(12) STANDARD PATENT (11) Application No. **AU 2017256517 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Improved methods for modification of target nucleic acids

(51) International Patent Classification(s)
C12N 9/22 (2006.01) ***C12N 15/90*** (2006.01)

(21) Application No: **2017256517** (22) Date of Filing: **2017.04.20**

(87) WIPO No: **WO17/186550**

(30) Priority Data

(31) Number	(32) Date	(33) Country
16167773.7	**2016.04.29**	**EP**
16167774.5	**2016.04.29**	**EP**
17156018.8	**2017.02.14**	**EP**

(43) Publication Date: **2017.11.02**
(44) Accepted Journal Date: **2023.09.28**

(71) Applicant(s)
BASF Plant Science Company GmbH

(72) Inventor(s)
Neuteboom, Leendert W.;McElver, John A.;De Pinho Barroco, Rosa;De Wilde, Chris;Felle, Max Fabian

(74) Agent / Attorney
Griffith Hack, Level 15, 376-390 Collins Street, Melbourne, VIC, 3000, AU

(56) Related Art
WO 2017/180711 A1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 November 2017 (02.11.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/186550 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/90* (2006.01)

(21) International Application Number:
PCT/EP2017/059331

(22) International Filing Date:
20 April 2017 (20.04.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

16167773.7	29 April 2016 (29.04.2016)	EP
16167774.5	29 April 2016 (29.04.2016)	EP
17156018.8	14 February 2017 (14.02.2017)	EP

(71) Applicant: **BASF PLANT SCIENCE COMPANY GMBH** [DE/DE]; Carl-Bosch-Str. 38, 67056 Ludwigshafen am Rhein (DE).

(72) Inventors: **NEUTEBOOM, Leendert W**; 10 Seabury Way, Youngsville, 27596 (US). **MCELVER, John A**; 722 Blackard Rd, Roxboro, 27574 (US). **DE PINHO BARROCO, Rosa**; Technologiepark 21, 9052 Gent (Zwijnaarde) (BE). **DE WILDE, Chris**; Pontweg 91, 9310 Herdersem-Aalst (BE). **FELLE, Max Fabian**; Valentin-Ostertagstr. 2, 67098 Bad Duerkheim (DE).

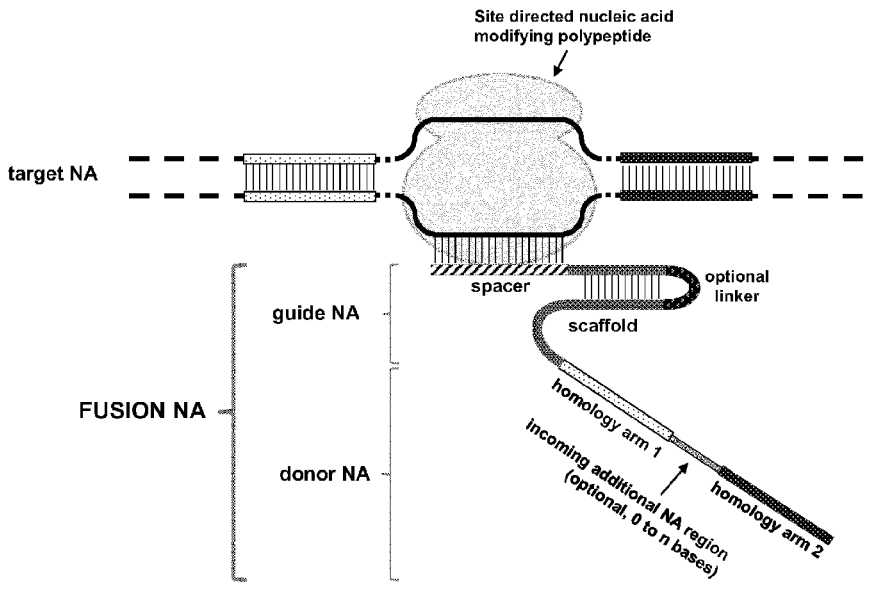
(74) Agent: **BASF IP ASSOCIATION**; G-FLP - C006, 67056 Ludwigshafen (DE).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

(54) Title: IMPROVED METHODS FOR MODIFICATION OF TARGET NUCLEIC ACIDS

Figure 1



(57) Abstract: Methods for modification of target nucleic acids. The method involves a construct in which guide RNA is covalently linked to donor RNA (fusion NA) to be introduced into the target nucleic acid by homologous recombination and is based on the introduction of a nuclease, e.g. CRISPR or TALEN, into the cell containing the target nucleic acid. The fusion NA may be introduced as a DNA vector.

WO 2017/186550 A1

[Continued on next page]

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

2017256517 31 Aug 2023

Improved methods for modification of target nucleic acids

Field of the Invention

The invention is directed to improved methods for modification of target nucleic acids.

Background

The CRISPR (clustered regularly interspaced short palindromic repeats) system was initially identified as an adaptive defense mechanisms of bacteria belonging to the genus of Streptococcus (WO2007/025097). Those bacterial CRISPR systems rely on guide RNA (gRNA) in complex with cleaving proteins to direct degradation of complementary sequences present within invading viral DNA. Cas9, the first identified protein of the CRISPR/Cas system, is a large monomeric DNA nuclease guided to a DNA target sequence adjacent to the PAM (protospacer adjacent motif) sequence motif by a complex of two noncoding RNAs: crRNA and trans-activating crRNA (tracrRNA). Later, a synthetic RNA chimera (single guide RNA or gRNA) created by fusing crRNA with tracrRNA was shown to be equally functional (Jinek et. al. 2012).

Several research groups have found that the CRISPR cutting properties could be used to disrupt genes in almost any organism's genome with unprecedented ease (Mali P, et al (2013) Science. 339(6121):819-823; Cong L, et al (2013) Science 339(6121)). Recently it became clear that providing a template for repair allowed for editing the genome with nearly any desired sequence at nearly any site, transforming CRISPR into a powerful gene editing tool (WO/2014/150624, WO/2014/204728).

Gene targeting refers to site specific gene modification by nucleic acid deletion, insertion or replacement via homologous recombination (HR). Targeting efficiency is highly promoted by a double-strand break (DSB) in the genomic target. Also, the direct presence of homology after DSB of chromosomal DNA seems to nearly eliminate non-homologous end joining (NHEJ) repair in favor of homologous recombination.

The invention at hand provides a guide nucleic acid fused to the appropriate donor for HR repair which interacts with a nucleic acid modifying polypeptide (fusion nucleic acid (fuNA) molecule comprising a guide nucleic acid (gNA) molecule covalently linked to at least one donor nucleic acid (doNA) molecule). To improve delivery of donor nucleic acid with homology to target DNA flanking a nuclease cutting site, a gene targeting strategy is presented where the donor nucleic acid (doNA) is covalently linked to the CRISPR components. In this way, the gene editing complex comprises not only the necessary recognition and cutting tools but

also the template for modification. Upon recognition by the guide nucleic acid (gNA), the nuclease cleaves the target region, and the immediate synchronized presence of incoming donor will facilitate the HR process, thereby increasing gene repair efficacy.

Many microbial systems lack an efficient NHEJ system [Standage-Beier K, Zhang Q, Wang X (2015) Targeted Large-Scale Deletion of Bacterial Genomes Using CRISPR-Nickases. ACS Synth Biol. 4(11): 1217–1225]. E.g. *Clostridium cellulolyticum* which is important for bioenergy research, cannot be engineered easily with CRISPR/Cas9 [Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, Van Nostrand JD, He Z, Zhou J (2015) Efficient Genome Editing in Clostridium cellulolyticum via CRISPR-Cas9 Nickase. Appl Environ Microbiol. 81(13):4423-31] and attempts to knockout genes by introducing DSBs result in cell death. Another example is *Escherichia coli* which also relies on homologous recombination (HR) for DSB repair. Current techniques for genome editing in such organisms focus on the use of CRISPR/Cas9-based nickases with subsequent inefficient repair via non-fused nucleic acid donors, inducible expression of CRISPR/Cas to coordinate nuclease action and the introduction of nucleic acid donors or recombinases which lack flexibility in addressing custom DNA sequences.

The application of the FusionCRISPR technology, which is the provision and/or application of a CRISPR/Cas or CRISPR/Cas like system comprising a fuNA molecule, will simplify the creation of knockouts in a wide variety of microbial species formerly not amenable to high throughput targeting. By offering the donor nucleic acids promptly upon induction of a double strand break (DSB), the HR repair machinery will successfully ligate the break with accompanying template. The introduction of FusionCRISPR is performed with common techniques otherwise used for the introduction of nicking versions of CRISPR/Cas9, recombinases or any other (trans)genes of interest. These techniques include, but are not limited to electroporation/heat shock of plasmids, viral transduction and conjugation.

Whereas genes in organisms with efficient NHEJ can be knocked out by introducing CRISPR/Cas9 in which the guide RNA consists of (1) spacer matching the target gene (2) essential sequences for correct guide RNA folding (tracrRNA:crRNA commonly combined in one single guide RNA (sgRNA)) with a typical length of ~99 nucleotides, knockouts in DSB-sensitive microbes can be achieved through a simple adaptation of the RNA resulting in the following composition: (1) spacer (2) essential sequences for correct secondary structure (3) at least 15 nucleotides matching the target (4) at least 15 nucleotides matching the target one or more bases downstream from the first at least 15 nucleotides matching the target. Two complementary ssDNA oligonucleotides representing a knockout FusionCRISPR cassette can readily be purchased commercially from any oligo synthesis company. Alternatively, it can be synthesized as dsDNA. Cloning of FusionCRISPR knockout cassettes can proceed similar to the most common method for cloning of 20 nucleotides spacers for an intended target in regular CRISPR cassettes with type IIS restriction enzymes. Thus, access, flexibility and ease of use of controlled gene knockouts through FusionCRISPR is comparable to regular CRISPR/Cas9 and the current invention expands high throughput knockouts to target species in which NHEJ is inefficient or absent.

FusionCRISPR can also be used for knock-ins in any organism. A selectable marker can be knocked in using standard introduction techniques and standard expression vehicles, including simple plasmids. This can be done to interrupt and knockout an endogenous gene while providing easy selection.

Genome-scale knockouts in human and other cells allows the discovery of genes involved in diseases, drug response and normal biological processes and the creation of disease models [Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science ;343(6166):84-7]. Knockouts are typically generated by introduction of CRISPR/Cas9 for a specific target and relying on NHEJ to create a "null" for the gene in question. Three major problems arise with this method: (1) Knockouts are often not created because NHEJ repair does not result in a frame shift (or alternative downstream start codons are being used resulting in a truncated gene product), (2) The outcome, i.e. the exact primary DNA sequence, is not known and needs to be determined for each DNA modifying "event" which is costly and time-consuming, (3) in diploid or polyploid organisms the NHEJ repair on each chromosome will occur differently, if successful targeting on all available "substrates" is even achieved in the first place, resulting in complex molecular analysis and/or a forced switch to haploid model lines [Wade M (2015) High-throughput silencing using the CRISPR-Cas9 system: A review of the benefits and challenges. Journal of Biomolecular Screening, Vol. 20(8):1027-39]. FusionCRISPR allows control of the deletion as described above, which leads to a predictable outcome. In diploid and polyploid cell systems any introduced modification on each of the chromosomes will for the majority be identical. The deletion can be designed to be large enough so that alternative transcription possibilities are reduced and by designing the deletion to consist of a number of bases that cannot be divided by three, the risk of still creating a functional protein which merely lacks a short stretch of amino acids will be eliminated.

Methods for the introduction of FusionCRISPR configurations designed to control knockouts in any cell system are identical to methods used for the introduction of CRISPR/Cas9 which is followed by NHEJ repair for creation of knockouts. This includes but is not limited to the use of AAV and lentiviral vectors for human cells to baculovirus expression system for insect cells. No further adaptation to current methods is required other than a short extension of the guideRNA as described above.

Genome editing can be used to treat a variety of genetic disorders. Many genetic disorders involve point mutations which could potentially be corrected by providing a site specific nuclease along with a corrective nucleic acid template containing the required correction(s). One example is sickle cell disease which results from a single DNA base mutation (A>T) in the sixth codon of the β-globin gene [Li C, Ding L, Sun CW, Wu LC, Zhou D, Pawlik KM, Khodadadi-Jamayran A, Westin E, Goldman FD, Townes TM (2016) Novel HDAd/EBV Reprogramming Vector and Highly Efficient Ad/CRISPR-Cas Sickle Cell Disease Gene Correction. Sci Rep.

6:30422]. Correction of mutations in cells by providing nuclease and corrective template separately will often result in uncoordinated initial break in the DNA and local arrival of the corrective template for HDR repair. To compensate for the lack of temporal/spatial coordination nuclease and corrective template concentrations need to be relatively high. Higher nuclease concentrations can lead to higher off-target cleavage with negative consequences (higher patient risk and/or higher costs in molecular analysis). FusionCRISPR will achieve correct gene correction at much lower concentration. Methods for applying FusionCRISPR are identical to current standards in the field [Maeder ML, Gersbach CA (2016) Genome-editing Technologies for Gene and Cell Therapy. Mol Ther. 24(3):430-46] with the sole difference being a slightly longer sgRNA encompassing the desired correction.

FusionCRISPR can be used to alter substrate or product specificity of a variety of enzymes in a variety of different organisms. For example, a single amino acid has been found to be a major determinant in triterpene synthases substrate and product specificity in various plant species [Salmon M, Thimmappa RB, Minto RE, Melton RE, Hughes RK, O'Maille PE, Hemmings AM, Osbourn A (2016) A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. Proc Natl Acad Sci U S A. 113(30):E4407-14]. Triterpenes are a diverse group of natural products with applications in pharmacy and biotechnology and the ability to influence product specificity opens the door to synthesis of novel or higher quantity biomolecules in a plant of interest. FusionCRISPR allows the modification of critical amino acids in substrate pockets or other domains of an enzyme that influence specificity. The desired amino acid is introduced by including the corresponding codon in the fused template flanked by nucleotide sequences that have homology with the sequences flanking the codon that needs to be replaced in the genome. One example of introducing the FusionCRISPR construct is by Agrobacterium-mediated T-DNA transformation in which the T-DNA contains a selectable marker, Cas9 driven by a strong promoter and the FusionCRISPR construct consisting of respectively: (1) an RNA polymerase III promoter, (2) spacer matching the intended target codon or a sequence adjacent to it (find the nearest PAM), (3) essential sequences for correct secondary structure of the guide, (4) homology arm matching one region flanking the break, (5) the novel codon(s) matching the amino acid(s) that affect specificity in the enzyme, (6) the homology arm on the other side of the break and (7) a terminator. Depending on where the cut is made, a silent mutation may have to be included to avoid having a PAM sequence in the donor. Homology arms, intended changes and sequences between intended changes will be contiguous for seamless incorporation of the donor. The selectable marker is used for obtaining plant cells that have stably integrated T-DNA. In these cells, throughout maintenance and regeneration of transformed plants, the FusionCRISPR components have the opportunity to alter the genome at the intended target. These alterations can make their way into the germline and the next generation of seedlings can be screened for the novel endogenous sequence and segregation of the T-DNA.

FusionCRISPR allows the insertion of epitope or other tags at endogenous genes in any organism. The tags can be used for tracking of endogenous gene products at the cellular and

2017256517 31 Aug 2023

subcellular level, identification of protein-protein interactions, ChIP and other molecular interactions and protein purification. Most applications will be in pathway discovery and identification of drug targets. The FusionCRISPR construct will have the nucleic acids encoding for the tags as payload between the stretches of homology as similar described above for introduction of substitutions. Introduction of FusionCRISPR either as DNA or transiently will follow the same protocols as currently used for regular CRISPR/Cas9 in the organisms of interest.

It is to be understood that if any prior art publication is referred to herein, such reference does not constitute an admission that the publication forms a part of the common general knowledge in the art in Australia or any other country.

Summary

Our yeast experiments have shown that the payload in FusionCRISPR can be at least 731 nucleotides without affecting the full function of the nuclease. A sequence of this or potentially larger size can comprise a strong promoter or promoter with different tissue or environmental cue specific activity compared to the natural promoter. FusionCRISPR can potentially displace the endogenous promoter and allow an alternative method of gene regulation including upregulation.

The invention at hand relates to simplifying the application of the CRISPR/Cas DNA repair system or CRISPR/Cas like DNA repair system.

The invention at hand also relates to enhancing the efficiency of homologous recombination in a target nucleic acid during DNA break repair.

Surprisingly this was achieved by covalently linking a donor nucleic acid molecule and a guide nucleic acid molecule the latter interacting with a site directed nucleic acid modifying polypeptide as elements of a CRISPR/Cas DNA repair system or a CRISPR/Cas like DNA repair system.

A first aspect provides a method for modification of a target nucleic acid (target NA) molecule in a cell comprising the steps of

a. providing a recombinant fusion nucleic acid (fuNA) molecule comprising a guide nucleic acid (gNA) molecule covalently linked to at least one donor nucleic acid (doNA) molecule, and

b. introducing said fuNA molecule into one or more cells comprising the target NA molecule , and

c. introducing a site directed nucleic acid modifying polypeptide into said one or more cells, and

d. incubating the one or more cells under conditions that allow for homologous recombination in said one or more cells, and optionally

2017256517 31 Aug 2023

e. isolating one or more cells in which homologous recombination occurred

wherein the fuNA consists of RNA,

wherein the doNA comprises two homology arms each homology arm comprising at least 15 bases which are 100% complementary to the same number of consecutive bases in the target NA molecule, and

wherein the gNA molecule comprises a spacer nucleic acid (spacer NA) molecule and a scaffold nucleic acid (scaffold NA) molecule, the scaffold NA molecule forming a secondary structure comprising at least one hairpin and wherein the spacer NA comprises at least 16 bases complementary to the target NA molecule.

A second aspect provides a recombinant fuNA molecule comprising a doNA molecule covalently linked to a gNA molecule, wherein the fuNA is consisting of RNA and the doNA comprises two homology arms each homology arm comprising at least 15 bases which are 100% complementary to the same number of consecutive bases in the target NA molecule, and wherein the gNA molecule comprises a spacer NA molecule and a scaffold NA molecule, the scaffold NA molecule forming a secondary structure comprising at least one hairpin and wherein the spacer NA comprises at least 16 bases complementary to the target NA molecule.

A third aspect provides a vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of the second aspect.

A fourth aspect provides a vector system comprising

a. the vector of the third aspect and
b. a vector encoding a site directed nucleic acid modifying polypeptide and optionally
c. a vector encoding a scaffold NA molecule.

A fifth aspect provides a system for modification of a target NA molecule in a cell comprising

a. the vector of the third aspect and
b. a vector encoding a site directed nucleic acid modifying polypeptide and
c. a cell comprising a target NA molecule and optionally
d. a vector encoding a scaffold NA molecule.

A sixth aspect provides a composition comprising

a. the vector of the third aspect and
b. a vector encoding a site directed nucleic acid modifying polypeptide and
c. a cell comprising a target NA molecule and optionally
d. a vector encoding a scaffold NA molecule.

A seventh aspect provides use of the vector of the third aspect, the vector system of the fourth aspect, the system of fifth aspect or the composition of the sixth aspect for modification of a target NA molecule in a cell.

2017256517 31 Aug 2023

An eighth aspect provides use of the recombinant fuNA molecule of the second aspect, the vector of the third aspect, the vector system of the fourth aspect, the system of the fifth aspect and/or the composition of the sixth aspect in the manufacture of a medicament for treating a genetic disorder.

A ninth aspect provides a method of treating a genetic disorder in a subject, the method comprising administering the recombinant fuNA molecule of the second aspect, the vector of the third aspect, the vector system of the fourth aspect, the system of the fifth aspect and/or the composition of the sixth aspect.

Detailed Description of the Invention

One embodiment of the invention at hand is a method for modification of a target nucleic acid (target NA) molecule in a cell or composition comprising the steps of

a. providing a recombinant fusion nucleic acid (fuNA) molecule comprising a guide nucleic acid (gNA) molecule covalently linked to at least one donor nucleic acid (doNA) molecule, and
b. introducing said fuNA molecule into one or more cells or composition comprising the target NA molecule, and
c. introducing a site directed nucleic acid modifying polypeptide into said one or more cells or composition, and
d. incubating the one or more cells or composition under conditions that allow for homologous recombination in said one or more cells or composition, and optionally
e. isolating one or more cells in which homologous recombination occurred.

Various preferred structures of the fusion nucleic acid molecule according to the invention are depicted in figure 1 to 12. The most preferred structure is depicted in figure 1.

The target nucleic acid may be modified by introducing a nucleic acid molecule into the target nucleic acid wherein the introduced nucleic acid molecule is heterologous to the target nucleic acid. The sequence between homology arm 1 and homology arm 2 of the donor nucleic acid molecule would in this case comprise the nucleic acid molecule which is supposed to be introduced into the target nucleic acid and which is not present between the regions in the target nucleic acid that are complementary to the homology arms.

The target nucleic acid may also be modified by deleting at least one base from the target nucleic acid. In that case, the sequence between homology arm 1 and homology arm 2 of the donor nucleic acid molecule would comprise a sequence complementary to the target nucleic acid molecule that lacks at least one base compared to the target nucleic acid.

The target nucleic acid may further be modified by replacing at least one base in the target nucleic acid with one or more bases heterologous to the target nucleic acid. In that case the sequence between homology arm 1 and 2 of the donor nucleic acid would comprise at least one mismatch compared to the complementary region in the target nucleic acid.

The target nucleic acid may comprise a “protospacer adjacent motif” (PAM) sequence adjacent to the targeted sequence in the target nucleic acid molecule which is required for some site directed nucleic acid modifying polypeptides for correct target site identification and binding. The sequence of the PAM is specific for the various site directed nucleic acid modifying polypeptides (Doudna and Charpentier, 2014, Science 346 (6213):1258096) and is known to the skilled person.

The method of the invention is preferably applied for target nucleic acid modification in living cells but may also be applied in in vitro systems.

The target nucleic acid molecule may be RNA or DNA, it may be single- or double-stranded. Preferably, the target nucleic acid molecule is DNA, more preferable the target nucleic acid molecule is double-stranded DNA.

The site directed nucleic acid modifying polypeptide may be introduced into the cell or composition as a polypeptide or may be introduced by introduction of an RNA molecule encoding said site directed nucleic acid modifying polypeptide or by introduction of an expression construct expressing said site directed nucleic acid modifying polypeptide wherein the expression construct is comprising a promoter functional in the respective cell or composition functionally linked to a gene encoding said site directed nucleic acid modifying polypeptide. Examples for such site directed nucleic acid modifying polypeptides are shown in Table 1. Further, any functional equivalent to such site directed nucleic acid modifying polypeptide may be used in the method of the invention.

Table 1 Examples for site directed nucleic acid modifying polypeptides

GenBank Acc No.	**Bacterium**
303229466	*Veillonella atypica* ACS-134-V-Col7a

34762592	*Fusobacterium nucleatum* subsp. *vincentii*
374307738	*Filifactor alocis* ATCC 35896
320528778	*Solobacterium moorei* F0204
291520705	*Coprococcus catus* GD-7
42525843	*Treponema denticola* ATCC 35405
304438954	*Peptoniphilus duerdenii* ATCC BAA-1640
224543312	*Catenibacterium mitsuokai* DSM 15897
24379809	*Streptococcus mutans* UA159
15675041	*Streptococcus pyogenes* SF370
16801805	*Listeria innocua* Clip11262
116628213	*Streptococcus thermophilus* LMD-9
323463801	*Staphylococcus pseudintermedius* ED99
352684361	*Acidaminococcus intestini* RyC-MR95
302336020	*Olsenella uli* DSM 7084
366983953	*Oenococcus kitaharae* DSM 17330
310286728	*Bifidobacterium bifidum* S17
258509199	*Lactobacillus rhamnosus* GG
300361537	*Lactobacillus gasseri* JV-V03
169823755	*Finegoldia magna* ATCC 29328
47458868	*Mycoplasma mobile* 163K
284931710	*Mycoplasma gallisepticum* str. F
363542550	*Mycoplasma ovipneumoniae* SC01
384393286	*Mycoplasma canis* PG 14
71894592	*Mycoplasma synoviae* 53
238924075	*Eubacterium rectale* ATCC 33656
116627542	*Streptococcus thermophilus* LMD-9
315149830	*Enterococcus faecalis* TX0012
315659848	*Staphylococcus lugdunensis* M23590
160915782	*Eubacterium dolichum* DSM 3991
336393381	*Lactobacillus coryniformis* subsp. *torquens*
310780384	*Ilyobacter polytropus* DSM 2926
325677756	*Ruminococcus albus* 8
187736489	*Akkermansia muciniphila* ATCC BAA-835
117929158	*Acidothermus cellulolyticus* 11B
189440764	*Bifidobacterium longum* DJO10A
283456135	*Bifidobacterium dentium* Bd1
38232678	*Corynebacterium diphtheriae* NCTC 13129
187250660	*Elusimicrobium minutum* Pei191
319957206	*Nitratifractor salsuginis* DSM 16511
325972003	*Sphaerochaeta globus* str. Buddy
261414553	*Fibrobacter succinogenes* subsp. *succinogenes*
60683389	*Bacteroides fragilis* NCTC 9343
256819408	*Capnocytophaga ochracea* DSM 7271
90425961	*Rhodopseudomonas palustris* BisB18
373501184	*Prevotella micans* F0438
294674019	*Prevotella ruminicola* 23
365959402	*Flavobacterium columnare* ATCC 49512
312879015	*Aminomonas paucivorans* DSM 12260
83591793	*Rhodospirillum rubrum* ATCC 11170
294086111	*Candidatus Puniceispirillum marinum* IMCC1322
121608211	*Verminephrobacter eiseniae* EF01-2
344171927	*Ralstonia syzygii* R24
159042956	*Dinoroseobacter shibae* DFL 12
288957741	*Azospirillum* sp- B510

92109262	*Nitrobacter hamburgensis* X14
148255343	*Bradyrhizobium* sp- BTAi1
34557790	*Wolinella succinogenes* DSM 1740
218563121	*Campylobacter jejuni* subsp. *jejuni*
291276265	*Helicobacter mustelae* 12198
229113166	*Bacillus cereus* Rock1-15
222109285	*Acidovorax ebreus* TPSY
189485225	uncultured Termite group 1
182624245	*Clostridium perfringens* D str.
220930482	*Clostridium cellulolyticum* H10
154250555	*Parvibaculum lavamentivorans* DS-1
257413184	*Roseburia intestinalis* L1-82
218767588	*Neisseria meningitidis* Z2491
15602992	*Pasteurella multocida* subsp. *multocida*
319941583	*Sutterella wadsworthensis* 3 1
254447899	*gamma proteobacterium* HTCC5015
54296138	*Legionella pneumophila* str. Paris
331001027	*Parasutterella excrementihominis* YIT 11859
34557932	*Wolinella succinogenes* DSM 1740
118497352	*Francisella novicida* U112
961512549	*Francisella tularensis subsp. novicida* U112
961512548	*Acidaminococcus sp.* BV3L6

The site directed nucleic acid modifying polypeptide may have a double-stranded nucleic acid digestion function or it may have a nickase function, cutting only one strand of a double-stranded nucleic acid molecule. The nucleic acid restriction or nickase capability of the site directed nucleic acid modifying polypeptide may also be inactivated and the recombinant site directed nucleic acid modifying polypeptide may be linked to other functional groups such as the DNA restriction region of FokI or of a homing endonulcease. Such recombinant site directed nucleic acid modifying polypeptides are for example described in Tsai et al (2014; Nat Biotechnol. 2014 32(6):569-76.) or Guilinger et al (2014; Nat Biotechnol. 2014 32(6):577-82).

The gNA molecule comprises a spacer nucleic acid (spacer NA) molecule comprising at least 12 bases 100% complementary to the target NA molecule. Preferably it comprises at least 13 bases, at least 14 bases or at least 15 bases complementary to the target NA molecule. More preferably it comprises at least 16 bases, at least 17 bases or at least 18 bases complementary to the target NA molecule. Even more preferably it comprises at least 19 bases or at least 20 bases complementary to the target NA.

The gNA molecule further comprises a scaffold nucleic acid (scaffold NA) molecule.
The scaffold NA may consist of one nucleic acid molecule, which comprises two regions each comprising at least eight bases being complementary to each other, capable to hybridize and form a hairpin structure. The scaffold NA may consist of two nucleic acid molecules each comprising at least one region of at least eight bases complementary to each other, capable to hybridize and form a double-stranded structure. If said regions are comprising more than eight complementary bases, each region comprises at least eight bases that are complementary to at least eight bases of the other region.

Preferably the scaffold NA consists of one molecule.

The scaffold NA molecule is covalently linked to the spacer NA molecule. In the event, the scaffold NA molecule consists of two independent molecules, at least one of these molecules of the scaffold NA is covalently linked to the spacer NA molecule.
In addition to the two regions comprising at least eight bases being complementary to each other, the scaffold NA molecule comprises a further region forming a secondary structure comprising at least one hairpin, preferably at least two hairpins.

The donor NA molecule comprises two homology arms. Each homology arm of the donor NA molecule comprises at least 15 bases and is at least 5%, preferably at least 10%, more preferably at least 15%, most preferably at least 20% of the size of the additional NA region spacing the homology arms. Homology arm 1 and 2 may have the same length or different length.

The homology arms each comprise at least 15 bases that are 100% complementary to the same number of consecutive bases in target NA molecule. In the event, a homology arm is larger than 15 bases it is preferably at least 60%, preferably at least 70%, more preferably at least 75%, more preferably 80%, more preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, even more preferably at least 98%, even more preferably at least 99% complementary to the target NA molecule. Most preferably each homology arm is 100% complementary to the target NA molecule.

The method of the invention may be applied in cells or compositions comprising the target NA molecule. Preferably the method is applied to cells, wherein the cell is a microbial, animal, human or plant cell, more preferably the method is applied to yeast or plant cells.

The methods of the invention may be applied to any plant cell, for example gymnosperm or angiosperm, preferably angiosperm, for example dicotyledonous or monocotyledonous plant cells. Preferred monocotyledonous plant cells are for example corn, wheat, rice, barley, sorghum, musa, sugarcane, miscanthus and brachypodium, especially preferred monocotyledonous plant cells are corn, wheat and rice. Preferred dicotyledonous plant cells are for example soy, rape seed, canola, linseed, cotton, potato, sugar beet, tagetes and Arabidopsis, especially preferred dicotyledonous plant cells are soy, rape seed, canola and potato.

The method of the invention may also be applied to any microorganism. The microorganism may be a bacteria, the bacterial cell may be any gram-positive bacterium or a gram negative bacterium. Gram positive bacteria include, but not limited to, Bacillus, Brevibacterium, Corynebacterium, Streptococcus, Streptomyces, Staphylococcus, Enterococcus, Lactobacillus, Lactococcus, Clostridium, Geobacillus, and Oceanobacillus.

Gram negative bacteria include, but not limited to, E. coli, Pseudomonas, Salmonella, Campylobacter, Helicobacter, Acetobacter, Flavobacterium, Fusobacterium, Gluconobacter. Preferably, the gram negative cell is an E. coli cell.

In the methods of the present invention, the bacterial cell may be any Bacillus cell. Bacillus cells useful in the practice of the present invention include, but are not limited to, Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus brevis, Bacillus circulans, Bacillus clausii, Bacillus coagulans, Bacillus firmus, Bacillus lautus, Bacillus lentus, Bacillus licheniformis, Bacillus megaterium, Bacillus pumilus, Bacillus stearothermophilus, Bacillus subtilis, and Bacillus thuringiensis cells. In a preferred aspect, the bacterial cell is a Bacillus amyloliquefaciens, Bacillus lentus, Bacillus licheniformis, Bacillus stearothermophilus or Bacillus subtilis cell. In a more preferred aspect, the bacterial cell is a Bacillus licheniformis cell or a Bacillus subtilis cell, preferably a Bacillus subtilis cell.

In the methods of the present invention, the bacterial host cell may be *Lactobacillus acidophilus Lactobacillus plantarum, Lactobacillus gasseri, Lactobacillus bulgaricusk,Lactobacillus reuteri, Staphylococcus aureus, Corynebacterium,* particularly the species *Corynebacterium glutamicum, Corynebacterium acetoglutamicum, Corynebacterium acetoacidophilum, Corynebacterium callunae, Corynebacterium ammoniagenes, Corynebacterium thermoaminogenes, Corynebacterium melassecola and Corynebacterium effiziens, Corynebacterium efficiens, Corynebacterium deserti, Brevibacterium flavum, Brevibacterium lactofermentum, Brevibacterium divarecatum, Pseudomonas putida, Pseudomonas syringae, Streptoymyces,* particularly the species *Streptomyces coelicolor, Streptomyces lividans, Streptomyces albus, Streptomyces avermitilis, Gluconobacter oxydans, Gluconobacter morbifer, Gluconobacter thailandicus, Acetobacter aceti, Clostridium acetobutylicum, Clostridium saccharobutylicum, Clostridium beijerinckii,* Streptococcus equisimilis, Streptococcus pyogenes, Streptococcus uberis, and Streptococcus equi subsp. Zooepidemicus. Another preferred bacteria is Basfia succiniciproducens.

The microorganism may be a eukaryotic cell. Suitable eukaryotic cells include yeast cells, as for example Saccharomyces spec, such as Saccharomyces cerevisiae, Hansenula spec, such as Hansenula polymorpha, Schizosaccharomyces spec, such as Schizosaccharomyces pombe, Kluyveromyces spec, such as Kluyveromyces lactis and Kluyveromyces marxianus, Yarrowia spec, such as Yarrowia lipolytica, Pichia spec, such as Pichia methanolica, Pichia stipites and Pichia pastoris, Zygosaccharomyces spec, such as Zygosaccharomyces rouxii and Zygosaccharomyces bailii, Candida spec, such as Candida boidinii, Candida utilis, Candida freyschussii, Candida glabrata and Candida sonorensis, Schwanniomyces spec, such as Schwanniomyces occidentalis, Arxula spec, such as Arxula adeninivorans, Ogataea spec such as Ogataea minuta, Klebsiella spec, such as Klebsiella pneumonia.

The target nucleic acid molecule may be endogenous to the cell or it may be heterologous to the cell, such as, for example a transgene or a viral nucleic acid molecule.

The doNA molecule and the gNA molecule are covalently bound to each other forming a fusion nucleic acid (fusion NA) molecule. The donor NA molecule may be covalently bound to the spacer NA part of the gNA molecule or to the scaffold NA part of the gNA molecule. In a preferred embodiment, the donor NA is covalently linked to the scaffold NA part of the gNA molecule.

Most preferably the fusion NA molecule is one molecule, preferably one continuous RNA molecule, wherein all elements (gNA, scaffold NA and doNA) are covalently linked.

The doNA molecule and guide NA molecule may consist of RNA, DNA, PNA. Preferably they consist of RNA or DNA. More preferably the doNA molecule consists of DNA and the guide NA molecule consists of RNA.

In a most preferred embodiment, both guide and donor NA consist of RNA, wherein at least the doNA and gNA are covalently linked to each other forming a fusion ribonucleic acid molecule (fuRNA).

The fuRNA molecule may be introduced into the cell or composition comprising the target NA molecule as an RNA molecule or as one or more expression constructs encoding said fuRNA molecules.

In another embodiment the doNA molecule and gNA molecule may consist of DNA, wherein the doNA and gNA are covalently linked to each other forming a fusion deoxyribonucleic acid molecule (fuDNA).

The fuDNA molecule may be introduced into the cell or composition comprising the target NA molecule as a DNA molecule by various methods as for example transfection, biolistics, electroporation, photoporation, whiskers, sonication, nanobodies or microfluids. It may also be introduced using an agrobacterium as a vehicle which is capable of transferring a T-DNA molecule into a cell but which is not capable to mediate the integration of said T-DNA molecule into the genomic DNA of the target cell. The T-DNA would comprise or consist of the fuDNA molecule.

A further embodiment of the invention is a recombinant fuNA molecule, for example a fuRNA or fuDNA molecule, comprising a doNA molecule covalently linked to a gNA molecule.

Another embodiment of the invention is a vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of the invention.

A further embodiment of the invention is a cell comprising a fuNA molecule of the invention and a nucleic acid modifying polypeptide.

A further embodiment of the invention is a vector system comprising

a. a first vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of the invention and
b. a second vector encoding a site directed nucleic acid modifying polypeptide and optionally
c. a third vector encoding one part of the scaffold NA molecule.

In a preferred embodiment, the vector under a. comprises an expression construct encoding a fuNA molecule comprising a spacerNA, scaffold NA and doNA.

The vector under c. is necessary and part of the vector system of the invention if the scaffold NA is consisting of two molecules and if the vector under a. is encoding a fusion NA molecule comprising only one molecule of the scaffold NA molecule and is not encoding the second molecule of the scaffold NA molecule.

A system for modification of a target NA in a cell comprising

A. a first vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of the invention and
B. a second vector encoding a site directed nucleic acid modifying polypeptide and
C. a cell comprising a target NA molecule and optionally
D. a third vector encoding one part of the scaffold NA molecule.

is another embodiment of the invention. In a preferred embodiment, the vector under A. comprises an expression construct encoding a fuNA molecule comprising a spacerNA, scaffold NA and doNA. The vector under D. is necessary and part of the system of the invention if the scaffold NA is consisting of two molecules and if the vector under A. is encoding a fusion NA molecule comprising only one molecule of the scaffold NA molecule and is not encoding the second molecule of the scaffold NA molecule.

Another embodiment of the invention is a composition comprising

a. a first vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of the invention and
b. a second vector encoding a site directed nucleic acid modifying polypeptide and
c. a cell comprising a target NA molecule and optionally
d. a third vector encoding one part of the scaffold NA molecule.

In a preferred embodiment, the vector under *a.* comprises an expression construct encoding a fuNA molecule comprising a spacerNA, scaffold NA and doNA. The vector under *d.* is necessary and part of the composition of the invention if the scaffold NA is consisting of two molecules and if the vector under *a.* is encoding a fusion NA molecule comprising only one molecule of the scaffold NA molecule and is not encoding the second molecule of the scaffold NA molecule.

The use of the vector of the invention, the vector system of the invention, the system of the invention and/or the composition of the invention for modification of a target NA molecule in a cell or composition is also an embodiment of the invention.

DEFINITIONS

It is to be understood that this invention is not limited to the particular methodology or protocols. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims. It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a vector" is a reference to one or more vectors and includes equivalents thereof known to those skilled in the art, and so forth. The term "about" is used herein to mean approximately, roughly, around, or in the region of. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 20 percent, preferably 10 percent up or down (higher or lower). As used herein, the word "or" means any one member of a particular list and also includes any combination of members of that list. The words "comprise," "comprising," "include," "including," and "includes" when used in this specification and in the following claims are intended to specify the presence of one or more stated features, integers, components, or steps, but they do not preclude the presence or addition of one or more other features, integers, components, steps, or groups thereof. For clarity, certain terms used in the specification are defined and used as follows:

Donor NA: the term “donor NA” or “doNA” means a nucleic acid comprising two homology arms each comprising at least 15 bases complementary to two different areas of at least 15 consecutive bases of the target NA, wherein said two homology arms are directly adjacent to each other or are separated by one or more additional bases.

The two different areas of the target NA to which the homology arms are complementary may be directly adjacent to each other or may be separated by additional bases of up to 20 kb,

preferably up to 10 kb, preferably up to 5 kb, more preferably up to 3 kb, more preferably up to 2,5 kb, more preferably up to 2 kb.

In the event a homology arm comprises more than 15 bases, it may be 100% complementary to the target NA or it may be at least 75% complementary, preferably at least 80% complementary, more preferably at least 85% complementary, more preferably at least 90% complementary, more preferably at least 95% complementary, more preferably at least 98% complementary to the target NA, wherein the homology arm comprises at least one stretch of at least 15 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, preferably the homology arm comprises at least one stretch of at least 18 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, more preferably the homology arm comprises at least one stretch of at least 20 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, even more preferably the homology arm comprises at least one stretch of at least 25 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, even more preferably the homology arm comprises at least one stretch of at least 50 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA.

The homology arms may have the same length and/or the same degree of complementarity to the target NA or may have different length and/or different degrees of complementarity to the target NA.

The homology arms may be directly adjacent to each other or may be separated by a nucleic acid molecule comprising at least one base not present between the regions in the target nucleic acid complementary to the homology arms.

Spacer NA: the term “spacer nucleic acid” or “spacer NA” means a nucleic acid comprising at least 12 bases 100% complementary to the target NA.

In the event the spacer NA comprises more than 12 bases, it may be at least 75% complementary to the target NA, preferably at least 80% complementary, more preferably at least 85% complementary, more preferably at least 90% complementary, more preferably at least 95% complementary, more preferably at least 98% complementary most preferably it is 100% complementary to the target NA, wherein the spacer NA comprises at least one stretch of at least 12 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, preferably the spacer NA comprises at least one stretch of at least 15 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, preferably the spacer NA comprises at least one stretch of at least 18 bases

that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, more preferably the spacer NA comprises at least one stretch of at least 20 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, even more preferably the spacer NA comprises at least one stretch of at least 25 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA, even more preferably the spacer NA comprises at least one stretch of at least 50 bases that are 100% complementary to a stretch of the same number of consecutive bases in the target NA.

The spacer NA is covalently linked to a scaffold NA. If the scaffold NA is consisting of two nucleic acid molecules, the spacer is covalently linked to one molecule of a scaffold NA.

Scaffold NA: the scaffold nucleic acid or scaffold NA comprises a nucleic acid forming a secondary structure comprising at least one hairpin, preferably at least two hairpins and/or a sequence that is/are bound by the site directed nucleic acid modifying polypeptide. Such site directed nucleic acid modifying polypeptides are known in the art, for example in WO/2014/150624; WO/2014/204728. The scaffold NA further comprises two regions each comprising at least eight bases being complementary to each other, hence capable to hybridize forming a double-stranded structure. If said regions of at least eight bases complementary to each other are comprising more than eight bases, each region comprises at least eight bases that are complementary to at least eight bases of the other region.
The two complementary regions of the scaffold NA may be covalently linked to each other via a linker molecule forming a hairpin structure or may consist of two independent nucleic acid molecules.

Guide NA: the guide nucleic acid or guide NA or gNA comprises a spacer nucleic acid and a scaffold nucleic acid wherein the spacer NA and the scaffold NA are covalently linked to each other. In the event the scaffold NA consists of two molecules, the spacer NA is covalently linked to one molecule of the scaffold NA whereas the other molecule of the scaffold NA molecule hybridizes to the first scaffold NA molecule. Hence, a guide NA molecule may consist of one nucleic acid molecule or may consist of two nucleic acid molecules. Preferably the guide NA consists of one molecule.

Fusion NA: the fusion nucleic acid comprises donor NA and guide NA, wherein the guide NA and the donor NA are covalently linked to each other.

Site directed nucleic acid modifying polypeptide: By "site directed nucleic acid modifying polypeptide" "nucleic acid-binding site directed nucleic acid modifying polypeptide" or "site directed polypeptide" it is meant a polypeptide that binds nucleic acids and is targeted to a specific nucleic acid sequence. A site-directed nucleic acid modifying polypeptide as described herein is targeted to a specific nucleic acid sequence in the target nucleic acid either by mechanism intrinsic to the polypeptide or, preferably by the nucleic acid molecule to which it is bound. The nucleic acid molecule bound by the polypeptide comprises a sequence that is complementary to a target sequence within the target nucleic acid, thus targeting the bound polypeptide to a specific location within the target nucleic acid (the target sequence).

Most site directed nucleic acid modifying polypeptides introduce dsDNA breaks, but they may be modified to have only nicking activity or the nuclease activity may be inactivated. The site directed nucleic acid modifying polypeptides may be bound to a further polypeptide having an activity such as fluorescence or nuclease activity such as the nuclease activity of the FokI polypeptide or a homing endonuclease polypeptide such as I-SceI.

Coding region: As used herein the term "coding region" when used in reference to a structural gene refers to the nucleotide sequences which encode the amino acids found in the nascent polypeptide as a result of translation of a mRNA molecule. The coding region is bounded, in eukaryotes, on the 5'-side by the nucleotide triplet "ATG" which encodes the initiator methionine, prokaryotes also use the triplets “GTG” and “TTG” as start codon. On the 3'-side it is bounded by one of the three triplets which specify stop codons (i.e., TAA, TAG, TGA). In addition a gene may include sequences located on both the 5'- and 3'-end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5'-flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3'-flanking region may contain sequences which direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

Complementary: "Complementary" or "complementarity" refers to two nucleotide sequences which comprise antiparallel nucleotide sequences capable of pairing with one another (by the base-pairing rules) upon formation of hydrogen bonds between the complementary base residues in the antiparallel nucleotide sequences. For example, the sequence 5'-AGT-3' is complementary to the sequence 5'-ACT-3'. Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases are not matched according to the base pairing rules. "Total" or "complete" complementarity between nucleic acid molecules is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid molecule strands has significant

effects on the efficiency and strength of hybridization between nucleic acid molecule strands. A "complement" of a nucleic acid sequence as used herein refers to a nucleotide sequence whose nucleic acid molecules show total complementarity to the nucleic acid molecules of the nucleic acid sequence.

Endogenous: An "endogenous" nucleotide sequence refers to a nucleotide sequence, which is present in the genome of a wild type microorganism.

Enhanced expression: “enhance” or “increase” the expression of a nucleic acid molecule in a microorganism are used equivalently herein and mean that the level of expression of a nucleic acid molecule in a microorganism is higher compared to a reference microorganism, for example a wild type. The terms "enhanced” or “increased" as used herein mean herein higher, preferably significantly higher expression of the nucleic acid molecule to be expressed. As used herein, an “enhancement” or “increase” of the level of an agent such as a protein, mRNA or RNA means that the level is increased relative to a substantially identical microorganism grown under substantially identical conditions. As used herein, “enhancement” or “increase” of the level of an agent, such as for example a preRNA, mRNA, rRNA, tRNA, expressed by the target gene and/or of the protein product encoded by it, means that the level is increased 50% or more, for example 100% or more, preferably 200% or more, more preferably 5 fold or more, even more preferably 10 fold or more, most preferably 20 fold or more for example 50 fold relative to a suitable reference microorganism. The enhancement or increase can be determined by methods with which the skilled worker is familiar. Thus, the enhancement or increase of the nucleic acid or protein quantity can be determined for example by an immunological detection of the protein. Moreover, techniques such as protein assay, fluorescence, Northern hybridization, densitometric measurement of nucleic acid concentration in a gel, nuclease protection assay, reverse transcription (quantitative RT-PCR), ELISA (enzyme-linked immunosorbent assay), Western blotting, radioimmunoassay (RIA) or other immunoassays and fluorescence-activated cell analysis (FACS) can be employed to measure a specific protein or RNA in a microorganism. Depending on the type of the induced protein product, its activity or the effect on the phenotype of the microorganism may also be determined. Methods for determining the protein quantity are known to the skilled worker. Examples, which may be mentioned, are: the micro-Biuret method (Goa J (1953) Scand J Clin Lab Invest 5:218-222), the Folin-Ciocalteau method (Lowry OH et al. (1951) J Biol Chem 193:265-275) or measuring the absorption of CBB G-250 (Bradford MM (1976) Analyt Biochem 72:248-254).

Expression: "Expression" refers to the biosynthesis of a gene product, preferably to the transcription and/or translation of a nucleotide sequence, for example an endogenous gene or a heterologous gene, in a cell. For example, in the case of a structural gene, expression involves transcription of the structural gene into mRNA and - optionally - the subsequent translation of mRNA into one or more polypeptides. In other cases, expression may refer only to the transcription of the DNA harboring an RNA molecule.

Foreign: The term "foreign" refers to any nucleic acid molecule (e.g., gene sequence) which is introduced into a cell by experimental manipulations and may include sequences found in that cell as long as the introduced sequence contains some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) and is therefore different relative to the naturally-occurring sequence.

Functional fragment: the term “functional fragment” refers to any nucleic acid and/or protein which comprises merely a part of the full length nucleic acid and/or full length polypeptide of the invention but still provides the same function, i.e. the function of an AAT enzyme catalyzing the reaction of acryloyl-CoA and butanol to n-BA and CoA. Preferably, the fragment comprises at least 50%, at least 60%, at least 70%, at least 80 %, at least 90 % at least 95%, at least 98 %, at least 99% of the sequence from which it is derived. Preferably, the functional fragment comprises contiguous nucleic acids or amino acids of the nucleic acid and/or protein from which the functional fragment is derived. A functional fragment of a nucleic acid molecule encoding a protein means a fragment of the nucleic acid molecule encoding a functional fragment of the protein.

Functional linkage: The term "functional linkage" or "functionally linked" is equivalent to the term “operable linkage” or “operably linked” and is to be understood as meaning, for example, the sequential arrangement of a regulatory element (e.g. a promoter) with a nucleic acid sequence to be expressed and, if appropriate, further regulatory elements (such as e.g., a terminator) in such a way that each of the regulatory elements can fulfill its intended function to allow, modify, facilitate or otherwise influence expression of said nucleic acid sequence. As a synonym the wording “operable linkage” or “operably linked” may be used. The expression may result depending on the arrangement of the nucleic acid sequences in relation to sense or antisense RNA. To this end, direct linkage in the chemical sense is not necessarily required. Genetic control sequences such as, for example, enhancer sequences, can also exert their function on the target sequence from positions which are further away, or indeed from other DNA molecules. Preferred arrangements are those in which the nucleic acid sequence to be expressed recombinantly is positioned behind the sequence acting as promoter, so that the two sequences are linked covalently to each other. In a preferred embodiment, the nucleic acid sequence to be transcribed is located behind the promoter in such a way that the transcription start is identical with the desired beginning of the chimeric RNA of the invention. Functional linkage, and an expression construct, can be generated by means of customary recombination and cloning techniques as described (e.g., Sambrook J, Fritsch EF and Maniatis T (1989); Silhavy et al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor (NY); Ausubel et al. (1987) Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley Interscience; Gelvin et al. (Eds) (1990) Plant Molecular Biology Manual; Kluwer Academic Publisher, Dordrecht, The Netherlands). However, further sequences, which, for example, act as a linker with specific cleavage sites for restriction enzymes, or as a signal peptide, may also be positioned between the two sequences.

The insertion of sequences may also lead to the expression of fusion proteins. Preferably, the expression construct, consisting of a linkage of a regulatory region for example a promoter and nucleic acid sequence to be expressed, can exist in a vector-integrated form or can be inserted into the genome, for example by transformation.

Gene: The term "gene" refers to a region operably linked to appropriate regulatory sequences capable of regulating the expression of the gene product (e.g., a polypeptide or a functional RNA) in some manner. A gene includes untranslated regulatory regions of DNA (e.g., promoters, enhancers, repressors, etc.) preceding (up-stream) and following (downstream) the coding region (open reading frame, ORF). The term "structural gene" as used herein is intended to mean a DNA sequence that is transcribed into mRNA which is then translated into a sequence of amino acids characteristic of a specific polypeptide.

Genome and genomic DNA: The terms “genome” or “genomic DNA” is referring to the heritable genetic information of a host organism. Said genomic DNA comprises the DNA of the nucleoid but also the DNA of the self-replicating plasmid.

Heterologous: The term "heterologous” with respect to a nucleic acid molecule or DNA refers to a nucleic acid molecule which is operably linked to, or is manipulated to become operably linked to, a second nucleic acid molecule to which it is not operably linked in nature, or to which it is operably linked at a different location in nature. A heterologous expression construct comprising a nucleic acid molecule and one or more regulatory nucleic acid molecule (such as a promoter or a transcription termination signal) linked thereto for example is a constructs originating by experimental manipulations in which either a) said nucleic acid molecule, or b) said regulatory nucleic acid molecule or c) both (i.e. (a) and (b)) is not located in its natural (native) genetic environment or has been modified by experimental manipulations, an example of a modification being a substitution, addition, deletion, inversion or insertion of one or more nucleotide residues. Natural genetic environment refers to the natural genomic locus in the organism of origin, or to the presence in a genomic library. In the case of a genomic library, the natural genetic environment of the sequence of the nucleic acid molecule is preferably retained, at least in part. The environment flanks the nucleic acid sequence at least at one side and has a sequence of at least 50 bp, preferably at least 500 bp, especially preferably at least 1,000 bp, very especially preferably at least 5,000 bp, in length. A naturally occurring expression construct - for example the naturally occurring combination of a promoter with the corresponding gene - becomes a transgenic expression construct when it is modified by non-natural, synthetic “artificial” methods such as, for example, mutagenization. Such methods have been described (US 5,565,350; WO 00/15815). For example a protein encoding nucleic acid molecule operably linked to a promoter, which is not the native promoter of this molecule, is considered to be heterologous with respect to the promoter. Preferably, heterologous DNA is not endogenous to or not naturally associated with the cell into which it is introduced, but has been obtained from another cell or has been synthesized.

Heterologous DNA also includes an endogenous DNA sequence, which contains some modification, non-naturally occurring, multiple copies of an endogenous DNA sequence, or a DNA sequence which is not naturally associated with another DNA sequence physically linked thereto. Generally, although not necessarily, heterologous DNA encodes RNA or proteins that are not normally produced by the cell into which it is expressed.

Hybridization: The term "hybridization" as used herein includes "any process by which a strand of nucleic acid molecule joins with a complementary strand through base pairing." (J. Coombs (1994) Dictionary of Biotechnology, Stockton Press, New York). Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acid molecules) is impacted by such factors as the degree of complementarity between the nucleic acid molecules, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acid molecules. As used herein, the term "Tm" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the Tm of nucleic acid molecules is well known in the art. As indicated by standard references, a simple estimate of the Tm value may be calculated by the equation: Tm=81.5+0.41(% G+C), when a nucleic acid molecule is in aqueous solution at 1 M NaCl [see e.g., Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization (1985)]. Other references include more sophisticated computations, which take structural as well as sequence characteristics into account for the calculation of Tm. Stringent conditions, are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.
Suitable hybridization conditions are for example hybridizing under conditions equivalent to hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 2 X SSC, 0.1% SDS at 50°C (low stringency) to a nucleic acid molecule comprising at least 50, preferably at least 100, more preferably at least 150, even more preferably at least 200, most preferably at least 250 consecutive nucleotides of the complement of a sequence. Other suitable hybridizing conditions are hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 1 X SSC, 0.1% SDS at 50°C (medium stringency) or 65°C (high stringency) to a nucleic acid molecule comprising at least 50, preferably at least 100, more preferably at least 150, even more preferably at least 200, most preferably at least 250 consecutive nucleotides of a complement of a sequence. Other suitable hybridization conditions are hybridization in 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4, 1 mM EDTA at 50°C with washing in 0,1X SSC, 0.1% SDS at 65°C (very high stringency) to a nucleic acid molecule comprising at least 50, preferably at least 100, more preferably at least 150, even more preferably at least 200, most preferably at least 250 consecutive nucleotides of a complement of a sequence.

"Identity": "Identity" when used in respect to the comparison of two or more nucleic acid or amino acid molecules means that the sequences of said molecules share a certain degree of sequence similarity, the sequences being partially identical.

For the determination of the percentage identity of two or more amino acids or of two or more nucleotide sequences several computer software programs have been developed. The identity of two or more sequences can be calculated with for example the software fasta, which presently has been used in the version fasta 3 (W. R. Pearson and D. J. Lipman, PNAS 85, 2444(1988); W. R. Pearson, Methods in Enzymology 183, 63 (1990); W. R. Pearson and D. J. Lipman, PNAS 85, 2444 (1988); W. R. Pearson, Enzymology 183, 63 (1990)). Another useful program for the calculation of identities of different sequences is the standard blast program, which is included in the Biomax pedant software (Biomax, Munich, Federal Republic of Germany). This leads unfortunately sometimes to suboptimal results since blast does not always include complete sequences of the subject and the query. Nevertheless as this program is very efficient it can be used for the comparison of a huge number of sequences. The following settings are typically used for such a comparisons of sequences:

-p Program Name [String]; -d Database [String]; default = nr; -i Query File [File In]; default = stdin; -e Expectation value (E) [Real]; default = 10.0; -m alignment view options: 0 = pairwise; 1 = query-anchored showing identities; 2 = query-anchored no identities; 3 = flat query-anchored, show identities; 4 = flat query-anchored, no identities; 5 = query-anchored no identities and blunt ends; 6 = flat query-anchored, no identities and blunt ends; 7 = XML Blast output; 8 = tabular; 9 tabular with comment lines [Integer]; default = 0; -o BLAST report Output File [File Out] Optional; default = stdout; -F Filter query sequence (DUST with blastn, SEG with others) [String]; default = T; -G Cost to open a gap (zero invokes default behavior) [Integer]; default = 0; -E Cost to extend a gap (zero invokes default behavior) [Integer]; default = 0; -X X dropoff value for gapped alignment (in bits) (zero invokes default behavior); blastn 30, megablast 20, tblastx 0, all others 15 [Integer]; default = 0; -I Show GI's in deflines [T/F]; default = F; -q Penalty for a nucleotide mismatch (blastn only) [Integer]; default = -3; -r Reward for a nucleotide match (blastn only) [Integer]; default = 1; -v Number of database sequences to show one-line descriptions for (V) [Integer]; default = 500; -b Number of database sequence to show alignments for (B) [Integer]; default = 250; -f Threshold for extending hits, default if zero; blastp 11, blastn 0, blastx 12, tblastn 13; tblastx 13, megablast 0 [Integer]; default = 0; -g Perfom gapped alignment (not available with tblastx) [T/F]; default = T; -Q Query Genetic code to use [Integer]; default = 1; -D DB Genetic code (for tblast[nx] only) [Integer]; default = 1; -a Number of processors to use [Integer]; default = 1; -O SeqAlign file [File Out] Optional; -J Believe the query defline [T/F]; default = F; -M Matrix [String]; default = BLOSUM62; -W Word size, default if zero (blastn 11, megablast 28, all others 3) [Integer]; default = 0; -z Effective length of the database (use zero for the real size) [Real]; default = 0; -K Number of best hits from a region to keep (off by default, if used a value of 100 is recommended) [Integer]; default = 0; -P 0 for multiple hit, 1 for single hit [Integer]; default = 0; -Y Effective length of the search space (use zero for the real size) [Real]; default = 0; -S Query strands to search against database (for blast[nx], and tblastx); 3 is both, 1 is top, 2 is bottom [Integer]; default = 3; -T Produce HTML output [T/F]; default = F; -l Restrict search of database to list of GI's [String] Optional; -U Use lower case filtering of FASTA sequence [T/F]

Optional; default = F; -y X dropoff value for ungapped extensions in bits (0.0 invokes default behavior); blastn 20, megablast 10, all others 7 [Real]; default = 0.0; -Z X dropoff value for final gapped alignment in bits (0.0 invokes default behavior); blastn/megablast 50, tblastx 0, all others 25 [Integer]; default = 0; -R PSI-TBLASTN checkpoint file [File In] Optional; -n MegaBlast search [T/F]; default = F; -L Location on query sequence [String] Optional; -A Multiple Hits window size, default if zero (blastn/megablast 0, all others 40 [Integer]; default = 0; -w Frame shift penalty (OOF algorithm for blastx) [Integer]; default = 0; -t Length of the largest intron allowed in tblastn for linking HSPs (0 disables linking) [Integer]; default = 0.

Results of high quality are reached by using the algorithm of Needleman and Wunsch or Smith and Waterman. Therefore programs based on said algorithms are preferred. Advantageously the comparisons of sequences can be done with the program PileUp (J. Mol. Evolution., 25, 351 (1987), Higgins et al., CABIOS 5, 151 (1989)) or preferably with the programs “Gap” and “Needle”, which are both based on the algorithms of Needleman and Wunsch (J. Mol. Biol. 48; 443 (1970)), and “BestFit”, which is based on the algorithm of Smith and Waterman (Adv. Appl. Math. 2; 482 (1981)). “Gap” and “BestFit” are part of the GCG software-package (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991); Altschul et al., (Nucleic Acids Res. 25, 3389 (1997)), “Needle” is part of the The European Molecular Biology Open Software Suite (EMBOSS) (Trends in Genetics 16 (6), 276 (2000)). Therefore preferably the calculations to determine the percentages of sequence identity are done with the programs “Gap” or “Needle” over the whole range of the sequences. The following standard adjustments for the comparison of nucleic acid sequences were used for “Needle”: matrix: EDNAFULL, Gap_penalty: 10.0, Extend_penalty: 0.5. The following standard adjustments for the comparison of nucleic acid sequences were used for “Gap”: gap weight: 50, length weight: 3, average match: 10.000, average mismatch: 0.000.

For example a sequence, which is said to have 80% identity with sequence SEQ ID NO: 1 at the nucleic acid level is understood as meaning a sequence which, upon comparison with the sequence represented by SEQ ID NO: 1 by the above program “Needle” with the above parameter set, has a 80% identity. Preferably the identity is calculated on the complete length of the query sequence, for example SEQ ID NO: 1.

Isolated: The term "isolated" as used herein means that a material has been removed by the hand of man and exists apart from its original, native environment and is therefore not a product of nature. An isolated material or molecule (such as a DNA molecule or enzyme) may exist in a purified form or may exist in a non-native environment such as, for example, in a transgenic host cell. For example, a naturally occurring nucleic acid molecule or polypeptide present in a living cell is not isolated, but the same nucleic acid molecule or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such nucleic acid molecules can be part of a vector and/or such nucleic acid molecules or polypeptides could be part of a composition, and would be isolated in that such a vector or composition is not part of its original environment. Preferably, the term "isolated" when used in

relation to a nucleic acid molecule, as in "an isolated nucleic acid sequence" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in its natural source. Isolated nucleic acid molecule is nucleic acid molecule present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acid molecules are nucleic acid molecules such as DNA and RNA, which are found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs, which encode a multitude of proteins. However, an isolated nucleic acid sequence comprising for example SEQ ID NO: 1 includes, by way of example, such nucleic acid sequences in cells which ordinarily contain SEQ ID NO: 1 where the nucleic acid sequence is in a genomic or plasmid location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid sequence may be present in single- or double-stranded form. When an isolated nucleic acid sequence is to be utilized to express a protein, the nucleic acid sequence will contain at a minimum at least a portion of the sense or coding strand (i.e., the nucleic acid sequence may be single-stranded). Alternatively, it may contain both the sense and anti-sense strands (i.e., the nucleic acid sequence may be double-stranded).

Non-coding: The term "non-coding" refers to sequences of nucleic acid molecules that do not encode part or all of an expressed protein. Non-coding sequences include but are not limited enhancers, promoter regions, 3' untranslated regions, and 5' untranslated regions.

Nucleic acids and nucleotides: The terms "nucleic acids" and "Nucleotides" refer to naturally occurring or synthetic or artificial nucleic acid or nucleotides. The terms “nucleic acids” and "nucleotides” comprise deoxyribonucleotides or ribonucleotides or any nucleotide analogue and polymers or hybrids thereof in either single- or double-stranded, sense or antisense form. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. The term "nucleic acid" is used inter-changeably herein with "gene", "cDNA, "mRNA", "oligonucleotide," and "nucleic acid molecule". Nucleotide analogues include nucleotides having modifications in the chemical structure of the base, sugar and/or phosphate, including, but not limited to, 5-position pyrimidine modifications, 8-position purine modifications, modifications at cytosine exocyclic amines, substitution of 5-bromo-uracil, and the like; and 2'-position sugar modifications, including but not limited to, sugar-modified ribonucleotides in which the 2'-OH is replaced by a group selected from H, OR, R, halo, SH, SR, NH2, NHR, NR2, or CN. Short hairpin RNAs (shRNAs) also can comprise non-natural elements such as non-natural bases, e.g., ionosin and xanthine, non-natural sugars, e.g., 2'-methoxy ribose, or non-natural phosphodiester linkages, e.g., methylphosphonates, phosphorothioates and peptides.

Nucleic acid sequence: The phrase "nucleic acid sequence" refers to a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5'- to the 3'-end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role. "Nucleic acid sequence" also refers to a consecutive list of abbreviations, letters, characters or words, which represent nucleotides. In one embodiment, a nucleic acid can be a "probe" which is a relatively short nucleic acid, usually less than 100 nucleotides in length. Often a nucleic acid probe is from about 50 nucleotides in length to about 10 nucleotides in length. A "target region" of a nucleic acid is a portion of a nucleic acid that is identified to be of interest. A "coding region" of a nucleic acid is the portion of the nucleic acid, which is transcribed and translated in a sequence-specific manner to produce into a particular polypeptide or protein when placed under the control of appropriate regulatory sequences. The coding region is said to encode such a polypeptide or protein.

Oligonucleotide: The term "oligonucleotide" refers to an oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics thereof, as well as oligonucleotides having non-naturally-occurring portions which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases. An oligonucleotide preferably includes two or more nucleomonomers covalently coupled to each other by linkages (e.g., phosphodiesters) or substitute linkages.

Overhang: An "overhang" is a relatively short single-stranded nucleotide sequence on the 5'- or 3'-hydroxyl end of a double-stranded oligonucleotide molecule (also referred to as an "extension," "protruding end," or "sticky end").

Polypeptide: The terms "polypeptide", "peptide", "oligopeptide", "polypeptide", "gene product", "expression product" and "protein" are used interchangeably herein to refer to a polymer or oligomer of consecutive amino acid residues.

Promoter: The terms "promoter", or "promoter sequence" are equivalents and as used herein, refer to a DNA sequence which when operably linked to a nucleotide sequence of interest is capable of controlling the transcription of the nucleotide sequence of interest into RNA. A promoter is located 5' (i.e., upstream), proximal to the transcriptional start site of a nucleotide sequence of interest whose transcription into mRNA it controls, and provides a site for specific binding by RNA polymerase and other transcription factors for initiation of transcription. The promoter does not comprise coding regions or 5´ untranslated regions. The promoter may for example be heterologous or homologous to the respective cell. A nucleic acid molecule sequence is "heterologous to" an organism or a second nucleic acid molecule sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to

a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is not naturally associated with the promoter (e.g. a genetically engineered coding sequence or an allele from a different ecotype or variety). Suitable promoters can be derived from genes of the host cells where expression should occur or from pathogens for this host.

Purified: As used herein, the term "purified" refers to molecules, either nucleic or amino acid sequences that are removed from their natural environment, isolated or separated. "Substantially purified" molecules are at least 60% free, preferably at least 75% free, and more preferably at least 90% free from other components with which they are naturally associated. A purified nucleic acid sequence may be an isolated nucleic acid sequence.

Significant increase: An increase for example in enzymatic activity, gene expression, productivity or yield of a certain product, that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 10% or 25% preferably by 50% or 75%, more preferably 2-fold or-5 fold or greater of the activity, expression, productivity or yield of the control enzyme or expression in the control cell, productivity or yield of the control cell, even more preferably an increase by about 10-fold or greater.

Significant decrease: A decrease for example in enzymatic activity, gene expression, productivity or yield of a certain product, that is larger than the margin of error inherent in the measurement technique, preferably a decrease by at least about 5% or 10%, preferably by at least about 20% or 25%, more preferably by at least about 50% or 75%, even more preferably by at least about 80% or 85%, most preferably by at least about 90%, 95%, 97%, 98% or 99%.

Substantially complementary: In its broadest sense, the term "substantially complementary", when used herein with respect to a nucleotide sequence in relation to a reference or target nucleotide sequence, means a nucleotide sequence having a percentage of identity between the substantially complementary nucleotide sequence and the exact complementary sequence of said reference or target nucleotide sequence of at least 60%, more desirably at least 70%, more desirably at least 80% or 85%, preferably at least 90%, more preferably at least 93%, still more preferably at least 95% or 96%, yet still more preferably at least 97% or 98%, yet still more preferably at least 99% or most preferably 100% (the later being equivalent to the term “identical” in this context). Preferably identity is assessed over a length of at least 19 nucleotides, preferably at least 50 nucleotides, more preferably the entire length of the nucleic acid sequence to said reference sequence (if not specified otherwise below). Sequence comparisons are carried out using default GAP analysis with the University of Wisconsin GCG, SEQWEB application of GAP, based on the algorithm of Needleman and Wunsch (Needleman and Wunsch (1970) J Mol. Biol. 48: 443-453; as defined above). A nucleotide sequence "substantially complementary " to a reference nucleotide sequence hybridizes to the reference nucleotide sequence under low stringency conditions, preferably medium stringency conditions, most preferably high stringency conditions (as defined above).

Transgene: The term "transgene" as used herein refers to any nucleic acid sequence, which is introduced into the genome of a cell by experimental manipulations. A transgene may be an "endogenous DNA sequence," or a "heterologous DNA sequence" (i.e., "foreign DNA"). The term "endogenous DNA sequence" refers to a nucleotide sequence, which is naturally found in the cell into which it is introduced so long as it does not contain some modification (e.g., a point mutation, the presence of a selectable marker gene, etc.) relative to the naturally-occurring sequence.

Transgenic: The term transgenic when referring to an organism means transformed, preferably stably transformed, with at least one recombinant nucleic acid molecule.

Vector: As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a genomic integrated vector, or "integrated vector", which can become integrated into the genomic DNA of the host cell. Another type of vector is an episomal vector, i.e., a plasmid or a nucleic acid molecule capable of extra-chromosomal replication. Vectors capable of directing the expression of genes to which they are operatively linked are referred to herein as "expression vectors". In the present specification, "plasmid" and "vector" are used interchangeably unless otherwise clear from the context.

Wild type: The term "wild type", "natural" or "natural origin" means with respect to an organism that said organism is not changed, mutated, or otherwise manipulated by man. With respect to a polypeptide or nucleic acid sequence, that the polypeptide or nucleic acid sequence is naturally occurring or available in at least one naturally occurring organism which is not changed, mutated, or otherwise manipulated by man.
A wild type of a microorganism refers to a microorganism whose genome is present in a state as before the introduction of a genetic modification of a certain gene. The genetic modification may be e.g. a deletion of a gene or a part thereof or a point mutation or the introduction of a gene.

The terms "production" or "productivity" are art-recognized and include the concentration of the fermentation product (for example, dsRNA) formed within a given time and a given fermentation volume (e.g., kg product per hour per liter). The term "efficiency of production" includes the time required for a particular level of production to be achieved (for example, how long it takes for the cell to attain a particular rate of output of a fine chemical).

The term "yield" or "product/carbon yield" is art-recognized and includes the efficiency of the conversion of the carbon source into the product (i.e., fine chemical). This is generally written as, for example, kg product per kg carbon source. By increasing the yield or production of the compound, the quantity of recovered molecules or of useful recovered molecules of that compound in a given amount of culture over a given amount of time is increased.

The term "*recombinant microorganism*" includes microorganisms which have been genetically modified such that they exhibit an altered or different genotype and/or phenotype (e. g., when the genetic modification affects coding nucleic acid sequences of the microorganism) as compared to the wild type microorganism from which it was derived. A recombinant microorganism comprises at least one recombinant nucleic acid molecule.

The term "*recombinant*" with respect to nucleic acid molecules refers to nucleic acid molecules produced by man using recombinant nucleic acid techniques. The term comprises nucleic acid molecules which as such do not exist in nature or do not exist in the organism from which the nucleic acid molecule is derived, but are modified, changed, mutated or otherwise manipulated by man. Preferably, a "recombinant nucleic acid molecule" is a non-naturally occurring nucleic acid molecule that differs in sequence from a naturally occurring nucleic acid molecule by at least one nucleic acid. A "recombinant nucleic acid molecules" may also comprise a "recombinant construct" which comprises, preferably operably linked, a sequence of nucleic acid molecules not naturally occurring in that order. Preferred methods for producing said recombinant nucleic acid molecules may comprise cloning techniques, directed or non-directed mutagenesis, gene synthesis or recombination techniques.

An example of such a recombinant nucleic acid molecule is a plasmid into which a heterologous DNA-sequence has been inserted or a gene or promoter which has been mutated compared to the gene or promoter from which the recombinant nucleic acid molecule derived. The mutation may be introduced by means of directed mutagenesis technologies known in the art or by random mutagenesis technologies such as chemical, UV light or x-ray mutagenesis or directed evolution technologies.

The term "directed evolution" is used synonymously with the term "metabolic evolution" herein and involves applying a selection pressure that favors the growth of mutants with the traits of interest. The selection pressure can be based on different culture conditions, ATP and growth coupled selection and redox related selection. The selection pressure can be carried out with batch fermentation with serial transferring inoculation or continuous culture with the same pressure.

The term "expression" or "gene expression" means the transcription of a specific gene(s) or specific genetic vector construct. The term "expression" or "gene expression" in particular means the transcription of gene(s) or genetic vector construct into mRNA. The process includes transcription of DNA and may include processing of the resulting RNA-product. The term "expression" or "gene expression" may also include the translation of the mRNA and therewith the synthesis of the encoded protein, i.e. protein expression.

Figures

Figures 1 to 12 depict preferred structures of the fusion nucleic acid molecules of the invention. A site directed nucleic acid modifying polypeptide is directed to the target sequence within the target double-stranded nucleic acid by a guide NA fused to a donor NA (which together form the fuNA molecule).

Figure 1

Fusion NA molecule comprising from 5' to 3': the guide NA (spacer followed by scaffold), homology arm 1 and 2 optionally separated by an additional nucleic acid region.

Figure 2

Fusion NA molecule comprising from 5' to 3': homology arm 2 and 1 optionally separated by an additional nucleic acid region, and the guide NA (scaffold followed by spacer)

Figure 3

Fusion NA molecule comprising from 5' to 3': the guide NA (spacer followed by scaffold), homology arm 2 and 1 optionally separated by an additional nucleic acid region.

Figure 4

Fusion NA molecule comprising from 5' to 3': homology arm 1 and 2 optionally separated by an additional nucleic acid region, and the guide NA (scaffold followed by spacer).

Figure 5

Fusion NA molecule comprising from 5' to 3': homology arm 1 and 2 optionally separated by an additional nucleic acid region, and the guide NA (spacer followed by scaffold).

Figure 6

Fusion NA molecule comprising from 5' to 3': the guide NA (scaffold followed by spacer), homology arm 1 and 2 optionally separated by an additional nucleic acid region.

Figure 7

Fusion NA molecule comprising from 5' to 3': homology arm 2 and 1 optionally separated by an additional nucleic acid region, and the guide NA (spacer followed by scaffold).

Figure 8

Fusion NA molecule comprising from 5' to 3': the guide NA (scaffold followed by spacer), homology arm 2 and 1 optionally separated by an additional nucleic acid region.

Figure 9

Fusion NA molecules comprising from 5' to 3': guide NA (comprising spacer and first molecule of the scaffold), homology arm 1 and 2 optionally separated by an additional nucleic acid region. The second molecule of the scaffold is hybridizing to the first molecule of the scaffold.

Figure 10

Fusion NA molecules comprising from 5' to 3': homology arm 1 and 2 optionally separated by an additional nucleic acid region, guide NA (comprising first molecule of the scaffold and spacer). The second molecule of the scaffold is hybridizing to the first molecule of the scaffold.

Figure 11

Fusion NA molecules comprising from 5' to 3': guide NA (comprising spacer and first molecule of the scaffold), homology arm 2 and 1 optionally separated by an additional nucleic acid region. The second molecule of the scaffold is hybridizing to the first molecule of the scaffold.

Figure 12

Fusion NA molecules comprising from 5' to 3': homology arm 2 and 1 optionally separated by an additional nucleic acid region, guide NA (comprising first molecule of the scaffold, spacer, and second molecule of the scaffold hybridizing to the first molecule of the scaffold).

Figure 13

Vector RWL121.

Figure 14

Vector Cas003.

Figure 15

Vector Cas018.

Figure 16

Vector Cas006.

Figure 17

Vector RWL137.

Figure 18

Vector Cas019.

Figure 19

Vector RLW138.

Figure 20

Vector Cas020.

Figure 21

Vector RLW139.

Figure 22

shows the amylase amyE locus of B. subtilis ATCC6051 strain (A) with the location of the homology regions HomA and HomB as indicated. The location of the protospacer sequence PS within in the amyE gene is indicated (A) and the sequence of the PS highlighted (black with white letters, B).

Figure 23

shows the vector map of the pCC004 plasmid – the derivative of the pJOE8999 plasmid carrying the amyE protospacer and the homology regions HomA and HomB of the region adjacent to the amyE gene. PS = protospacer; PvanP* = hemisynthetic promoter; 'gRNA = guideRNA; lambda T0 terminator; PmanP = promoter of the manP gene B. subtilis; Cas9 = endonuclease from S. pyrogenes; KanR = kanamycin resistance gene; origin of pUC for replication in E. coli, origin of pE194 for replication in Bacillus.

Figure 24

shows the schematic drawing of the EcoRI/XbaI fragment of the various plasmids used in this study (as exemplified in Figure 2 with plasmid pCC004). The Cas9 endonuclease, the PmanP promoter and the vector backbone with pUC replication origin, pE194 replication origin, kanamycin resistance gene are not shown. The promoter (Pro) driving the transcription of downstream genetic elements is indicated. PS = Protospacer; gRNA = guide RNA consisting of crRNA-loop-tracrRNA, T= lambda T0 terminator; homology region A and homology region B are depicted as arrows indicative of the orientation of the amyE gene. Detailed description of the plasmid genetic elements by J. Altenbuchner (Altenbuchner J. 2016. Editing of the Bacillus subtilis genome by the CRISPR-Cas9 system. Appl Environ Microbiol 82:5421–5).

Figure 25

The gene knockout efficiency as exemplified for the amylase gene for each gene deletion construct (pJOE8999, pCC005-pCC008) relative to pCC004 is plotted against the deletion constructs as indicated.

Figure 26

shows 0.8% agarose gels of PCR reactions with oligonucleotides Seq ID NO: 60 and 61 on genomic DNA of 13 individual clones from gene deletion reactions with indicated plasmids pCC004, pCC005, pCC006, pCC007, pCC008. The amplification of a DNA fragment of 1.4kb indicates gene knockout by recombination whereas a DNA fragment of 3.4 kb indicates amylase gene inactivation by rather a SOS repair mechanism. The 3.4kb band for WT indicates wildtype amylase locus of B-.subtilis WT ATCC6051. C denotes water control with no genomic DNA added. M indicates DNA ladder ‘Perfect plus 1 kb DNA ladder’ (roboklon) with the size of three bands indicated (1.0 kb, 1.5 kb. 4.0 kb).

EXAMPLES

Chemicals and common methods

Unless indicated otherwise, cloning procedures carried out for the purposes of the present invention including restriction digest, agarose gel electrophoresis, purification of nucleic acids, ligation of nucleic acids, transformation, selection and cultivation of bacterial cells are performed as described (Sambrook J, Fritsch EF and Maniatis T (1989)). Sequence analyses of recombinant DNA are performed with a laser fluorescence DNA sequencer (Applied Biosystems, Foster City, CA, USA) using the Sanger technology (Sanger et al., 1977). Unless described otherwise, chemicals and reagents are obtained from Sigma Aldrich (Sigma Aldrich, St. Louis, USA), from Promega (Madison, WI, USA), Duchefa (Haarlem, The Netherlands) or Invitrogen (Carlsbad, CA, USA). Restriction endonucleases are from New England Biolabs (Ipswich, MA, USA) or Roche Diagnostics GmbH (Penzberg, Germany). Oligonucleotides are synthesized by Eurofins MWG Operon (Ebersberg, Germany).

Introduction to experimental procedures

A yeast codon-optimized version of the Cas9 protein bearing a C-terminus SV40 nuclear localization signal (SEQ ID NO: 1) was synthetized and cloned into a yeast expression vector. The same vector included one or more guide RNAs (gRNAs) expressed from the *Saccharomyces cerevisiae* SNR52 polymerase III promoter.

Cas9 binds DNA and cleaves both strands upon recognition of a target sequence by the gRNA, but only if the correct protospacer-adjacent motif (PAM) is present at the 3' end. Theoretically, any sequence of the form GN20GG can be targeted. So, a second vector was constructed for co-expression in yeast of a reporter system (GAL4-UAS (SEQ ID NO: 7)) to be targeted by the designed CRISPR system. gRNA-donor fusions (fusion NA) were used to target and repair several non-functional Gal4 targets (SEQ ID NOs: 9-15).

Gal4 (SEQ ID NO: 8) is a yeast transcriptional activator consisting of two-components: the DNA binding domain located N-terminus and the region for transcriptional activation at C-terminus. Gal4 binds to the specific recognition sequence UAS (upstream activating sequence) of marker genes in the yeast genome, activating their transcription. The MaV203 yeast strain contains single copies of each of three reporter genes (HIS3, URA3 and lacZ) that are stably integrated at different loci in the yeast genome. The promoter regions of URA3, HIS3, and lacZ are unrelated (except for the presence of GAL4 binding sites).

Several non-functional (deleted and/or disrupted by insertion of a STOP codon) versions of Gal4 were synthesized (SEQ ID NOs: 9-15) and transformed into yeast cells, so that they could be targeted and repaired by the co-expressed CRISPR machinery. Restoration of the full-length Gal4 by homologous recombination (HR) with the appropriate repair donor sequence provided with the CRISPR components results in activation of lacZ and HIS3 reporter genes. Gal4 gene repair and consequent transcription activation can be monitored by cell growth on plates lacking histidine, whereas induction of the lacZ gene results in a blue color when assayed with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside).

The employed yeast strain contains two additional auxotrophic mutations (leu2 and trp1) to allow selection for both expression constructs.

To verify repair efficacy increase of the fusion system disclosed here, all experiments were performed in parallel with non-fused cassettes, in which donor and guide RNA are transcribed separately.

Yeast strain, media and cultivation conditions

The *Saccharomyces cerevisiae* strain used in the examples described is MaV203 (MATα, leu2-3,112, trp1-901, his3Δ200, ade2-101, gal4Δ, gal80Δ, SPAL10::URA3, GAL1::lacZ, HIS3UAS GAL1::HIS3@LYS2, can1R, cyh2R), commercialized by Life Technologies. Yeast was grown in Synthetic Minimal Media (SD Media) based upon Yeast Nitrogen Base supplemented with 2% glucose and lacking the appropriate auxotrophic compounds (ForMedium, United Kingdom). Cultures were grown at 30°C, either in a shaker or incubation oven.

Escherichia coli was used as propagation microorganism for all the plasmids used in our experiments, as well as for further propagation and maintenance of the modified targets. E. coli was grown according standard microbiological practices (Molecular Cloning: A Laboratory Manual, 3rd ed., Vols 1,2 and 3. J.F. Sambrook and D.W. Russell, ed., Cold Spring Harbor Laboratory Press, 2001). Plasmids containing the Cas9, guide RNA and donor NA included a pUC-based replication origin and ampicillin resistance gene for replication and maintenance in E. coli. Whereas GAL4 target plasmids contained a gentamicin resistance gene (Gmr).

Example 1 Plasmid Construction

The Cas9 gene was a yeast codon-optimized version of the Streptococcus pyogenes Cas9 (SpCas9; WO2007/025097) originally constructed for expression in eukaryotic cells (Mali et al (2013) Science 339(6121); Cong et al (2013) Science 339(6121)). This Cas9 gene was tagged with a SV40 nuclear localization signal at both ends and synthesised. Also, the gRNA and donor expression cassette containing the SNR52 promoter for in vivo RNA synthesis were synthesised.

The GAL4-AD coding sequence in pDEST22 (Life Technologies) was replaced by the synthetic Cas9 via Seamless Cloning (Life Technologies). This vector contains the constitutive moderate-strength promoter and transcription terminator of the yeast Alcohol Dehydrogenase gene (ADH1) for expression in yeast as well as a TRP1 gene for selection in yeast on medium lacking tryptophan.

The same vector contains two recombination sites, attR1 and attR2, flanking a chloramphenicol resistance gene (Cmr) and a ccdB gene, allowing the designed gRNA and donor expression cassettes (as fusion or dual molecule) to be introduced in the same expression vector via Gateway Cloning (Life Technologies). Following the LR recombination reaction, the Cmr and ccdB genes were replaced by the fusion NA cassette or non-fused donor and guide expression cassettes.

Modified GAL4 coding sequences used as targets for CRISPR repair in yeast were synthesized. The pDEST32 plasmid for expression in yeast (Life Technologies) was cut with HindIII and SacII and the backbone, containing the ADH1 promoter and terminator, was gel purified. The GAL4 synthesized inserts were assembled into the vector using Seamless Cloning. This

vector included a LEU2 gene for selection in yeast on medium lacking tryptophan.

Target-sites for recognition by Cas9 in the GAL4 sequence were empirically selected by choosing 20-mer regions preceding potential PAM (NGG) sequences within the GAL4 gene (Sternberg et al (2014); Nature 507(7490)).

To facilitate Cas9 binding and R-loop formation, we chose a single guide RNA design with the secondary structure containing a dangling spacer, an extended hairpin region and a long 3′ end, as initially designed by Jinek et al (2012) Science;337(6096)).

Example 2 Yeast Transformation

Simultaneous transformation of the CRISPR editing tools (Cas9 enzyme and fusion NA expressing cassette) and GAL4 target plasmid was performed by heat-shock as described in the manufacture's protocol (Life Technologies) and propagated in the appropriate synthetic complete (SC) media lacking the auxotrophic compounds complemented by the plasmids being introduced (leucine and tryptophan). The transformed cells were allowed to propagate overnight and equal amounts of transformants (according to OD measurement) were transferred to solid plates containing synthetic complete (SC) media lacking histidine with 100mM 3-Amino-1,2,4-triazole (3-AT; ForMedium, UK). Expression of HIS3 (for allowing yeast grow in medium without histidine) is GAL4-dependent and therefore transformants are only able to grow if GAL4 repair had occurred. More above, 3-AT is a competitive inhibitor of the product of the HIS3 gene, by applying 3-AT to the yeast transformants which are dependent upon HIS3 to produce histidine, an increased level of HIS3 expression is required in order for the yeast cells to survive.

Additionally the yeast strain used contained a lacZ marker gene under the control of GAL4, which allowed for blue/white selection of GAL4-repaired transformants. Induction of the lacZ gene results in a blue color when assayed with X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside).

Example 3 X-gal Assay

Transformants growing in plates lacking histidine were replica plated onto a nitrocellulose membrane (Hybond, GE Healthcare) placed on the surface of a plate with YPAD medium (Complex yeast media containing a homogeneous blend of Peptone, Yeast Extract and Glucose; ForMedium, UK). Assay was performed after 18-24 h incubation of the YPAD plates containing a membrane. For each membrane, 5mg X-gal were dissolved in 50µl DMF and combined with 30µl 2-mercaptoethanol and 5ml Z buffer. This solution was used to saturate two round filter papers (Whatman 541) in a 15-cm petri dish. Using forceps, the membrane was carefully remove from the surface of the YPAD plate and completely immersed in liquid nitrogen for about 20 seconds. The frozen membrane was placed on top of the soaked Whatman filters (colony side up). The plates were tightly covered and incubate at 37°C. Appearance of blue color was monitored after 24 hours.

Example 4 Sequencing of Target (CRISPR Repaired) Plasmids

Four each experiment, at least eight GAL4-repaired positive transformants (colonies able to

grow in medium without histidine) were sub-cultured overnight in liquid medium and the GAL4 containing plasmid was isolated (using Zymoprep Yeast Plasmid Miniprep II, Zymo Research). The isolated plasmids were introduced in E. coli for further propagation and commercial sequencing. GAL4 sequencing allowed verification of the sequence repair and assembly with the donor molecules.

Sanger sequencing of Gal4 gene in the positive clones further validated the sequence specificity of this targeting process, and showed no difference in repair of cells expressing the donor and gRNA as fusion or non-fused, even if cells transformed with fusion NA exhibit a much higher number of successful HR events.

Example 5 Deleting 1nt with 15bp homology arms

Fusion of donor (donor 1; SEQ ID NO: 26) to the guide RNA resulted in repaired transformants (able to grow on medium lacking histidine), whereas no growth was observed for transformants with non-fused guide and donor RNA. The low efficiency of gene repair is consistent with the reduced sequence overlap available for homologous recombination.

Example 6 Deleting 1nt with 50bp homology arms

Fusion of donor (donor 2; SEQ ID NO: 27) to the guide RNA resulted in at least 50 times more transformants than with non-fused donor and guide NA.

Sequencing of the Gal4 gene in the positive clones showed that repair results only or very largely from HR (no evidence of NHEJ for all sequenced clones).

Example 7 Inserting 20nt with 50/26bp homology arms

The same fusion NA as above was used to repair a similar target where 20nt were removed (target 3; SEQ ID NO: 11), and as a consequence one homology arm was reduced. Fusion resulted in about 5 times more transformants than with non-fused donor and guide NA. Sequencing of the Gal4 gene in the positive clones showed that repair results exclusively from homologous recombination.

Example 8 Inserting missing 400bp with 50bp homology arms (while testing two target sequences 3nt apart)

We tested for simultaneous targeting of two sequences (spacer 2 and spacer 3; SEQ ID NOs: 20 and 21) located in close proximity (3nt gap between the two 20 nt target), both independently and together (multiplexed targeting). The multiplex fusion cassette consisted of promoter followed by two tandem fusion NA sequences, resulting in production of a single molecule composed of two gRNAs and repair templates. Our experiments clearly showed that fusion NA is also amenable for targeting two sequences simultaneously.

For both targets repair in the presence of the donor-guide fusion was largely more efficient than with non-fused version (up to 10 times more for targeting with space 2 and five times for spacer 3).

Example 9 Inserting full GAL4 gene (960bp) except HR ends with 120 bp homology arms

In order to test if fusion CRISPR could be effective for introduction full length coding sequences, we have tested introducing the full length GAL4 gene (SEQ ID NO: 7). As example, we have selected for 120bp homology arms as to keep the ratio of donor/homology arm length already found to be effective in example 4. Insertion of full-length GAL4 gene is about four times more effective with Fusion construct.

Our results show that targeted editing is at least 50 times more efficient when the repair donor sequence was fused to the gRNA. The experiments performed indicate a broad Fusion-related improved effectiveness from a single base removal up to full gene insertion. The examples reported show that this CRISPR fusion system is suitable to carry relatively large Donor molecules fused to the guide RNA.

Example 10a Constructs for Expression in Rice

To accommodate the CRISPR/Cas system to Agrobacterium-mediated plant transformation, Gateway binary T-DNA vectors have been designed for co-expression of Cas9 nuclease and guide RNA-donor expression cassette (either as single or dual RNA molecules). A version of the Streptococcus pyogenes Cas9 (SpCas9) codon-optimized for expression in rice (*Oryza sativa*), attached to SV40 nuclear localization signals (NLS) at both ends (Seq ID NO: 6), was synthesized The synthesized cassette includes the maize polyubiquitin (Ubi) promoter (Seq ID NO: 32) for constitutive expression located upstream the Cas9, and the nopaline synthase (nos) terminator (Seq ID NO: 33) at the 3'-end. This gene cassette has been cloned via Seamless into a vector, which contained as functional elements within the T-DNA borders: a plant selectable marker; a screenable marker expression cassette; and a Gateway cassette intended for LR recombination with the gRNA-donor expression cassette in an entry clone.
Three gRNA have been designed, which targeted the rice Protoporphyrinogen Oxidase (PPO) gene (WO/2015/092706; WO/2015/022640 (Seq ID NO: 35), resulting in genomic double-strand cleavage at selected target sites (spacer 8, spacer 9 and spacer 10 (Seq ID NO: 36, 37 and 38)). Modifications aim two amino acid substitutions (L419F, F442V; single site mutations and double site mutation), which have been previously identified as potential hotspots for Saflufenacil survival.
The RNA expressing cassette (including gene-specific spacer sequences for the selected locations in the PPO gene) containing either fusion or non-fusion NA were synthesized and cloned into entry vectors, which was cloned (via Gateway) into the destination vector containing the CAS9 expression cassette. RNA expression of gRNA and donor is driven by pol III type promoter of U3 snRNA.
After the LR recombination step, the resulting expression vector is transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

Example 10b Constructs for Expression in Rice

An identical vector as described in Example 10a was synthesized with the exception that the NLS derived from SV40 was replaced with plant nuclear localization signals (NLS)

(MSERKRREKL, SEQ ID NO: 71) at the N-terminal end and importin NLS (KRPAAT-KKAGQAKKKK SEQ ID NO: 72) at the C-terminal end and the promoter driving the RNA expression of gRNA and donor was rice pol III type promoter of U3 snRNA (SEQ ID NO: 73).

The RNA expressing cassette (including gene-specific spacer sequences for the selected locations in the PPO gene) containing either fusion or non-fusion NA were synthesized and cloned into entry vectors, which was cloned (via Gateway) into the destination vector containing the CAS9 expression cassette.

The vector used as non fusion control contains PRO0231::U3 RNA pol III promoter::spacer::sgRNA scaffold::TTTTTTTT terminator::U3 RNA pol III promoter::template::TTTTTTTT terminator.

After the LR recombination step, the resulting expression vector is transformed into Agrobacterium strain LBA4044 according to methods well known in the art.

Example 11 Rice Transformation and Selection of Herbicide-tolerant Calli

The Agrobacterium containing the expression vector is used to transform scutellum-derived callus of indica rice (*Oryza sativa* L.). Sterilization of mature seeds has been carried out by incubating for one minute in 70% ethanol, followed by 40 minutes in 6% sodium hypochlorite, followed by a 3 to 5 times wash with sterile MQ water. The sterilized seeds are then germinated on a medium containing 2,4-D (callus induction medium). After 6 days of incubation in the light, scutellum-derived calli are incubated for 90 seconds in bacterial solution (OD_{600} = 0,1), drained, dried on sterile filter paper and then co-cultured with bacteria for 3 days in the dark at 25°C. The co-cultivated calli are transferred to selection medium containing G418 for 4 weeks in the light at 32°C. Antibiotic-resistant callus pieces are transferred to selection medium containing 25 or 50 µM saflufenacil (Kixor™) for 2 weeks in the light at 32°C. These herbicide selection conditions have been established through the analysis of tissue survival in kill curves with saflufenacil. After transfer of herbicide-resistant material to a regeneration medium and incubation in the light, the embryogenic potential is released and shoots developed in the next four to five weeks. Shoots are excised from the calli and incubated for 2 to 3 weeks on an auxin-containing medium until shoots are well rooted for transfer to soil. Hardened shoots are grown under high humidity and short days in a greenhouse.

Example 12 Molecular Characterization of the Herbicide Tolerant Transformants

Leaf tissue collected from each individual plant transformant is used for copy number analysis and molecular characterization of PPO gene sequence mutations. Genomic DNA is extracted using a Wizard 96 Magnetic DNA Plant System kit (Promega, US Patent Nos. 6,027,945 & 6,368,800) as directed by the manufacturer. Isolated DNA was PCR amplified using the appropriate probe, together with forward and reverse primers. Following this quantitative PCR analysis to verify copy number of the T-DNA insert, only low copy transgenic plants that exhibit tolerance to the selection agent are kept for harvest of T1 seeds. Seeds are then harvested three to five months after transplanting.

PCR amplification of PPO genomic sequences is performed using Fusion Taq DNA Polymerase (Thermo Scientific) using thermocycling program as follows: 96°C for 15 min, followed by 35 cycles (96°C, 30 sec; 58°C, 30 sec; 72°C, 3 min and 30 sec), 10 min at 72°C. PCR products are verified for concentration and fragment size via agarose gel electrophoresis, and send for sequencing using the PCR primers. Sequence analysis is performed on the representative chromatogram trace files and corresponding AlignX alignment with default settings and edited to call secondary peaks.

Mutations identified in several individuals, based on sequence information, show that the technology described in this invention, which involves fusion of NA to the CRISPR components, is applicable to plant organisms. Homologous recombination repair with the provided donors confers tolerance to Saflufenacil (single site mutation and multiple site mutation).

Example 13 Controlled gene knockout in *Escherichia coli*

In this example FusionCRISPR is being used to knockout target gene RecA in *E. coli* strain K-12 substr. MG1655. The bacterial strain is inoculated in 10 ml SOB in a 100 ml Erlenmeyer flask and grown overnight at 37˚C (SOB: 2% bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$). 3 ml of the overnight culture is diluted in 250 ml SOB in a 1 liter Erlenmeyer flask and grown at 18˚C with vigorous shaking (200-250 rpm) until the OD_{660nm} is 0.6. Subsequently the culture is transfered to precooled 50 ml tubes and centrifuge at 5000 rpm for 5 min at 4˚C. The pellet is resuspend in 1/3 of the original volume of ice-cold TB (TB: 250 mM KCl, 10 mM PIPES free acid, 15 mM $CaCl_2$•$2H_2O$, 55 mM $MnCl_2$•$2H_2O$) and incubated on ice for 10 min. The cells are centrifuged at 5000 rpm for 5 min at 4˚C and the pellet resuspended in 1/12 of the original volume of ice-cold TB. DMSO is added with gentle mixing to a final concentration of 7%. The competent cells are alliquoted in 200 µl portions and freezed in liquid nitrogen. One aliquot of competent cells is added together with 0.1-0.5 µg of plasmid containing a chloramphenicol selectable marker and Cas9 expression cassette as present in pCas9 [Jiang W, Bikard D, Cox D, Zhang F, Marraffini L (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol] and a cassette for expression of fusion RNA [Zhao D, Yuan S, Xiong B, Sun H, Ye L, Li J, Zhang X, Bi C. (2016) Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. Microb Cell Fact. 15(1):205] with the following FusionCRISPR sequence and RecA spacer:

gatgtggaaaccatctctacGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAG-TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTCCATGGATGTGGAAACC ATCGCTTTCACTGGATATCGCG (SEQ ID NO: 42)

in which the spacer recognizing RecA is highlighted, the essential sequences for the sgRNA is capitalized, not underlined, homology arm 1 with RecA is double underlined and homology arm 2 with RecA is single underlined. Promoters and terminators for the FusionCRISPR construct and Cas9 can be chosen from http://parts.igem.org/Promoters/Catalog/Constitutive

and http://parts.igem.org/Terminators/Catalog. The targeted RecA gene has the following sequence:

ATGGCTATCGACGAAAACAAACAGAAAGCGTTGGCGGCAGCACTGGGCCAGATTGA-
GAAACAATTTGGTAAAGGCTCCATCATGCGCCTGGGTGAAGACCGTTCCATGGATGTG
GAAACCATCTCTAC*CGG*TTCGCTTTCACTGGATATCGCGCTTGGGG-
CAGGTGGTCTGCCGATGGGCCGTATCGTCGAAATCTACGGACCGGAATCTTCCGGTA
AAACCACGCTGACGCTGCAGGTGATCGCCGCAGCGCAGCGTGAAGGTAAAAC-
CTGTGCGTTTATCGATGCTGAACACGCGCTGGACCCAATCTACGCACGTAAACTGGGC
GTCGATATCGACAACCTGCTGTGCTCCCAGCCGGACACCGGCGAGCAGGCACTG-
GAAATCTGTGACGCCCTGGCGCGTTCTGGCGCAGTAGACGTTATCGTCGTTGACTCC
GTGGCGGCACTGACGCCGAAAGCG-
GAAATCGAAGGCGAAATCGGCGACTCTCACATGGGCCTTGCGGCACGTATGATGAGC
CAGGCGATGCGTAAGCTGGCGGGTAACCTGAAGCAGTCCAACAC-
GCTGCTGATCTTCATCAACCAGATCCGTATGAAAATTGGTGTGATGTTCGGTAACCCG
GAAACCACTACCGGTGGTAACGCGCTGAAATTCTAC-
GCCTCTGTTCGTCTCGACATCCGTCG-
TATCGGCGCGGTGAAAGAGGGCGAAAACGTGGTGGGTAGCGAAACCCGCGTGAAAG
TGGTGAAGAACAAAATCGCTGCGCCGTTTAAACAGGCTGAATTCCAGATCCTCTAC-
GGCGAAGGTATCAACTTCTACGGCGAACTGGTTGACCTGGGCGTAAAAGAGAAGCTG
ATCGAGAAAGCAGGCGCGTGGTACAGCTACAAAGGTGAGAAGATCGGTCAGGGTAAA-
GCGAATGCGACTGCCTGGCTGAAAGATAACCCGGAAACCGCGAAAGAGATCGAGAAG
AAAGTACGTGAGTTGCTGCTGAGCAACCCGAACTCAACGCCGGATTTCTCTG-
TAGATGATAGCGAAGGCGTAGCAGAAACTAACGAAGATTTTTAA (SEQ ID NO: 43)

in which the PAM sequence is in italics, homology arm 1 is double underlined, homology arm 2 is single underlined, and the protospacer is highlighted. DNA and cells are kept on ice for 30 minutes prior to a 90 seconds heat shock at 42˚C. Cells and DNA are transferred to ice and 1 ml LB is added after 1 minute (LB: 1% tryptone, 1% NaCl, 0.5% yeast extract, pH 7.0). Cells are allowed to recover for 1 hour at 37˚C. The recovery phase can be extended to 16 hours to allow the FusionCRISPR components more time to edit the *E. coli* genome. 25 µg/ml chloramphenicol should be added after 1 hour to prevent loss of the plasmid. Cells are plated on LB medium with 25 µg/ml chloramphenicol and incubated at 37˚C for 1 day. Single colonies are selected from plate and grown overnight in LB with chloramphenicol at 37˚C after which genomic DNA is extracted [He, F. (2011) E. coli Genomic DNA Extraction. Bio-protocol Bio101: e97]. PCR with a forward primer upstream from the first homology arm (ATGGCTATCGACGAAAACAAA) (SEQ ID NO: 44) and reverse primer downstream from the second homology arm (CGTCAGCGTGGTTTTACCGGA) (SEQ ID NO: 45) is performed to identify colonies in which the 11 nucleotides shown in bold in the RecA sequence (SEQ ID 43) are no longer present due to homologous recombination repair with the FusionCRISPR template. PCR fragments can be sequenced (expected size 220 bp) or, in this case, subjected to *Age*I digestions (the deleted sequence around PAM contains the *Age*I recognition site ACCGGT)

to verify modification of the locus after standard gel electrophoresis. Deletion of 11 nucleotides ensures a disruption of the open reading frame.

Example 14 Controlled knockout of the PRDM9 gene in human-induced pluripotent stem cells (hiPSCs) and HEK293 cells

Cell culture maintenance, plasmid construction, transfection methods and molecular analysis of genome editing in hiPSCs and HEK293 cells are described in great detail in Yang L, Yang JL, Byrne S, Pan J, Church G (2014) CRISPR/Cas9-directed genome editing of cultured cells. Current Protocols in Molecular Biology 31.1.1-31.1.17. For knockout of the PRDM9 gene, all steps are followed as described therein, with only a minor change in the gRNA plasmid design. The synthesized gRNA should have the following sequence:

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTAC-
CAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGAT
ACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAG-
TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTAT
GTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTT-
GGCTTTA-
*TATATCTTGTGGAAAGGACGAAACACC*ggcatccctcaggctgggctGTTTTAGAGCTAGAAATA
GCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG-
TCGGTGCTGGCCATCAGGCATCCCTCAGTATGGAATGAGGCATCTGATtttttt (SEQ ID NO: 46)

in which the U6 promoter is shown in italics, the spacer recognizing PRDM9 exon ENSE00001804383 is highlighted, the essential sequences for the sgRNA is capitalized not underlined or in italics, homology arm 1 with PRDM9 is double underlined, homology arm 2 with PRDM9 is single underlined and the terminator is in small case. The targeted PRDM9 exon ENSE00001804383 has the sequence:

ATTGTGAGATGTGTCAGAACTTCTTCATT-
GACAGCTGTGCTGCCCATGGGCCCCCTACATTT-
GTAAAGGACAGTGCAGTGGACAAGGGGCACCCCAACCGTTCAGCCCTCAGTCTGCCC
CCAGGGCTGAGAATTGGGCCATCAGGCATCCCTCAG**GCTGGGCT***TGG***AG-**
TATGGAATGAGGCATCTGATCTGCCGCTGGGTCTGCACTTTGGCCCTTATGAGGGCCG
AATTACAGAAGACGAAGAGGCAGCCAACAATGGATACTCCTGGCTGTGG (SEQ ID NO: 47)

in which the PAM sequence is in italics, homology arm 1 is double underlined, homology arm 2 is single underlined, and the protospacer is highlighted. The nucleotides shown in boldface are deleted upon homologous recombination with the FusionCRISPR construct resulting in a frame shift as shown using PCR amplifying the respective genomic region from genomic DNA and subsequent sequencing of the resulting PCR products.

Example 15: Introduction of point mutations in Rice plants leading to cyclohexanedione (DIM) and/or aryloxyphenoxypropionate (FOP) in rice.

Mutations I1781L and G2096S in plastidic Acetyl Coenzyme A Carboxylase (ACCase) are known to confer tolerance to DIM and FOP herbicides. These mutations can be introduced at the endogenous ACCase locus using the following vectors.

Vector RLW137 SEQ ID NO: 66

The backbone of this vector is the gateway-enabled construct RLW121 SEQ ID NO: 62.

ENTR vectors for RLW137 are vectors CC003 SEQ ID NO: 63 (selectable marker for the incoming T-DNA), CC018 SEQ ID NO: 64 (producing the FusionCRISPR construct which targets and introduces G2096S after cutting upstream from the DNA that corresponds to G2096) and CC006 SEQ ID NO: 65 (providing Cas9).

CC018 (short for CRISPRCas018) contains ~300 nt homology arms flanking the incoming nucleotides (in this case encoding G2096S). Additional mutations are co-introduced to avoid self-cleavage of the T-DNA (mutated PAM, alternatively or in addition the spacer could include many mutations which are preferably silent in parts that correspond with exons and do not affect intron/exon borders if present) and early termination of transcription on long stretches of T present either in the homology arms or incoming nucleotides.

A control vector is synthesized which is identical except that the donor molecule is expressed as separate molecule which is not linked to the guide RNA.

Vector RLW137 and the control vector are transformed into rice using the protocol described above. Initial selection is for the presence of the ZmAHAS A122T S553N marker. Analysis of the transformed plants is performed as described in example 12.

Similar to the procedure described above for RLW137, RLW138 introduces the same mutation, but this time using an alternative, downstream protospacer site. RLW138 consists of the RLW121 backbone with the CC003, CC019 (SEQ ID NO: 67) and CC006.

The mutation I1781L is introduced by RLW139 (SEQ ID NO: 70) consisting of RLW121, CC003, CC020 (SEQ ID NO: 69) and CC006.

Example 16 Application of Fusion CRISPR in Bacillus

Electrocompetent Bacillus subtilis cells and electroporation

Transformation of DNA into B. subtilis ATCC 6051 is performed via electroporation. Preparation of electrocompetent B. subtilis ATCC 6051 cells and transformation of DNA is performed as essentially described by Xue et al (Xue, G.-P., 1999, Journal of Microbiological Methods 34, 183-191) with the following modification: Upon transformation of DNA, cells are recovered in 1ml LBSPG buffer and incubated for 60min at 37°C (Vehmaanperä J., 1989, FEMS Microbio. Lett., 61: 165-170) following plating on selective LB-agar plates. For plasmids containing the temperature- sensitive pE194 replication origin, cells are recovered for 3h at 33°C.

Plasmid Isolation

Plasmid DNA was isolated from Bacillus and E. coli cells by standard molecular biology methods described in (Sambrook,J. and Russell,D.W. Molecular cloning. A laboratory manual, 3rd

ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 2001.) or the the alkaline lysis method (Birnboim, H. C., Doly, J. (1979). Nucleic Acids Res 7(6): 1513-1523). Bacillus cells were in comparison to E. coli treated with 10mg/ml lysozyme for 30min at 37°C prior to cell lysis.

Annealing of oligonucleotides to form oligonucleotide-duplexes.

Oligonucleotides were adjusted to a concentration of 100µM in water. 5µl of the forward and 5µl of the corresponding reverse oligonucleotide were added to 90µl 30mM Hepes-buffer (pH 7.8). The reaction mixture was heated to 95°C for 5min following annealing by ramping from 95°C to 4°C with decreasing the temperature by 0.1°C/sec (Cobb, R. E., Wang, Y., & Zhao, H. (2015). High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. ACS Synthetic Biology, 4(6), 723–728).

Molecular biology methods and techniques

Plasmid pJOE8999:

Altenbuchner J. 2016. Editing of the Bacillus subtilis genome by the CRISPR-Cas9 system. Appl Environ Microbiol 82:5421–5.

Plasmid pCC001

The pJOE8999 and the synthetic gene fragment Seq ID 048 provided in a standard E. coli cloning vector (pUC derivative) are cut with AvrII and XbaI following isolation of the pJOE8999 plasmid backbone and the smaller AvrII/XbaI fragment of Seq ID 048. The two fragments are ligated using with T4-DNA ligase (NEB) following transformation into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid was recovered and named pCC001.

Plasmid pCC002

The pJOE8999 and the synthetic gene fragment Seq ID 049 provided in a standard E. coli cloning vector (pUC derivative) are cut with AvrII and XbaI following isolation of the pJOE8999 plasmid backbone and the smaller AvrII/XbaI fragment of Seq ID 049. The two fragments are ligated using with T4-DNA ligase (NEB) following transformation into E. coli XL1-Blue competent cells (Stratagene). The correct plasmid was recovered and named pCC002.

Plasmid pCC003

The oligonucleotides SeqID 050 and Seq ID 051 with 5' phosphorylation are annealed to form an oligonucleotide duplex encoding for the protospacer sequence targeting the amylase gene amyE of B.subtilis ATCC6051. The plasmid pJOE8999 is cut with BsaI following ligation of the oligonucleotide duplex to recover plasmid pCC003.

Plasmid pCC004

The 5'homology region (also referred to as HomA) and the 3' homology region (also referred to as HomB) adjacent to the amylase amyE gene of B. subtilis ATCC6051 were PCR-amplified on isolated genomic DNA with oligonucleotides Seq ID NO: 52, Seq ID NO: 53 and SeqID NO: 54, Seq ID NO: 55 respectively. The two homology regions HomA and HomB were fused and amplified using overlap PCR with oligonucleotides Seq ID NO: 52 and Seq ID NO: 55 to recover the HomAB PCR fragment of the homology regions of the amyE gene. The plasmid pCC003 and the HomAB-amyE PCR fragment were cut with SfiI following ligation with T4-

DNA ligase (NEB). The reaction mixture was transformed into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid containing the amyE protospacer and the HomAB of amyE was recovered and named pCC004 (Figure 23).

Plasmid pCC005

The plasmid pCC001 was cut with BsaI following cloning of the amyE protospacer oligonucleotide duplex (SeqID 050/Seq ID051) as described for pCC003. The resulting plasmid and the PCR-fragment of the homology regions HomAB of the amyE gene as described for construction of pCC004 were cut with SfiI following ligation with T4-DNA ligase (NEB). The reaction mixture was transformed into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid containing the amyE protospacer and the HomAB of amyE was recovered and named pCC005.

Plasmid pCC006

The 5' homology region (also referred to as HomA) and the 3' homology region (also referred to as HomB) adjacent to the amylase amyE gene of B. subtilis ATCC6051 were PCR-amplified on isolated genomic DNA with oligonucleotides Seq ID NO: 56, Seq ID NO: 57 and SeqID NO: 58, Seq ID NO: 59 respectively. The two homology regions HomA and HomB were fused and amplified using overlap PCR with oligonucleotides Seq ID NO: 56 and Seq ID NO: 59 to recover the HomAB PCR fragment of the homology regions of the amyE gene. The plasmid pCC001 was cut with BsaI following ligation of the amyE protospacer oligonucleotide duplex (SeqID NO: 50/Seq ID NO: 51) with T4-DNA ligase (NEB) as described for pCC003. The resulting plasmid and the PCR-fragment of the homology regions HomAB of the amyE gene were cut with SfiI following ligation with T4-DNA ligase (NEB). The reaction mixture was transformed into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid containing the amyE protospacer and the HomAB of amyE in reverse orientation compared to pCC005 was recovered and named pCC006.

Plasmid pCC007

The plasmid pCC002 was cut with BsaI following ligation of the amyE protospacer oligonucleotide duplex (SeqID NO: 50/Seq ID NO: 51) with T4-DNA ligase (NEB) as described for pCC003. The resulting plasmid and the PCR-fragment of the homology regions HomAB of the amyE gene as described for construction of pCC004 were cut with SfiI following ligation with T4-DNA ligase (NEB). The reaction mixture was transformed into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid containing the the HomAB of the amyE gene and the amyE protospacer was recovered and named pCC007.

Plasmid pCC008

The plasmid pCC002 was cut with BsaI following ligation of the amyE protospacer oligonucleotide duplex (SeqID NO: 50/Seq ID NO: 51) with T4-DNA ligase (NEB) as described for pCC003. The resulting plasmid and the PCR-fragment of the homology regions HomAB of the amyE amplified with oligonucleotides Seq ID NO: 56 and Seq ID NO: 59 as described for pCC006 were cut with SfiI following ligation with T4-DNA ligase (NEB). The reaction mixture was transformed into E.coli XL1-Blue competent cells (Stratagene). The correct plasmid containing the the HomAB of amyE in reverse orientation compared to pCC007and the amyE protospacer was recovered and named pCC008.

Gene deletion using Fusion-CRISPR.

Electrocompetent B. subtilis ATCC6051 cells were transformed with 1µg each of plasmids pJOE8999, pCC004, pCC005, pCC006, pCC007, pCC008 as essentially described by Xue et al (Xue, G.-P., 1999, Journal of Microbiological Methods 34, 183-191) with the following modification: Upon transformation of DNA, cells were recovered in 1ml LBSPG buffer and incubated for 3h at 33°C (Vehmaanperä J., 1989, FEMS Microbio. Lett., 61: 165-170) following plating on LB-Lennox plates supplemented with 20µg/ml kanamycin and 0.2% D-Mannose for Cas9 induction. Plates were incubated for 20-22h at 33°C. Up to 10 clones from each plasmid transformation were picked and transferred onto a fresh preheated LB-Lennox-plate following incubation at 50°C for 18h. From each large grown colony, cells were picked and 3 strokes on fresh LB-Lennox plates performed to yield single colonies after 7-8h incubation at 45°C. Single colonies were transferred onto LB-Lennox plates and LB-Lennox plates supplemented with 20µg/ml kanamycin, following incubation for 16-18 hours at 30°C. Kanamycin-sensitive clones, indicative of plasmid loss, were plated on LB-Lennox plates supplemented with 1% soluble starch following incubation for 20 hours at 30°C. Inactivation of the amylase amyE gene was visualized by covering the plates with iodine containing Lugols solution and analyzed for the presence or absence of a light halo, the latter indicating a successful inactivation.

Table 2 summarizes the amount of total clones after plasmid curing, amount of clones with inactivated amylase, the percentage of clones with inactivated amylase relative to total clones and the relative knockout efficiency with the indicated plasmids relative to pCC004 (Figure 25)

Table 2

Construct	Subclones total	Subclones Amy. neg.	Subclones Amy. neg. [%]	Relative to pCC004
pJOE8999	90	0	0	0
pCC004	177	42	24	100
pCC005	197	113	57	242
pCC006	192	79	41	173
pCC007	117	95	81	342
pCC008	146	116	79	335

2017256517 05 Jun 2023

What is claimed is:

1. A method for modification of a target nucleic acid (target NA) molecule in a cell comprising the steps of
    a. providing a recombinant fusion nucleic acid (fuNA) molecule comprising a guide nucleic acid (gNA) molecule covalently linked to at least one donor nucleic acid (doNA) molecule, and
    b. introducing said fuNA molecule into one or more cells comprising the target NA molecule , and
    c. introducing a site directed nucleic acid modifying polypeptide into said one or more cells, and
    d. incubating the one or more cells under conditions that allow for homologous recombination in said one or more cells, and optionally
    e. isolating one or more cells in which homologous recombination occurred

    wherein the fuNA consists of RNA,

    wherein the doNA comprises two homology arms each homology arm comprising at least 15 bases which are 100% complementary to the same number of consecutive bases in the target NA molecule, and

    wherein the gNA molecule comprises a spacer nucleic acid (spacer NA) molecule and a scaffold nucleic acid (scaffold NA) molecule, the scaffold NA molecule forming a secondary structure comprising at least one hairpin and wherein the spacer NA comprises at least 16 bases complementary to the target NA molecule.

2. The method of claim 1 wherein the spacer NA comprises at least 18 bases which are 100% complementary to a stretch of the same number of consecutive bases in the target NA molecule.

3. The method of claim 1 or claim 2 wherein said two homology arms are directly adjacent to each other or are separated by at least one additional base.

4. The method of claim 3 wherein the scaffold NA molecule is covalently bound to the gNA molecule.

5. The method of any of claim 1 to 4 wherein the cell is a microbial, animal, human or plant cell.

6. The method of any of claim 1 to 5 wherein the site directed nucleic acid modifying polypeptide is a nucleic acid guided nucleic acid modifying polypeptide or a functional equivalent thereof.

2017256517 05 Jun 2023

7. The method of any of claim 1 to 6 wherein the fuNA molecule is introduced as one or more expression constructs encoding said fuNA molecule.

8. A recombinant fuNA molecule comprising a doNA molecule covalently linked to a gNA molecule, wherein the fuNA is consisting of RNA and the doNA comprises two homology arms each homology arm comprising at least 15 bases which are 100% complementary to the same number of consecutive bases in the target NA molecule, and wherein the gNA molecule comprises a spacer NA molecule and a scaffold NA molecule, the scaffold NA molecule forming a secondary structure comprising at least one hairpin and wherein the spacer NA comprises at least 16 bases complementary to the target NA molecule.

9. A vector comprising an expression construct comprising a promoter functionally linked to a DNA molecule encoding the fuNA molecule of claim 8.

10. A vector system comprising
    a. the vector of claim 9 and
    b. a vector encoding a site directed nucleic acid modifying polypeptide and optionally
    c. a vector encoding a scaffold NA molecule.

11. A system for modification of a target NA molecule in a cell comprising
    a. the vector of claim 9 and
    b. a vector encoding a site directed nucleic acid modifying polypeptide and
    c. a cell comprising a target NA molecule and optionally
    d. a vector encoding a scaffold NA molecule.

12. A composition comprising
    a. the vector of claim 9 and
    b. a vector encoding a site directed nucleic acid modifying polypeptide and
    c. a cell comprising a target NA molecule and optionally
    d. a vector encoding a scaffold NA molecule.

13. A use of the vector of claim 9, the vector system of claim 10, the system of claim 11 or the composition of claim 12 for modification of a target NA molecule in a cell.

14. The method of any one of claim 1 to 7 using the recombinant fuNA molecule of claim 8, the vector of claim 9, the vector system of claim 10, the system of claim 11 or the composition of claim 12.

2017256517 31 Aug 2023

15. Use of the recombinant fuNA molecule of claim 8, the vector of claim 9, the vector system of claim 10, the system of claim 11 and/or the composition of claim 12 in the manufacture of a medicament for treating a genetic disorder.

16. A method of treating a genetic disorder in a subject, the method comprising administering the recombinant fuNA molecule of claim 8, the vector of claim 9, the vector system of claim 10, the system of claim 11 and/or the composition of claim 12.

17. The use of claim 15 or the method of claim 16, wherein the genetic disorder is sickle cell disease.

Figure 1

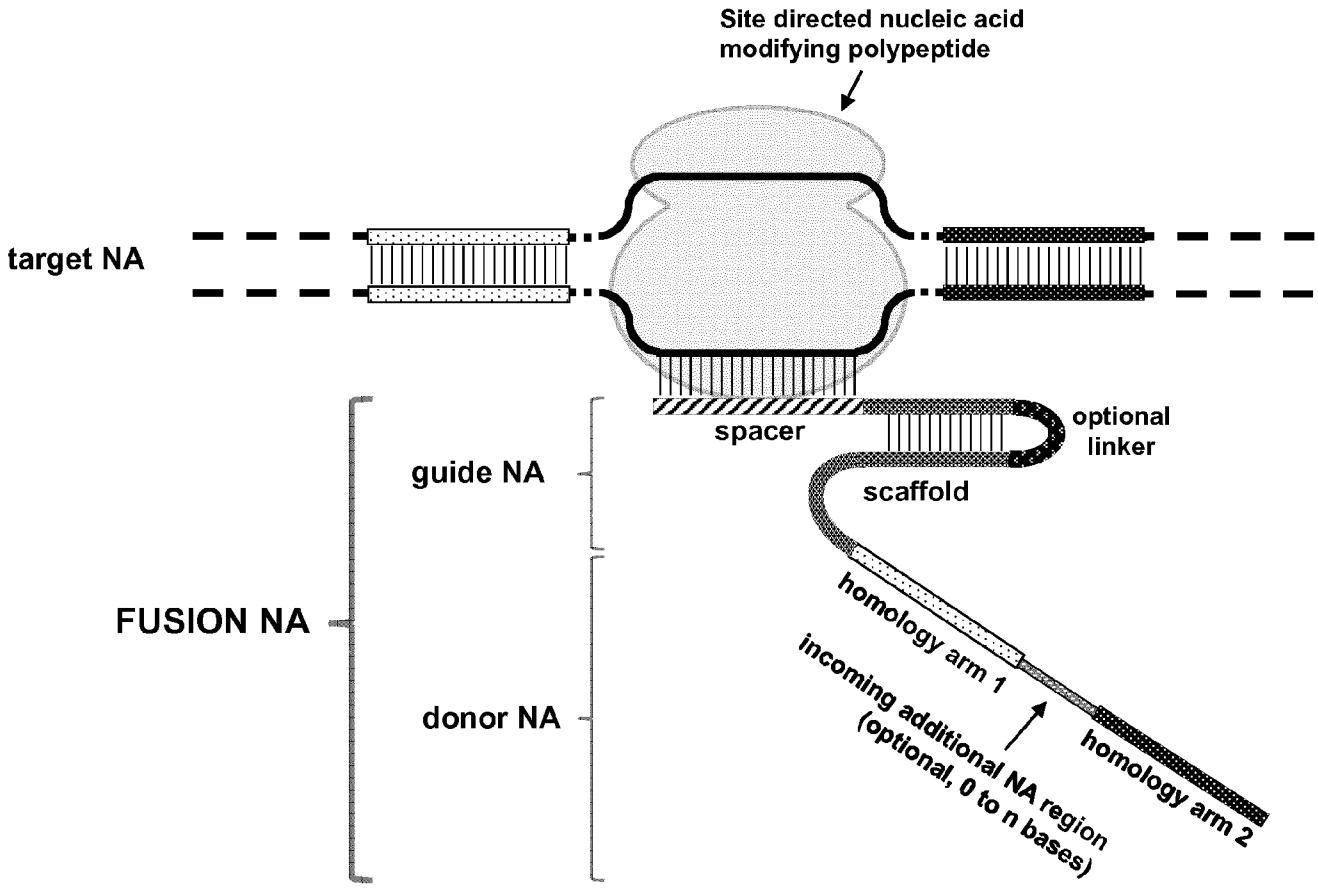

Site directed nucleic acid modifying polypeptide
target NA
spacer
optional linker
guide NA
scaffold
FUSION NA
homology arm 1
donor NA
incoming additional NA region (optional, 0 to n bases)
homology arm 2

Figure 2

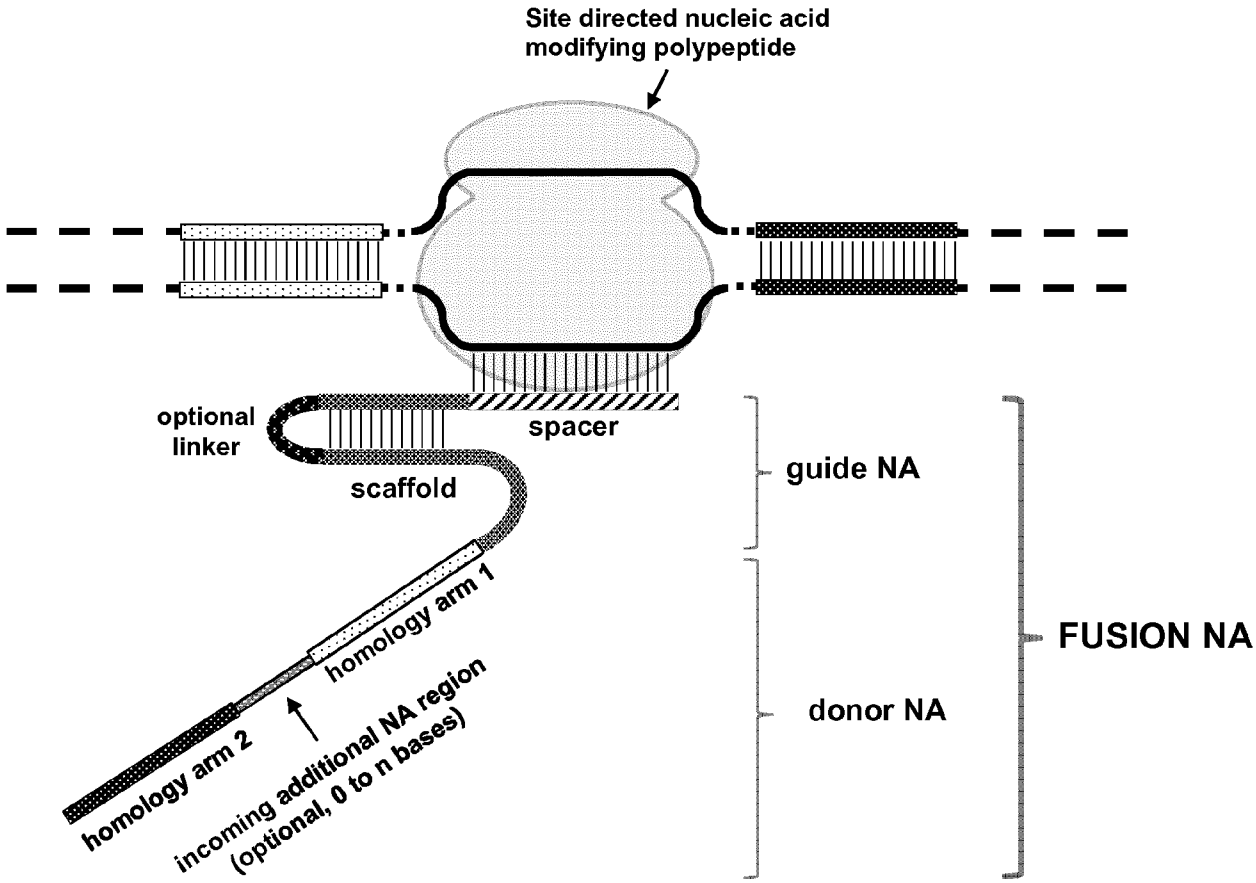

Site directed nucleic acid modifying polypeptide
optional linker
spacer
scaffold
guide NA
homology arm 1
FUSION NA
donor NA
homology arm 2
incoming additional NA region (optional, 0 to n bases)

Figure 3

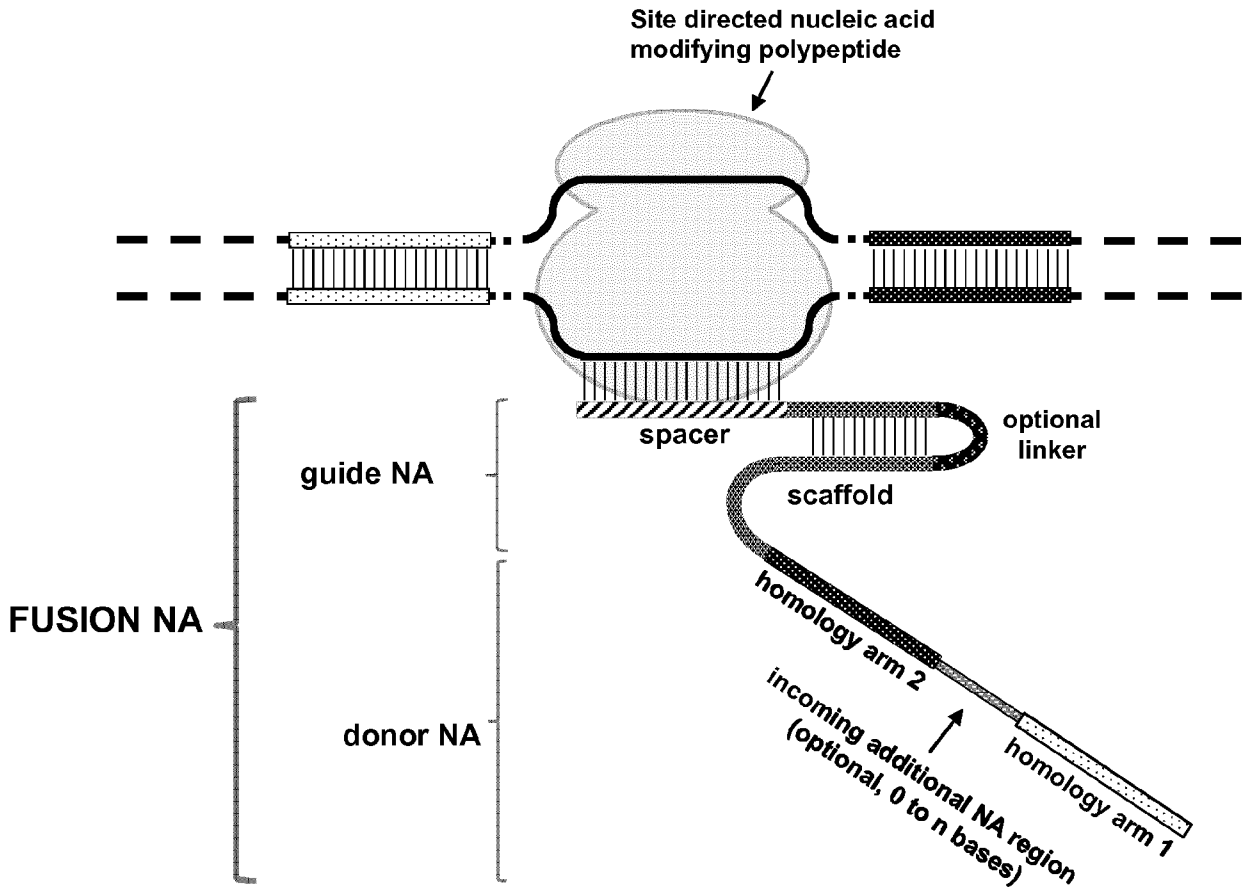

Site directed nucleic acid modifying polypeptide
spacer
optional linker
scaffold
guide NA
FUSION NA
donor NA
homology arm 2
incoming additional NA region (optional, 0 to n bases)
homology arm 1

Figure 4

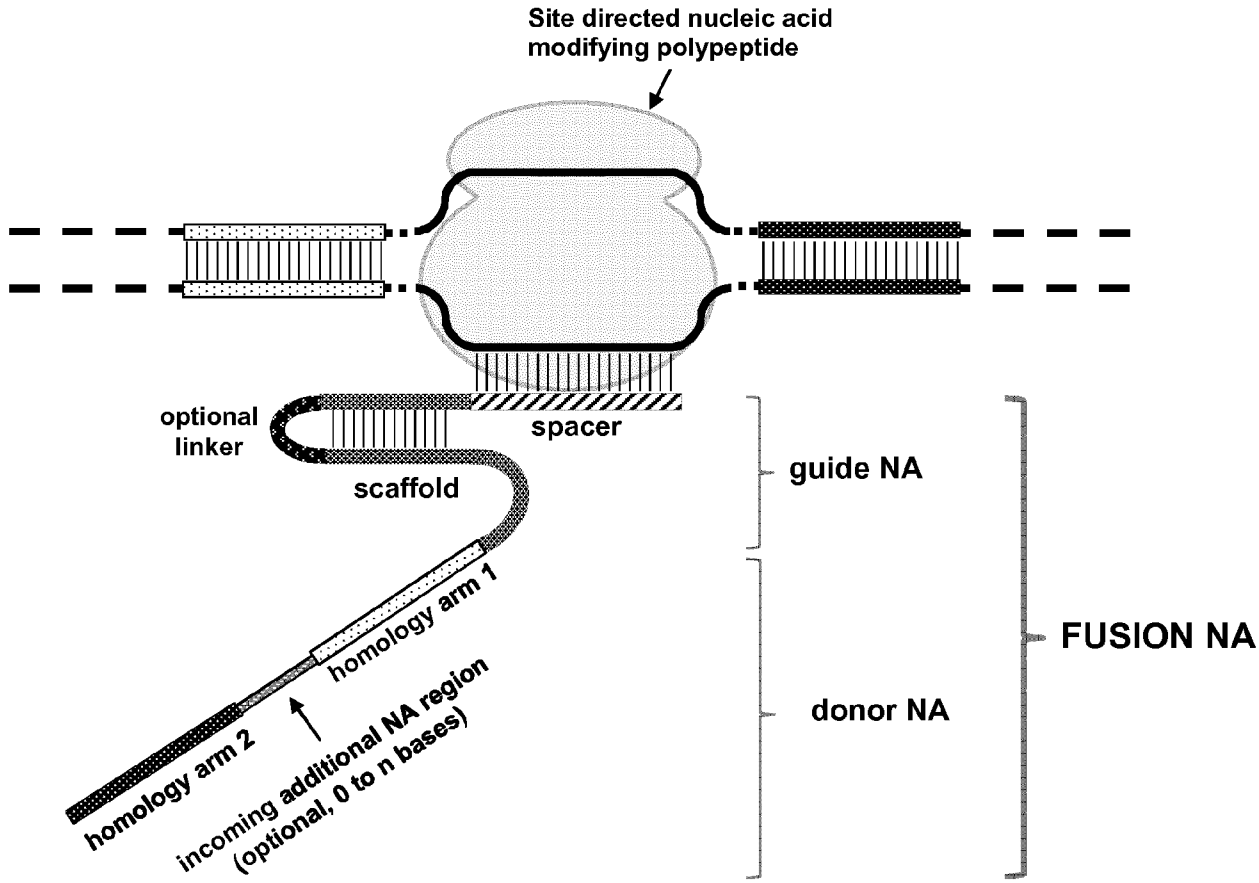

Site directed nucleic acid modifying polypeptide
optional linker
spacer
scaffold
guide NA
homology arm 1
homology arm 2
incoming additional NA region (optional, 0 to n bases)
donor NA
FUSION NA

Figure 5

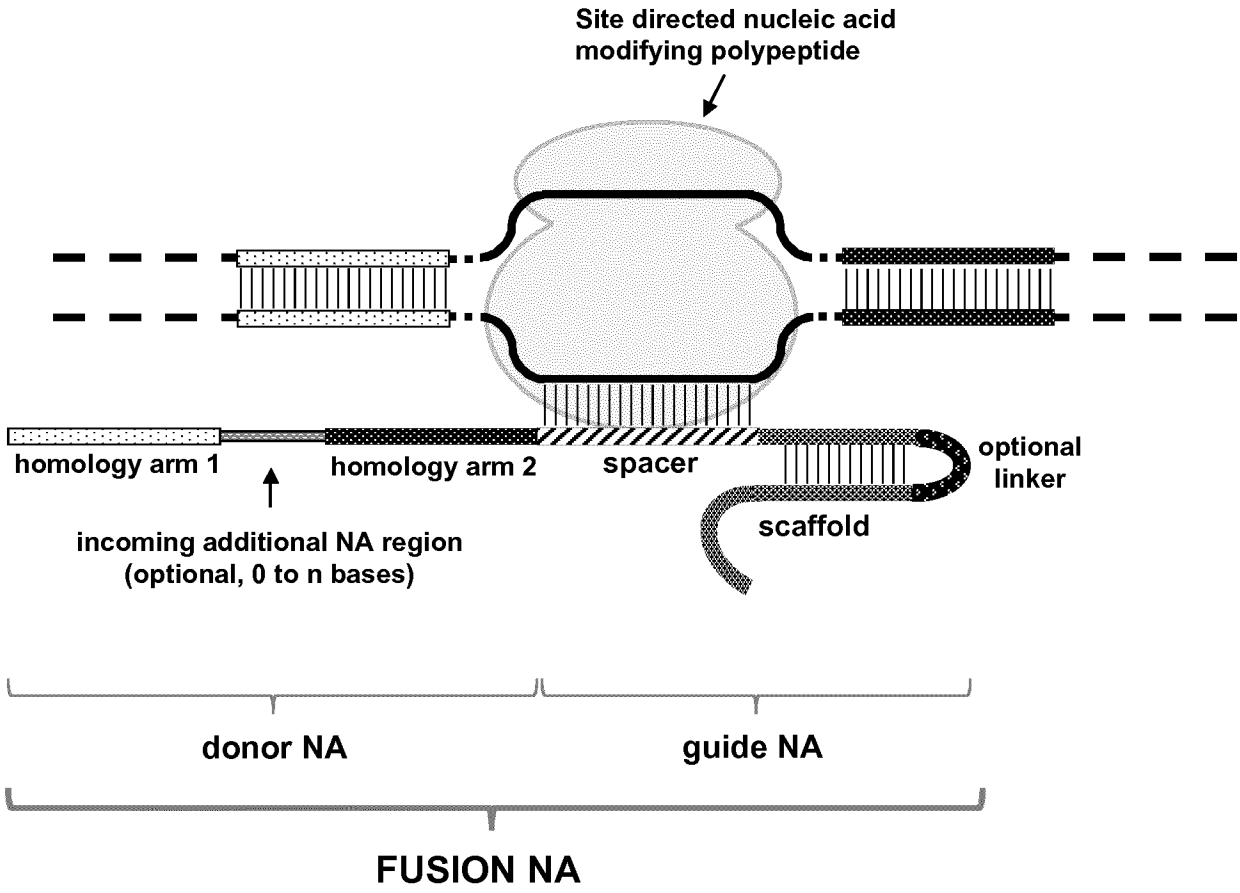

Site directed nucleic acid modifying polypeptide
homology arm 1
homology arm 2
spacer
optional linker
scaffold
incoming additional NA region (optional, 0 to n bases)
donor NA
guide NA
FUSION NA

Figure 6

5

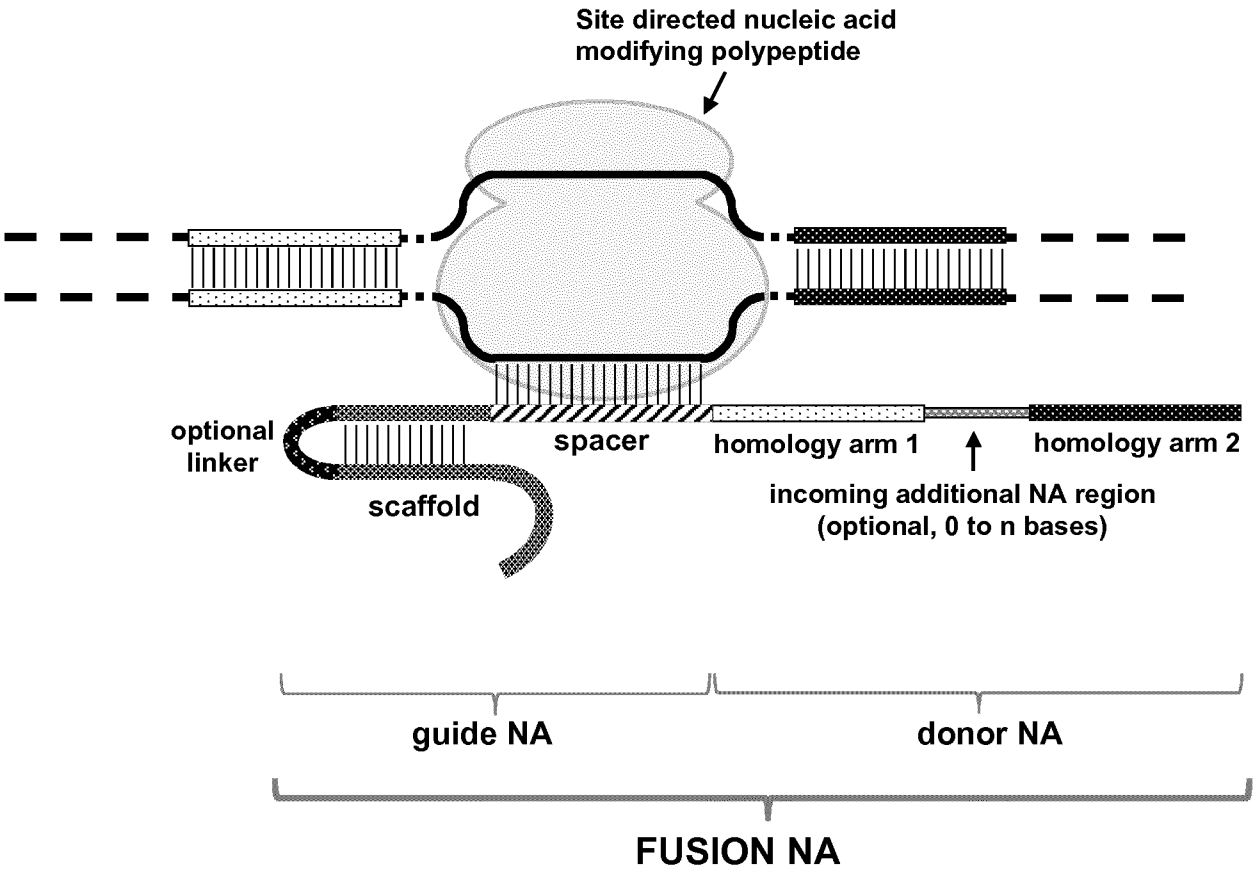

Site directed nucleic acid modifying polypeptide
optional linker
scaffold
spacer
homology arm 1
homology arm 2
incoming additional NA region (optional, 0 to n bases)
guide NA
donor NA
FUSION NA

Figure 7

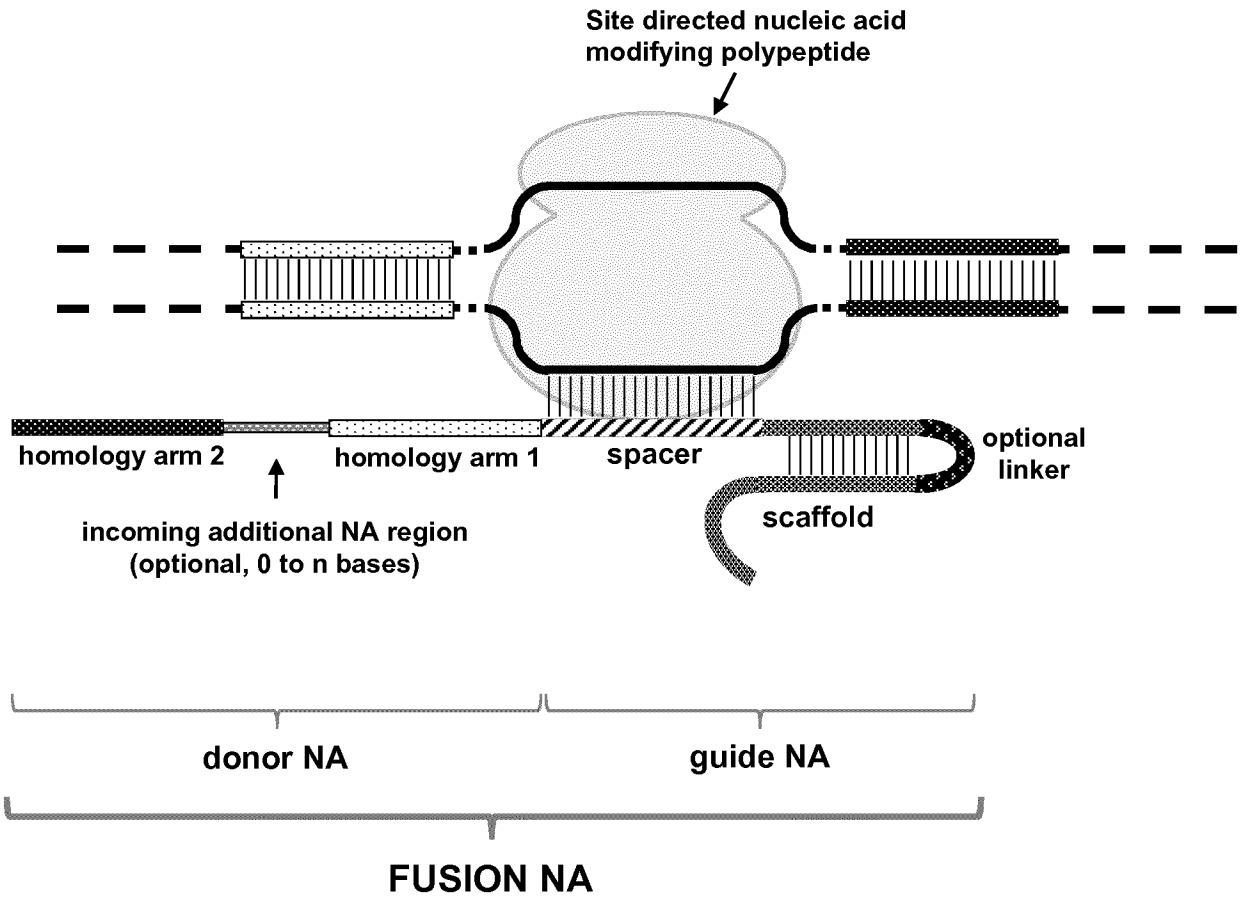

Site directed nucleic acid modifying polypeptide
homology arm 2
homology arm 1
spacer
optional linker
scaffold
incoming additional NA region (optional, 0 to n bases)
donor NA
guide NA
FUSION NA

Figure 8

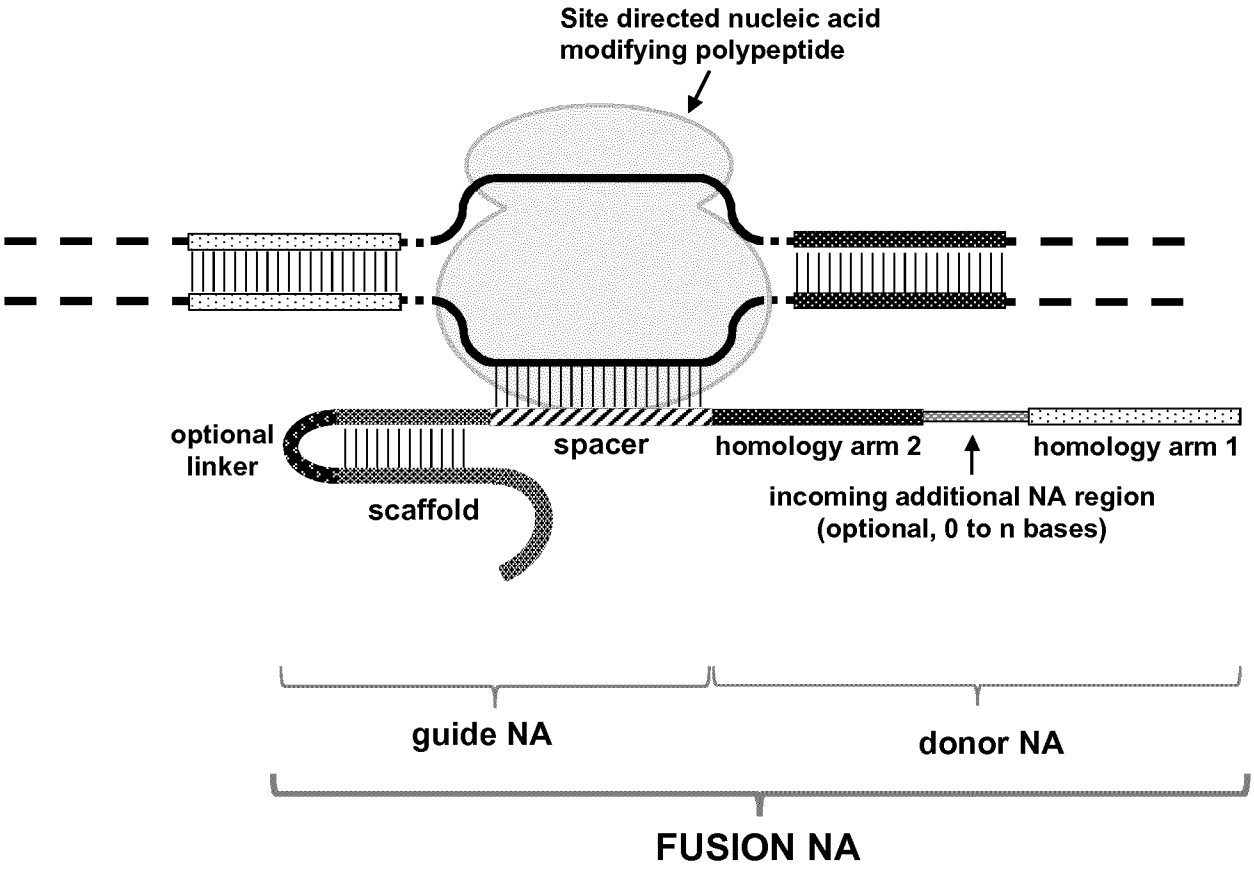

Site directed nucleic acid modifying polypeptide
optional linker
scaffold
spacer
homology arm 2
homology arm 1
incoming additional NA region (optional, 0 to n bases)
guide NA
donor NA
FUSION NA

Figure 9

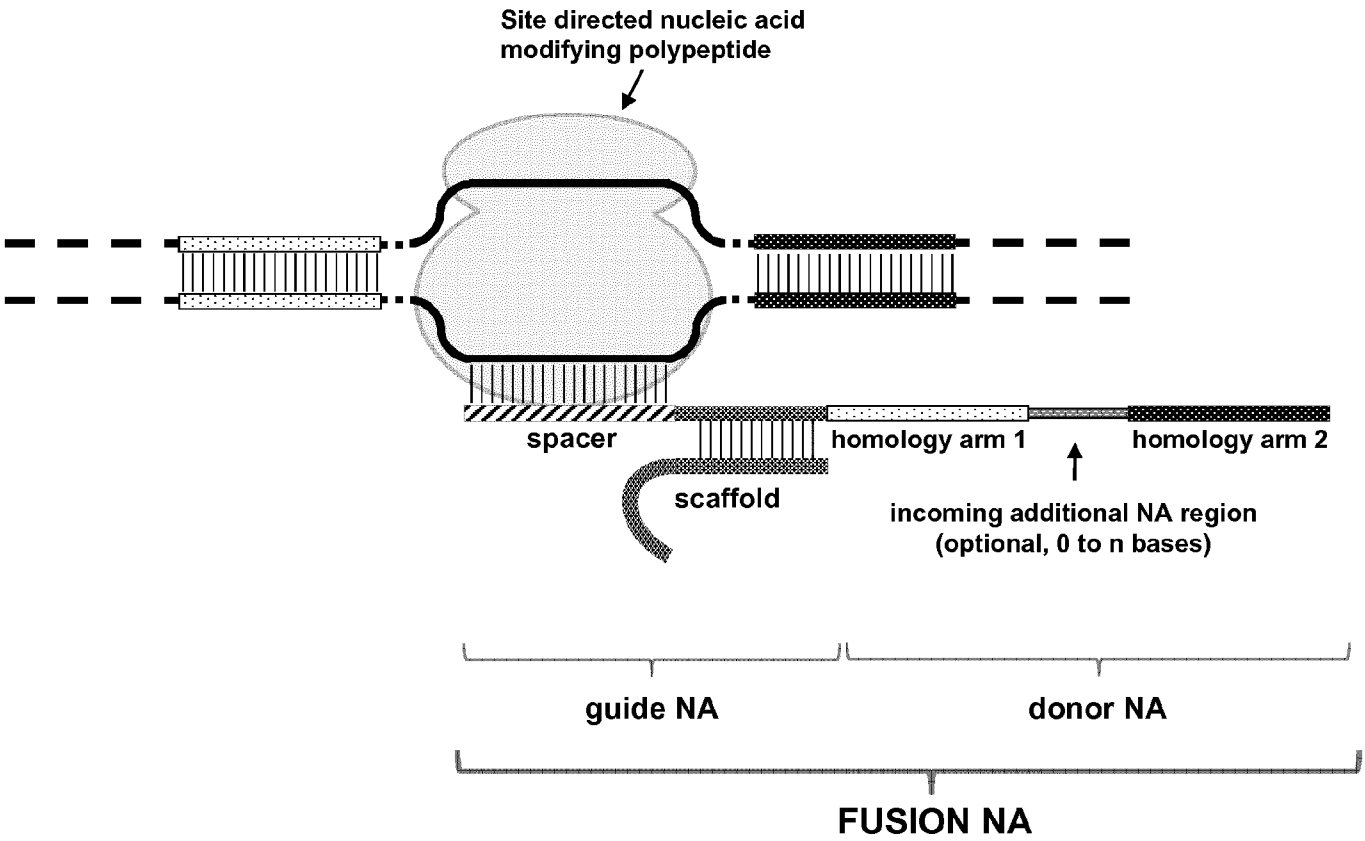

Site directed nucleic acid modifying polypeptide
spacer
homology arm 1
homology arm 2
scaffold
incoming additional NA region (optional, 0 to n bases)
guide NA
donor NA
FUSION NA

Figure 10

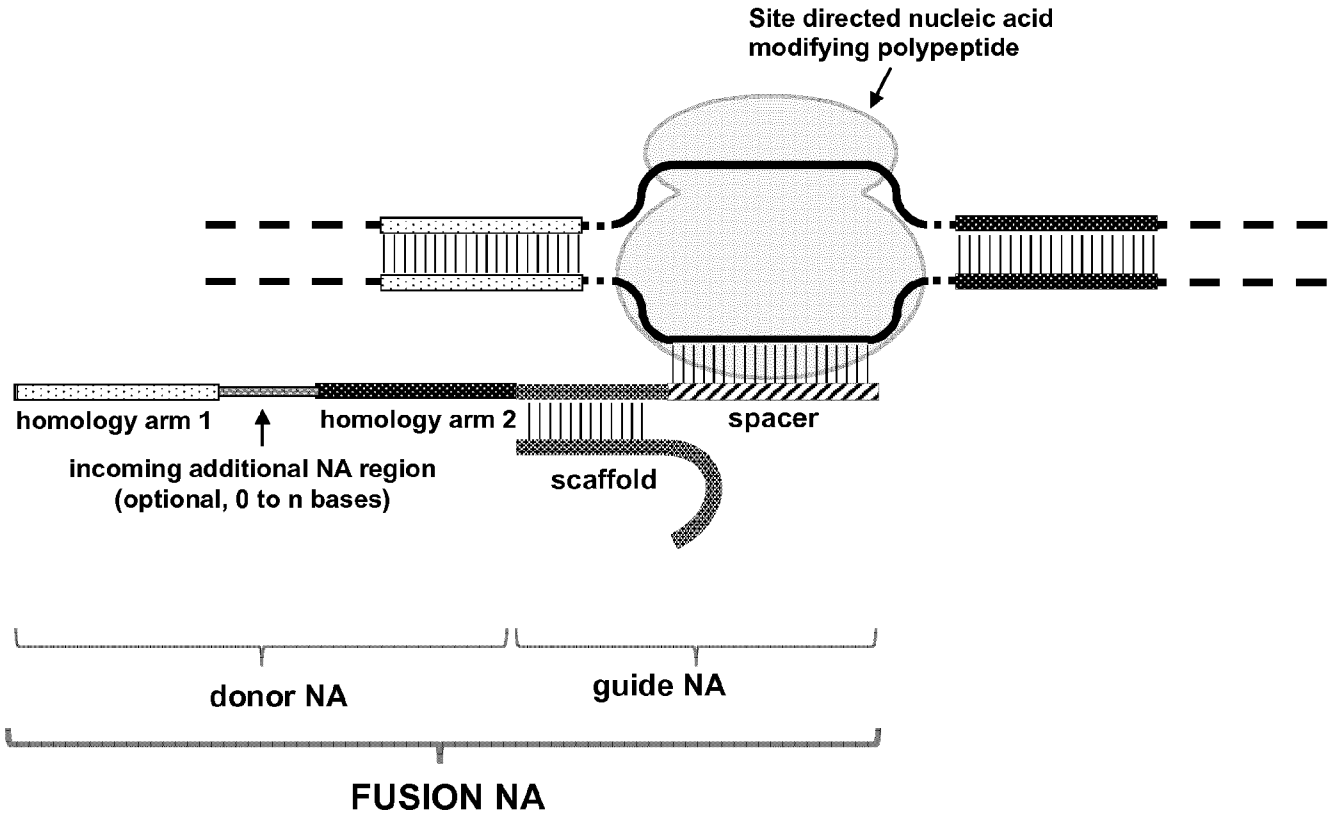

Site directed nucleic acid modifying polypeptide
homology arm 1
homology arm 2
spacer
incoming additional NA region (optional, 0 to n bases)
scaffold
donor NA
guide NA
FUSION NA

Figure 11

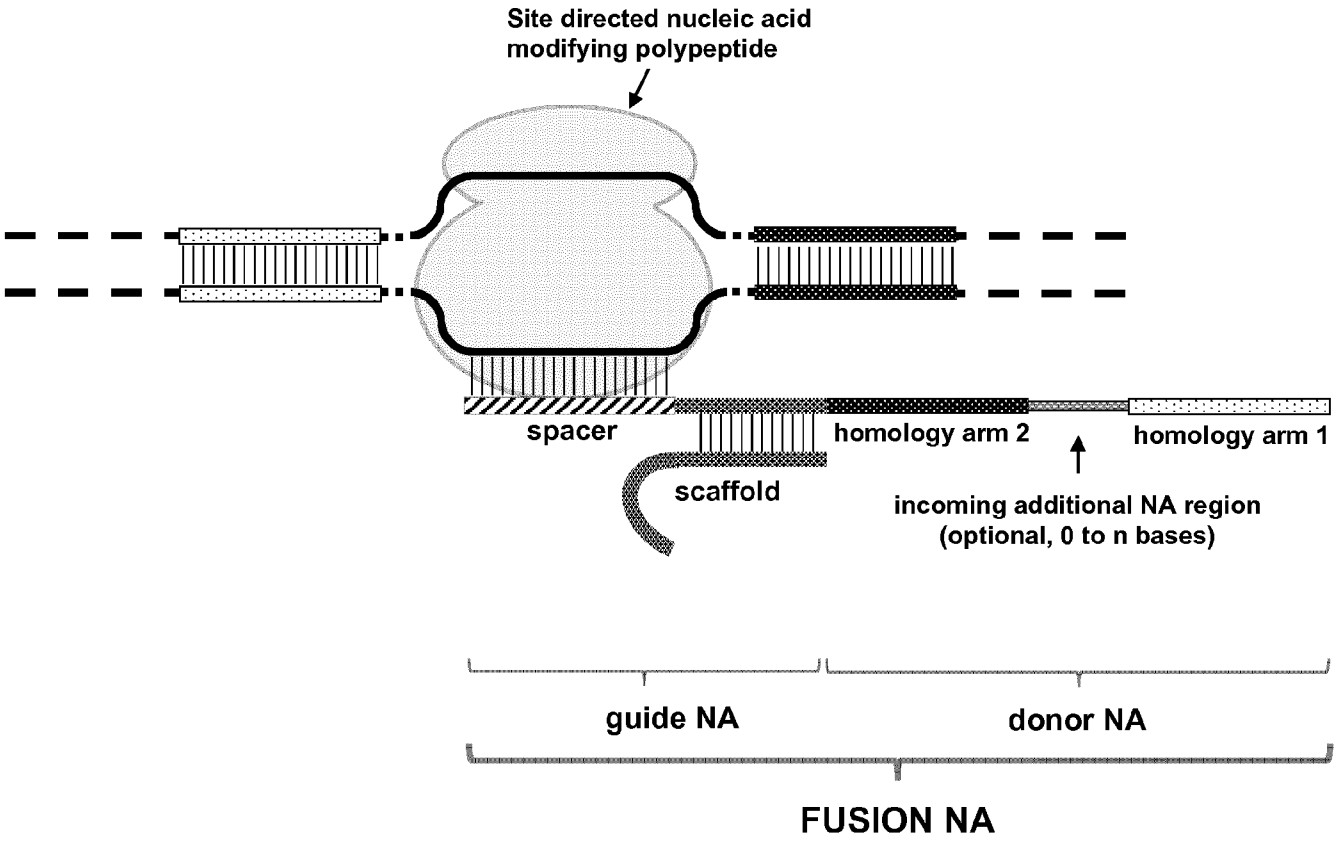

Site directed nucleic acid modifying polypeptide
spacer
scaffold
homology arm 2
homology arm 1
incoming additional NA region (optional, 0 to n bases)
guide NA
donor NA
FUSION NA

Figure 12

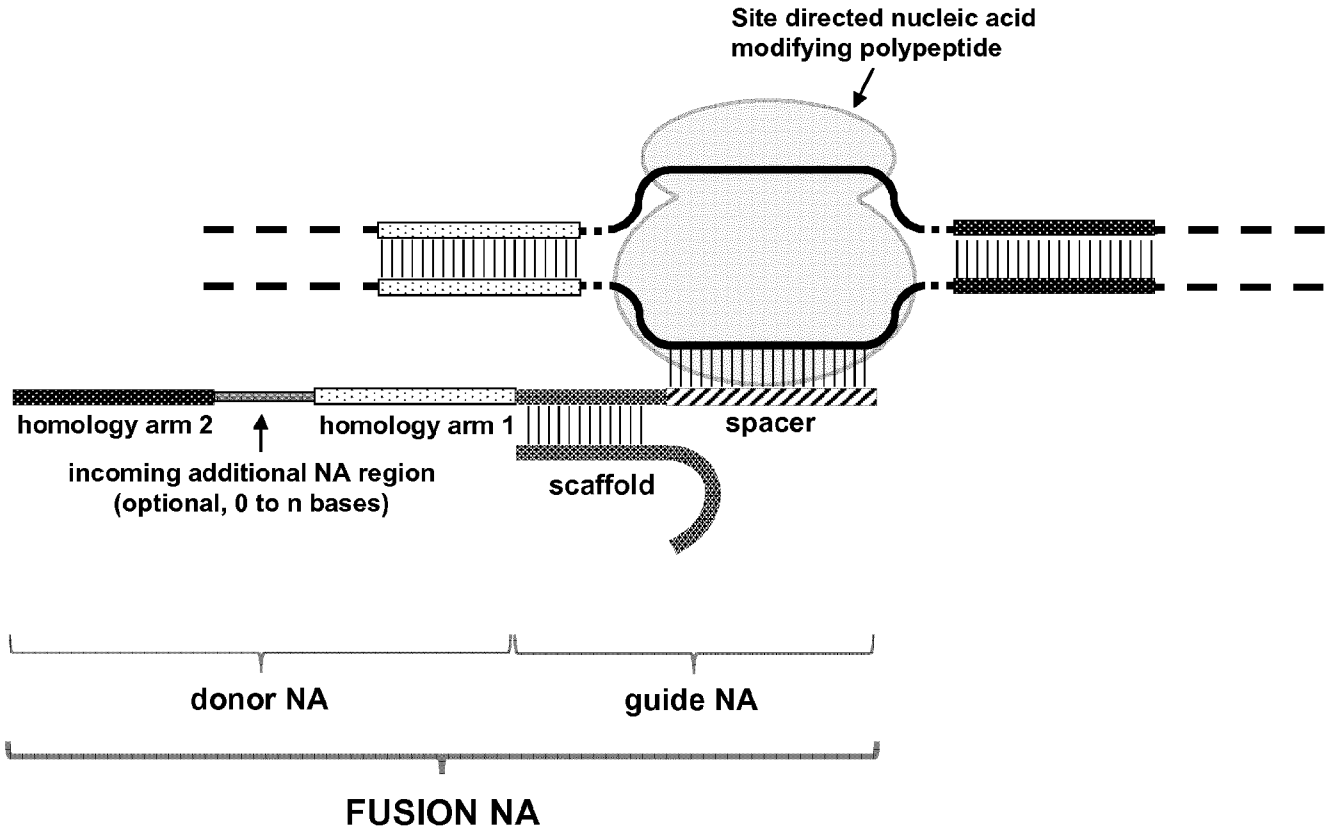

Site directed nucleic acid modifying polypeptide
homology arm 2
homology arm 1
spacer
incoming additional NA region (optional, 0 to n bases)
scaffold
donor NA
guide NA
FUSION NA

Figure 13

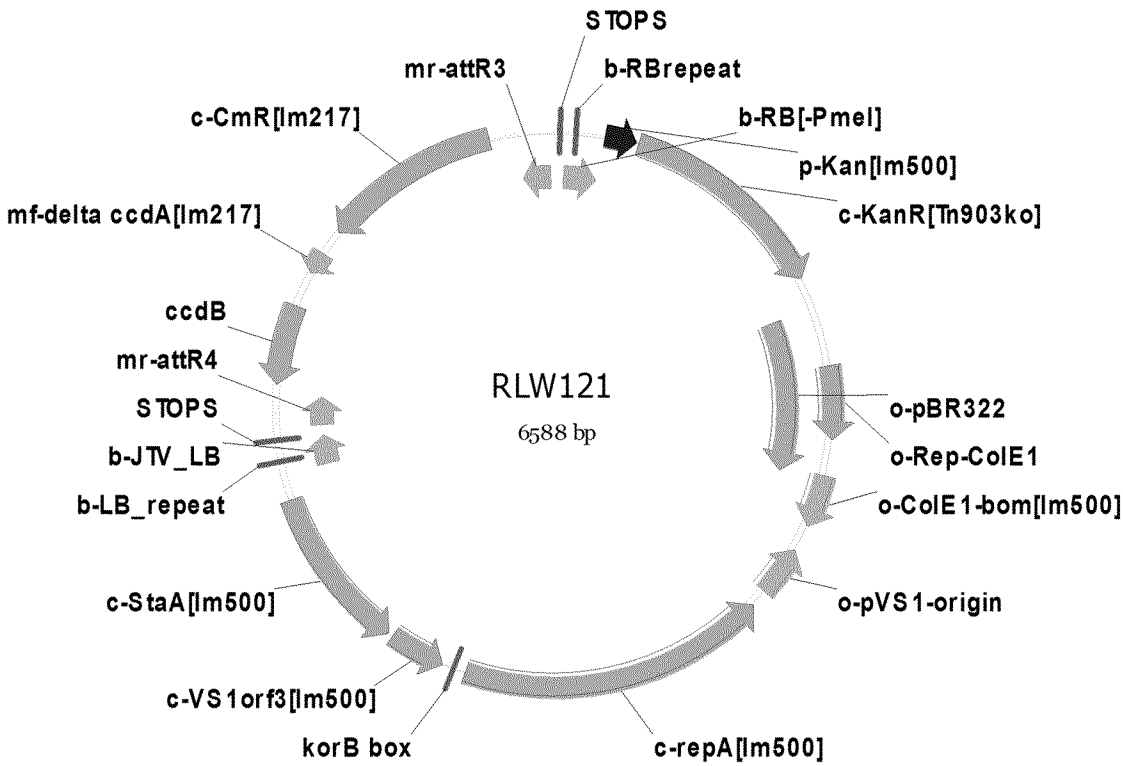

STOPS
mr-attR3
b-RBrepeat
c-CmR[Im217]
b-RB[-PmeI]
p-Kan[Im500]
mf-delta ccdA[Im217]
c-KanR[Tn903ko]
ccdB
mr-attR4
RLW121
6588 bp
STOPS
o-pBR322
b-JTV_LB
o-Rep-ColE1
b-LB_repeat
o-ColE1-bom[Im500]
c-StaA[Im500]
o-pVS1-origin
c-VS1orf3[Im500]
korB box
c-repA[Im500]

Figure 14

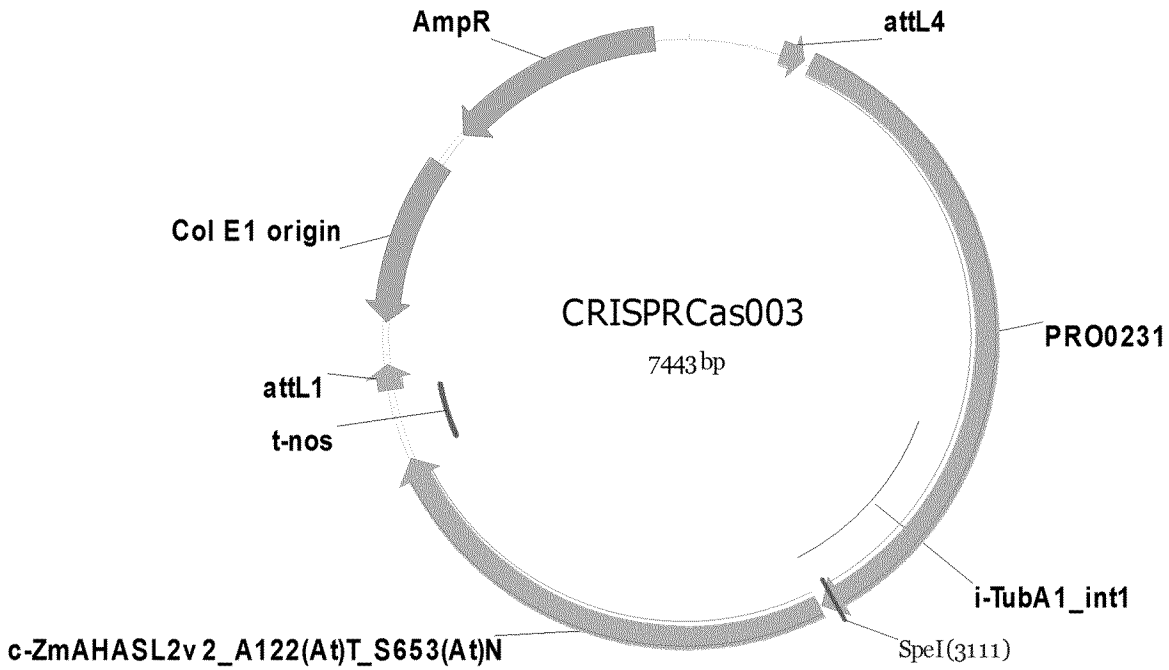

AmpR
attL4
Col E1 origin
CRISPRCas003
7443 bp
PRO0231
attL1
t-nos
i-TubA1_int1
c-ZmAHASL2v 2_A122(At)T_S653(At)N
SpeI(3111)

Figure 15

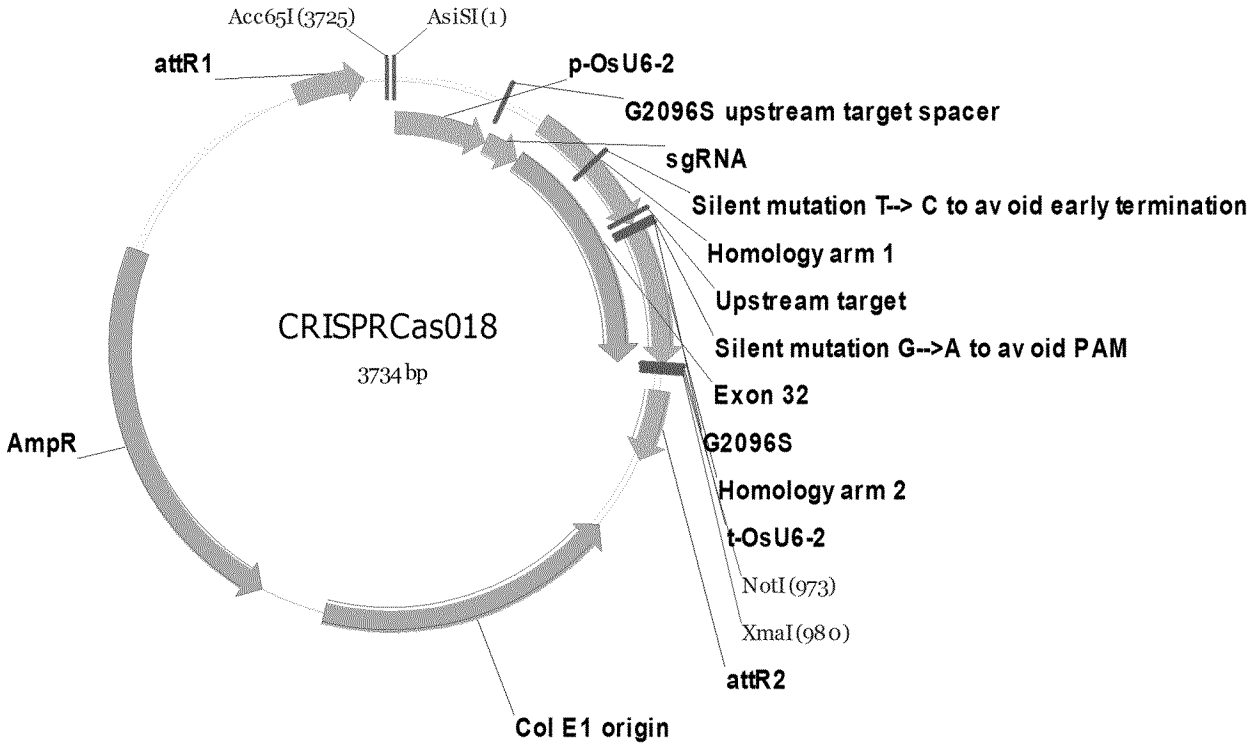

Acc65I (3725)
AsiSI (1)
attR1
p-OsU6-2
G2096S upstream target spacer
sgRNA
Silent mutation T--> C to av oid early termination
Homology arm 1
Upstream target
Silent mutation G-->A to av oid PAM
Exon 32
G2096S
Homology arm 2
t-OsU6-2
NotI (973)
XmaI (980)
attR2
CRISPRCas018
3734 bp
AmpR
Col E1 origin

Figure 16

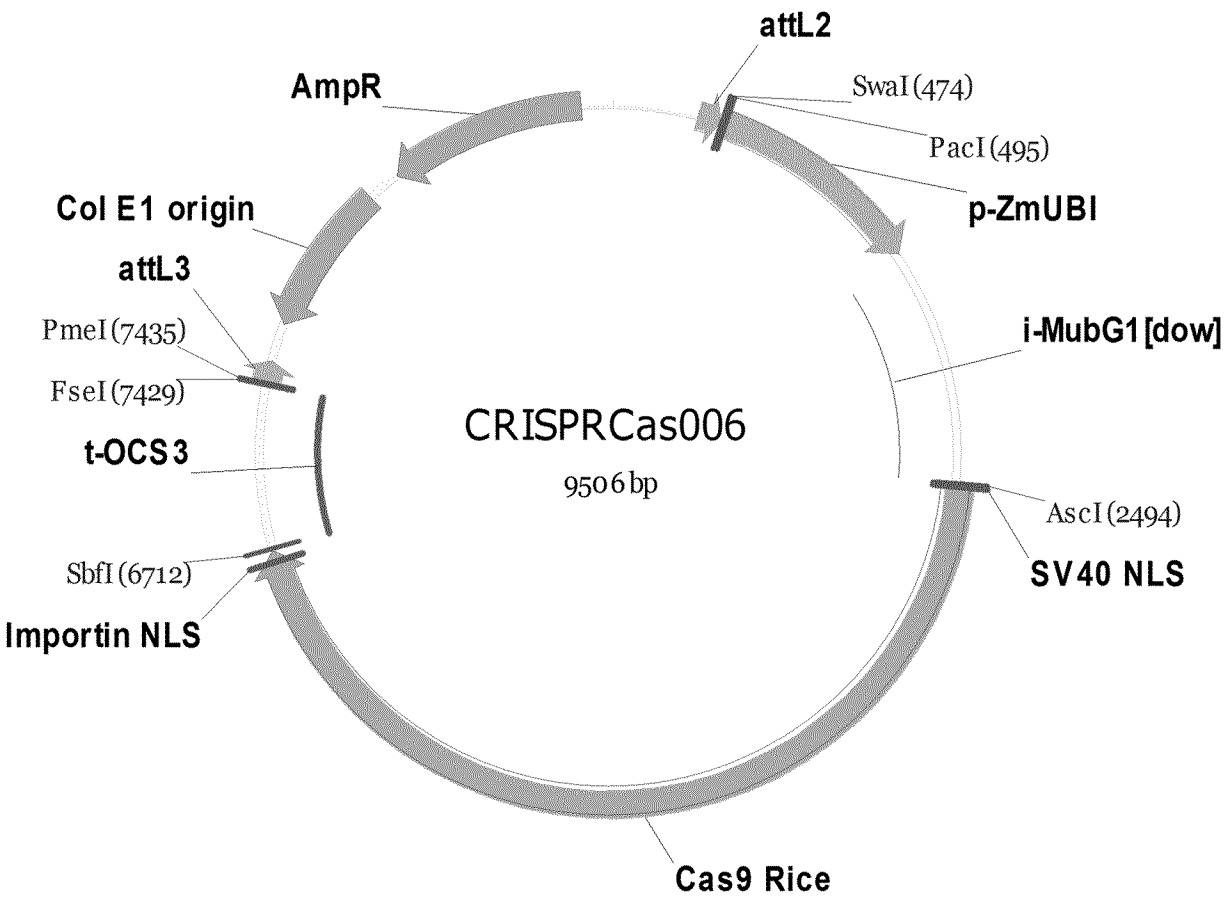

attL2
SwaI(474)
PacI(495)
p-ZmUBI
i-MubG1[dow]
AscI(2494)
SV40 NLS
Cas9 Rice
SbfI(6712)
Importin NLS
t-OCS3
FseI(7429)
PmeI(7435)
attL3
Col E1 origin
AmpR
CRISPRCas006
9506 bp

Figure 17

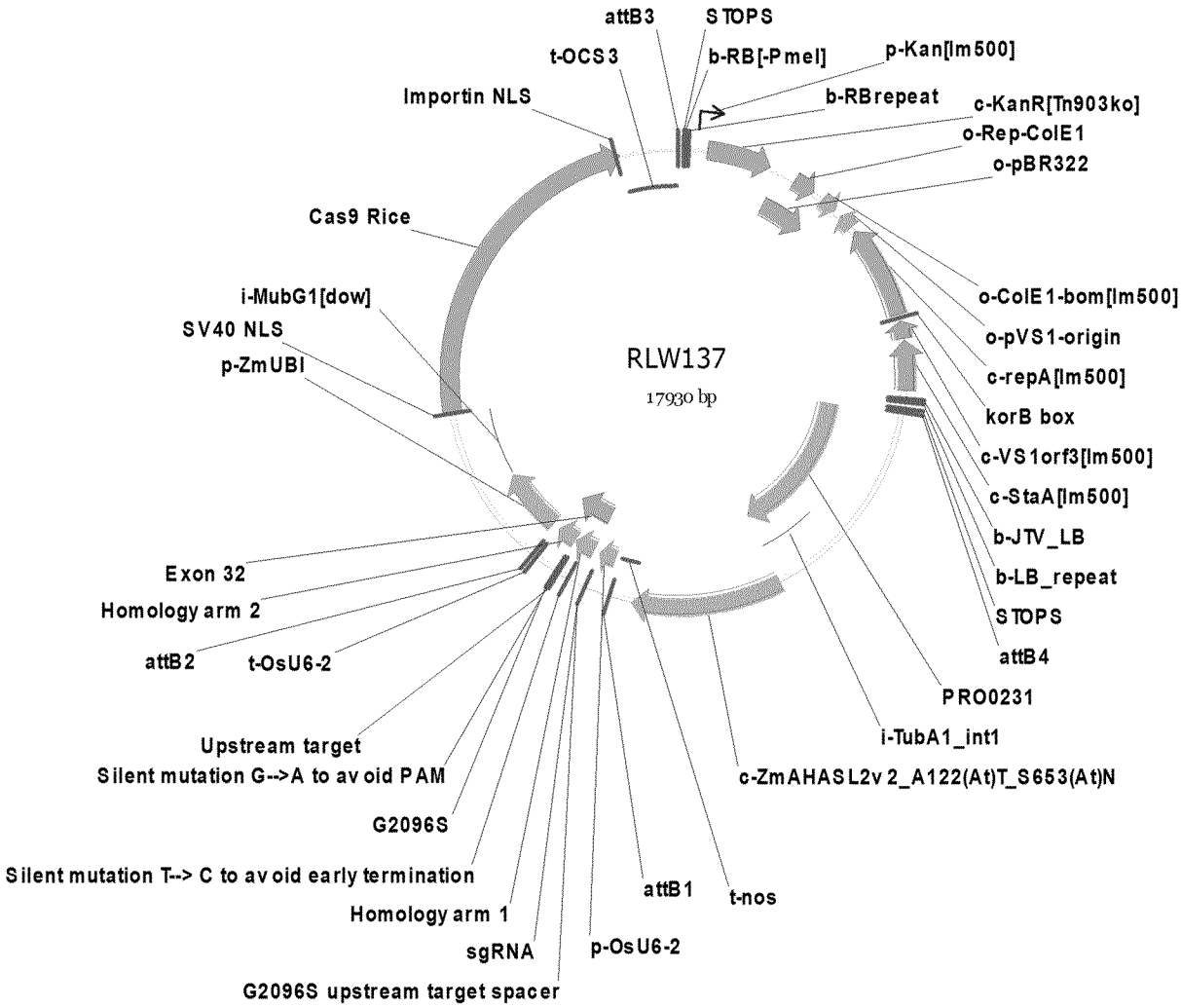

attB3
STOPS
t-OCS3
b-RB[-PmeI]
p-Kan[Im500]
Importin NLS
b-RBrepeat
c-KanR[Tn903ko]
o-Rep-ColE1
o-pBR322
Cas9 Rice
o-ColE1-bom[Im500]
i-MubG1[dow]
SV40 NLS
p-ZmUBI
o-pVS1-origin
c-repA[Im500]
RLW137
17930 bp
korB box
c-VS1orf3[Im500]
c-StaA[Im500]
b-JTV_LB
b-LB_repeat
Exon 32
Homology arm 2
attB2
t-OsU6-2
STOPS
attB4
PRO0231
i-TubA1_int1
Upstream target
Silent mutation G-->A to avoid PAM
c-ZmAHASL2v2_A122(At)T_S653(At)N
G2096S
Silent mutation T--> C to avoid early termination
attB1
t-nos
Homology arm 1
sgRNA
p-OsU6-2
G2096S upstream target spacer

Figure 18

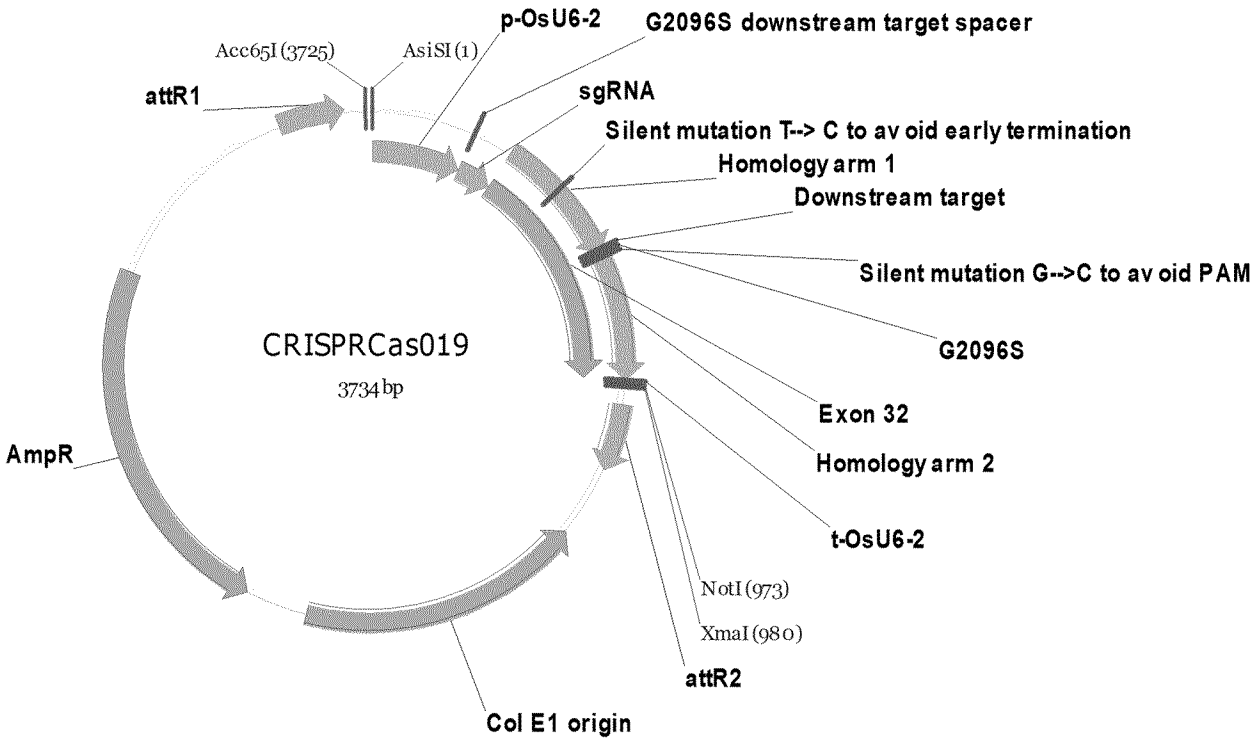

p-OsU6-2
G2096S downstream target spacer
Acc65I (3725)
AsiSI (1)
attR1
sgRNA
Silent mutation T--> C to av oid early termination
Homology arm 1
Downstream target
Silent mutation G-->C to av oid PAM
CRISPRCas019
3734 bp
G2096S
Exon 32
Homology arm 2
AmpR
t-OsU6-2
NotI (973)
XmaI (980)
attR2
Col E1 origin

Figure 19

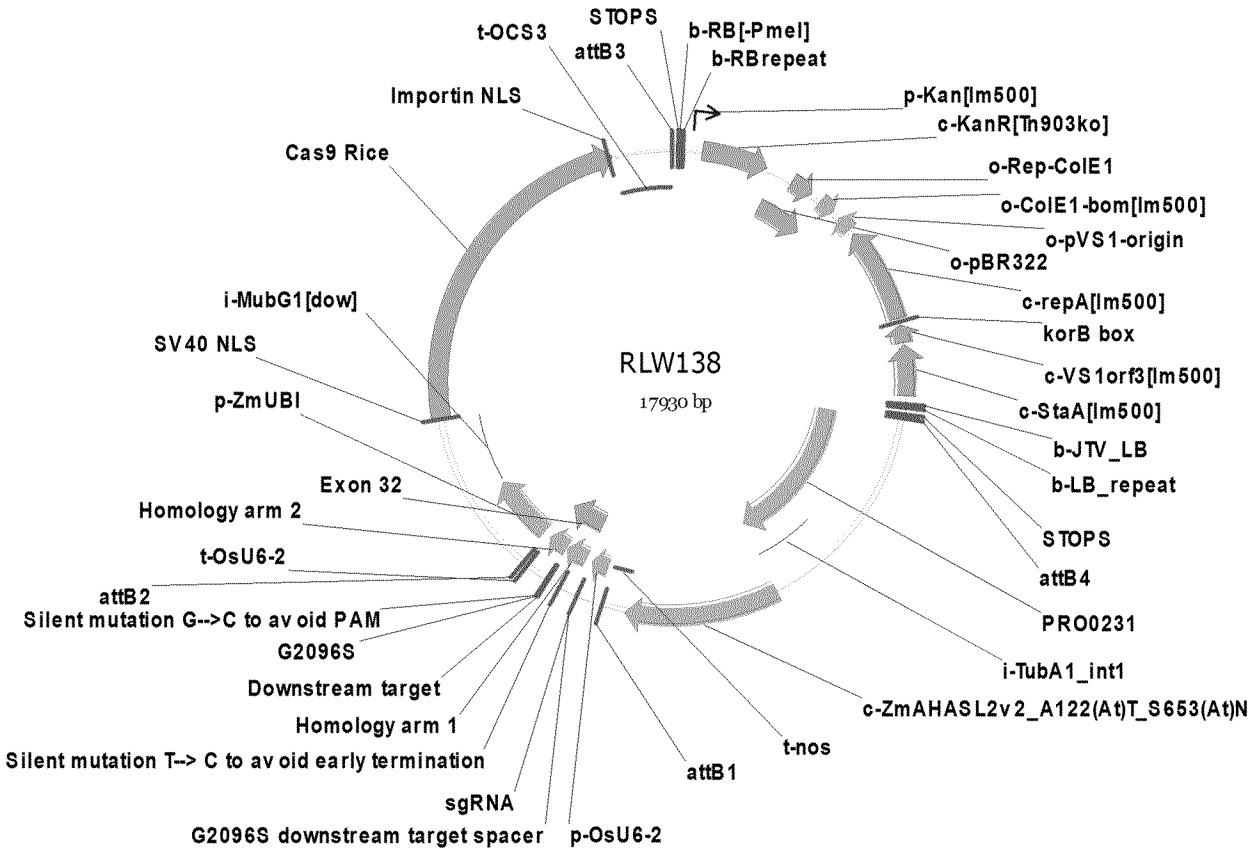

RLW138
17930 bp
t-OCS3
STOPS
b-RB[-PmeI]
attB3
b-RBrepeat
Importin NLS
p-Kan[Im500]
c-KanR[Tn903ko]
Cas9 Rice
o-Rep-ColE1
o-ColE1-bom[Im500]
o-pVS1-origin
o-pBR322
i-MubG1[dow]
c-repA[Im500]
korB box
SV40 NLS
c-VS1orf3[Im500]
p-ZmUBI
c-StaA[Im500]
b-JTV_LB
b-LB_repeat
Exon 32
Homology arm 2
STOPS
t-OsU6-2
attB4
attB2
PRO0231
Silent mutation G-->C to av oid PAM
i-TubA1_int1
G2096S
c-ZmAHASL2v2_A122(At)T_S653(At)N
Downstream target
t-nos
Homology arm 1
attB1
Silent mutation T--> C to av oid early termination
sgRNA
G2096S downstream target spacer
p-OsU6-2

Figure 20

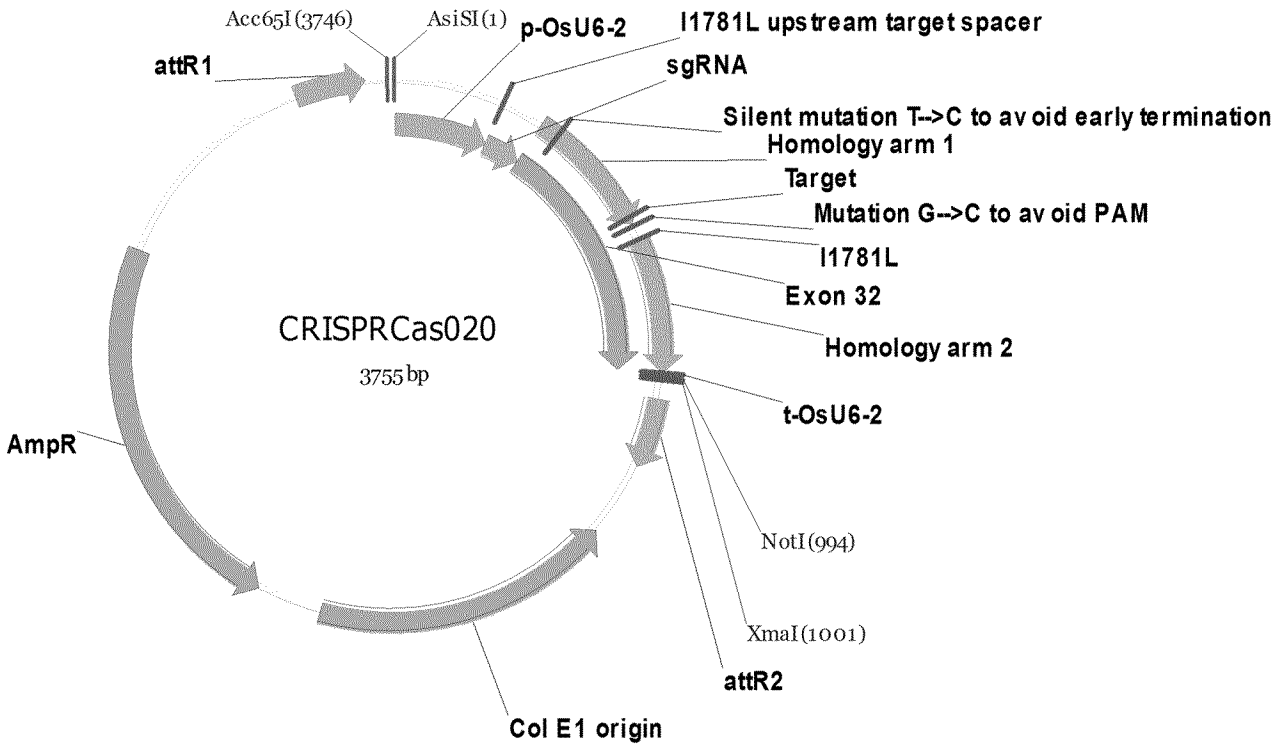

Acc65I(3746)
AsiSI(1)
p-OsU6-2
I1781L upstream target spacer
attR1
sgRNA
Silent mutation T-->C to avoid early termination
Homology arm 1
Target
Mutation G-->C to avoid PAM
I1781L
Exon 32
CRISPRCas020
3755 bp
Homology arm 2
t-OsU6-2
AmpR
NotI(994)
XmaI(1001)
attR2
Col E1 origin

Figure 21

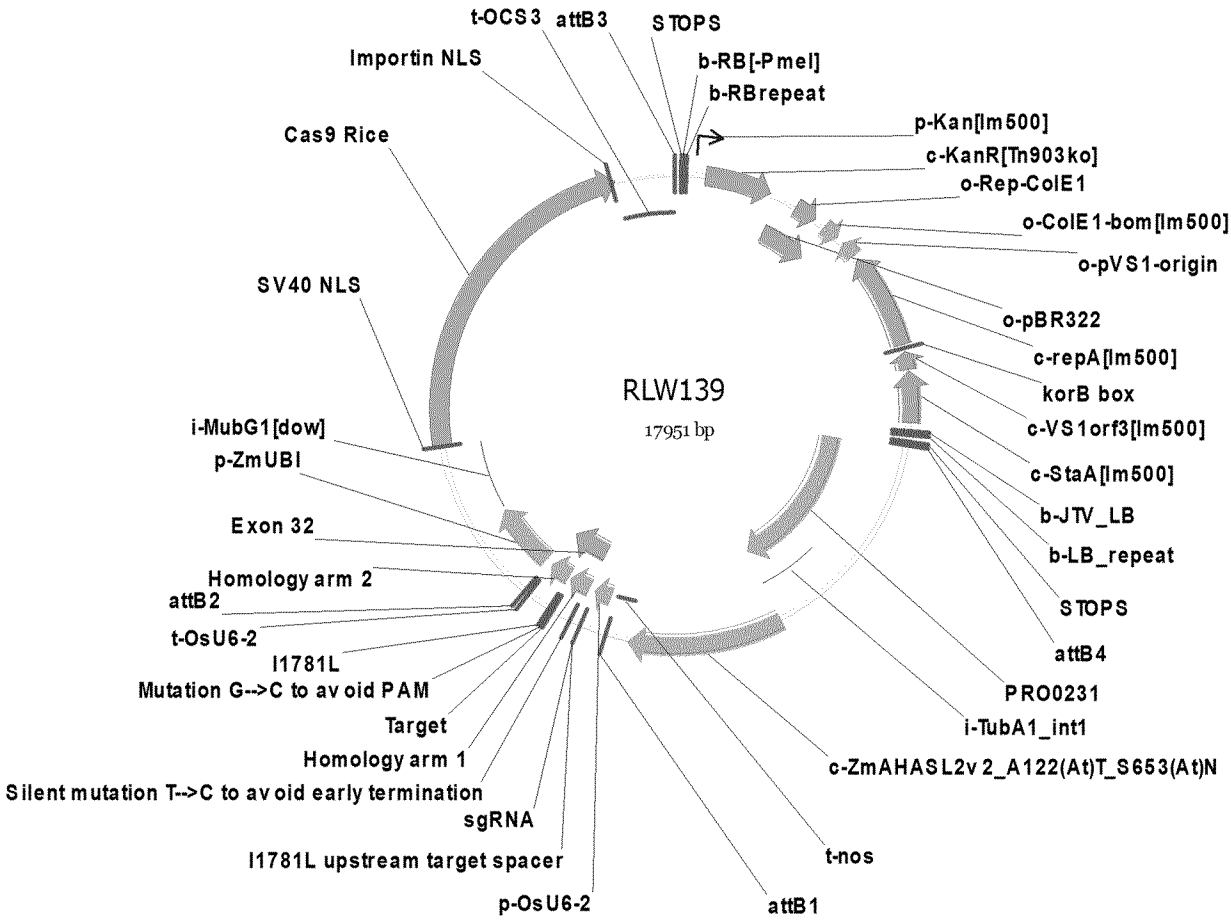

RLW139
17951 bp
t-OCS3
attB3
STOPS
Importin NLS
b-RB[-PmeI]
b-RBrepeat
p-Kan[Im500]
Cas9 Rice
c-KanR[Tn903ko]
o-Rep-ColE1
o-ColE1-bom[Im500]
o-pVS1-origin
SV40 NLS
o-pBR322
c-repA[Im500]
korB box
i-MubG1[dow]
c-VS1orf3[Im500]
p-ZmUBI
c-StaA[Im500]
b-JTV_LB
Exon 32
b-LB_repeat
Homology arm 2
STOPS
attB2
attB4
t-OsU6-2
PRO0231
I1781L
i-TubA1_int1
Mutation G-->C to avoid PAM
c-ZmAHASL2v2_A122(At)T_S653(At)N
Target
Homology arm 1
Silent mutation T-->C to avoid early termination
sgRNA
t-nos
I1781L upstream target spacer
p-OsU6-2
attB1

Figure 22

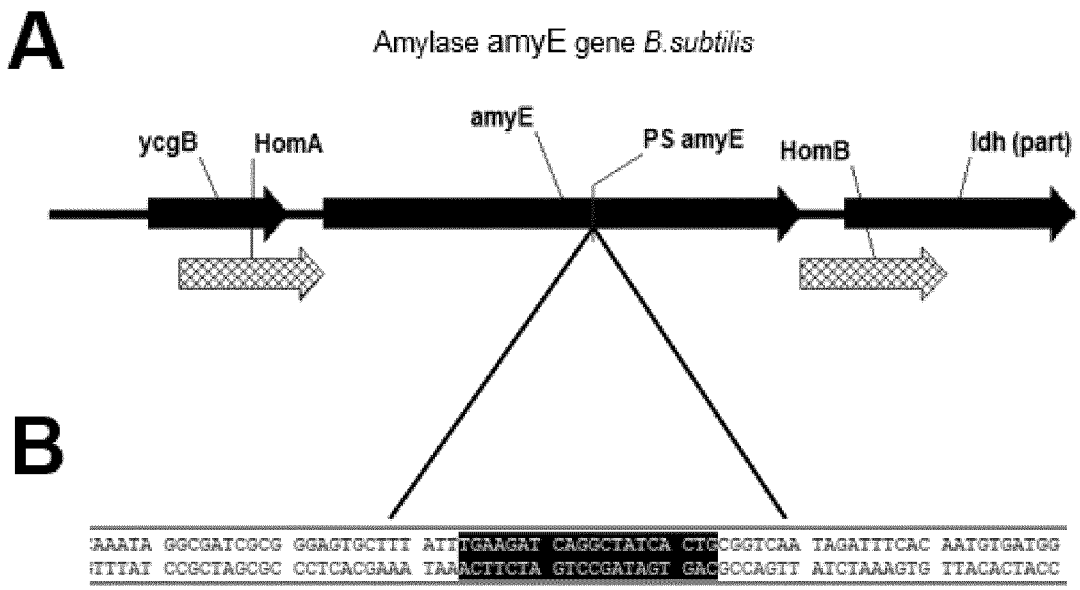

A
Amylase amyE gene B.subtilis
ycgB
HomA
amyE
PS amyE
HomB
ldh (part)
B

Figure 23

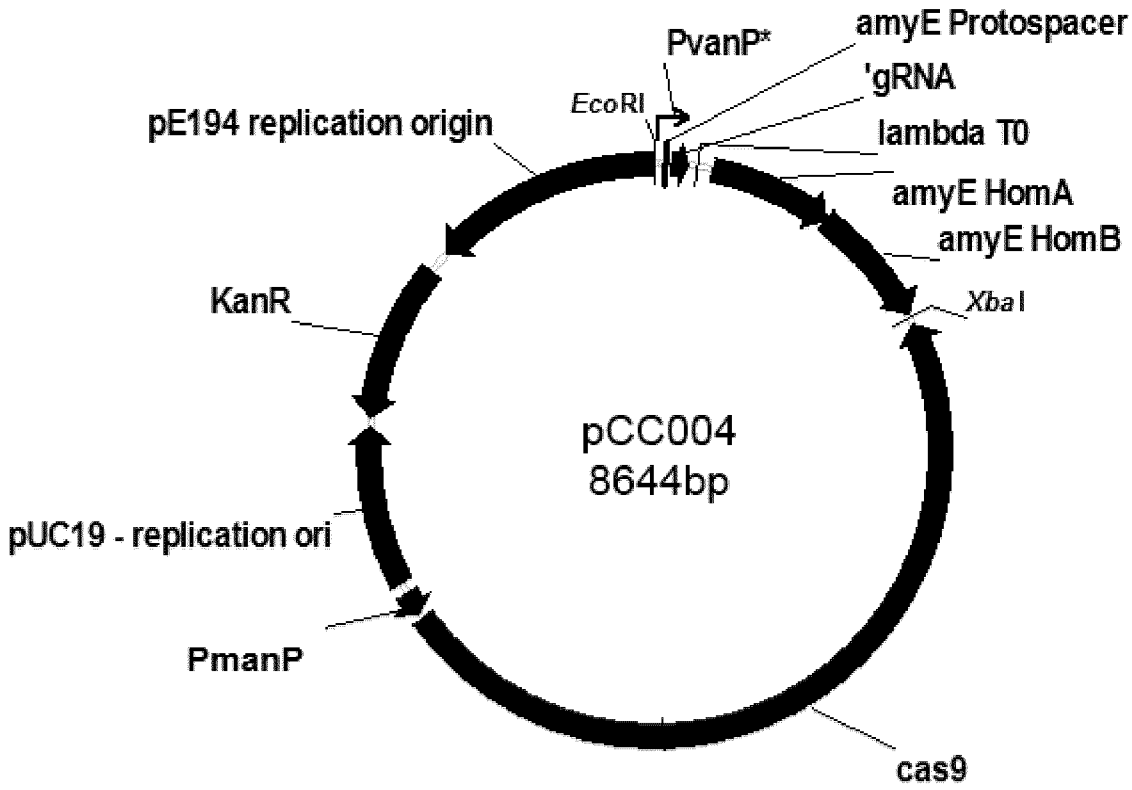

PvanP*
amyE Protospacer
'gRNA
EcoRI
pE194 replication origin
lambda T0
amyE HomA
amyE HomB
KanR
XbaI
pCC004
8644bp
pUC19 - replication ori
PmanP
cas9

Figure 24

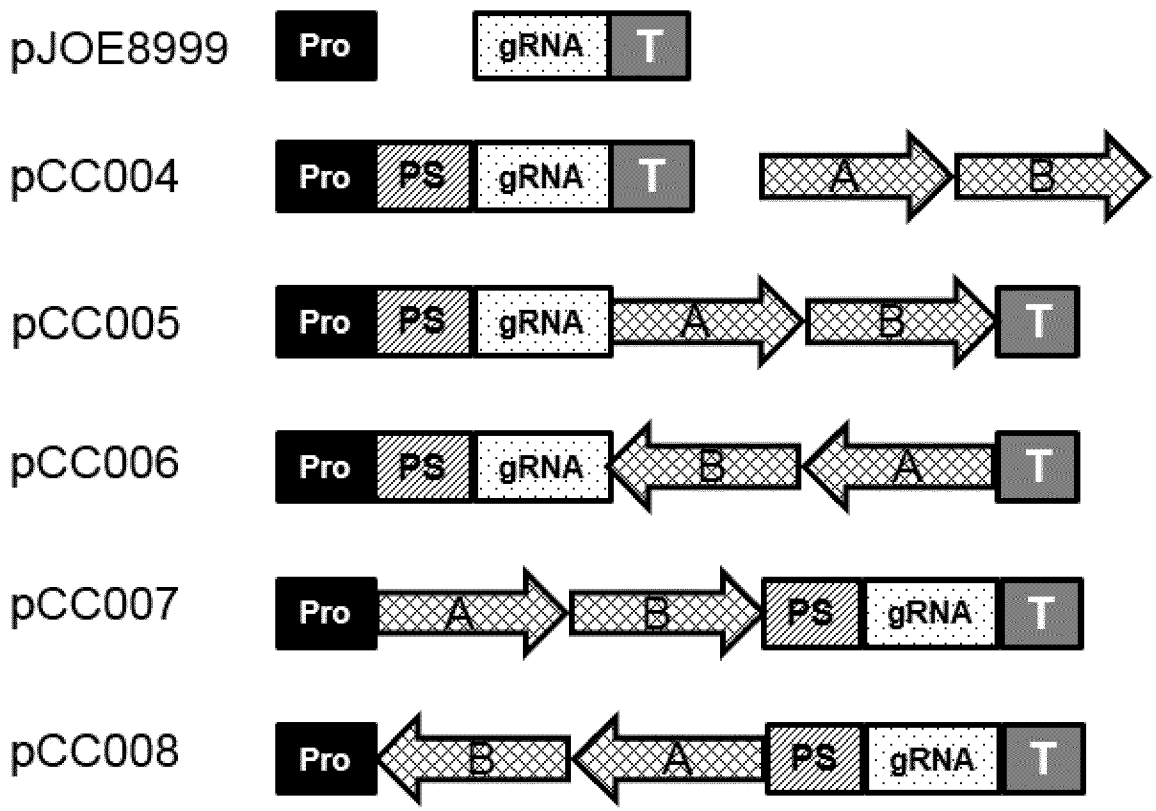

pJOE8999
Pro
gRNA
T
pCC004
Pro
PS
gRNA
T
A
B
pCC005
Pro
PS
gRNA
A
B
T
pCC006
Pro
PS
gRNA
B
A
T
pCC007
Pro
A
B
PS
gRNA
T
pCC008
Pro
B
A
PS
gRNA
T

Figure 25

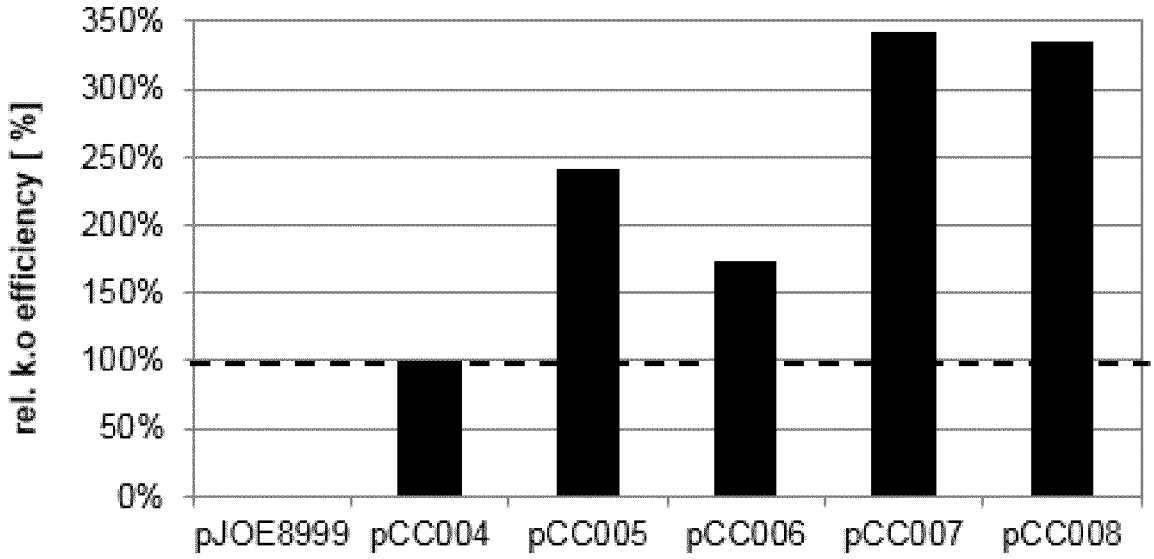
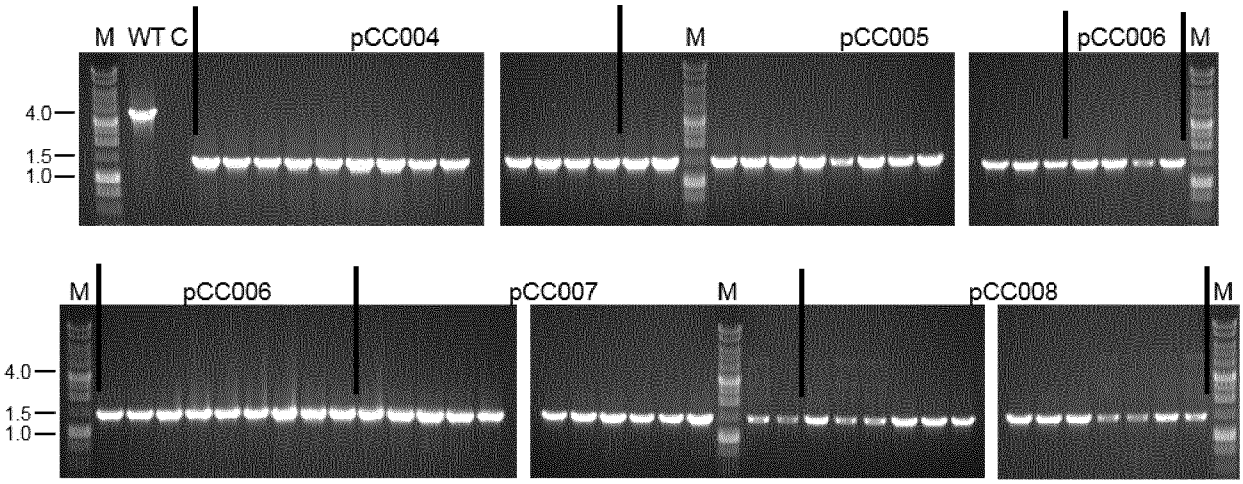

rel. k.o efficiency [%]
350%
300%
250%
200%
150%
100%
50%
0%
pJOE8999
pCC004
pCC005
pCC006
pCC007
pCC008

Figure 26


M
WT C
pCC004
M
pCC005
pCC006
M
4.0
1.5
1.0
M
pCC006
pCC007
M
pCC008
M
4.0
1.5
1.0