The functional protein domains of the sequences were predicted with the web-based InterProScan tool (Jones et al., 2014).

Gene accession numbers

The GenBank RefSeq accession numbers of the targeted genes are Os06g0154500 (*MPK1*), Os08g0157000 (*MPK2*), Os03g0285800 (*MPK5*), and Os10g0533600 (*MPK6*).

TABLES

Table 4.1: Design and efficiency of PTG/Cas9 gene constructs to target four closely related *MPK* genes.

Binary vector	# of gRNAs	Target genes	Efficiency ²
PTGb3	2	<i>MPK1</i>	100% (4)
PTGb4	2	<i>MPK2</i>	100% (13)
PTGb5	2	<i>MPK6</i>	100% (2)
PTGb6 ¹	2	<i>MPK5</i>	86% (17)
PTGb2	1	<i>MPK5</i>	100% (9)
PTGb7 ¹	4	<i>MPK5/1</i>	100% / 89% (17)
PTGb8	4	<i>MPK6/2</i>	83% / 66% (12)
PTGb9 ¹	8	<i>MPK5/1/6/2</i>	86% / 86% / 86% / 86% (14)

¹ Previously described in Xie et al., *PNAS* 2015

² Percentage of genome-edited T0 lines over total number of tested T0 lines, Number of total tested lines in parentheses.

Table 4.2: Summary of PTG/Cas9 induced mutations types observed in all sequenced PCR products from the mutant plants..

Target gene	x bp insertion	x bp deletion	x bp conversion
<i>MPK1</i>	4, 1	15, 3, 1	1
<i>MPK2</i>	2, 1	11, 2, 1	-
<i>MPK5</i>	1	74, 9, 8, 5, 4, 3, 1	3, 2, 1
<i>MPK6</i>	2, 1	48, 21, 9, 6, 4, 3, 1	1

-: not detected.

5

Table 4.3: Oligonucleotides used in the study and their purposes.

Oligo name	Sequence (5'-3')	Purpose
MPK1-F2	CCTCGTGTGTTGGTGTGCTG	Genotyping of <i>MPK1</i>
MPK1-R2	TGCGCCTAAAAATCGAGGGT	
MPK2-nbF	TTTGGGAAGCATGTATGAAGC	Genotyping of <i>MPK2</i>
MPK2-R2	TATGCCAGCCAATGAGCCAA	
MPK5-F256	GCCACCTTCCTTCCTCATCCG	Genotyping of <i>MPK5</i>
MPK5-nbR	ACTCCGTCATCATGTCGCTC	
MPK6-F	TTGACGCCCCAACATAAATAA	Genotyping of <i>MPK6</i>
MPK6-R	TGTTGCTGCCGCTTTTCT	
UGW-U3-F	GACCATGATTACGCCAAGCTTAAGGAAT	Detection of <i>U3:PTG</i>
	CTTTAAACATACG	
UGW-gRNA-R	GGACCTGCAGGCATGCACGCGCTAAAAA CGGACTAGC	Detection of <i>Cas9</i>
gtCas9-F	GCTTGTGCGTTTCGATTTGA	
gtCas9-R	CCGCTCGTGCTTCTTATCCT	Control for DNA quality
control-F	CGCTGCCACTCTCCACTGA	
control-R	AGCTGCTTCCACTCGTTCCA	

10

MPK1-F2 (SEQ ID NO:174); MPK1-R2 (SEQ ID:175); MPK2-nbF (SEQ ID NO:176); MPK2-R2 (SEQ ID NO:177); MPK5-F256 (SEQ ID NO:178); MPK5-nbR (SEQ ID NO:179); MPK6-F (SEQ ID NO:180); MPK6-R (SEQ ID NO:181); UGW-U3-F (SEQ ID NO:182); UGW-gRNA-R (SEQ ID NO:183); gtCas9-F (SEQ ID NO:184); gtCas9-R (SEQ

ID NO:185); control-F (SEQ ID NO:186); control-R (SEQ ID NO:187).

LITERATURE CITED

- 5 Asai, T., Tena, G., Plotnikova, J., Willmann, M.R., Chiu, W.-L., Gomez-Gomez, L., Boller, T., Ausubel, F.M., and Sheen, J. (2002). MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* 415, 977–983.
- Barbezier, N., Canino, G., Rodor, J., Jobet, E., Saez-Vasquez, J., Marchfelder, A., and Echeverría, M. (2009). Processing of a dicistronic tRNA-snoRNA precursor: combined
10 analysis in vitro and in vivo reveals alternate pathways and coupling to assembly of snoRNP. *Plant Physiol.* 150, 1598–1610.
- Beckers, G.J.M., Jaskiewicz, M., Liu, Y., Underwood, W.R., He, S.Y., Zhang, S., and Conrath, U. (2009). Mitogen-activated protein kinases 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana*. *Plant Cell* 21, 944–953.
- 15 Brooks, C., Nekrasov, V., Lippman, Z.B., and Eck, J. Van (2014). Efficient Gene Editing in Tomato in the First Generation Using the Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-Associated9 System 1. 166, 1292–1297.
- Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., Han, T., and Hou, W. (2015). CRISPR/Cas9-Mediated Genome Editing in Soybean Hairy Roots. *PLoS One* 10,
20 e0136064.
- Canino, G., Bocian, E., Barbezier, N., Echeverría, M., Forner, J., Binder, S., and Marchfelder, A. (2009). *Arabidopsis* encodes four tRNase Z enzymes. *Plant Physiol.* 150, 1494–1502.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W.,
25 Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* (80-.). 339, 819–824.
- Endo, M., Mikami, M., and Toki, S. (2014). Multigene Knockout Utilizing Off-Target Mutations of the CRISPR/Cas9 System in Rice. *Plant Cell Physiol.* 56, 41–47.
- Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D.-L., Wang, Z., Zhang, Z., Zheng, R., Yang, L., et al. (2014). Multigeneration analysis reveals the inheritance, specificity, and
30 patterns of CRISPR/Cas-induced gene modifications in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4632–4637.

- Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P., and Xia, Q. (2015). CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Mol. Biol.* 87, 99–110.
- 5 Gutmann, B., Gobert, A., and Giegé, P. (2012). PRORP proteins support RNase P activity in both organelles and the nucleus in *Arabidopsis*. *Genes Dev.* 26, 1022–1027.
- Hiei, Y., and Komari, T. (2008). *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. *Nat. Protoc.* 3, 824–834.
- Hyun, Y., Kim, J.-S.J., Cho, S.W., Choi, Y., Kim, J.-S.J., and Coupland, G. (2014). Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. *Planta* 241, 271–284.
- 10 Jia, H., and Wang, N. (2014). Targeted genome editing of sweet orange using Cas9/sgRNA. *PLoS One* 9, e93806.
- Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B., and Weeks, D.P. (2013). Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Res.* 41, e188.
- 15 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. a, and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–1240.
- 20 Kishi-Kaboshi, M., Okada, K., Kurimoto, L., Murakami, S., Umezawa, T., Shibuya, N., Yamane, H., Miyao, A., Takatsuji, H., Takahashi, A., et al. (2010). A rice fungal MAMP-responsive MAPK cascade regulates metabolic flow to antimicrobial metabolite synthesis. *Plant J.* 63, 599–612.
- 25 Kurusu, T., Yagala, T., Miyao, A., Hirochika, H., and Kuchitsu, K. (2005). Identification of a putative voltage-gated Ca²⁺ channel as a key regulator of elicitor-induced hypersensitive cell death and mitogen-activated protein kinase activation in rice. *Plant J.* 42, 798–809.
- 30 Li, J.-F., Norville, J.E., Aach, J., McCormack, M., Zhang, D., Bush, J., Church, G.M., and Sheen, J. (2013). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana bethamiana* using guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691.

- Li, W., Cowley, a., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* 43, 580–584.
- Lowder, L.G., Paul, J.W., Baltes, N.J., Voytas, D.F., Zhang, Y., Zhang, D., Tang, X.,
5 Zheng, X., Hsieh, T.-F., and Qi, Y. (2015). A CRISPR/Cas9 toolbox for multiplexed plant genome editing and transcriptional regulation. *Plant Physiol.* doi:10.1104/pp.15.00636.
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., et al. (2015a). A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol. Plant* 8, 1274–1284.
- 10 Ma, X., Chen, L., Zhu, Q., Chen, Y., and Liu, Y.-G. (2015b). Rapid Decoding of Sequence-Specific Nuclease-Induced Heterozygous and Biallelic Mutations by Direct Sequencing of PCR Products. *Mol. Plant* 8, 1285–1287.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science* (80-.).
15 339, 823–827.
- Miao, J., Guo, D., Zhang, J., Huang, Q., Qin, G., Zhang, X., Wan, J., Gu, H., and Qu, L.-J. (2013). Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 23, 1233–1236.
- Mikami, M., Toki, S., and Endo, M. (2015). Parameters affecting frequency of CRISPR/Cas9 mediated targeted mutagenesis in rice. *Plant Cell Rep.* 34, 1807–1815.
- 20 Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. *Genes Dev.* 24, 1832–1860.
- Pitzschke, A., Schikora, A., and Hirt, H. (2009). MAPK cascade signalling networks in plant defence. *Curr. Opin. Plant Biol.* 12, 421–426.
- Reyna, N.S., and Yang, Y. (2006). Molecular analysis of the rice MAP kinase gene family
25 in relation to Magnaporthe grisea infection. *Mol. Plant. Microbe. Interact.* 19, 530–540.
- Schiffer, S., Rösch, S., and Marchfelder, A. (2002). Assigning a function to a conserved group of proteins: The tRNA 3'-processing enzymes. *EMBO J.* 21, 2769–2777.
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas
30 system. *Nat. Biotechnol.* 31, 686–688.

- Shi, J., Wang, E., Milazzo, J.P., Wang, Z., Kinney, J.B., and Vakoc, C.R. (2015). Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* 33, 661–667.
- 5 Sugano, S.S., Shirakawa, M., Takagi, J., Matsuda, Y., Shimada, T., Hara-Nishimura, I., and Kohchi, T. (2014). CRISPR/Cas9-mediated targeted mutagenesis in the liverwort *Marchantia polymorpha* L. *Plant Cell Physiol.* 55, 475–481.
- Voytas, D.F. (2013). Plant genome engineering with sequence-specific nucleases. *Annu. Rev. Plant Biol.* 64, 327–350.
- 10 Xie, K., and Yang, Y. (2013). RNA-Guided Genome Editing in Plants Using a CRISPR-Cas System. *Mol. Plant* 6, 1975–1983.
- Xie, K., Zhang, J., and Yang, Y. (2014a). Genome-Wide Prediction of Highly Specific Guide RNA Spacers for the CRISPR-Cas9-Mediated Genome Editing in Model Plants and Major Crops. *Mol. Plant* 7, 923–926.
- 15 Xie, K., Minkenberg, B., and Yang, Y. (2014b). Targeted Gene Mutation in Rice Using a CRISPR-Cas9 System. *Bio-Protocol* 4, e1225.
- Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* 112, 3570–3575.
- 20 Xu, R.-F., Li, H., Qin, R.-Y., Li, J., Qiu, C.-H., Yang, Y.-C., Ma, H., Li, L., Wei, P.-C., and Yang, J.-B. (2015). Generation of inheritable and “transgene clean” targeted genome-modified rice in later generations using the CRISPR/Cas9 system. *Sci. Rep.* 5, 11491 doi: 10.1038/srep11491.
- 25 Zhang, H., Gou, F., Zhang, J., Liu, W., Li, Q., Mao, Y., Botella, J.R., and Zhu, J.-K. (2015). TALEN-mediated targeted mutagenesis produces a large variety of heritable mutations in rice. *Plant Biotechnol. J.* doi: 10.1111/pbi.12372.
- Zhang, H.H., Zhang, J., Wei, P., Zhang, B., Gou, F., Feng, Z., Mao, Y., Yang, L., Zhang, H.H., Xu, N., et al. (2014). The CRISPR/Cas9 system produces specific and homozygous targeted gene editing in rice in one generation. *Plant Biotechnol. J.* 12, 797–807.
- 30 Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H., and Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* 42, 10903–10914.

TABLE OF SEQUENCES

	SEQ	TYPE	DESCRIPTION
	SEQ ID NO:1	DNA	sgRNA1
	SEQ ID NO:2	DNA	sgRNA2
	SEQ ID NO:3	DNA	gRNA1
5	SEQ ID NO:4	DNA	gRNA2
	SEQ ID NO:5	DNA	gRNA3
	SEQ ID NO:6	DNA	gRNA4
	SEQ ID NO:7	DNA	gRNA5
	SEQ ID NO:8	DNA	gRNA6
10	SEQ ID NO:9	DNA	gRNA7
	SEQ ID NO:10	DNA	gRNA8
	SEQ ID NO:11	DNA	gRNA9
	SEQ ID NO:12	DNA	gRNA10
	SEQ ID NO:13	DNA	pre-tRNA
15	SEQ ID NO:14	DNA	Last 10 bp of U3p
	SEQ ID NO:15	DNA	gRNA-tRNA in pGTR plasmid
	SEQ ID NO:16	DNA	PTG1
	SEQ ID NO:17	DNA	PTG2
	SEQ ID NO:18	DNA	PTG1.1
20	SEQ ID NO:19	DNA	PTG2.1
	SEQ ID NO:20	DNA	PTG3
	SEQ ID NO:21	DNA	PTG4
	SEQ ID NO:22	DNA	PTG5
	SEQ ID NO:23	DNA	PTG6
25	SEQ ID NO:24	DNA	PTG7
	SEQ ID NO:25	DNA	PTG8
	SEQ ID NO:26	DNA	PTG9
	SEQ ID NO:27	DNA	PTG10
	SEQ ID NO:28	DNA	Primer gRNA1-R
30	SEQ ID NO:29	DNA	Primer gRNA2-R
	SEQ ID NO:30	DNA	Primer Bsa-gRNA-F
	SEQ ID NO:31	DNA	Primer gRNA1-F
	SEQ ID NO:32	DNA	Primer gRNA2-F
	SEQ ID NO:33	DNA	Primer gRNA-R
35	SEQ ID NO:34	DNA	Primer UBI-qF
	SEQ ID NO:35	DNA	Primer UBI-qR
	SEQ ID NO:36	DNA	Primer MPK5-qF

TABLE OF SEQUENCES

	SEQ	TYPE	DESCRIPTION
40	SEQ ID NO:37	DNA	Primer g-tRNA-F
	SEQ ID NO:38	DNA	Primer tRNA-R
	SEQ ID NO:39	DNA	Primer UBI-F
	SEQ ID NO:40	DNA	Primer UBI-R

	SEQ ID NO:41	DNA	Primer UGW-U3-F
	SEQ ID NO:42	DNA	Primer UGW-gRNA-R
	SEQ ID NO:43	DNA	Primer MPK1-F
	SEQ ID NO:44	DNA	Primer MPK1-R2
5	SEQ ID NO:45	DNA	Primer MPK2-F
	SEQ ID NO:46	DNA	Primer MPK2-R2
	SEQ ID NO:47	DNA	Primer MPK5-F256
	SEQ ID NO:48	DNA	Primer MPK5-R611
	SEQ ID NO:49	DNA	Primer MPK6-F
10	SEQ ID NO:50	DNA	Primer MPK6-R
	SEQ ID NO:51	DNA	Primer PDS-F
	SEQ ID NO:52	DNA	Primer PDS-R
	SEQ ID NO:53	DNA	Primer L5AD5-F
	SEQ ID NO:54	DNA	Primer L3AD5-R
15	SEQ ID NO:55	DNA	Primer S5AD5-F
	SEQ ID NO:56	DNA	Primer S3AD5-R
	SEQ ID NO:57	DNA	Primer gR1-F
	SEQ ID NO:58	DNA	Primer gR1-R
	SEQ ID NO:59	DNA	Primer gR2-F
20	SEQ ID NO:60	DNA	Primer gR2-R
	SEQ ID NO:61	DNA	Primer gR3-F
	SEQ ID NO:62	DNA	Primer gR3-R
	SEQ ID NO:63	DNA	Primer gR4-F
	SEQ ID NO:64	DNA	Primer gR4-R
25	SEQ ID NO:65	DNA	Primer gR5-F
	SEQ ID NO:66	DNA	Primer gR5-R
	SEQ ID NO:67	DNA	Primer gR6-F
	SEQ ID NO:68	DNA	Primer gR6-R
	SEQ ID NO:69	DNA	Primer gR7-F
30	SEQ ID NO:70	DNA	Primer gR7-R
	SEQ ID NO:71	DNA	Primer gR8-F
	SEQ ID NO:72	DNA	Primer gR8-R
	SEQ ID NO:73	DNA	Primer Ln-gR7-F
	SEQ ID NO:74	DNA	Primer Ln-gR7-R
35			

TABLE OF SEQUENCES

	SEQ	TYPE	DESCRIPTION
40	SEQ ID NO:75	DNA	Primer gR9-F
	SEQ ID NO:76	DNA	Primer gR9-R
	SEQ ID NO:77	DNA	Primer gR10-F
	SEQ ID NO:78	DNA	Primer gR10-R

	SEQ ID NO:79	DNA	Wild type MPK1
	SEQ ID NO:80	PRT	Wild type MPK1
	SEQ ID NO:81	DNA	Wild type MPK2
	SEQ ID NO:82	DNA	Wild type MPK5
5	SEQ ID NO:83	PRT	Wild type MPK5
	SEQ ID NO:84	DNA	Wild type MPK6
	SEQ ID NO:85	DNA	WT UBI
	SEQ ID NO:86	DNA	WT PSD3
	SEQ ID NO:87	RNA	tRNA-gRNA-tRNA scheme FIG 2B
10	SEQ ID NO:88	RNA	gRNA from PTG gene FIG 2C
	SEQ ID NO:89	RNA	gRNA from sgRNA gene FIG 2H
	SEQ ID NO:90	DNA	MPK1 #1 indel FIG 3C
	SEQ ID NO:91	DNA	MPK1 #2 indel FIG 3C
	SEQ ID NO:92	DNA	MPK2 #1 indel FIG 3C
15	SEQ ID NO:93	DNA	MPK2 #2 indel FIG 3C
	SEQ ID NO:94	DNA	MPK2 #3 indel FIG 3C
	SEQ ID NO:95	DNA	MPK5 #1 indel FIG 3C
	SEQ ID NO:96	DNA	MPK5 #2 indel FIG 3C
	SEQ ID NO:97	DNA	MPK6 #1 indel FIG 3C
20	SEQ ID NO:98	DNA	MPK6 #2 indel FIG 3C
	SEQ ID NO:99	DNA	MPK6 #3 indel FIG 3C
	SEQ ID NO:100	DNA	MPK5 indel FIG 4B
	SEQ ID NO:101	DNA	MPK1 indel FIG 4B
	SEQ ID NO:102	DNA	MPK5 target sense strand FIG 5A
25	SEQ ID NO:103	DNA	MPK5 target antisense strand FIG 5A
	SEQ ID NO:104	DNA	MPK1 target sense strand FIG 5B
	SEQ ID NO:105	DNA	MPK1 target antisense strand FIG 5B
	SEQ ID NO:106	DNA	MPK2 target sense strand FIG 5C
	SEQ ID NO:107	DNA	MPK2 target antisense strand FIG 5C
30	SEQ ID NO:108	DNA	MPK6 target sense strand FIG 5D
	SEQ ID NO:109	DNA	MPK6 target antisens strand FIG 5D
	SEQ ID NO:110	DNA	Multiple cloning site sense FIG 6C
	SEQ ID NO:111	DNA	Multiple cloning site antisense FIG 6C
35	SEQ ID NO:112	DNA	sgRNA1 #1 indel FIG 8

TABLE OF SEQUENCES

	SEQ	TYPE	DESCRIPTION
40	SEQ ID NO:113	DNA	sgRNA1 #2 indel FIG 8
	SEQ ID NO:114	DNA	sgRNA1 #3 indel FIG 8
	SEQ ID NO:115	DNA	PTG1 indel FIG 8
	SEQ ID NO:116	DNA	sgRNA2 indel FIG 8

	SEQ ID NO:117	DNA	PTG2 indel FIG 8
	SEQ ID NO:118	DNA	PTG1.1 indel FIG 9A
	SEQ ID NO:119	DNA	PTG2.1 indel FIG 9B
	SEQ ID NO:120	DNA	MPK1 #1 indel FIG 16
5	SEQ ID NO:121	DNA	MPK1 #10 indel FIG 16
	SEQ ID NO:122	DNA	MPK1 #11 indel FIG 16
	SEQ ID NO:123	DNA	MPK2 #1 indel FIG 16
	SEQ ID NO:124	DNA	MPK2 #10 indel FIG 16
	SEQ ID NO:125	DNA	MPK2 #11 indel FIG 16
10	SEQ ID NO:126	DNA	MPK5 #1 indel FIG 16
	SEQ ID NO:127	DNA	MPK5 #10 indel FIG 16
	SEQ ID NO:128	DNA	MPK5 #11 indel FIG 16
	SEQ ID NO:129	DNA	MPK6 #1 indel FIG 16
	SEQ ID NO:130	DNA	MPK6 #10 indel FIG 16
15	SEQ ID NO:131	DNA	MPK6 #11 indel FIG 16
	SEQ ID NO:132	DNA	PDS target sense strand FIG 17B
	SEQ ID NO:133	DNA	PDS indel 1 FIG 17E
	SEQ ID NO:134	DNA	PDS indel 2 FIG 17E
	SEQ ID NO:135	DNA	MPK1 1-2a FIG 19B
20	SEQ ID NO:136	DNA	MPK1 1-2b FIG 19B
	SEQ ID NO:137	DNA	MPK1 2-1 FIG 19B
	SEQ ID NO:138	DNA	MPK1 3-2a FIG 19B
	SEQ ID NO:139	DNA	MPK1 3-2b FIG 19B
	SEQ ID NO:140	DNA	MPK1 4-2 FIG 19B
25	SEQ ID NO:141	DNA	MPK2 1-2 FIG 19B
	SEQ ID NO:142	DNA	MPK2 2-1 FIG 19B
	SEQ ID NO:143	DNA	MPK2 3-1 FIG 19B
	SEQ ID NO:144	DNA	MPK2 4-2a FIG 19B
	SEQ ID NO:145	DNA	MPK2 4-2b FIG 19B
30	SEQ ID NO:146	DNA	MPK5 1-2 FIG 19B
	SEQ ID NO:147	DNA	MPK5 2-1 FIG 19B
	SEQ ID NO:148	DNA	MPK5 3-2 FIG 19B
	SEQ ID NO:149	DNA	MPK5 4-2 FIG 19B

TABLE OF SEQUENCES

35	SEQ	TYPE	DESCRIPTION
	SEQ ID NO:150	DNA	MPK6 1-2a FIG 19B
	SEQ ID NO:151	DNA	MPK6 1-2b FIG 19B
	SEQ ID NO:152	DNA	MPK6 2-1 FIG 19B
	SEQ ID NO:153	DNA	MPK6 3-2 FIG 19B
40	SEQ ID NO:154	DNA	MPK6 4-2a FIG 19B
	SEQ ID NO:155	DNA	MPK6 4-2b FIG 19B
	SEQ ID NO:156	PRT	MPK1 1-2a FIG 22C
	SEQ ID NO:157	PRT	MPK1 2-1/3-2/4-2 FIG 22C

	SEQ ID NO:158	PRT	MPK1 1-2b FIG 22C
	SEQ ID NO:159	PRT	MPK5 1-2 FIG 22D
	SEQ ID NO:160	PRT	MPK5 4-2 FIG 22D
	SEQ ID NO:161	PRT	MPK5 3-2 FIG 22D
5	SEQ ID NO:162	PRT	MPK5 2-1 FIG 22D
	SEQ ID NO:163	DNA	MPK5 4Δ FIG 28
	SEQ ID NO:164	DNA	HDAC1-sg1
	SEQ ID NO:165	DNA	HDAC1-sg2
	SEQ ID NO:166	DNA	HDAC2-sg1
10	SEQ ID NO:167	DNA	HDAC2-sg2
	SEQ ID NO:168	DNA	HDAC3-sg1
	SEQ ID NO:169	DNA	HDAC3-sg2
	SEQ ID NO:170	DNA	HDAC4-sg1
	SEQ ID NO:171	DNA	HDAC4-sg2
15	SEQ ID NO:172	DNA	HDAC6-sg1
	SEQ ID NO:173	DNA	HDAC6-sg2
	SEQ ID NO:174	DNA	MPK1-F2
	SEQ ID NO:175	DNA	MPK1-R2
	SEQ ID NO:176	DNA	MPK2-nbF
20	SEQ ID NO:177	DNA	MPK2-R2
	SEQ ID NO:178	DNA	MPK5-F256
	SEQ ID NO:179	DNA	MPK5-nbR
	SEQ ID NO:180	DNA	MPK6-F
	SEQ ID NO:181	DNA	MPK6-R
25	SEQ ID NO:182	DNA	UGW-U3-F
	SEQ ID NO:183	DNA	UGW-gRNA-R
	SEQ ID NO:184	DNA	gtCas9-F
	SEQ ID NO:185	DNA	gtCas9-R
	SEQ ID NO:186	DNA	control-F
30	SEQ ID NO:187	DNA	control-R
	SEQ ID NO:188	DNA	pretRNA sequence form PTG7
	SEQ ID NO:189	DNA	pretRNA sequence form PTG10

What is claimed is:

1. A use of a polynucleotide construct for multiplex RNA mediated genetic manipulation in a recipient cell
wherein the polynucleotide construct encodes two or more RNA mediated genetic manipulation sequences in tandem array with one or more tRNA cleavage sequence,
5 wherein said tRNA cleavage sequence comprises a sequence recognized and excised by a tRNA splicing endonuclease, and
wherein said polynucleotide construct is suitable for introduction to said recipient cell so that said recipient cell's tRNA processing system cleaves the transcript produced from
10 the transcribed polynucleotide construct at the tRNA cleavage sequences.
2. The use according to claim 1 wherein said RNA mediated genetic manipulation sequences target multiple sites within a single gene in said recipient cell.
- 15 3. The use according to claim 1 wherein said RNA mediated genetic manipulation sequences target multiple different genes in said recipient cell.
4. The use according to any one of claims 1-3 wherein said RNA mediated genetic manipulation is RNA guided genome editing, wherein said RNA guided genome editing
20 comprises targeted mutation, homology-dependent repair, transcriptional activation and repression, epigenome editing, and/or genomic labeling.
5. The use according to any one of claims 1-3 wherein said RNA mediated genetic manipulation comprises RNA interference.
- 25 6. The use according to claim 5 wherein said RNA interference is microRNA and/or

small RNA mediated RNA interference.

7. The use according to claim 4 wherein said RNA guided genome editing is CRISPR/Cas mediated genome editing.

5

8. The use according to any one of claims 1-7 wherein said tRNA cleavage sequence comprises an active site for one or more of RNase P and/or RNase Z and/or RNase E.

9. The use according to any one of claims 1-8 wherein said tRNA cleavage sequence
10 comprises a pretRNA acceptor stem, a D-loop arm and a TΨC-loop arm.

10. The use according to any one of claims 1-9 wherein said tRNA cleavage sequence is SEQ ID NO: 13, 188 or 189.

15 11. The use according to any one of claims 1-4 or 7-10 wherein said polynucleotide construct comprises a tRNA-guideRNA-in tandem alignment.

12. The use according to claim 11 wherein said guideRNA comprises a spacer RNA and a scaffold RNA sequence.

20

13. The use according to claim 11 wherein said polynucleotide construct comprises a tRNA-guideRNA-tRNA.

14. The use according to any one of claims 1-13 wherein said recipient cell is a plant
25 cell.

15. The use according to any one of claims 1-13 wherein said recipient cell is an animal cell.

30 16. The use according to any one of claims 1-13 wherein said recipient cell is a

microbial cell.

17. The use according to any one of claims 1-13 wherein said recipient cell is a human cell.

5

18. A nucleic acid construct for multiplex RNA mediated genome manipulation comprising: two or more RNA mediated genetic manipulation sequences and a tRNA cleavage sequence, wherein said tRNA cleavage sequence comprises a sequence recognized and excised by a tRNA splicing endonuclease.

10

19. The nucleic acid construct of claim 18 wherein said two or more RNA mediated genetic manipulation sequences target multiple sites within a single gene.

20. The nucleic acid construct of claim 18 wherein said RNA mediated genetic manipulation sequences target multiple different genes.

15

21. The nucleic acid construct of any one of claims 18-20 wherein said RNA mediated genetic manipulation sequence is an interfering RNA sequence.

22. The nucleic acid construct of claim 21 wherein said RNA interference sequence is microRNA and/or small RNA mediated RNA interference.

20

23. The nucleic acid construct of any one of claims 18-22 wherein said tRNA cleavage sequence comprises an active site for one or more of RNase P and/or RNase Z and/or RNase E.

25

24. The nucleic acid construct of any one of claims 18-23 wherein said tRNA cleavage sequence comprises a pretRNA acceptor stem, a D-loop arm and a TΨC-loop arm.

25. The nucleic acid construct of any one of claims 18-24 wherein said tRNA cleavage sequence is SEQ ID NO: 13, 188 or 189.

5 26. The nucleic acid construct of any one of claims 18-20 or 23-25 wherein said nucleic acid construct comprises a tRNA-guideRNA arrayed in tandem.

27. The nucleic acid construct of claim 26 wherein said guideRNA comprises a spacer RNA and a scaffold RNA sequence.

10

28. The nucleic acid construct of claim 26 wherein said nucleic acid construct comprises a tRNA-guideRNA-tRNA.

15 29. The nucleic acid construct of any one of claims 18-28 further comprising a regulatory element for expression of said nucleic acid construct.

30. The nucleic acid construct of any one of claims 18-29 wherein said nucleic acid construct comprises a promoter sequence operably linked to said RNA mediated genetic manipulation sequences and said tRNA cleavage sequence.

20

31. The nucleic acid construct of any one of claims 18-30 further comprising a terminator sequence.

25 32. The nucleic acid construct of claim 30 wherein said promoter sequence is a Pol III promoter.

33. The nucleic acid construct of claim 31 wherein said terminator sequence is a Pol III terminator.

34. An expression cassette comprising the nucleic acid construct of any one of claims 18-33.

35. A vector comprising the expression cassette of claim 34.

5

36. A recipient cell comprising the nucleic acid construct of any one of claims 18-33, the expression cassette of claim 34, or the vector of claim 35.

37. The cell of claim 36, wherein the cell comprises a genomic insertion or deletion.

10

38. The cell of claim 36 or claim 37 wherein said cell is a plant cell.

39. The cell of claim 36 or claim 37 wherein said cell is an animal cell.

15

40. The cell of claim 36 or claim 37 wherein said cell is a microbial cell.

41. The cell of claim 39 wherein said cell is a human cell.

20

42. A use of a nucleic acid construct for modifying a target site in the genome of a cell, the nucleic acid construct having a two or more guide RNAs and a tRNA cleavage sequence, the nucleic acid construct for use in the presence of a Cas endonuclease, wherein said tRNA cleavage sequence comprises a sequence recognized and excised by a tRNA processing system splicing endonuclease, wherein said guide RNA is for cleavage by said cell and one or more of said cleaved guide RNAs and Cas endonuclease form a complex that enables the Cas endonuclease to introduce a double strand break at said target site.

25

43. The use according to claim 42, wherein said nucleic acid construct further comprises a sequence encoding the Cas endonuclease.

44. The use according to claim 42 or claim 43, wherein the nucleic acid construct is for introduction to said cell by particle bombardment, protoplast transfection or *Agrobacterium* transformation.

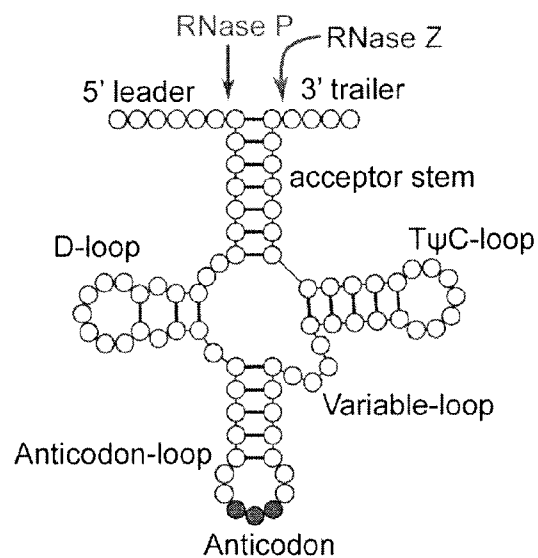


FIG. 1A

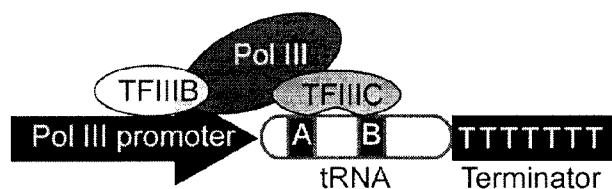


FIG. 1B

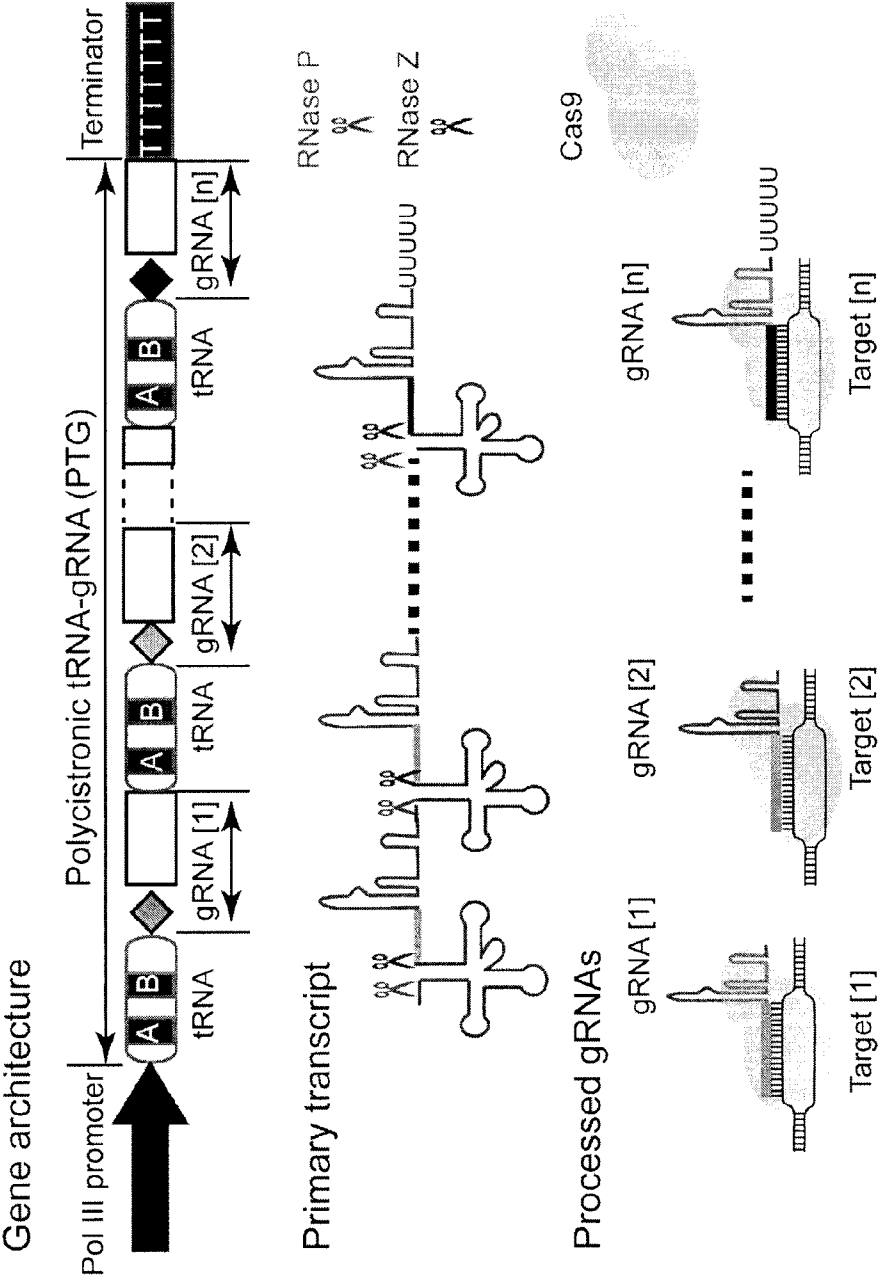
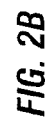
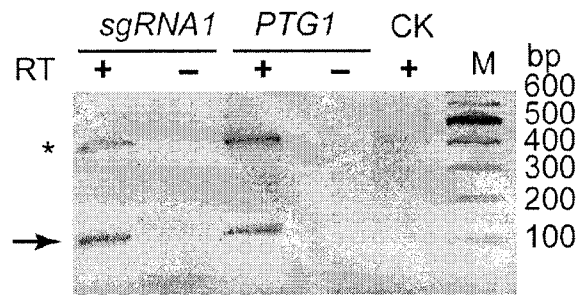
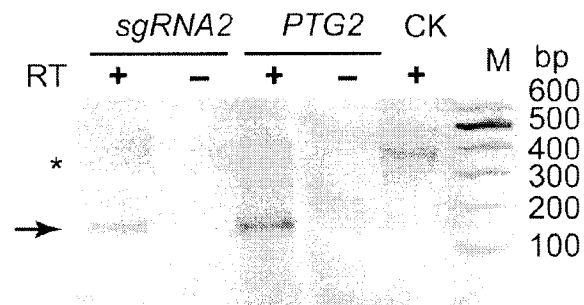
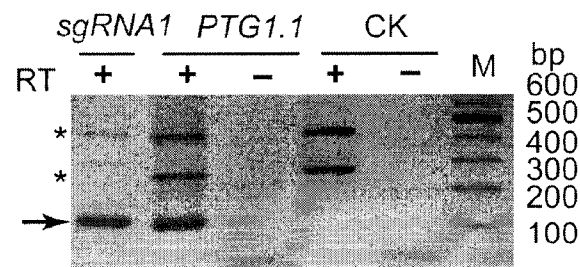
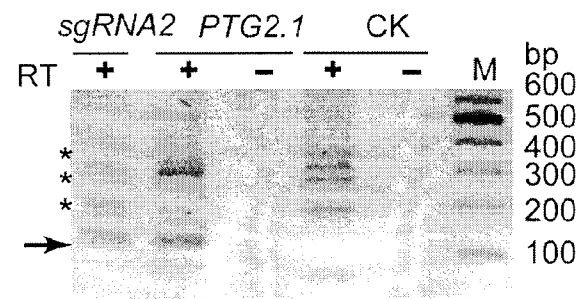
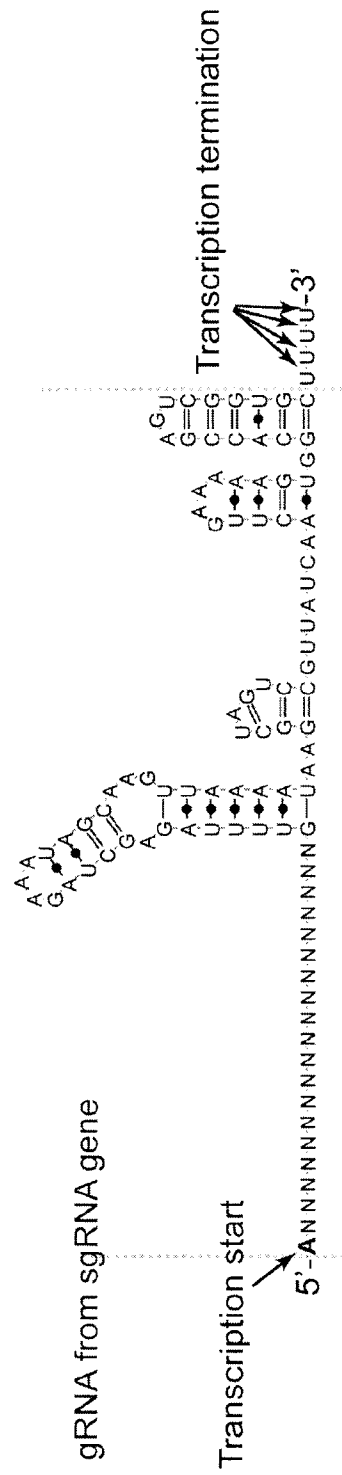
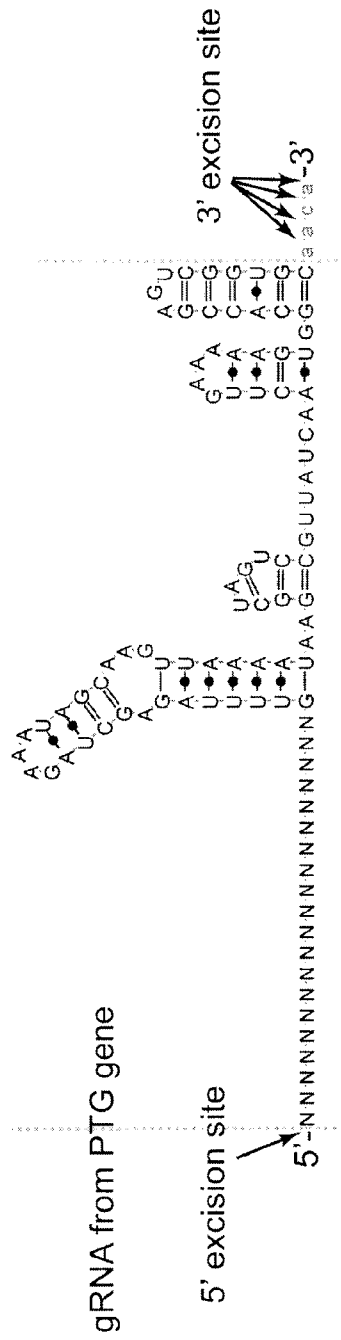


FIG. 1C

FIG. 2A



**FIG. 2C****FIG. 2D****FIG. 2E****FIG. 2F**



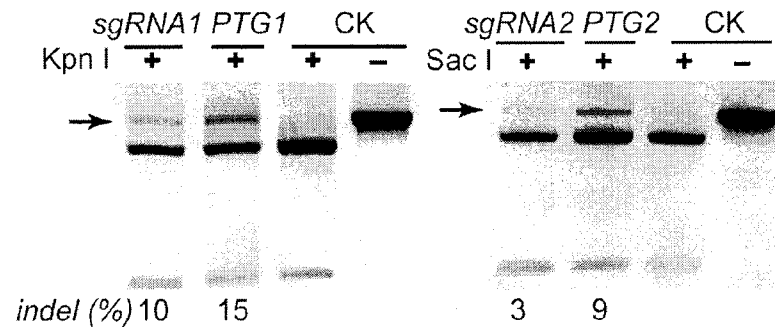


FIG. 2I

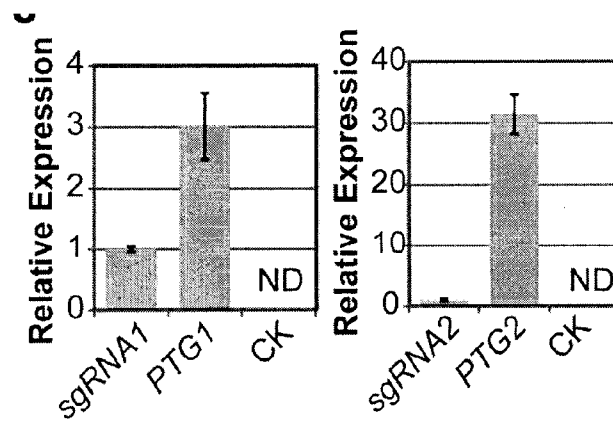
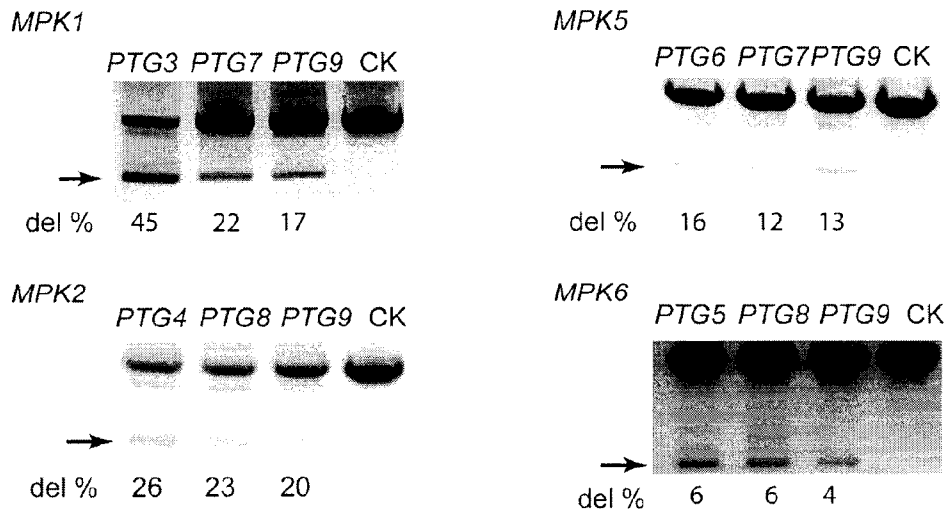


FIG. 2J

Gene ID	# of gRNA	Targeted loci	Gene architecture
<i>PTG3</i>	2	<i>MPK1</i>	tRNA-gRNA3-tRNA-gRNA4
<i>PTG4</i>	2	<i>MPK2</i>	tRNA-gRNA5-tRNA-gRNA6
<i>PTG5</i>	2	<i>MPK6</i>	tRNA-gRNA7-tRNA-gRNA8
<i>PTG6</i>	2	<i>MPK5</i>	tRNA-gRNA1-tRNA-gRNA2
<i>PTG7</i>	4	<i>MPK5</i> and <i>MPK1</i>	<i>PTG6-PTG3</i>
<i>PTG8</i>	4	<i>MPK6</i> and <i>MPK2</i>	<i>PTG5-PTG4</i>
<i>PTG9</i>	8	<i>MPK5</i> , <i>MPK1</i> , <i>MPK6</i> and <i>MPK2</i>	<i>PTG7-PTG8</i>

FIG. 3A**FIG. 3B**

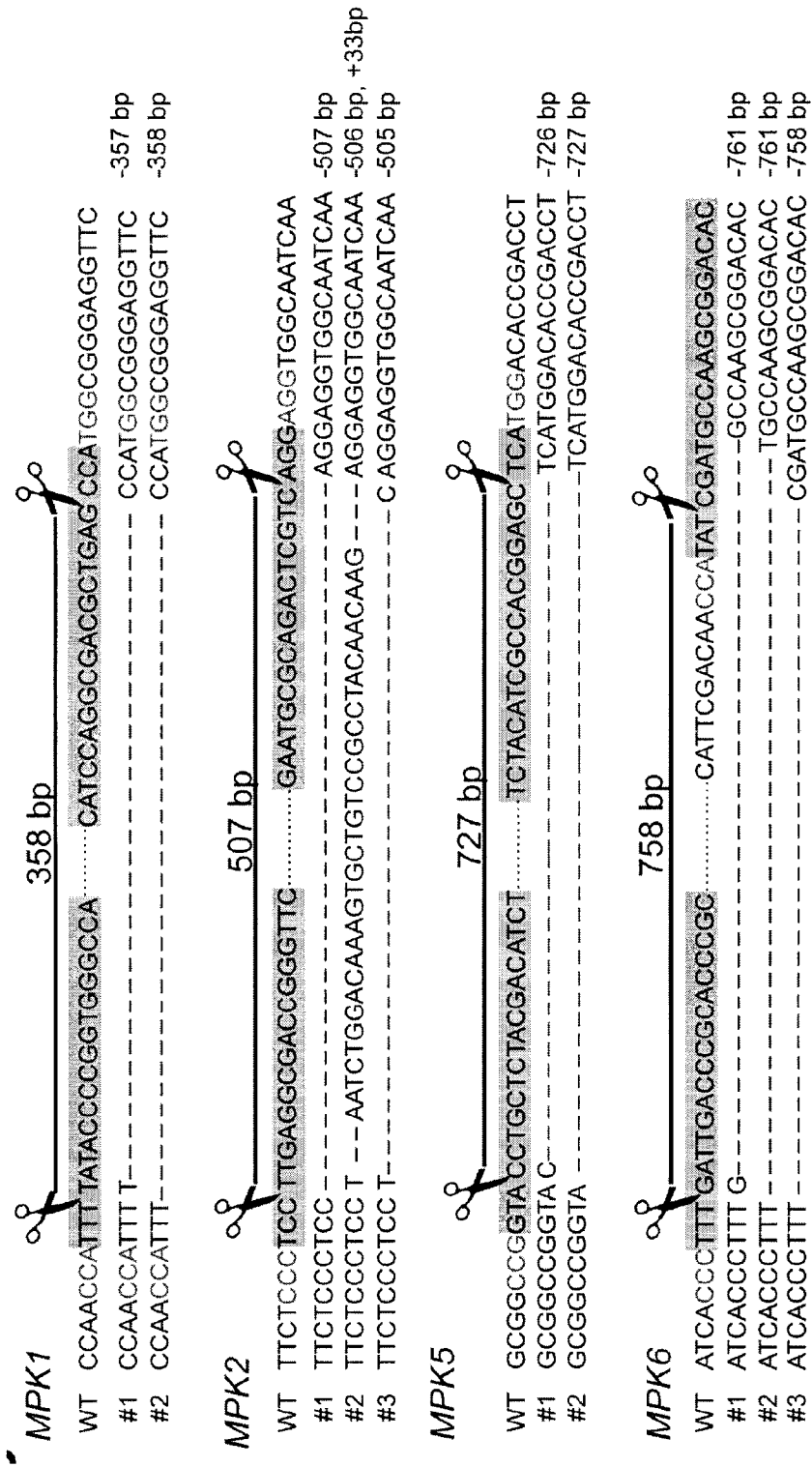


FIG. 3C

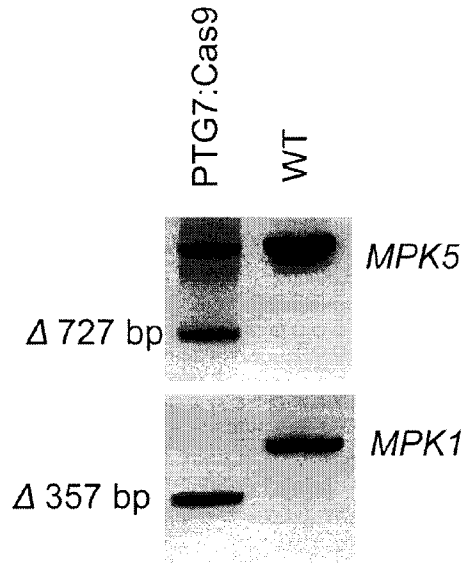


FIG. 4A

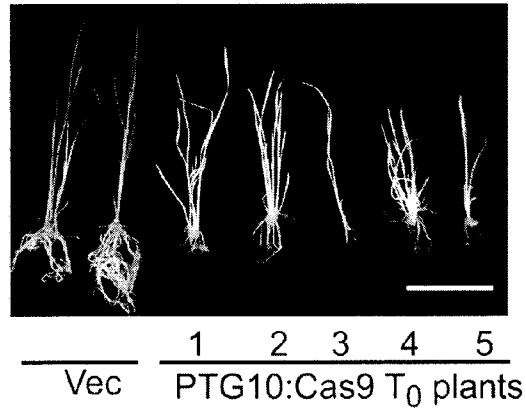


FIG. 4B

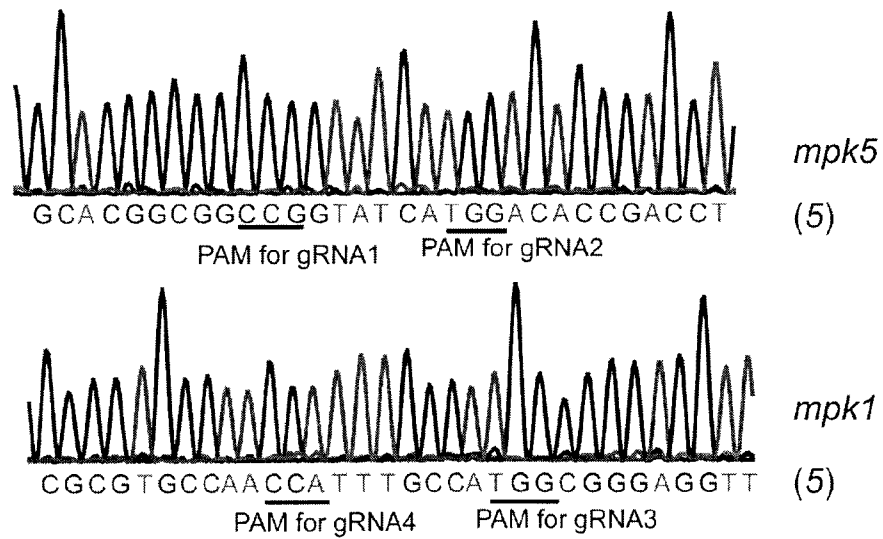
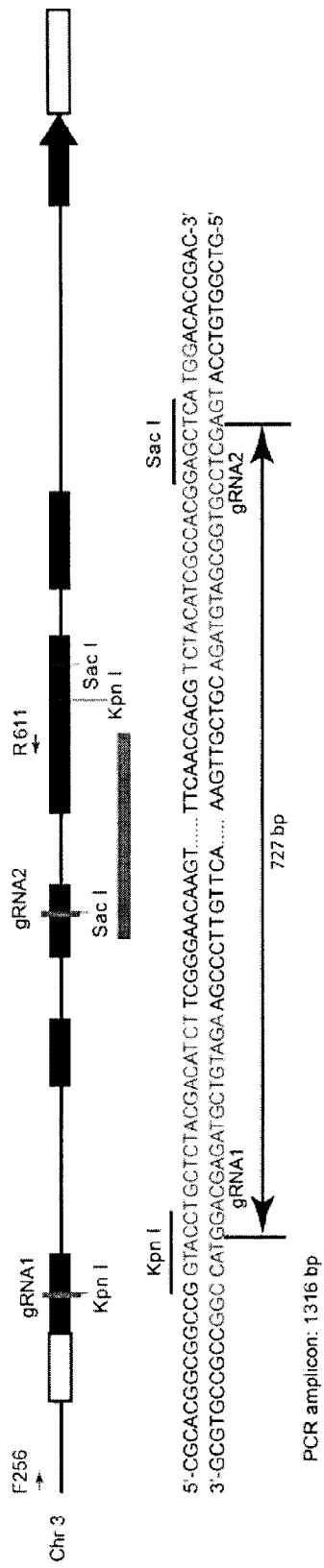


FIG. 4C

MPK5



10/50

FIG. 5A

MPK1

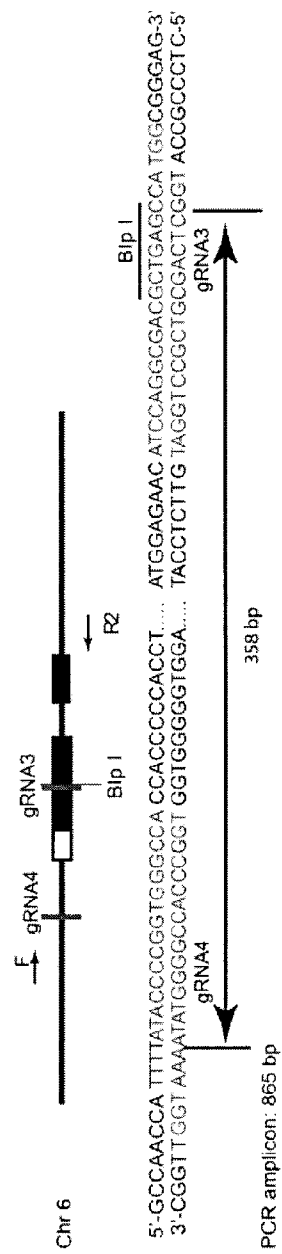


FIG. 5B

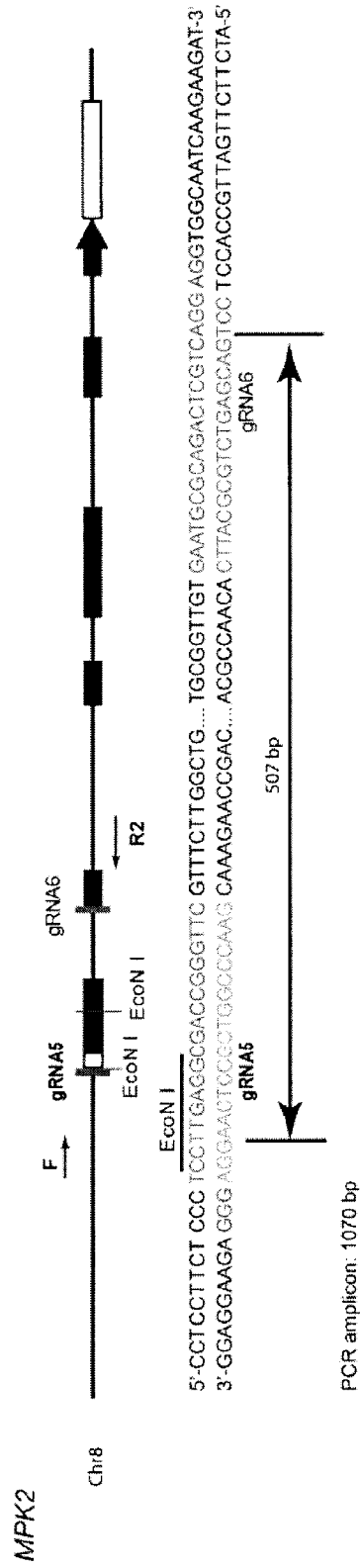


FIG. 5C

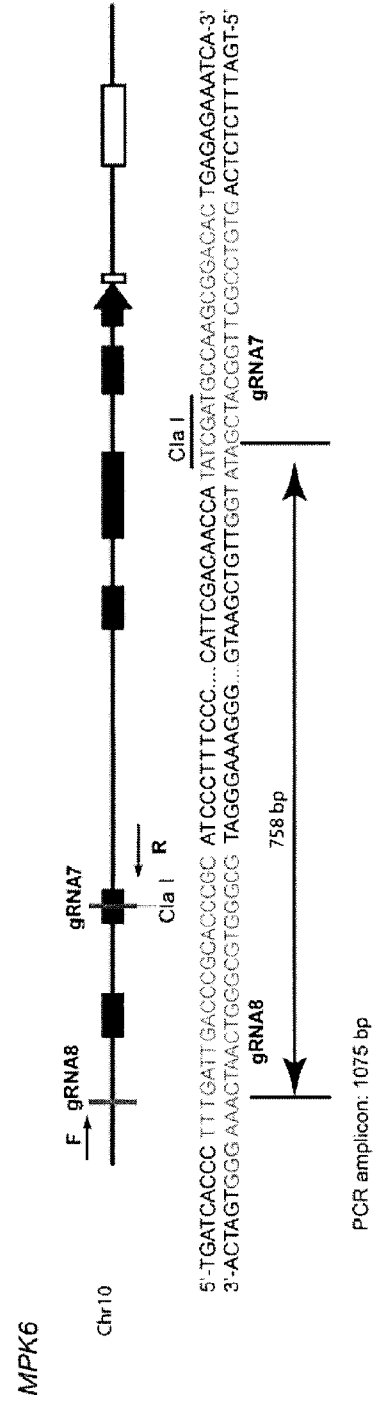


FIG. 5D

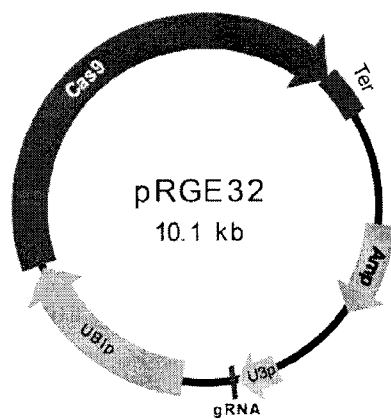


FIG. 6A

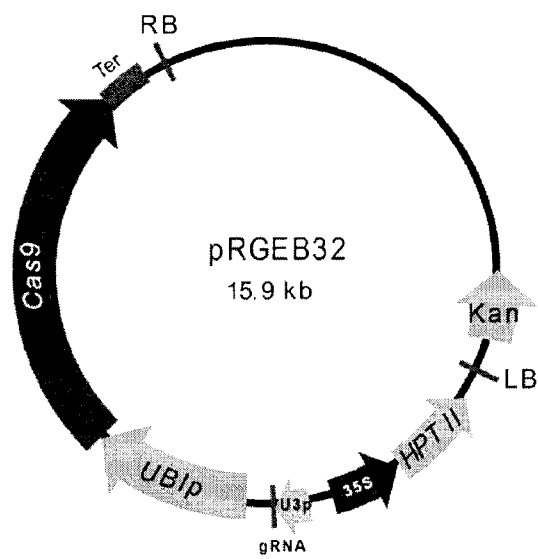


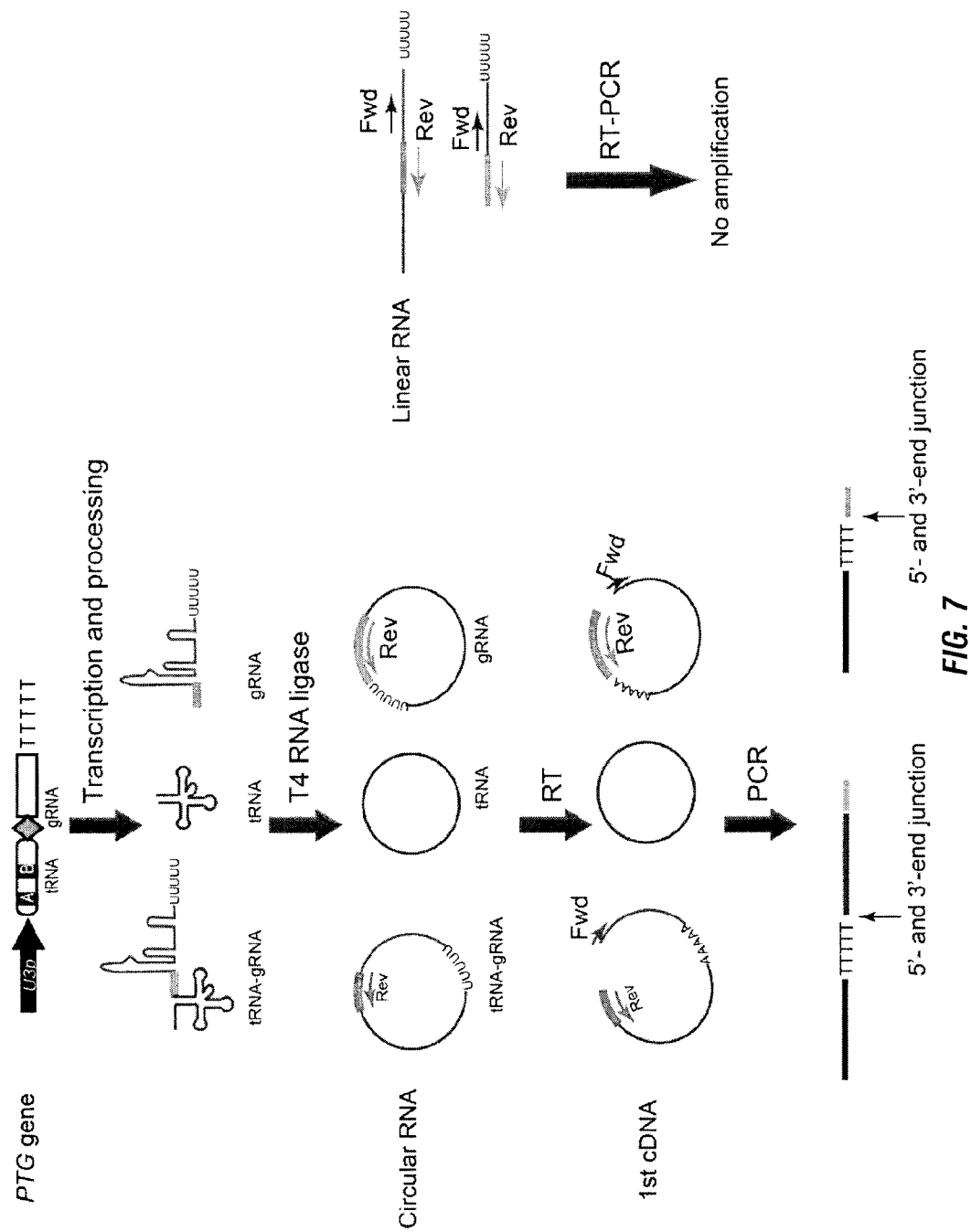
FIG. 6B

Diagram illustrating the structure of the PTG (Protein Terminator Gene) construct, which is a linear DNA molecule.

Construct Features:

- 5' End:** 5'-AAGGAATCTTTAAACA.....TGATTAAAGAGTTGTGCAGATGATCCGT GGCAGGAGACCCGAGGTCTCGGTTT TAGAGCTAGAAA.....CCGAGTCGGTGC TTTTTT -3'
- 3' End:** 3'-TTCTTAGAAATTGT.....AC_TAAATTTT CTCAACACGTCCTACTAGGCA CCGTCCCTCTGGCTCCAGAGC CAAA ATCTCGATCTTT.....GGCTCAGCCACG AAAAAA -5'
- TATA Box:** Located near the 5' end.
- Transcription start:** Indicated by an arrow pointing to the beginning of the coding sequence.
- U3p:** A region near the 3' end.
- Bsa I Restriction Sites:** Two sites are marked with vertical lines.
- gRNA scaffold (76 bp):** A region between the Bsa I sites.
- oligo duplex to construct sgRNA:**
 - 5'-GGCANNNNNNNNNNNNNNNNNNNN-3'
 - 3'-NNNNNNNNNNNNNNNNNNNCAAA-5'
- PTG Region:** A box containing the PTG sequence, flanked by a TTTT terminator and a CAAA-5' end.
- In vitro assembled PTG with Pol III terminator:** A note indicating the assembly method.

FIG. 6C



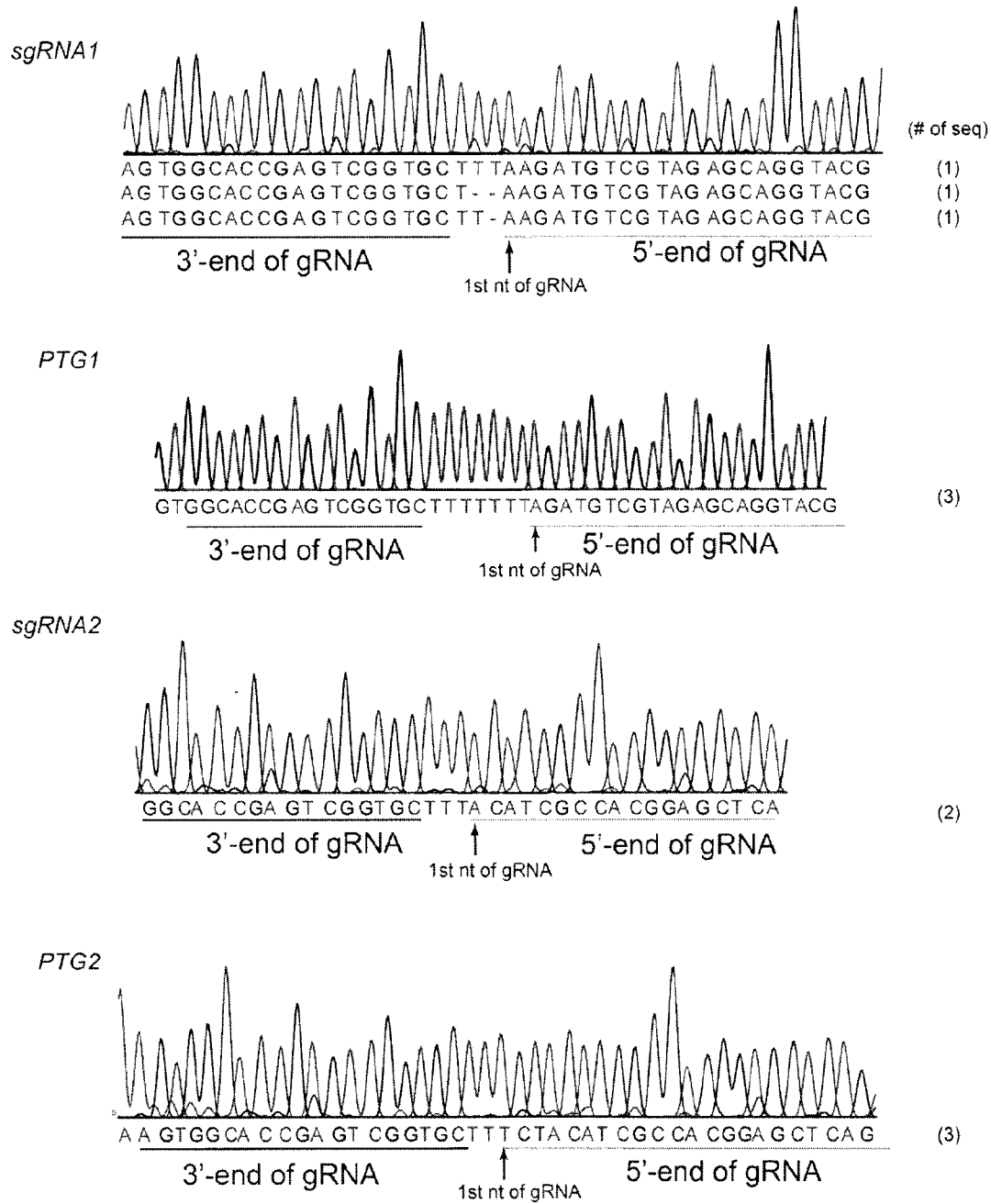


FIG. 8

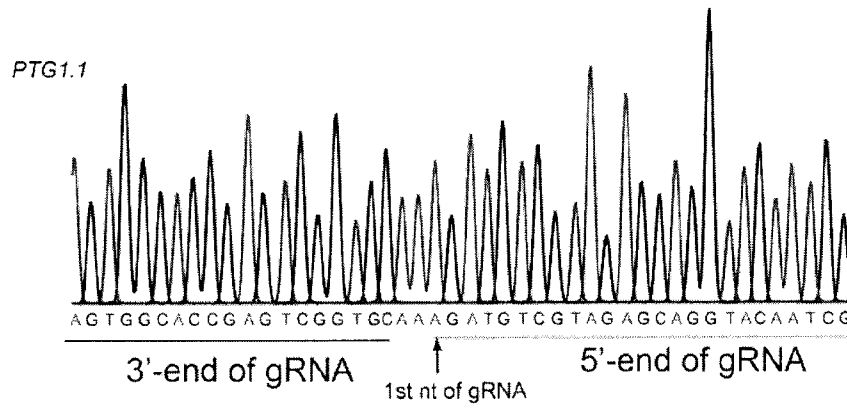


FIG. 9A

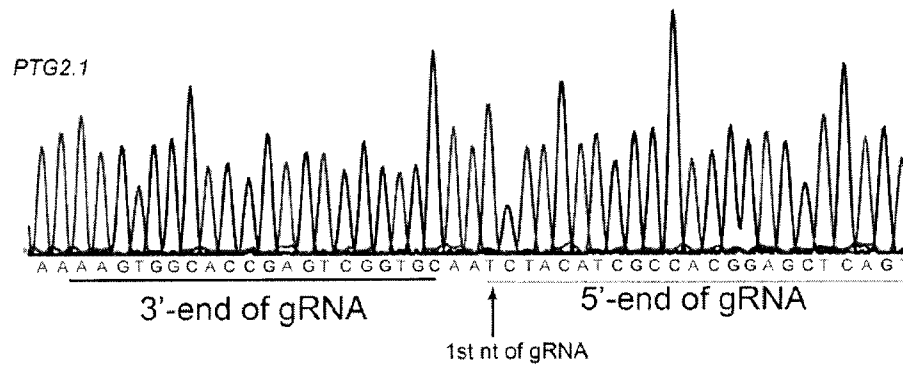


FIG. 9B

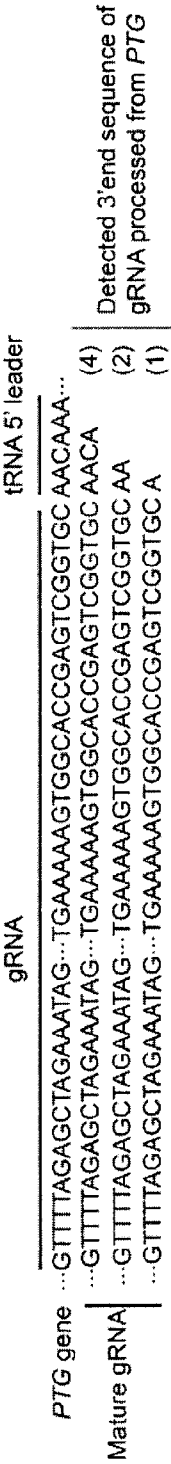


FIG. 9C

Design primers to synthesize assembly parts

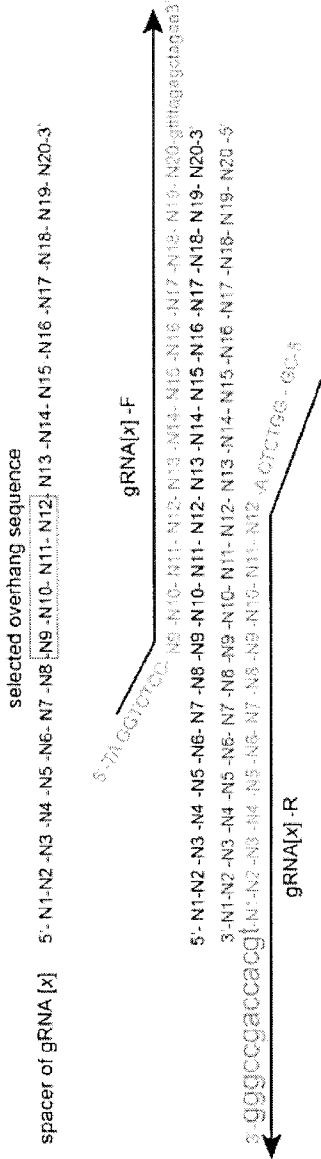


FIG. 10A

Ligate two parts with GG assembly

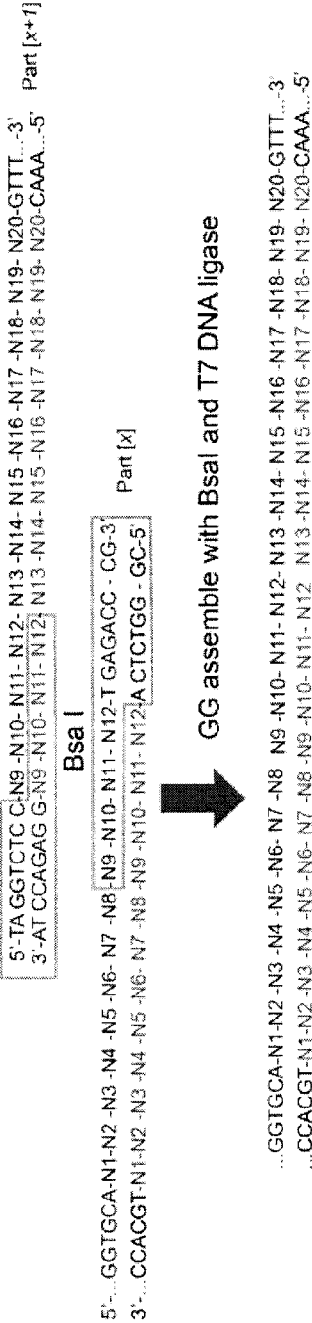


FIG. 10B

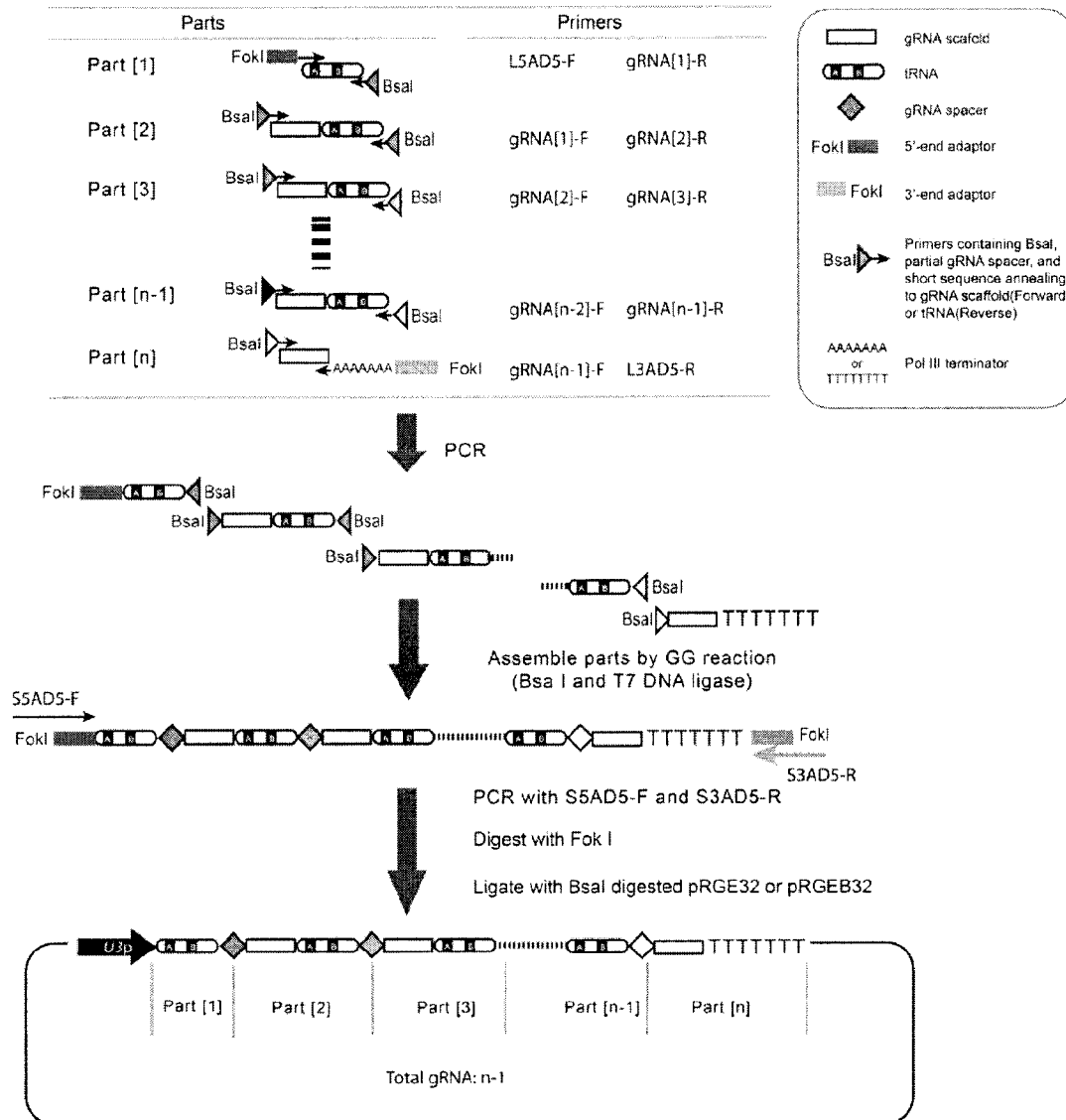


FIG. 10C

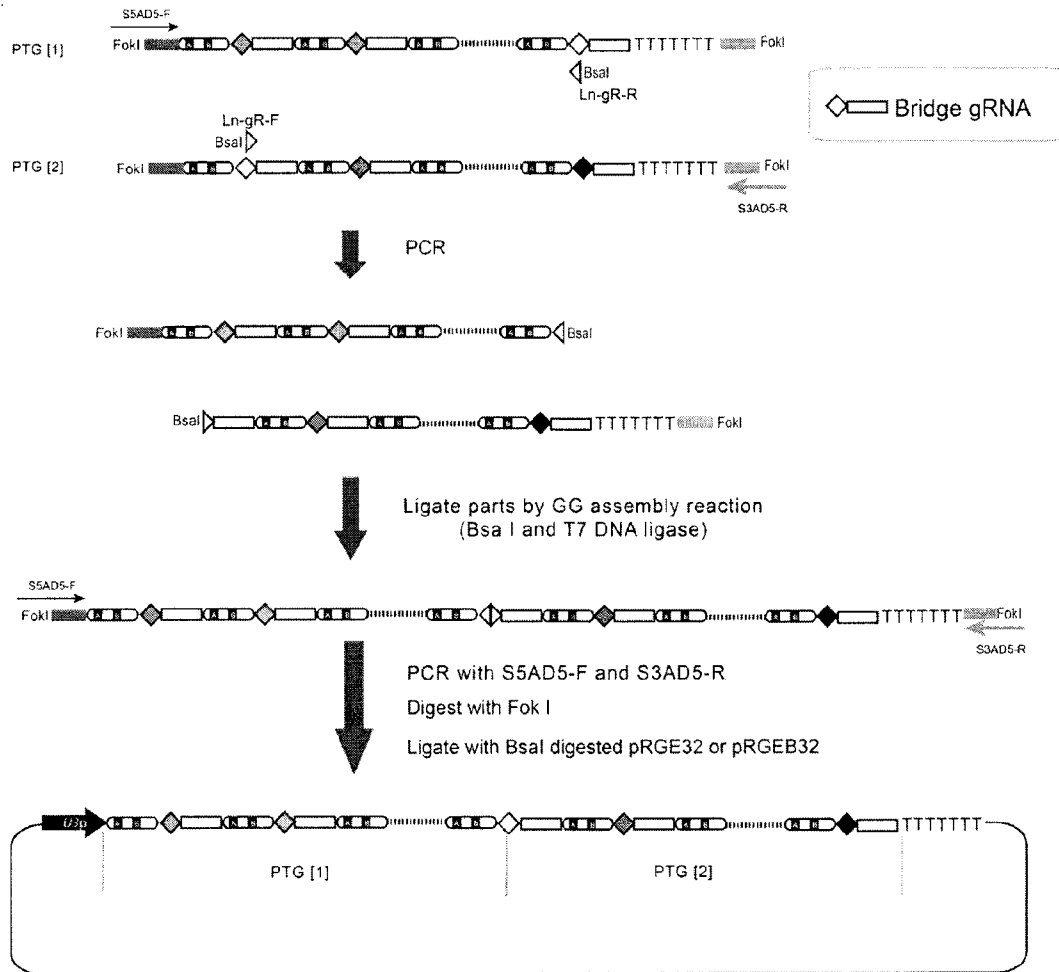


FIG. 11

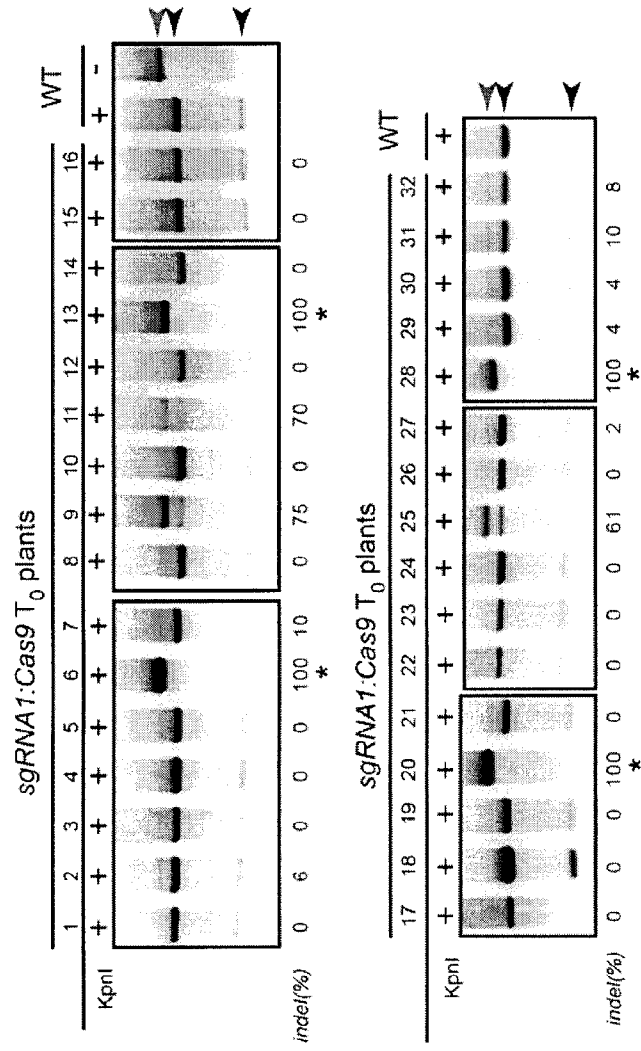
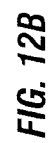


FIG. 12A



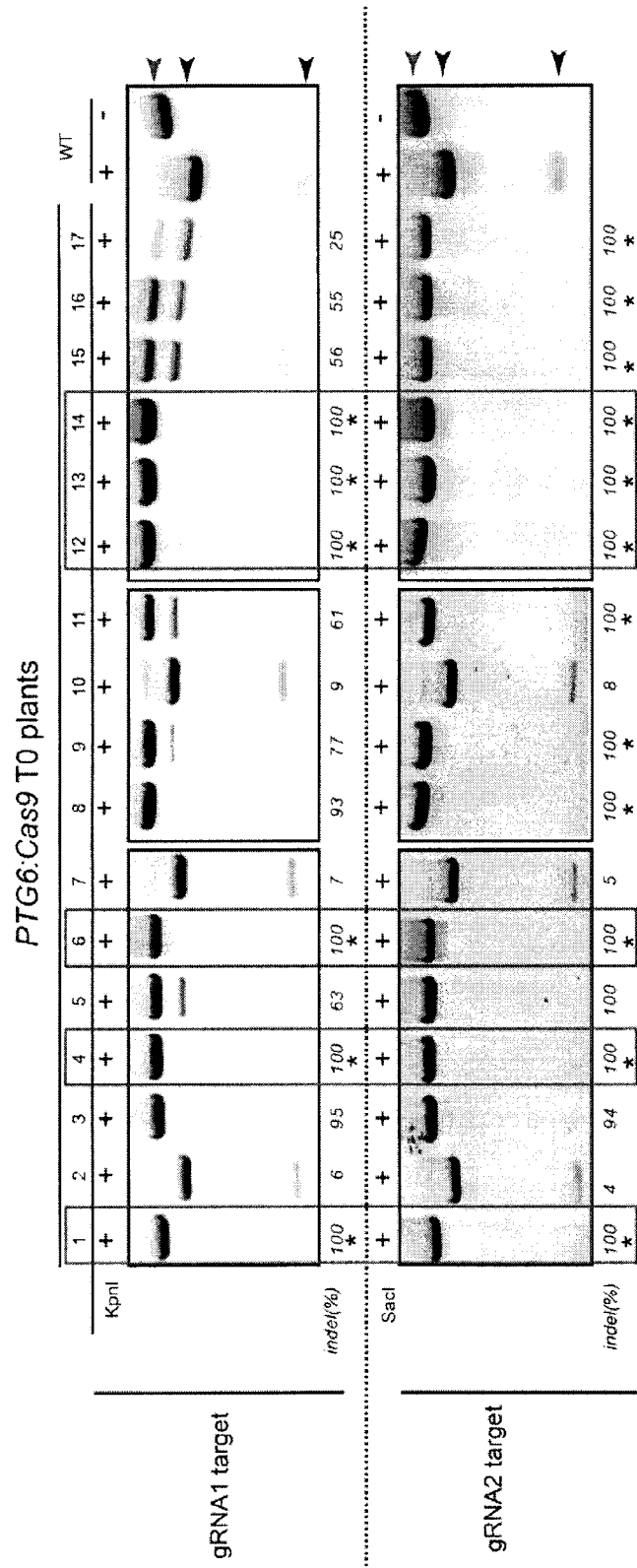


FIG. 13

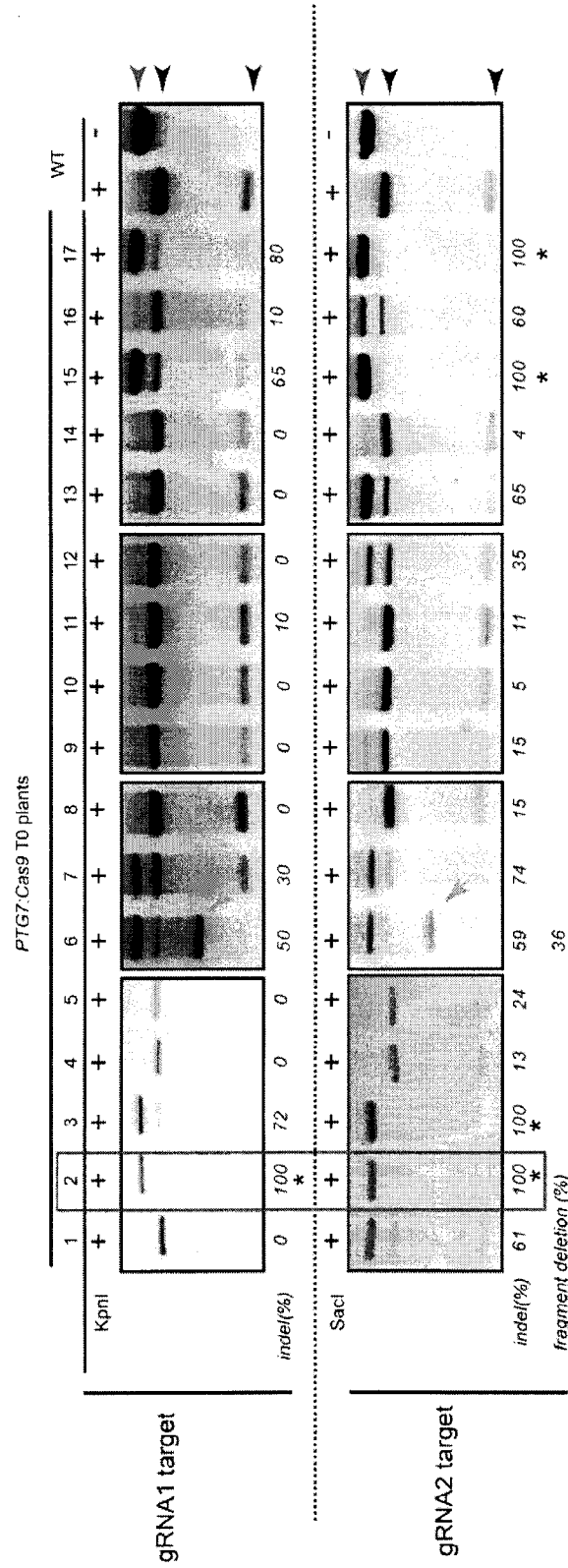


FIG. 14

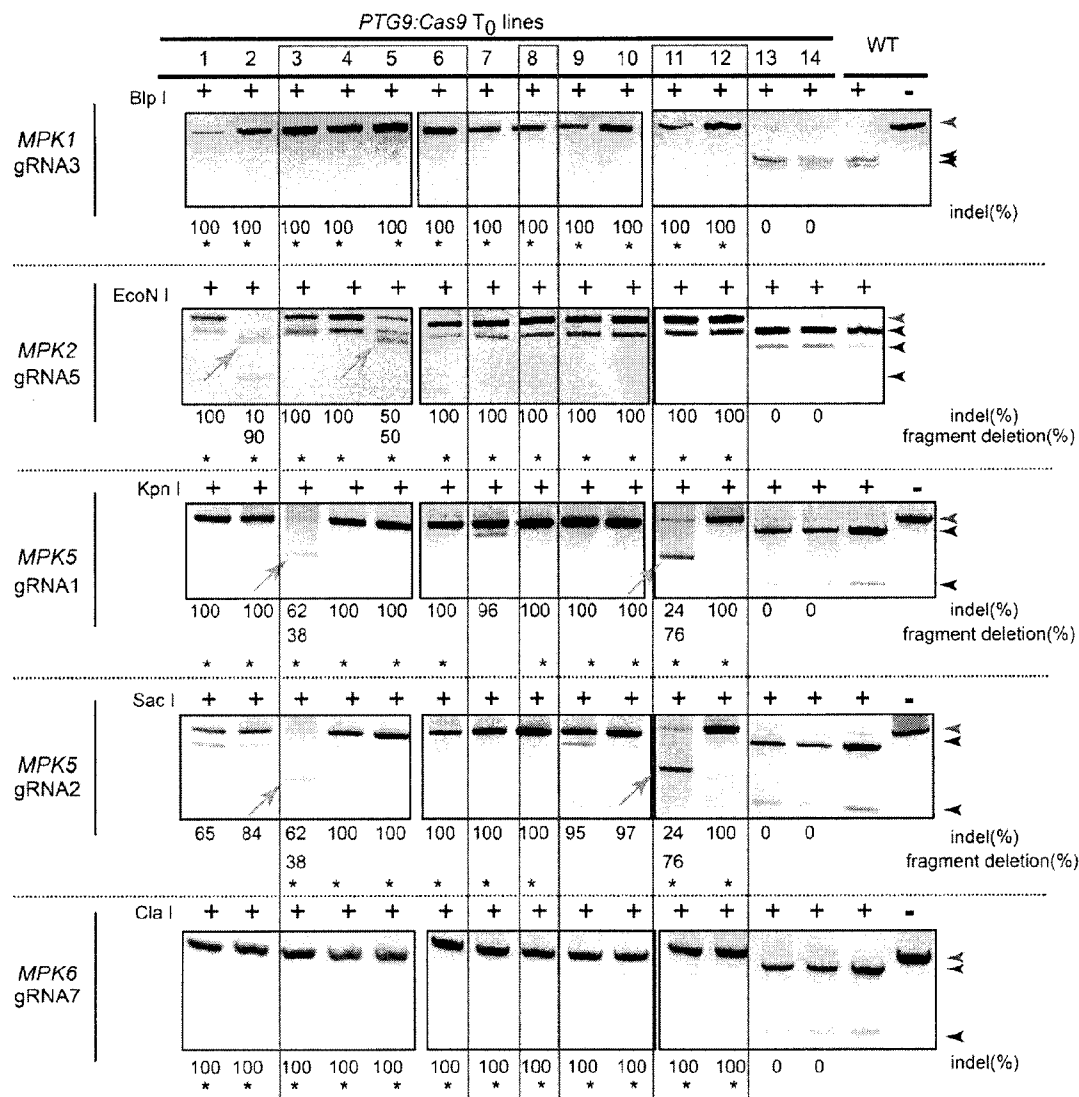


FIG. 15

MPK1	WT	CGCGTGCCCAACCAATTTATACCCCGGTGGGCCA	gRNA4	CATCCAGGCGACGCTGAG	gRNA3	CCATGGCGGGAGGTTTC	gRNA4 site	gRNA3 site
	#1	CGCGTGCCCAACCAATTTATACCCCGGTGGGCCA		CATCCAGGCGACGCTGAG		CCATGGCGGGAGGTTTC	-1 bp	G→A, +1 bp
	#10	CGCGTGCCCAACCAATTTATACCCCGGTGGGCCA		CATCCAGGCGACGCT---		CCATGGCGGGAGGTTTC	-1 bp	-3 bp
	#11	CGCGTGCCCAACCAATTTATACCCCGGTGGGCCA		CATCCAGGCGACGCTGA-		CCATGGCGGGAGGTTTC	-1 bp	-1 bp
MPK2	WT	TCCTCCTTCTCCCTCCCTGAGGCGGACCGGGTTC	gRNA5	GAATGGCGAGACTCGTC	gRNA6	AGGAGGTGGCAATCAA	gRNA5 site	gRNA6 site
	#1	TCCTCCTTCTCCCTCCCTGAGGCGGACCGGGTTC		GAATGGCG		AGGAGGTGGCAATCAA	-1 bp	-42 bp
	#10	TCCTCCTTCTCCCTCCCTGAGGCGGACCGGGTTC		GAATGGCG		AGGAGGTGGCAATCAA	+1 bp	+1 bp
	#11	TCCTCCTTCTCCCTCCCTGAGGCGGACCGGGTTC		GAATGGCGAGACTCGTC		AGGAGGTGGCAATCAA	-11 bp	+1 bp
MPK5	WT	CGCACGGCGGCGCGGTA CCTGCTCTACGACATCT	gRNA1	TCTACATCGCCACCGGAGC	gRNA2	TCA TGGACACCGACCT	gRNA1 site	gRNA2 site
	#1	CGCACGGCGGCGCGGTA CCTGCTCTACGACATCT		TCTACATCGCCACCGGAGC		TCA TGGACACCGACCT	-7 bp	+1 bp
	#10	CGCACGGCGGCGCGGTA CCTGCTCTACGACATCT		TCTACATCGCCACCGGAGC		TCA TGGACACCGACCT	-23 bp	+1 bp
	#11	CGCACGGCGGCGCGGTA CCTGCTCTACGACATCT		TCTACATCGCCACCGGAGC		TCA TGGACACCGACCT	delete 727 bp	
MPK6	WT	GGTGTGATCACCCCTTGATTGACCGCGCACCCGCG	gRNA8	CATTGACAAACCATAT	gRNA7	CGATGCCAAGCGGACAC	gRNA8 site	gRNA7 site
	#1	GGTGTGATCACCCCTTGATTGACCGCGCACCCGCG		CATTGACAAACCATAT		CGATGCCAAGCGGACAC	+1bp, -5 bp	-3 bp
	#10	GGTGTGATCACCCCTTGATTGACCGCGCACCCGCG		CATTGACAAACCATAT		CGATGCCAAGCGGACAC	-2 bp	-6 bp
	#11	GGTGTGATCACCCCTTGATTGACCGCGCACCCGCG		CATTGACAAACCATAT		CGATGCCAAGCGGACAC	-3 bp	+1 bp

FIG. 16

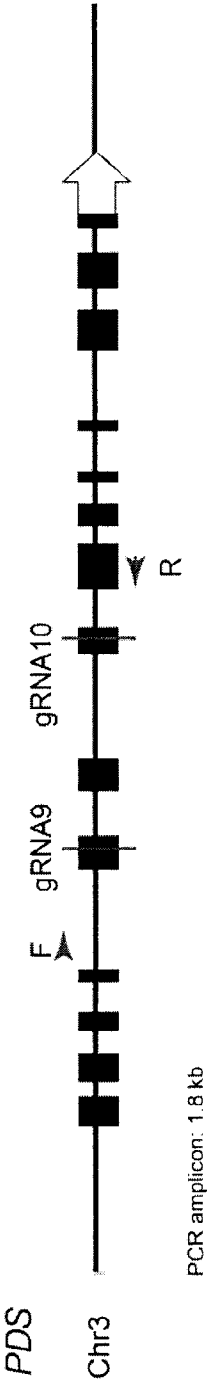


FIG. 17A

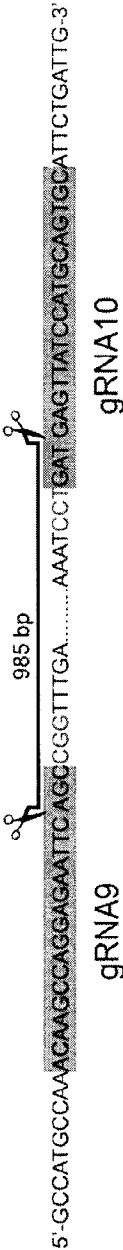


FIG. 17B

Gene ID	Gene architecture
<i>PTG10</i>	tRNA-gRNA9-tRNA-gRNA10

FIG. 17C

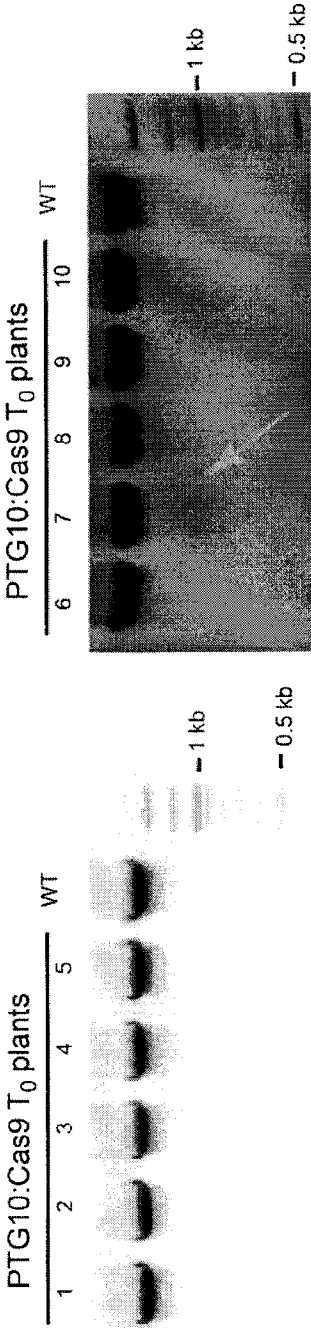


FIG. 17D

PTG10:Cas9 T₀ line 1

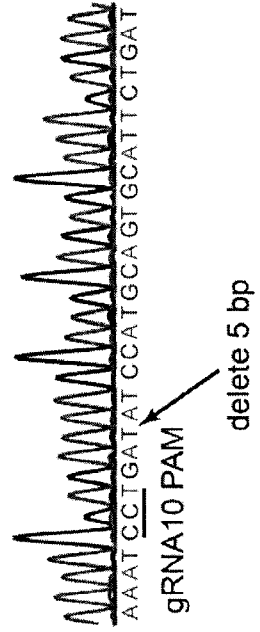


FIG. 17E

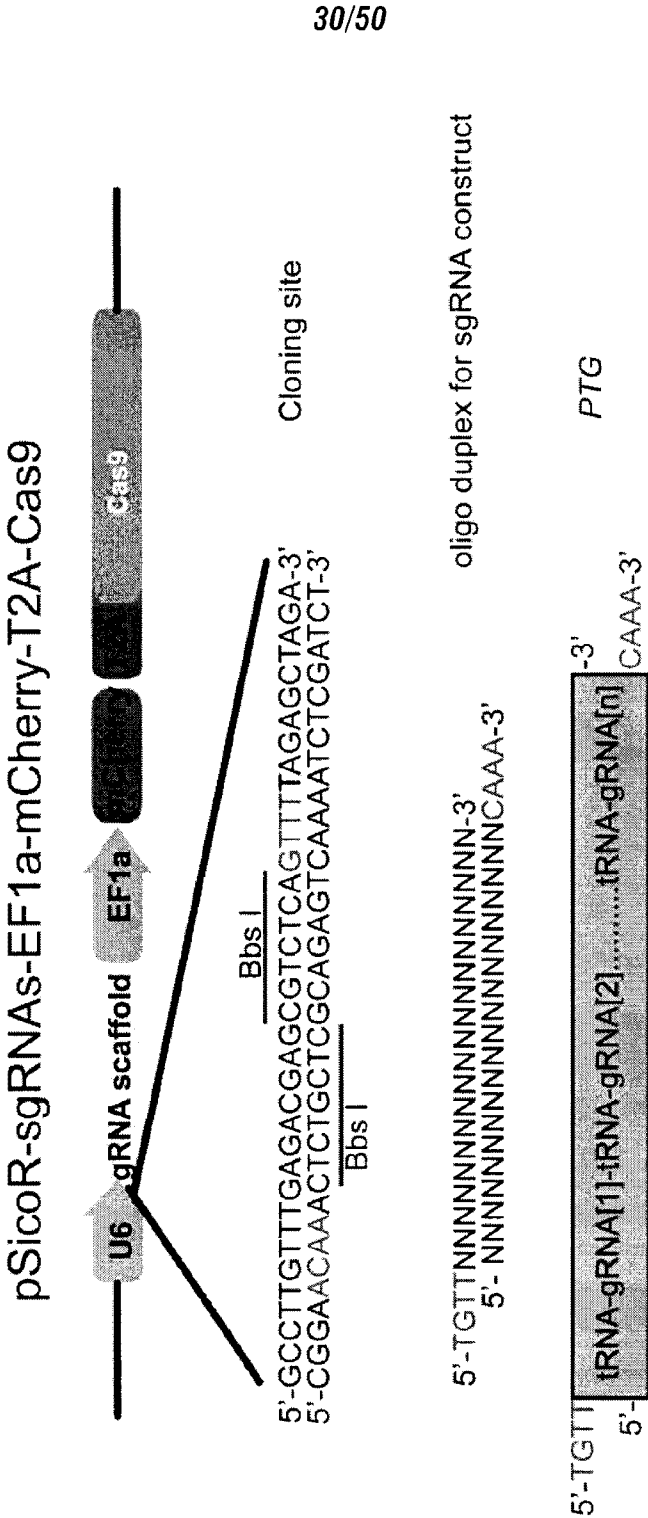


FIG. 18

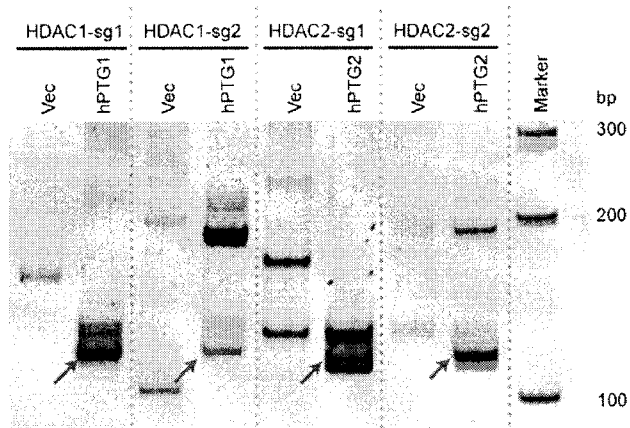


FIG. 19A

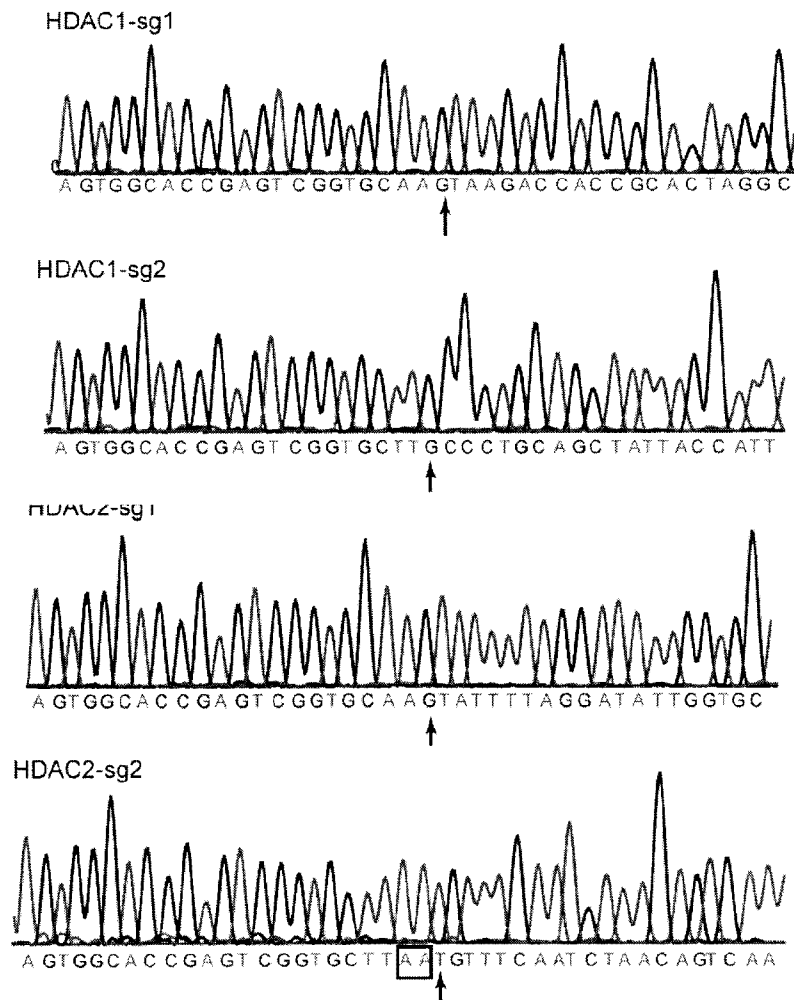
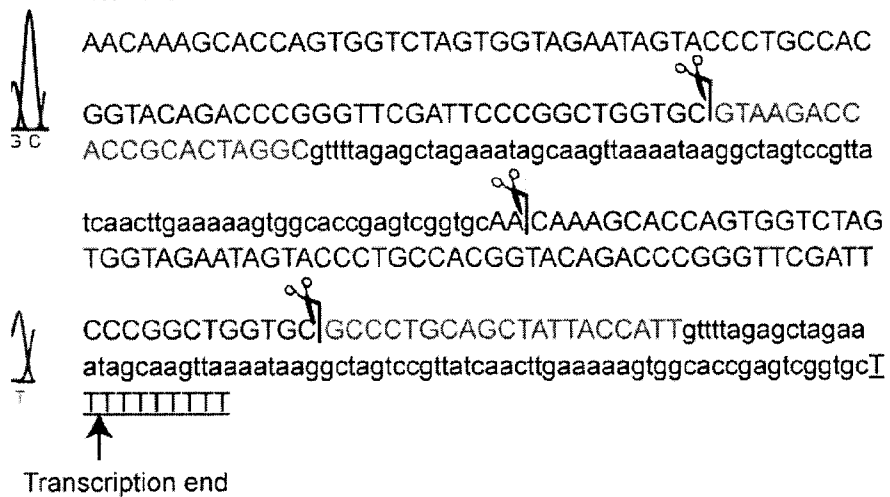


FIG. 19B

hPTG1



hPTG2

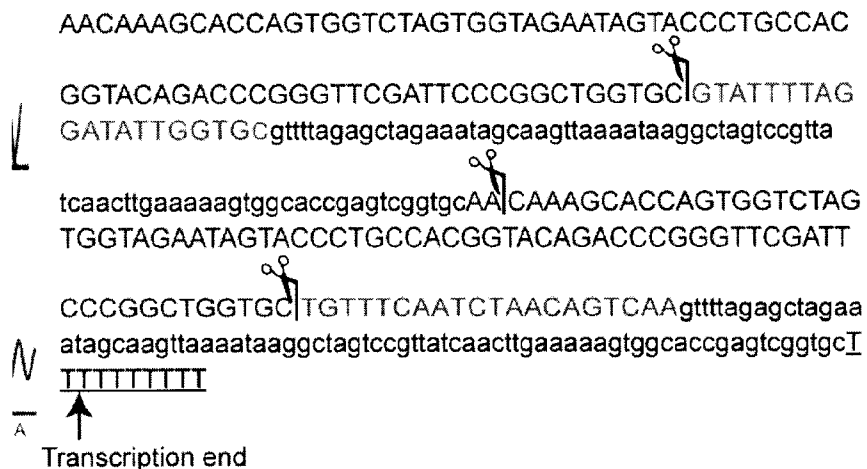


FIG. 19C

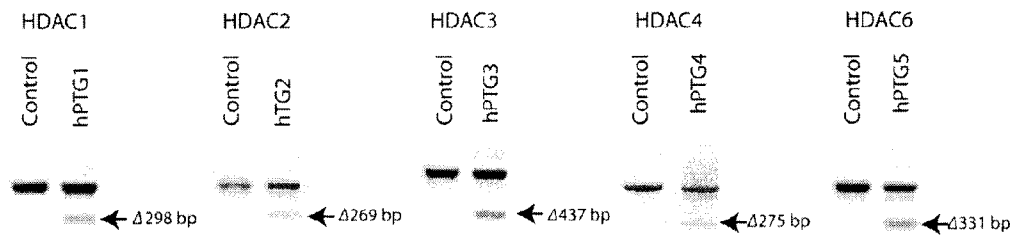


FIG. 20A

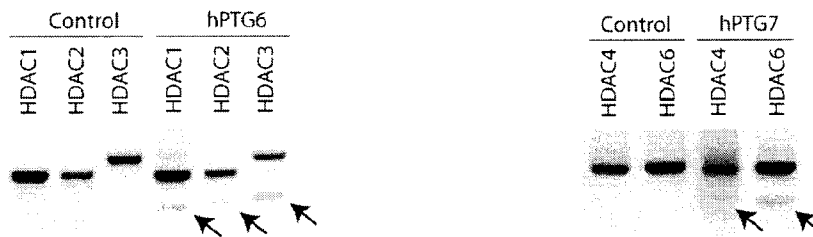


FIG. 20B

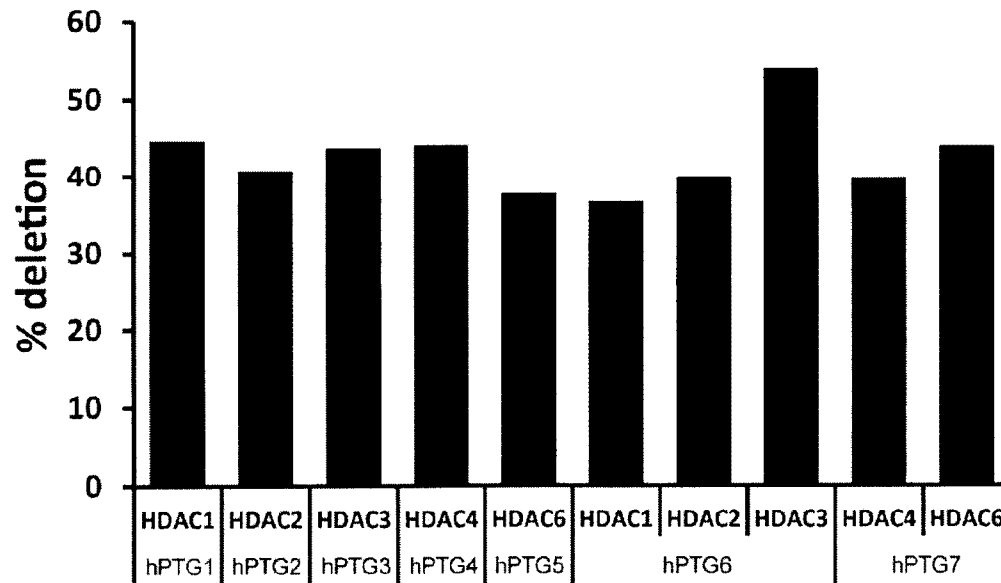


FIG. 20C

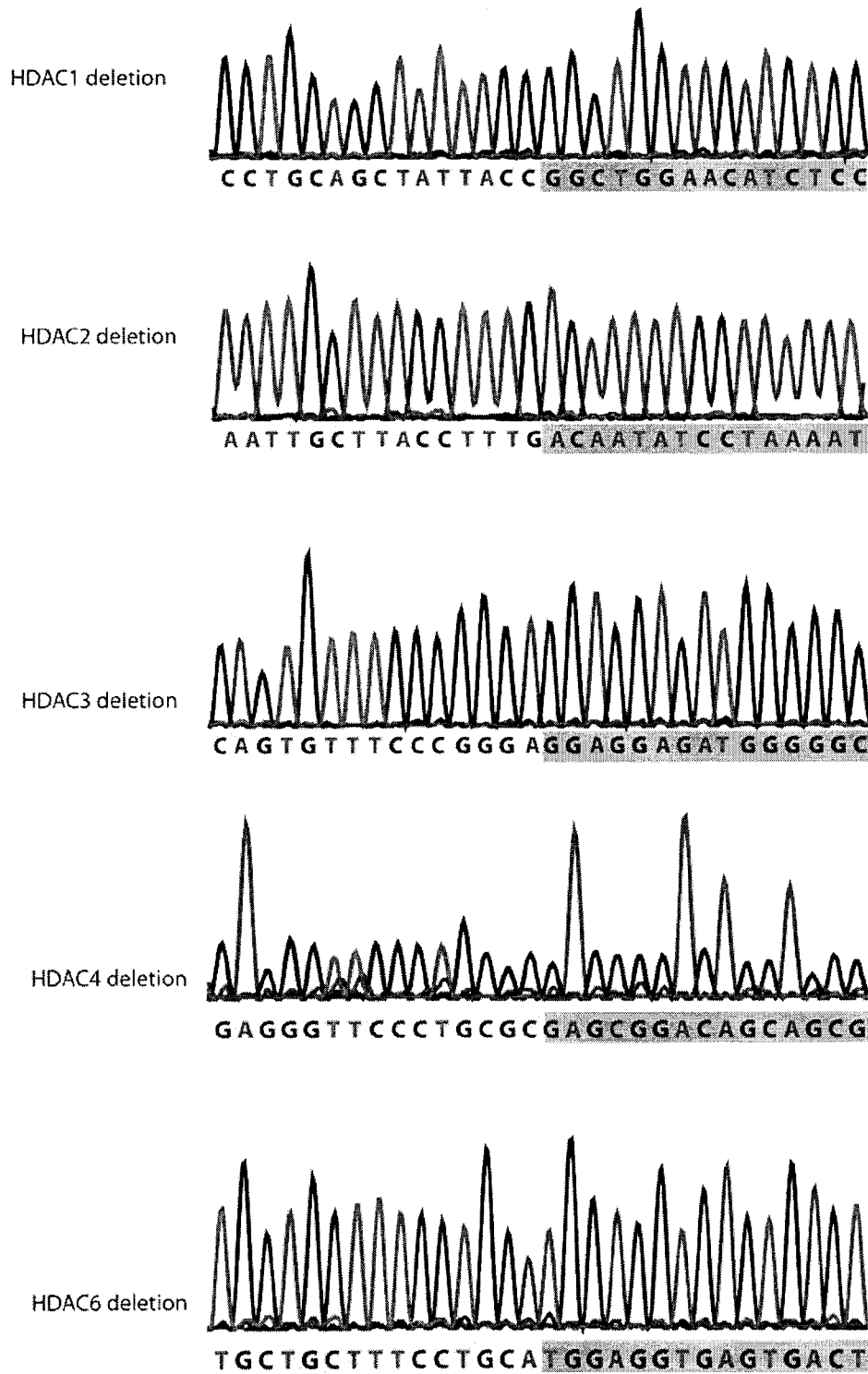
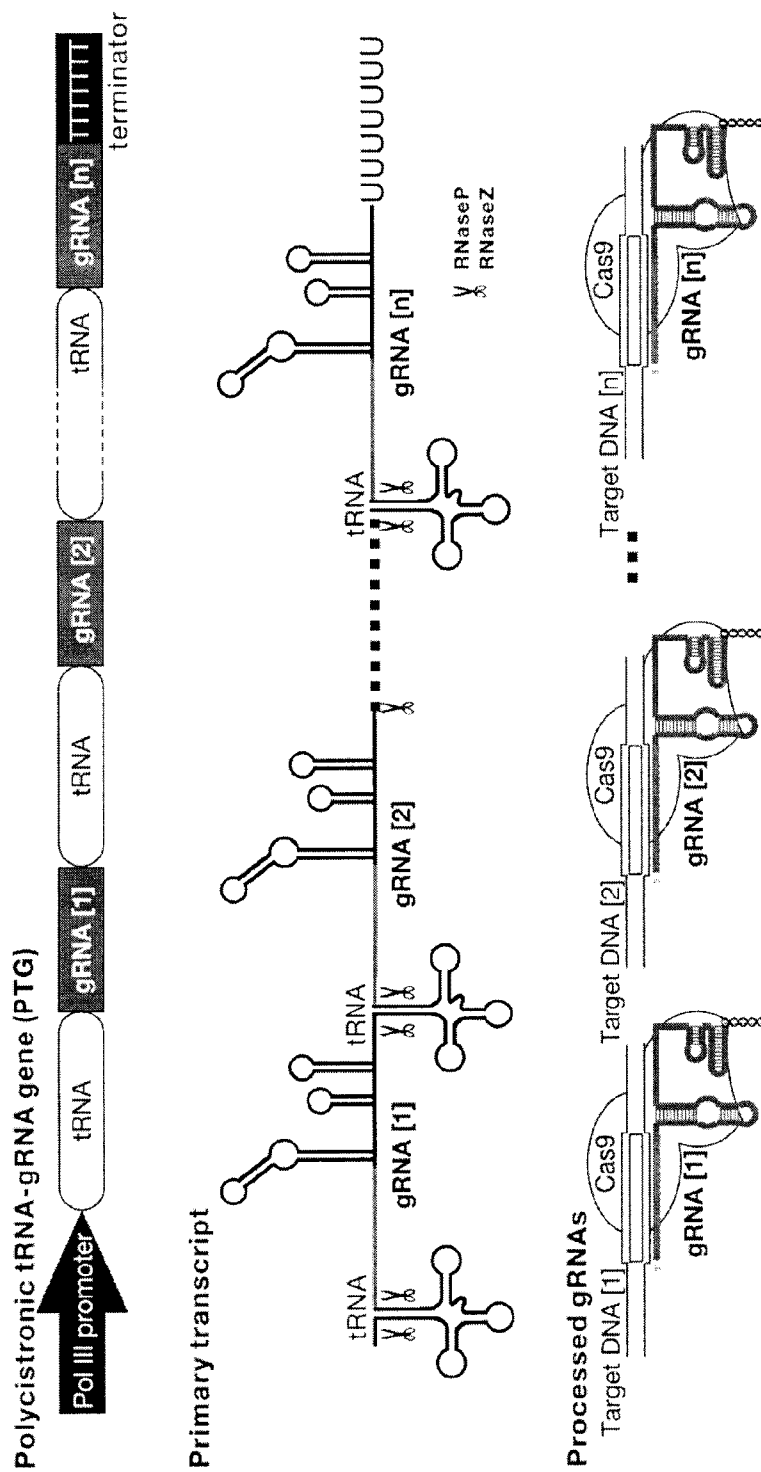


FIG. 21



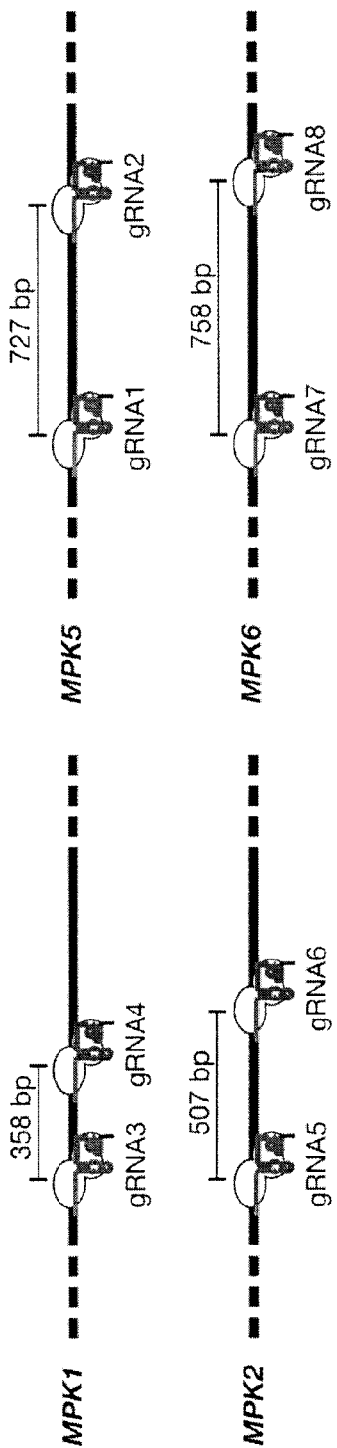


FIG. 22B

- | | |
|-------|-----------------------|
| PTG2: | tRNA-gRNA2 |
| PTG3: | tRNA-gRNA3-tRNA-gRNA4 |
| PTG4: | tRNA-gRNA5-tRNA-gRNA6 |
| PTG5: | tRNA-gRNA7-tRNA-gRNA8 |
| PTG6: | tRNA-gRNA1-tRNA-gRNA2 |
| PTG7: | PTG6 + PTG3 (4gRNAs) |
| PTG8: | PTG4 + PTG5 (4gRNAs) |
| PTG9: | PTG7 + PTG8 (8gRNAs) |

FIG. 22C

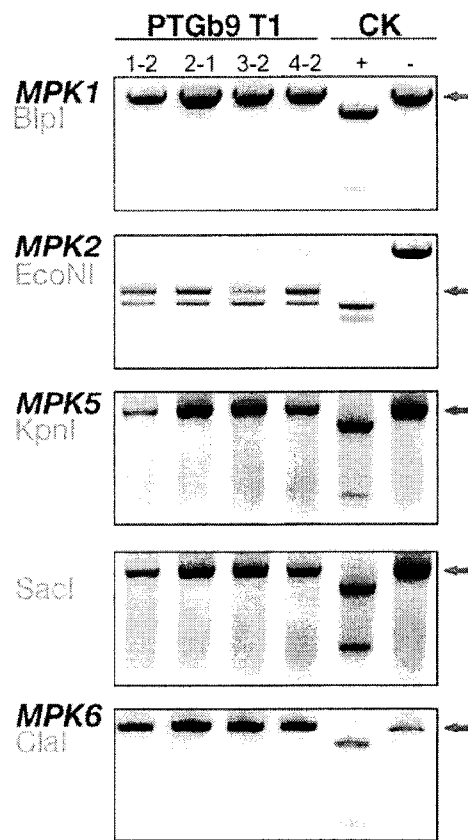


FIG. 23A

MPK1

WT CGCGTGCCAAACCA...TTT-TATACCCCGGTGGGCCA...CATCCAGGCGACGCTGAGC...CATGGCGGGAGGTTTC -1bp, -15bp
 1-2aCGCGTGCCAAACCA...TTT--ATACCCCGGTGGGCCA...CATCC-----ATGGCGGGAGGTTTC -3bp, -3bp
 1-2bCGCGTGCCAAACCA...TTT-T-ACCCCGGTGGGCCA...CATCCAGGCGACGCTGA--ATGGCGGGAGGTTTC -4bp, -3bp
 2-1 CGCGTGCCAAACCAaacaTTT-TATACCCCGGTGGGCCA...CATCCAGGCGACGCT---CCATGGCGGGAGGTTTC +1bp, -3bp
 3-2aCGCGTGCCAAACCA...TTTaTATACCCCGGTGGGCCA...CATCCAGGCGACGCT---CCATGGCGGGAGGTTTC 1c, -3bp
 3-2bCGCGTGCCAAACCA...TTT-TATAICCCCGGTGGGCCA...CATCCAGGCGACGCT---CCATGGCGGGAGGTTTC -1bp, -3bp
 4-2 CGCGTGCCAAACCA...TTT--ATACCCCGGTGGGCCA...CATCCAGGCGACGCT---CCATGGCGGGAGGTTTC

MPK2

WT TCCTCCTTCTCCCTCC...TTGAGGGCGACCCGGGTTTC...GAATGCGCAGACTCGTC-AGGAGGTGGCAATCAA -2bp, +2bp
 1-2 TCCTCCTTCTCCCTCC...GAGGCGACCCGGGTTTC...GAATGCGCAGACTCGTCaaAGGAGGTGGCAATCAA +1bp, +1bp
 2-1 TCCTCCTTCTCCCTCCtTTGAGGCGACCCGGGTTTC...GAATGCGCAGACTCGTCT-AGGAGGTGGCAATCAA +1bp, +1bp
 3-2 TCCTCCTTCTCCCTCCtTTGAGGCGACCCGGGTTTC...GAATGCGCAGACTCGTCC-AGGAGGTGGCAATCAA -11bp, -1bp
 4-2aTCCTCCTT...AGGCGACCCGGGTTTC...GAATGCGCAGACTCGTCT-AGGAGGTGGCAATCAA -11bp, +1bp
 4-2bTCCTCCTT...AGGCGACCCGGGTTTC...GAATGCGCAGACTCGTCT-AGGAGGTGGCAATCAA

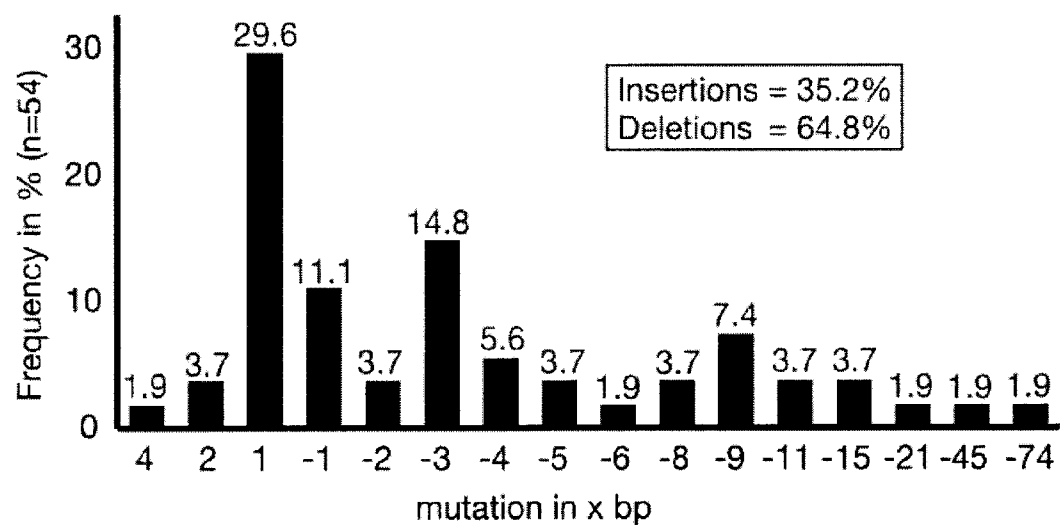
MPK5

WT CGCACGGGCGGCGCGGTA...CCTGCTCTACGACATCTTC...TCTACATGCCACCGGAG-C-TCATGGACACCGACCT +1bp, +1bp
 1-2 CGCACGGGCGGCGCGGTAICCTGCTCTACGACATCTTC...TCTACATGCCACCGGAGcCT-CATGGACACCGACCT 1c/-4bp, -8bp
 2-1 CGCACGGGCGGCGCGGTI-----CTCTACGACATCTTC...TCTACATCG-----CT-CATGGACACCGACCT -5bp, -8bp
 3-2 CGCACGGGCGGCGCGGTA-----TCTACGACATCTTC...TCTACATCG-----CT-CATGGACACCGACCT -5bp, +1bp
 4-2 CGCACGGGCGGCGCGGTA-----TCTACGACATCTTC...TCTACATGCCACCGGAG-CTtCATGGACACCGACCT

MPK6

WT GGTGTGATCACCCCTTTTGATTGACCCCGCACCCCGC...ATTGACAACCATAT-CGATGCCAAGCGGACAC -6bp, 1c/-1bp
 1-2aGGTGTGATCACCC-----TTGACCCCGCACCCCGC...ATTGACAACCAT-c-CGATGCCAAGCGGACAC -48bp, -21bp
 1-2b-----ACCCGACCCCGC...ATTGACAACCATAT-----TGCCAAGCGGACAC -4bp, -3bp
 2-1 GGTGTGATCACCCCTTT-A-----ACCCGACCCCGC...ATTGACAACCATAT-----TGCCAAGCGGACAC -8bp, -3bp
 3-2 GGTGTGATCACCCCTTT-A-----CACCCGC...ATTGACAACCATAT-----TGCCAAGCGGACAC -9 bp, -3bp
 4-2aGGTGTGATCACCCCTTT-A-----CACCCGC...ATTGACAACCATATaCGATGCCAAGCGGACAC -9 bp, +1bp
 4-2bGGTGTGATCACCCCTTT-A-----CACCCGC

FIG. 23B

**FIG. 24**

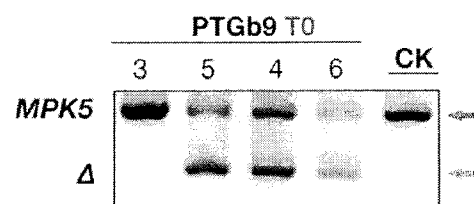


FIG. 25A

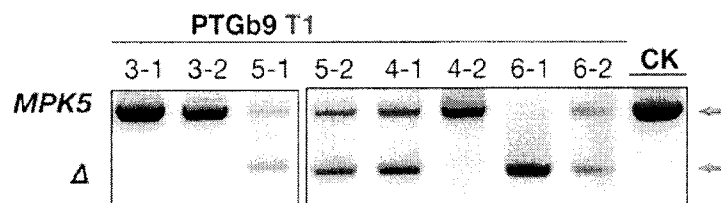


FIG. 25B

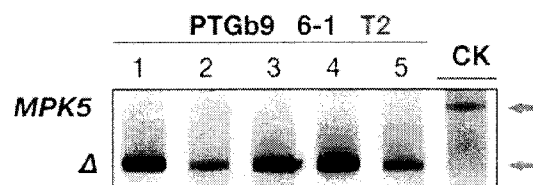


FIG. 25C

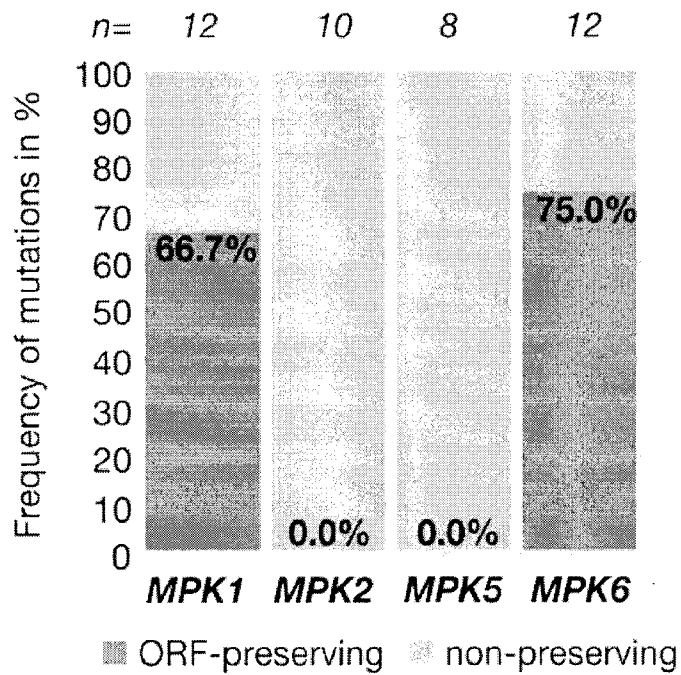


FIG. 26A

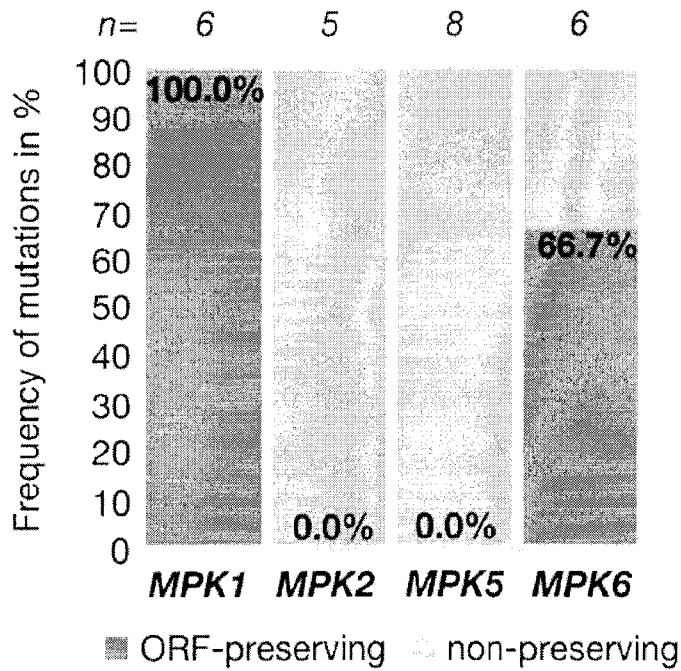


FIG. 26B

MPK1
mpk1 1-2a
mpk1 2-1 / 3-2 / 4-2
mpk1 1-2b

MDAGAQPDPDTEMAEAGGGQQPPAAAAAAGAGAGAGMMENIQATLSHGGRFIQYNIFGNVF 60
MDAGAQPDPDTEMAEAGGGQQPPAAAAAAGAGAGAGMMENI-----HGGRFIQYNIFGNVF 55
MDAGAQPDPDTEMAEAGGGQQPPAAAAAAGAGAGAGMMENIQAT-LHGGRFIQYNIFGNVF 59
MDAGAQPDPDTEMAEAGGGQQPPAAAAAAGAGAGAGMMENIQAT-LNGGRFIQYNIFGNVF 59
*****;*****

FIG. 26C

mpk5 1-2
mpk5 4-2
mpk5 3-2
mpk5 2-1
MPK5

MDGAPVAEFRPTMTHGGRYPA----- 21
MDGAPVAEFRPTMTHGGRY----- 19
MDGAPVAEFRPTMTHGGRY----- 19
MDGAPVAEFRPTMTHGGRF----- 19
MDGAPVAEFRPTMTHGGRYLLYDIFGNKFEVTNKYQPPIMPIGRGAYGIVCSVMNFETRE 60
*****;

FIG. 26D

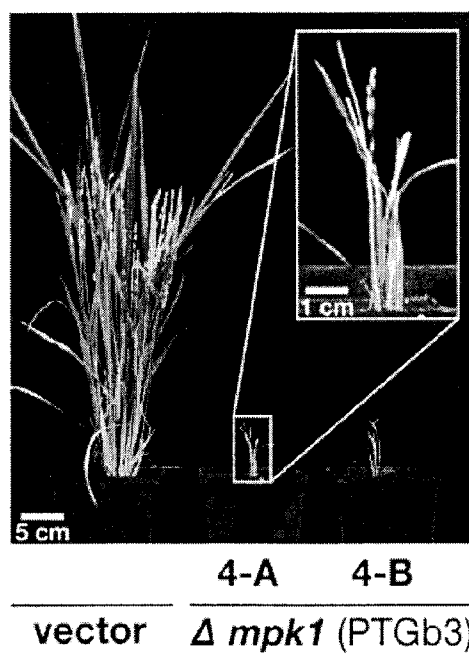


FIG. 26E

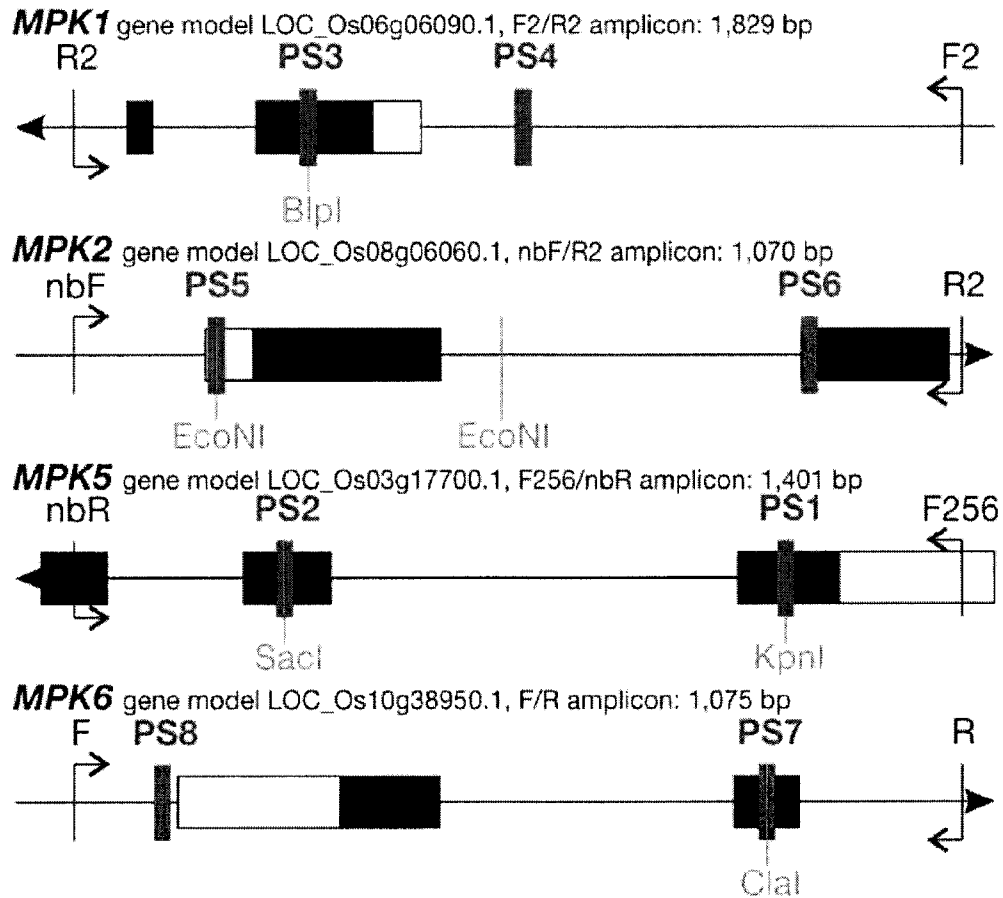


FIG. 27

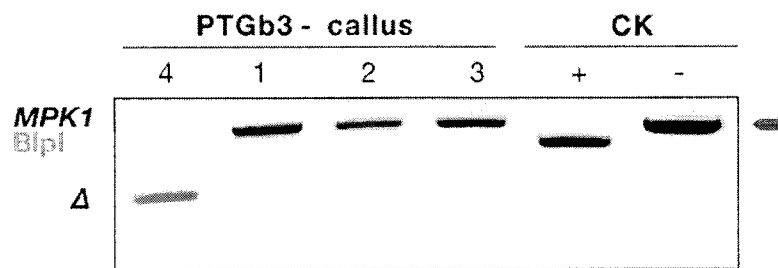


FIG. 28A

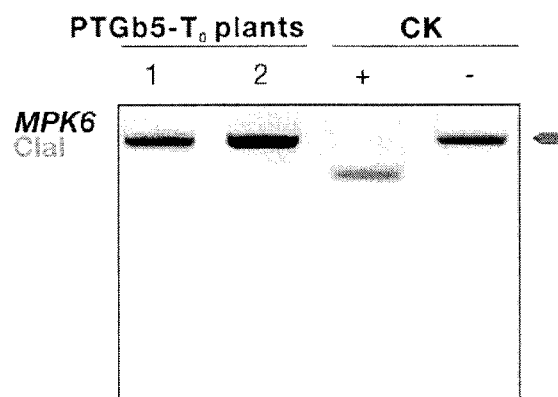


FIG. 28B

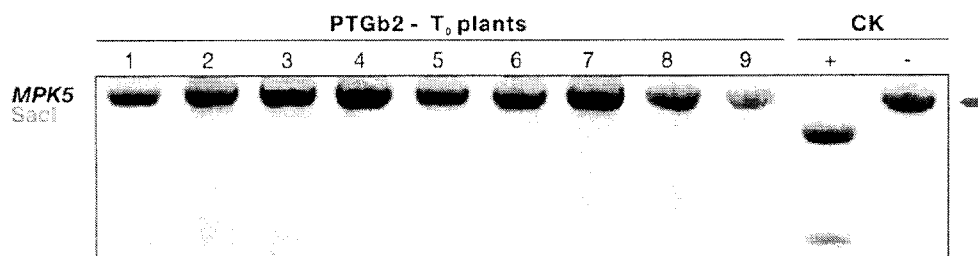


FIG. 29A

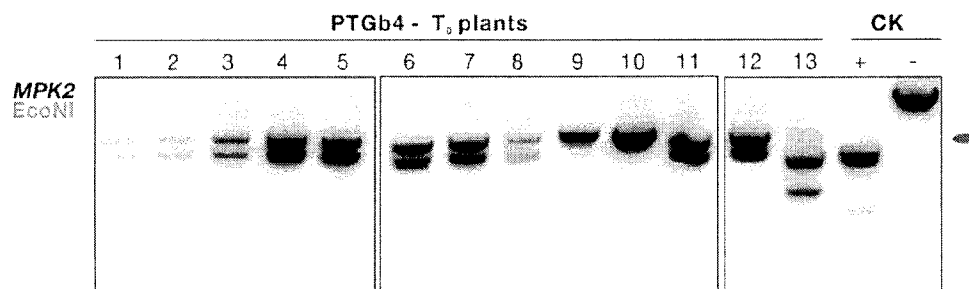


FIG. 29B

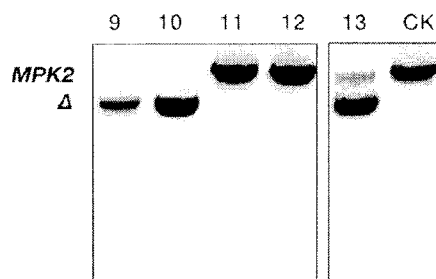
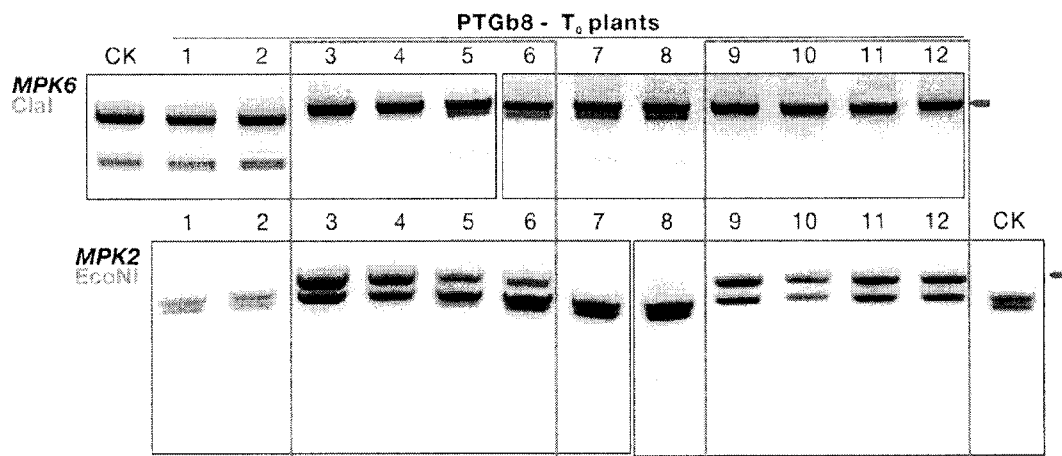


FIG. 29C

**FIG. 30**

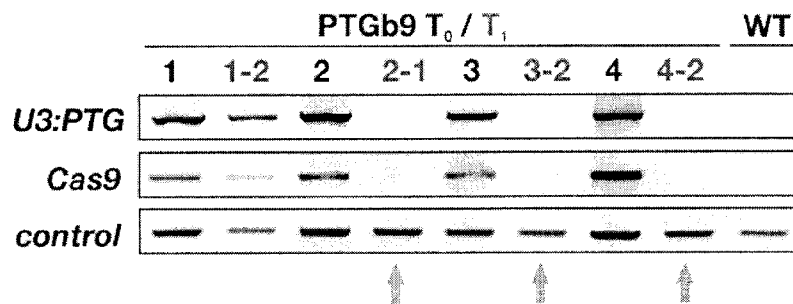


FIG. 31

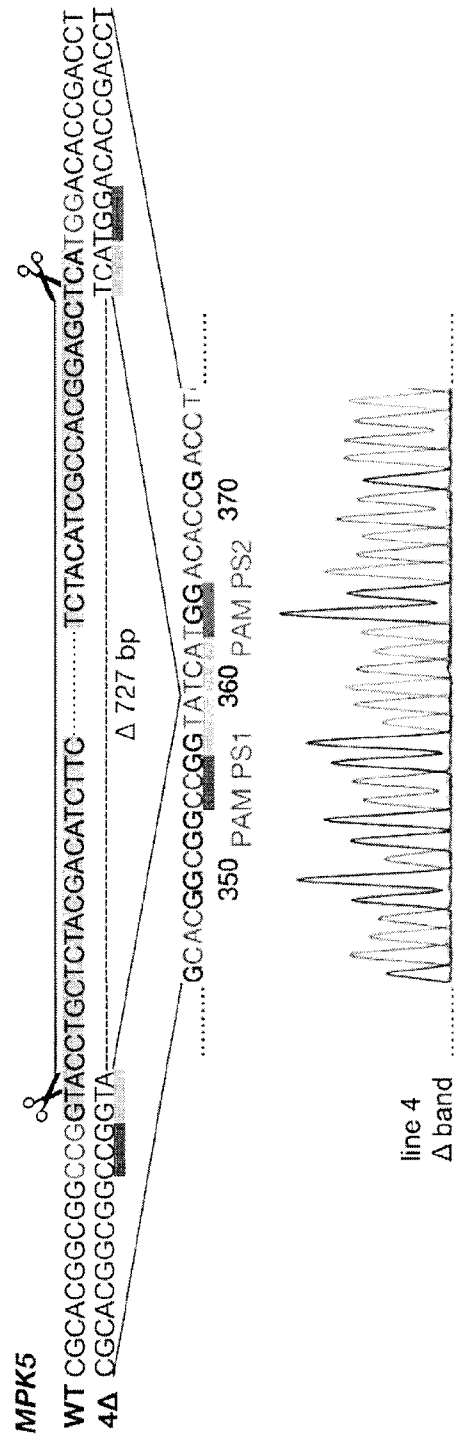


FIG. 32

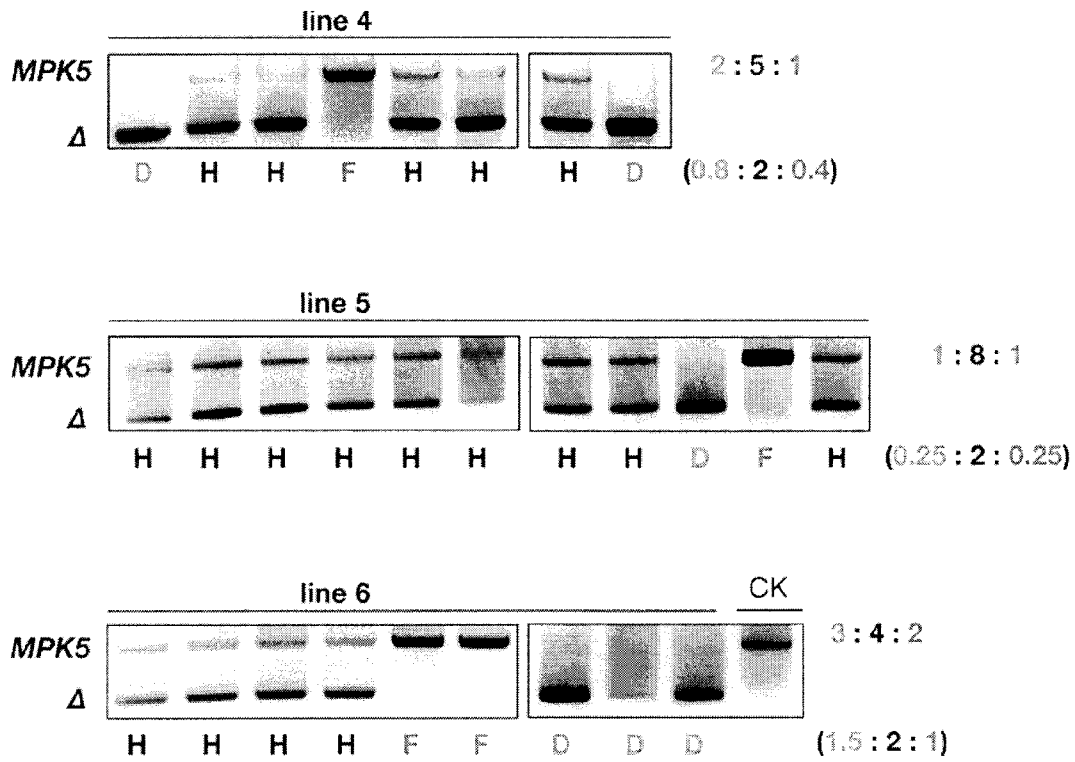
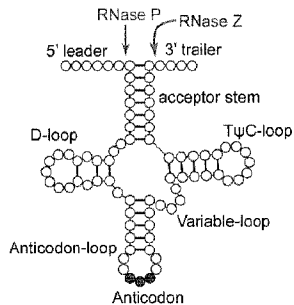
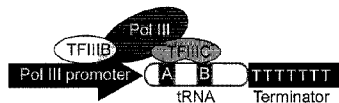


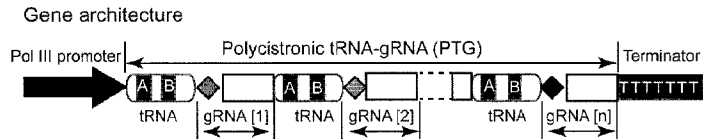
FIG. 33



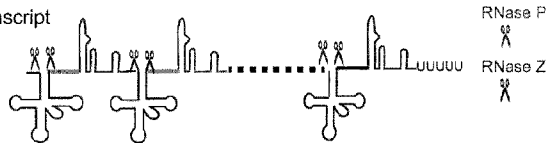
A



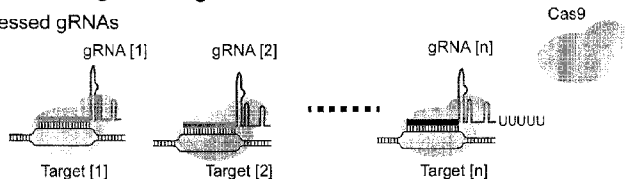
B



Primary transcript



Processed gRNAs



C