



(86) Date de dépôt PCT/PCT Filing Date: 2015/12/16
(87) Date publication PCT/PCT Publication Date: 2016/06/23
(85) Entrée phase nationale/National Entry: 2017/06/15
(86) N° demande PCT/PCT Application No.: US 2015/066192
(87) N° publication PCT/PCT Publication No.: 2016/100568
(30) Priorités/Priorities: 2014/12/16 (CN PCT/CN2014/093918);
2014/12/16 (CN PCT/CN2014/093916);
2014/12/16 (CN PCT/CN2014/093914)

(51) Cl.Int./Int.Cl. *C12N 15/113* (2010.01),
C12N 1/15 (2006.01), *C12N 15/09* (2006.01),
C12N 15/10 (2006.01), *C12N 15/11* (2006.01),
C12N 15/80 (2006.01), *C12N 15/90* (2006.01),
C12N 9/22 (2006.01)

(71) Demandeur/Applicant:
DANISCO US INC, US

(72) Inventeurs/Inventors:
BOWER, BENJAMIN S., US;
CHAN, JIMMY, US;
GE, JING, US;
GU, XIAOGANG, US;
KIM, STEVEN SUNGJIN, US;

(54) Titre : SYSTEMES DE MODIFICATION DU GENOME DE CHAMPIGNONS ET PROCEDES D'UTILISATION
DESDITS SYSTEMES

(54) Title: FUNGAL GENOME MODIFICATION SYSTEMS AND METHODS OF USE

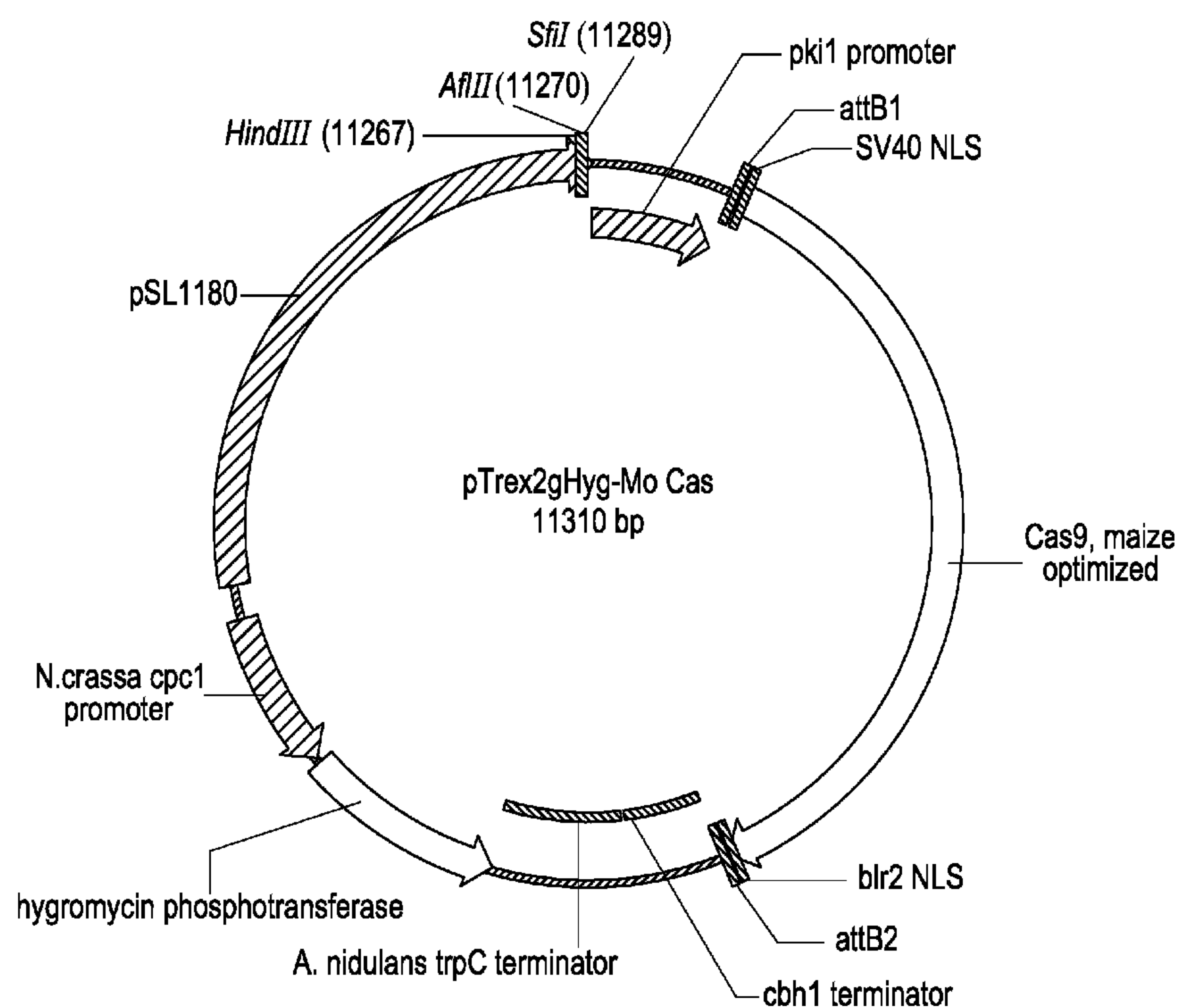


FIG. 2

(57) **Abrégé/Abstract:**

Compositions and methods are provided for genome modification at a target site in the genome of a filamentous fungal cell. The methods and compositions are drawn to a guide polynucleotide/ Cas endonuclease system for promoting modification of the DNA sequence at a target site in a filamentous fungal host cell genome.



(21) **2 971 247**

(13) **A1**

(72) **Inventeurs(suite)/Inventors(continued)**: MADRID, SUSAN MAMPUSTI, US; SONG, DANFENG, US;
SONG, MINGMIN, US; WARD, MICHAEL, US

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

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
23 June 2016 (23.06.2016)



(10) International Publication Number
WO 2016/100568 A1

- (51) **International Patent Classification:**
C12N 15/80 (2006.01) *C12N 9/22* (2006.01)
- (21) **International Application Number:**
PCT/US2015/066192
- (22) **International Filing Date:**
16 December 2015 (16.12.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PCT/CN2014/093914
16 December 2014 (16.12.2014) CN
PCT/CN2014/093916
16 December 2014 (16.12.2014) CN
PCT/CN2014/093918
16 December 2014 (16.12.2014) CN
- (71) **Applicant:** **DANISCO US INC** [US/US]; 925 Page Mill Road, Palo Alto, California 94304 (US).
- (72) **Inventors:** **BOWER, Benjamin S.**; 925 Page Mill Road, Palo Alto, California 94304 (US). **CHAN, Jimmy**; 925 Page Mill Road, Palo Alto, California 94304 (US). **GE,**

Jing; 925 Page Mill Road, Palo Alto, California 94304 (US). **GU, Xiaogang**; 925 Page Mill Road, Palo Alto, California 94304 (US). **KIM, Steven Sungjin**; 925 Page Mill Road, Palo Alto, California 94304 (US). **MADRID, Susan Mampusti**; 925 Page Mill Road, Palo Alto, California 94304 (US). **SONG, Danfeng**; 925 Page Mill Road, Palo Alto, California 94304 (US). **SONG, Mingmin**; 925 Page Mill Road, Palo Alto, California 94304 (US). **WARD, Michael**; 925 Page Mill Road, Palo Alto, California 94304 (US).

(74) **Agent:** **HONG, Mei**; DANISCO US INC, 925 Page Mill Road, Palo Alto, California 94304 (US).

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

[Continued on next page]

(54) **Title:** FUNGAL GENOME MODIFICATION SYSTEMS AND METHODS OF USE

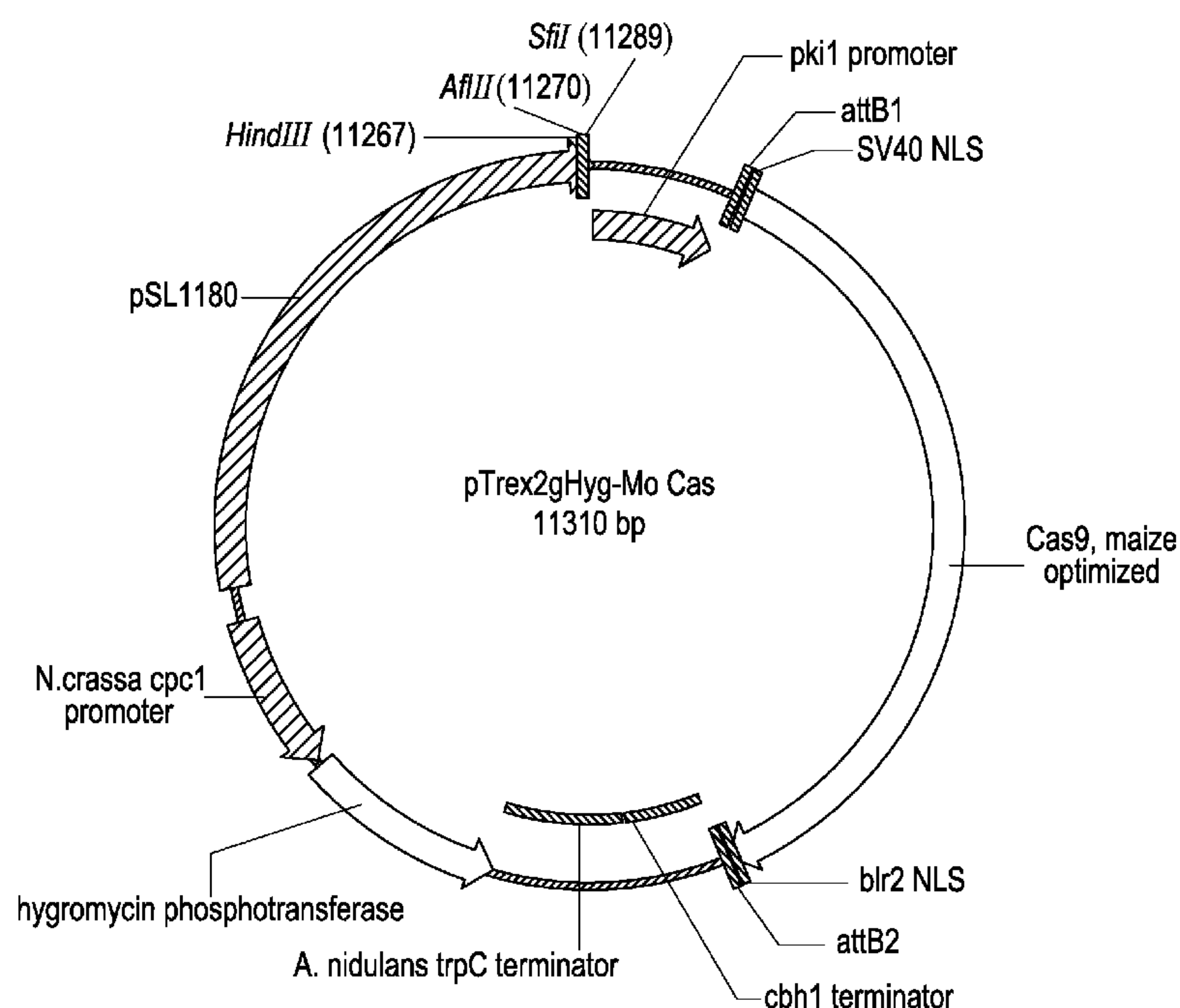


FIG. 2

(57) **Abstract:** Compositions and methods are provided for genome modification at a target site in the genome of a filamentous fungal cell. The methods and compositions are drawn to a guide polynucleotide/ Cas endonuclease system for promoting modification of the DNA sequence at a target site in a filamentous fungal host cell genome.

WO 2016/100568 A1

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

 - *with international search report (Art. 21(3))*
 - *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
 - *with sequence listing part of description (Rule 5.2(a))*

FUNGAL GENOME MODIFICATION SYSTEMS AND METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATIONS

5 **[01]** The present application claims priority to PCT Patent Appln. Ser. Nos. PCT/CN2014/093918, PCT/CN2014/093916, and PCT/CN2014/093914, all filed December 16, 2014, which are hereby incorporated by reference in their entireties.

SEQUENCE LISTING

10 **[02]** The sequence listing submitted via EFS, in compliance with 37 C.F.R. §1.52(e), is incorporated herein by reference. The sequence listing text file submitted via EFS contains the file "40532-WO-PCT-6_2015-868_Final_ST25.txt" created on December 13, 2015, which is 151 kilobytes in size.

BACKGROUND

15 **[03]** Bacteria and archaea have evolved adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems that can introduce double strand breaks in DNA in a sequence-specific manner. Cas systems perform their functions through the activity of a ribonucleoprotein complex
20 that includes short RNA sequences (tracrRNA and crRNA) and an RNA dependent endonuclease (Cas endonuclease) that targets a specific DNA sequence (through homology to a portion of the crRNA, called the variable targeting domain) and generates double strand breaks in the target. CRISPR loci were first recognized in *E. coli* (Ishino et al. (1987) *J. Bacteriol.* 169:5429-5433; Nakata et al. (1989) *J. Bacteriol.*
25 171:3553-3556), with similar interspersed short sequence repeats being subsequently identified in a number of bacterial species, including but not limited to *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, and *Mycobacterium tuberculosis* (Groenen et al. (1993) *Mol. Microbiol.* 10:1057-1065; Hoe et al. (1999) *Emerg. Infect. Dis.* 5:254-263; Masepohl et al. (1996) *Biochim. Biophys. Acta* 1307:26-30; Mojica et al.
30 (1995) *Mol. Microbiol.* 17:85-93).

[04] It is well known that inducing cleavage at a specific target site in genomic DNA can be used to introduce modifications at or near that site. For example, homologous recombination for gene targeting has been shown to be enhanced when the targeted

DNA site contains a double-strand break (see, e.g., Rudin et al., Genetics 122:519-534; Smih et al., Nucl. Acids Res. 23:5012-5019). Given the site-specific nature of Cas systems, genome modification/engineering technologies based on these systems have been described, including in mammalian cells (see, e.g., Hsu et al.; Cell vol. 157, p1262–1278, 5 June 2014 entitled “Development and Applications of CRISPR-Cas9 for Genome Engineering”). The power of the Cas-based genome engineering comes from the ability to target virtually any specific location within a complex genome by designing a recombinant crRNA (or equivalently functional polynucleotide) in which the DNA-targeting region (variable targeting domain) of the crRNA is homologous to the desired target site in the genome and combining it with a Cas endonuclease (through any convenient means) into a functional complex in a host cell.

[05] Although Cas-based genome engineering technologies have been applied to a number of different host cell types, the efficient use of such systems in fungal cells has proven to be difficult. Thus, there still remains a need for developing efficient and effective Cas-based genome engineering methods and compositions for modifying/altering a genomic target site in a fungal cell.

BRIEF SUMMARY

[06] Compositions and methods are provided that relate to employing a guide RNA/Cas endonuclease system for modifying the DNA sequence at a target site in the genome of a fungal cell, e.g., a filamentous fungal cell.

[07] Aspects of the present disclosure are drawn to methods for modifying the DNA sequence at a target site in the genome of a fungal cell. In some embodiments, the method includes: a) introducing into a population of fungal cells a Cas endonuclease and a guide RNA, wherein the Cas endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cells; and b) identifying at least one fungal cell from the population that has a modification of the DNA sequence at the target site, where the Cas endonuclease, the guide RNA, or both are introduced transiently into the population of fungal cells.

[08] In one aspect, the present disclosure are drawn to a method for modifying the DNA sequence at a target site in the genome of a fungal cell, the method including: a) introducing into a fungal cell a Cas endonuclease and a guide RNA, wherein the Cas

endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cell; and b) identifying if a modification of the DNA sequence at the target site has occurred in the fungal cell, where the Cas endonuclease, the guide RNA, or both are introduced transiently into the fungal cell.

[09] In another aspect, the present disclosure is drawn to methods for modifying the DNA sequence at a target site in the genome of a fungal cell. In some embodiments, the method includes: a) introducing into a population of fungal cells a Cas endonuclease and a guide RNA, wherein the Cas endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cells; and b) identifying at least one fungal cell from the population that has a modification of the DNA sequence at the target site, where both the Cas endonuclease and the guide RNA are introduced non-transiently into the population of fungal cells.

[010] In yet another aspect, the present disclosure are drawn to a method for modifying the DNA sequence at a target site in the genome of a fungal cell, the method including: a) introducing into a fungal cell a Cas endonuclease and a guide RNA, wherein the Cas endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cell; and b) identifying if a modification of the DNA sequence at the target site has occurred in the fungal cell, where both the Cas endonuclease and the guide RNA are introduced non-transiently into the fungal cell.

[011] In certain embodiments of the methods described herein, the modification of the DNA sequence at said target site is selected from the group consisting of a deletion of one or more nucleotides, an insertion of one or more nucleotides, a substitution of one or more nucleotides, and any combination thereof.

[012] In certain embodiments, the identifying step comprises culturing the population of fungal cells or the fungal cell from step (a) under conditions to select for or screen for the modification of the DNA sequence at the target site. In certain embodiments, the identifying step comprises culturing the population of fungal cells or the fungal cell from step (a) under conditions to screen for unstable transformants

[013] Several different types of CRISPR-Cas systems have been described and can be classified as Type I, Type II, and Type III CRISPR-Cas systems (see, e.g., the description in Liu and Fan, CRISPR-Cas system: a powerful tool for genome editing. *Plant Mol Biol* (2014) 85:209-218). In certain embodiments, the Cas endonuclease or variant thereof is a Cas9 endonuclease of the Type II CRISPR-Cas system. The Cas9 endonuclease may be any convenient Cas9 endonuclease, including but not limited to Cas9 endonucleases, and functional fragments thereof, from the following bacterial species: *Streptococcus* sp. (e.g., *S. pyogenes*, *S. mutans*, and *S. thermophilus*), *Campylobacter* sp. (e.g., *C. jejuni*), *Neisseria* sp. (e.g., *N. meningitides*), *Francisella* sp. (e.g., *F. novicida*), and *Pasteurella* sp. (e.g., *P. multocida*). Numerous other species of Cas9 can be used. For example, functional Cas9 endonucleases or variants thereof containing an amino acid sequence that has at least 70% identity to any one of SEQ ID NOs:1 to 7 may be employed, e.g., at least 80% identity, at least 90% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, and including up to 100% identity to any one of SEQ ID NOs:1 to 7. In other embodiments, the Cas endonuclease or variant thereof is a Cpf1 endonuclease of the Type II CRISPR-Cas system. Cpf1 mediates robust DNA interference with features distinct from Cas9. Cpf1 lacks tracrRNA and utilizes a T-rich protospacer-adjacent motif. It cleaves DNA via a staggered DNA double-stranded break. See, e.g., Zetsche et al., *Cell* (2015) 163:759–771.

[014] Introducing the Cas endonuclease or the guide RNA into the population of fungal cells can be achieved using any convenient method, including: transfection, transduction, transformation, electroporation, particle bombardment (biolistic particle delivery), and cell fusion techniques.

[015] In certain embodiments, introducing the Cas endonuclease and/or the guide RNA into the fungal cells includes introducing one or more DNA constructs comprising expression cassettes for the Cas endonuclease, the guide RNA, or both into the fungal cells. The one or more DNA constructs, once in the fungal cells, express the Cas endonuclease and/or the guide RNA. In certain embodiments, the DNA construct is a linear DNA construct. In certain embodiments, the DNA construct is a circular DNA construct. In certain embodiments, the DNA construct is a recombinant DNA construct.

[016] In certain embodiments, the introducing step includes directly introducing a Cas endonuclease polypeptide, a guide RNA, or both into the fungal cells. Any combination

of direct introduction and using DNA constructs can be employed (e.g., introducing a DNA construct with an expression cassette for a Cas endonuclease into the fungal cell and directly introducing a guide RNA into the cell, either simultaneously or sequentially as desired).

5 **[017]** In certain embodiments of the methods described herein, the Cas expression cassette in the DNA construct includes a Cas endonuclease encoding gene that is optimized for expression in the fungal cell. For example, a Cas endonuclease encoding gene that is optimized for expression in filamentous fungal cells includes a sequence that has at least 70% sequence identity to SEQ ID NO:8 (encoding Cas9 from *S.*
10 *pyogenes*; SEQ ID NO:1), e.g., at least 80% identity, at least 90% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, and including up to 100% identity to SEQ ID NO:8.

[018] In some instances, the Cas endonuclease is operably linked to one or more nuclear targeting signal (also referred to as a nuclear localization signal/sequence;
15 NLS). SEQ ID NO:9 and SEQ ID NO:10 provide an example of a filamentous fungal cell optimized Cas9 gene with NLS sequences at the N- and C-termini and the encoded amino acid sequence, respectively. Many different NLSs are known in eukaryotes. They include monopartite, bipartite and tripartite types. Any convenient NLS can be used, the monopartite type being somewhat more convenient with examples including
20 the SV40 NLS, a NLS derived from the *T. reesei* blr2 (blue light regulator 2) gene, or a combination of both. In some embodiments, the DNA construct is a recombinant one and comprises a promoter operably linked to a filamentous fungal cell optimized polynucleotide sequence encoding a Cas9 endonuclease or variant thereof.

[019] In certain embodiments of the methods described herein, a DNA construct or an
25 expression cassette comprising a guide RNA-encoding sequence and capable of expressing the guide RNA, is introduced into the population of fungal cells or the fungal cell. In some embodiments, the DNA construct or the expression cassette comprises a RNA polymerase III dependent promoter functional in a *Euascomycete* or *Pezizomycete*, wherein the promoter is operably linked to the guide RNA-encoding
30 sequence. In some embodiments, the promoter is derived from a *Trichoderma* U6 snRNA gene. In certain embodiments, the promoter comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 11 or 12. In specific embodiments, the promoter comprises the

sequence of SEQ ID NO: 11 or 12. In some embodiments, the DNA construct or the expression cassette for the guide RNA comprises a guide RNA-encoding DNA with an intron sequence from a *Trichoderma* U6 snRNA gene. In some embodiments, the intron sequence derived from *Trichoderma* U6 snRNA gene comprises a nucleotide
 5 sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 90. In specific embodiments, the intron sequence derived from *Trichoderma* U6 snRNA gene comprises the sequence of SEQ ID NO: 90.

[020] In certain embodiments, the modification of the DNA sequence at the target site in the genome of the fungal cells or the fungal cell is caused by non-homologous end
 10 joining (NHEJ), either without the presence of a donor DNA or in the presence of a donor DNA that is also introduced into the fungal cells or the fungal cell. In certain other embodiments, the modification of the DNA sequence at the target site is caused by homologous recombination, optionally through the presence of a donor DNA that is also introduced into the fungal cell(s). In some embodiments, the modification (e.g., a
 15 deletion of one or more nucleotides, an insertion of one or more nucleotides, insertion of an expression cassette encoding a protein of interest, or a substitution of one or more nucleotides) is originally present in the donor DNA. In some embodiments, the donor DNA has a sequence homologous to a region of the chromosomal DNA on each side of, or at or near, the target site of the Cas/guide RNA complex over at least . In
 20 some other embodiments, the donor DNA does not have a sequence homologous to a region of the chromosomal DNA on each side of, or at or near, the target site of the Cas/guide RNA complex. In certain embodiments, the donor DNA comprises an expression cassette encoding a protein of interest. In certain embodiments, the protein of interest encoded by the expression cassette is an enzyme. In particular
 25 embodiments, the protein of interest is a hemicellulase, a peroxidase, a protease, a cellulase, a xylanase, a lipase, a phospholipase, an esterase, a cutinase, a pectinase, a keratinase, a reductase, an oxidase, a phenol oxidase, a lipoxygenase, a ligninase, a pullulanase, a tannase, a pentosanase, a mannanase, a beta-glucanase, an arabinosidase, a hyaluronidase, a chondroitinase, a laccase, an amylase, a
 30 glucoamylase, a variant thereof, a functional fragment thereof, or a hybrid or mixture of two or more thereof. In yet other particular embodiments, the protein of interest is a peptide hormone, a growth factor, a clotting factor, a chemokine, a cytokine, a lymphokine, an antibody, a receptor, an adhesion molecule, a microbial antigen, a

variant thereof, a functional fragment thereof, or a hybrid or mixture of two or more thereof.

[021] In certain embodiments where homologous recombination between the donor DNA and the genome of the fungal cell(s) is desired, the NHEJ pathway in the fungal cell(s) is non-functional (inactivated) or reduced, e.g., where one or more components of the NHEJ pathway are inactivated, nonfunctional, or have reduced activity (e.g., ku80, ku70, rad50, mre11, xrs2, lig4, xrs, or combinations thereof). For example, the fungal cell can have an inactivated/reduced activity form of ku80. In certain other embodiments, the NHEJ pathway in the fungal cell(s) is functional.

[022] Fungal cells that find use in the subject methods can be filamentous fungal cell species. In certain embodiments, the fungal cell is a Eumycotina or Pezizomycotina fungal cell. In some embodiments, the fungal cell is selected from Trichoderma, Penicillium, Aspergillus, Humicola, Chrysosporium, Fusarium, Neurospora, Myceliophthora, Thermomyces, Hypocrea, and Emericella. The filamentous fungi Trichoderma reesei, P. chrysogenum, M. thermophila, Thermomyces lanuginosus, A. oryzae and A. niger are of particular interest. Other fungal cells, including species of yeast, can also be employed.

[023] The target site selected by a user of the disclosed methods can be located within a region of a gene of interest selected from the group consisting of: an open reading frame, a promoter, a regulatory sequence, a terminator sequence, a regulatory element sequence, a splice site, a coding sequence, a polyubiquitination site, an intron site, and an intron enhancing motif. Examples of genes of interest include genes encoding acetyl esterases, aminopeptidases, amylases, arabinases, arabinofuranosidases, carboxypeptidases, catalases, cellulases, chitinases, cutinase, deoxyribonucleases, epimerases, esterases, α -galactosidases, β -galactosidases, α -glucanases, glucan lysases, endo- β -glucanases, glucoamylases, glucose oxidases, α -glucosidases, β -glucosidases, glucuronidases, hemicellulases, hexose oxidases, hydrolases, invertases, isomerases, laccases, lipases, lyases, mannosidases, oxidases, oxidoreductases, pectate lyases, pectin acetyl esterases, pectin depolymerases, pectin methyl esterases, pectinolytic enzymes, peroxidases, phenoloxidases, phytases, polygalacturonases, proteases, rhamno-galacturonases, ribonucleases, transferases, transport proteins, transglutaminases, xylanases, hexose oxidases, and combinations thereof. Target genes encoding regulatory proteins such as transcription factors,

repressors, proteins that modify other proteins such as kinases, proteins involved in post-translational modification (e.g., glycosylation) can be subjected to Cas mediated editing as well as genes involved in cell signaling, morphology, growth rate, and protein secretion. No limitation in this regard is intended.

5 **[024]** In some embodiments of the methods, the step of identifying a fungal cell having a genomic modification at the site of interest includes culturing the population of cells from step (a) under conditions to select for or screen for the modification at the target site. Such conditions include antibiotic selection conditions, conditions that select for or screen for auxotrophic cells, and the like.

10 **[025]** In certain embodiments, the introducing step includes: (i) obtaining a parental fungal cell population that stably expresses the Cas endonuclease, and (ii) transiently introducing the guide RNA into the parental fungal cell population. Conversely, the introducing step can include: (i) obtaining a parental fungal cell population that stably expresses the guide RNA, and (ii) transiently introducing the Cas endonuclease into the
15 parental fungal cell population.

[026] Aspects of the present disclosure are drawn to recombinant fungal cells produced by the methods described above as well as those for use as parental host cells in performing the methods.

[027] Aspects of the present disclosure further include an engineered nucleic acid,
20 e.g., a recombinant DNA construct that can be used in the methods described above or disclosed herein. In one aspect, the engineered nucleic acid encodes a Cas endonuclease or variant thereof. In some embodiments, the Cas endonuclease or variant thereof encoded by the engineered nucleic acid comprises an amino acid sequence that has at least 70%, 75%, 80%, 85%, 90%, or 95% identity to any one of
25 SEQ ID NOs:1 to 7. In some embodiments, the engineered nucleic acid comprises a polynucleotide sequence that is codon-optimized for expression in filamentous fungi. In some embodiments, the engineered nucleic acid comprises a polynucleotide sequence that is at least 70% 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:8. In a particular embodiment, the nucleic acid comprises the sequence of
30 SEQ ID NO:8. In some embodiments, the engineered nucleic acid comprises a promoter for expression of the Cas endonuclease or variant thereof.

[028] In another aspect, the engineered nucleic acid encodes a guide RNA. In some embodiments, the nucleic acid encoding the guide RNA comprises a RNA polymerase

III dependent promoter functional in a filamentous fungal cell, a Eufungiomycete or a
 Pezizomycete. In some embodiments, the promoter is derived from a Trichoderma U6
 snRNA gene. In particular embodiments, the promoter comprises a nucleotide
 sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%,
 5 or 99% identity to SEQ ID NO: 11 or 12 or a functional fragment thereof. In particular
 embodiments, the nucleic acid comprises the sequence of SEQ ID NO: 11 or 12. In
 some embodiments, the guide RNA-encoding nucleic acid has a promoter operably
 linked to at least one heterologous sequence or guide RNA-encoding sequence, where
 the promoter functions in a filamentous fungal cell as an RNA polymerase III (pol III)
 10 dependent promoter to express the heterologous sequence and includes a
 polynucleotide sequence that has at least 80% sequence identity to SEQ ID NO:11 or
 12 (e.g., 80%, 85%, 90%, 95%, 98%, 99%, 100%, or any value there between) or a
 functional fragment thereof. In certain embodiments, the heterologous sequence or
 guide RNA-encoding sequence comprises an intron sequence derived from a
 15 Trichoderma U6 snRNA gene. In particular embodiments, the heterologous sequence
 or guide RNA-encoding sequence includes an intron that contains a U6 B-Box
 sequence, e.g., a B-Box sequence having the polynucleotide sequence of
 GTTCGTTTC. The intron can have a polynucleotide sequence with at least 60%, 65%,
 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 90.
 20 In particular embodiments, the intron comprises a polynucleotide sequence with at least
 80% sequence identity to SEQ ID NO:90. In particular embodiments, the nucleic acid
 comprises the sequence of SEQ ID NO:90. In some embodiments, the guide RNA-
 encoding nucleic acid comprises both the RNA polymerase III dependent promoter and
 the intron sequence derived from Trichoderma U6 snRNA gene as described herein. In
 25 some embodiments, the engineered nucleic acid or the recombinant DNA construct
 further includes a transcriptional terminator sequence downstream of the heterologous
 sequence, e.g., the sequence set forth in SEQ ID NO:91 or its derivative.

[029] In certain embodiments, the promoter comprised in the Cas endonuclease-
 encoding engineered nucleic acid and/or the guide RNA-encoding engineered nucleic
 30 acid is derived from a filamentous fungal cell. The filamentous fungal cell can be
 selected from any of a wide variety of filamentous fungal cells, with specific examples
 including *T. reesei* and *A. niger*. In some cases, the promoter is derived from a
 ribosomal RNA (rRNA) promoter.

[030] The recombinant DNA construct operably linked to promoter may encode a functional RNA. In certain aspects, for example, the heterologous sequence encodes a guide RNA polynucleotide, e.g., a guide RNA that includes (i) a first nucleotide sequence domain that is complementary to a polynucleotide sequence in a target DNA (variable targeting domain); and (ii) a second nucleotide sequence domain that interacts with a Cas endonuclease (CER domain).

[031] Aspects of the present disclosure include a vector having the recombinant DNA construct having a promoter operably linked to at least one heterologous sequence as described herein. The vector can further include an expression cassette for a Cas endonuclease.

[032] The present disclosure further provides a filamentous fungal cell containing a recombinant DNA constructs having a promoter operably linked to at least one heterologous sequence as described herein. Methods of expressing a heterologous sequence in a filamentous fungal cell by a) introducing the recombinant DNA construct having a promoter operably linked to at least one heterologous sequence (e.g., as an vector) into a filamentous fungal cell, and b) culturing the filamentous fungal cell of step a) under conditions to allow expression of the heterologous sequence in the recombinant DNA construct (or vector).

[033] Additional embodiments of the methods and compositions of the present disclosure are shown herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[034] The disclosure can be more fully understood from the following detailed description and the accompanying drawings, which form a part of this application.

[035] **FIG. 1** depicts the nucleotide sequence of a putative *T. reesei* U6 gene (SEQ ID NO:22). Elements of interest are indicated, including the TATA box (underlined), the transcriptional start site (downward arrow), the A –box (underlined), the Intron (forward arrow), the B-box (underlined; within the Intron of the gene), the sequences that are identical to the human U6 gene (in bold italics), and the terminator (underlined).

[036] **FIG. 2** shows a schematic of the pTrex2gHyg-Mo Cas plasmid.

[037] **FIG. 3** shows a schematic of the p219M plasmid.

[038] FIG. 4 shows a schematic of the *T. reesei* ad3A gene with PCR primer sites and intronic regions shown.

[039] FIG. 5 shows a schematic of the *T. reesei* glucoamylase gene (TrGA) with PCR primer and intronic regions shown.

5 **[040] FIG. 6** shows a schematic of the pTrex2gHygMoCasgPyr2TS6 plasmid which includes telomere sequences.

[041] FIG. 7. Plasmid map of pET30a-SpyCas9.

[042] FIG. 8A shows a plasmid map for pSM1guide which is used for flexible cloning of any potential guide RNA variable targeting (VT) domain matching the sequence
 10 pattern GGN18NGG or GN19NGG. **FIG. 8B** is a more detailed map of the single molecule guide RNA expression cassette region of the pSM1guide plasmid in panel A and shows the configuration of the T7 promoter, the transcriptional start site, the type II restriction endonuclease sites of Bsa1 (used to insert the desired VT domain, e.g., using annealed oligos), the CER domain (which includes the transcriptional terminator
 15 sequence TTTTT; not shown), and the full region encoding the single molecule guide RNA. Restriction enzyme DRA1 is used to linearize this plasmid before in vitro transcription. When transcribed, the CER domain of the guide RNA will form a hairpin structure that is able to bind to a cognate Cas9 polypeptide, thus generating a functional Cas9/guide RNA complex that can induce a double strand break at a DNA
 20 target site (one having a sequence complementary to the VT domain and the appropriate PAM site).

[043] FIG. 9A shows a map of the pXA3 plasmid which was used for creating linearized DNA substrate. This plasmid contains the coding sequence for the *xyr1* gene (SEQ ID NO:89) and was linearized by digestion with the restriction enzyme NdeI to
 25 produce the DNA substrate.

[044] FIG. 9B shows the results of guide RNA/Cas9 cleavage assay (visualized by ethidium bromide staining). Agarose gel analysis of *xyr1*-specific in vitro cleavage assay is shown in this figure. Lane 1 shows molecular weight markers; Lane 2 shows linearized plasmid substrate (containing the *xyr1* gene) in the absence of Cas9 and
 30 guide RNA; Lane 3 shows cleavage of the plasmid substrate in the presence of Cas9 and a guide RNA with the *xyr1*Ta VT domain; Lane 4 shows cleavage of the plasmid substrate in the presence of Cas9 and a guide RNA with the *xyr1*Tc VT. Positions of the linearized plasmid substrate and the cleaved products are indicated at the right.

[045] FIG. 10. Sequence analysis of the of the pyr4 gene from strains that are resistant to FOA and requires uridine for growth. Alignment with the wild type sequence (K21 control T4; SEQ ID NO:68) revealed the presence of sequence modifications at the target site in the pyr4 gene (insertions of a few (1-2 bps) or many (68bp) nucleotides).

5 SEQ ID NOs: 69 to 77 are the sequences for strains T4 4-3, T4 4-13, T4 4-11, T4 4-12, T4 4-18, T4 4-20, T4 4-19, T4 4-4, and T4 4-7, respectively. Strains T4 4-13 (SEQ ID NO: 70) and T4 4-12 (SEQ ID NO: 72) have no changes from the wild type sequence at the target site.

[046] FIGS. 11A and 11B. DNA sequence modification at a target site by uptake of in vitro formed Cas9/guide RNA complex. **FIG. 11A** shows agarose gel analysis of pyr4 specific PCR products (encompassing the target site) of two strains (T4 2.2. and T4 4.1) resistant to FOA and that require uridine for growth isolated after direct introduction of in vitro formed Cas9/guide RNA complex followed by growth on Vogel's Uridine/FOA plates. Strain T4 2.2 (Lane 2) showed a PCR product that is of lower molecular weight than the T4 4.1 clone (Lane 3; which is equivalent to the control, shown in Panel B, Lane 2), indicating a large deletion in the pyr4 gene. **FIG. 11B** shows a similar PCR/agarose gel analysis as in FIG. 11A, but showing T4 strains 4.1, 4.2, 4.3, and 4.4, all of which are resistant to FOA and that require uridine for growth. Strain 4.3 (Lane 5) showed PCR product of the pyr4 gene that is of lower molecular weight than the control (C+; Lane 2).

[047] FIG. 12. Sequence analysis of the pyr4 genes derived from clones T4 2.2 (shown in FIG. 11A) and T4 2.4. Sequence analysis shows that the T4 2.2 clone (top alignment) has a deletion of 611 base pairs at the target site of the introduced Cas9/guide RNA complex. The sequence corresponding to the VT domain sequence of the guide RNA is boxed and the PAM site is circled. The bottom alignment shows a 1 base pair insertion in the pyr4 gene at the target site of the isolated T4 2.4 strain (a "G" residue). The sequence corresponding to the VT domain sequence of the guide RNA is indicated with a line over the alignment and the PAM site is circled. SEQ ID NOs:78 to 81 are the sequences for 9-96 (T4 2.2 strain), Pyr Tr (wild type sequence), Query (wild type sequence), and Sbjct (T4 2.4 strain), respectively.

[048] FIG. 13. Sequence analysis of the pyr4 genes derived from clones T4 4.1 and 4.2 (top alignment), 4.3 (bottom alignment) and 4.4 (middle alignment) (which are shown in FIG. 11B). The wild type pyr4 sequence is the first sequence (top) in all

alignments and a consensus is shown on the bottom of all alignments (SEQ ID NO:82). The top alignment shows that the T4 4.1 clone (third sequence in the alignment; SEQ ID NO:84) has an insertion of a T nucleotide while the T4 4.2 clone (second sequence in the alignment; SEQ ID NO:83) has an insertion of a G nucleotide at the target site in the pyr4 gene. (The consensus sequence in this alignment is the same as SEQ ID NO:82.) The middle alignment shows that the T4 4.4 clone (second sequence in the alignment; SEQ ID NO:85) has a deletion of an A nucleotide at the target site in the pyr4 gene. (The consensus sequence in this alignment is the same as SEQ ID NO:85.) The bottom alignment shows that the pyr4 gene sequence in the T4 4.3 clone (second sequence in the alignment; SEQ ID NO: 86) diverges abruptly at the target site. (The consensus sequence in this alignment is SEQ ID NO:87; spaces in the consensus sequence in FIG. 13 are represented by "N" in SEQ ID NO:87.) Further alignment analysis (not shown) confirmed that the T4 4.3 clone has a deletion of 988 base pairs at the target site for the introduced Cas9/guide RNA complex.

DETAILED DESCRIPTION

[049] The present disclosure includes compositions and methods that find use in modifying the DNA sequence at a target site in the genome of a fungal cell. The methods employ a functional guide RNA/Cas endonuclease complex which recognizes a desired target site and introduces a double strand break at the site. Repair of this double-strand break can introduce modifications to the DNA sequence at the target site.

[050] Before the present compositions and methods are described in greater detail, it is to be understood that the present compositions and methods are not limited to particular embodiments described, and as such may, of course, vary. 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 be limiting, since the scope of the present compositions and methods will be limited only by the appended claims.

[051] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the present compositions and methods. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also encompassed within the present

compositions and methods, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the present compositions and methods.

5 **[052]** Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrequited
10 number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number. For example, in connection with a numerical value, the term "about" refers to a range of -10% to +10% of the numerical value, unless the term is otherwise specifically defined in context. In another example, the phrase a "pH value of about 6" refers to pH values of from 5.4 to 6.6,
15 unless the pH value is specifically defined otherwise.

[053] The headings provided herein are not limitations of the various aspects or embodiments of the present compositions and methods which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

20 **[054]** The present document is organized into a number of sections for ease of reading; however, the reader will appreciate that statements made in one section may apply to other sections. In this manner, the headings used for different sections of the disclosure should not be construed as limiting.

[055] Unless defined otherwise, all technical and scientific terms used herein have the
25 same meaning as commonly understood by one of ordinary skill in the art to which the present compositions and methods belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present compositions and methods, representative illustrative methods and materials are now described.

30 **[056]** All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the

publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present compositions and methods are not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication
5 dates which may need to be independently confirmed.

[057] In accordance with this detailed description, the following abbreviations and definitions apply. Note that the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “an enzyme” includes a plurality of such enzymes, and reference to “the dosage”
10 includes reference to one or more dosages and equivalents thereof known to those skilled in the art, and so forth.

[058] It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as “solely,” “only” and the like in connection with the recitation of
15 claim elements, or use of a “negative” limitation.

[059] As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the
20 present compositions and methods described herein. Any recited method can be carried out in the order of events recited or in any other order which is logically possible.

Definitions

[060] As used herein, a polypeptide referred to as a “Cas endonuclease” or having “Cas endonuclease activity” relates to a CRISPR associated (Cas) polypeptide encoded by a Cas gene where the Cas protein is capable of cutting a target DNA sequence when functionally coupled with one or more guide polynucleotides (see, e.g., US Patent 8697359 entitled “CRISPR-Cas systems and methods for altering
30 expression of gene products”). Variants of Cas endonucleases that retain guide polynucleotide directed endonuclease activity are also included in this definition. The Cas endonucleases employed in the donor DNA insertion methods detailed herein are endonucleases that introduce double-strand breaks into the DNA at the target site. A

Cas endonuclease is guided by the guide polynucleotide to recognize and cleave a specific target site in double stranded DNA, e.g., at a target site in the genome of a cell.

[061] As used herein, the term “guide polynucleotide” relates to a polynucleotide sequence that can form a complex with a Cas endonuclease and enables the Cas endonuclease to recognize and cleave a DNA target site. The guide polynucleotide can be a single molecule or a double molecule. The guide polynucleotide sequence can be a RNA sequence, a DNA sequence, or a combination thereof (a RNA-DNA combination sequence). Optionally, the guide polynucleotide can comprise at least one nucleotide, phosphodiester bond or linkage modification such as, but not limited, to Locked Nucleic Acid (LNA), 5-methyl dC, 2,6-Diaminopurine, 2'-Fluoro A, 2'-Fluoro U, 2'-O-Methyl RNA, phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer 18 (hexaethylene glycol chain) molecule, or 5' to 3' covalent linkage resulting in circularization. A guide polynucleotide that solely comprises ribonucleic acids is also referred to as a “guide RNA”.

[062] The guide polynucleotide can be a double molecule (also referred to as duplex guide polynucleotide) comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide sequence domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. The CER domain of the double molecule guide polynucleotide comprises two separate molecules that are hybridized along a region of complementarity. The two separate molecules can be RNA, DNA, and/or RNA-DNA-combination sequences. In some embodiments, the first molecule of the duplex guide polynucleotide comprising a VT domain linked to a CER domain is referred to as “crDNA” (when composed of a contiguous stretch of DNA nucleotides) or “crRNA” (when composed of a contiguous stretch of RNA nucleotides), or “crDNA-RNA” (when composed of a combination of DNA and RNA nucleotides). The crNucleotide can comprise a fragment of the crRNA naturally occurring in Bacteria and Archaea. In one embodiment, the size of the fragment of the crRNA naturally occurring in Bacteria and Archaea that is present in a crNucleotide disclosed herein can range from, but is not limited to, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides. In some embodiments the second molecule of the duplex guide polynucleotide comprising a CER domain is referred to as “tracrRNA” (when composed of a contiguous stretch of RNA nucleotides) or “tracrDNA” (when composed of a

contiguous stretch of DNA nucleotides) or “tracrDNA-RNA” (when composed of a combination of DNA and RNA nucleotides). In certain embodiments, the RNA that guides the RNA/Cas9 endonuclease complex is a duplexed RNA comprising a duplex crRNA-tracrRNA.

5 **[063]** The guide polynucleotide can also be a single molecule comprising a first nucleotide sequence domain (referred to as Variable Targeting domain or VT domain) that is complementary to a nucleotide sequence in a target DNA and a second nucleotide domain (referred to as Cas endonuclease recognition domain or CER domain) that interacts with a Cas endonuclease polypeptide. By “domain” it is meant a
 10 contiguous stretch of nucleotides that can be RNA, DNA, and/or RNA-DNA-combination sequence. The VT domain and / or the CER domain of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA-combination sequence. In some embodiments the single guide polynucleotide comprises a crNucleotide (comprising a VT domain linked to a CER domain) linked to a
 15 tracrNucleotide (comprising a CER domain), wherein the linkage is a nucleotide sequence comprising a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. The single guide polynucleotide being comprised of sequences from the crNucleotide and tracrNucleotide may be referred to as “single guide RNA” (when composed of a contiguous stretch of RNA nucleotides) or “single guide DNA” (when
 20 composed of a contiguous stretch of DNA nucleotides) or “single guide RNA-DNA” (when composed of a combination of RNA and DNA nucleotides). In one embodiment of the disclosure, the single guide RNA comprises a crRNA or crRNA fragment and a tracrRNA or tracrRNA fragment of the type II CRISPR/Cas system that can form a complex with a type II Cas endonuclease, wherein the guide RNA/Cas endonuclease
 25 complex can direct the Cas endonuclease to a fungal cell genomic target site, enabling the Cas endonuclease to introduce a double strand break into the genomic target site.

[064] One aspect of using a single guide polynucleotide versus a duplex guide polynucleotide is that only one expression cassette needs to be made to express the single guide polynucleotide in a target cell.

30 **[065]** The term “variable targeting domain” or “VT domain” is used interchangeably herein and includes a nucleotide sequence that is complementary to one strand (nucleotide sequence) of a double strand DNA target site. The % complementation between the first nucleotide sequence domain (VT domain) and the target sequence is

at least 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 63%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or is 100% complementary. The VT domain
 5 can be at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 or 30 nucleotides in length. In some embodiments, the VT domain comprises a contiguous stretch of 12 to 30 nucleotides. The VT domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA sequence, or any combination thereof.

10 **[066]** The term “Cas endonuclease recognition domain” or “CER domain” of a guide polynucleotide is used interchangeably herein and includes a nucleotide sequence (such as a second nucleotide sequence domain of a guide polynucleotide), that interacts with a Cas endonuclease polypeptide. The CER domain can be composed of a DNA sequence, a RNA sequence, a modified DNA sequence, a modified RNA
 15 sequence (see for example modifications described herein), or any combination thereof.

[067] The nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a RNA sequence, a DNA sequence, or a RNA-DNA combination sequence. In one embodiment, the nucleotide sequence linking
 20 the crNucleotide and the tracrNucleotide of a single guide polynucleotide can be at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94,
 25 95, 96, 97, 98, 99 or 100 nucleotides in length. In another embodiment, the nucleotide sequence linking the crNucleotide and the tracrNucleotide of a single guide polynucleotide can comprise a tetraloop sequence, such as, but not limiting to a GAAA tetraloop sequence.

[068] Nucleotide sequence modification of the guide polynucleotide, VT domain and/or
 30 CER domain can be selected from, but not limited to, the group consisting of a 5' cap, a 3' polyadenylated tail, a riboswitch sequence, a stability control sequence, a sequence that forms a dsRNA duplex, a modification or sequence that targets the guide polynucleotide to a subcellular location, a modification or sequence that provides for

tracking, a modification or sequence that provides a binding site for proteins, a Locked Nucleic Acid (LNA), a 5-methyl dC nucleotide, a 2,6-Diaminopurine nucleotide, a 2'-Fluoro A nucleotide, a 2'-Fluoro U nucleotide; a 2'-O-Methyl RNA nucleotide, a phosphorothioate bond, linkage to a cholesterol molecule, linkage to a polyethylene glycol molecule, linkage to a spacer molecule, a 5' to 3' covalent linkage, or any combination thereof. These modifications can result in at least one additional beneficial feature, wherein the additional beneficial feature is selected from the group of a modified or regulated stability, a subcellular targeting, tracking, a fluorescent label, a binding site for a protein or protein complex, modified binding affinity to complementary target sequence, modified resistance to cellular degradation, and increased cellular permeability.

[069] As used herein, the term “guide polynucleotide/Cas endonuclease system” (and equivalents) includes a complex of a Cas endonuclease and a guide polynucleotide (single or double) that is capable of introducing a double strand break into a DNA target sequence. The Cas endonuclease unwinds the DNA duplex in close proximity of the genomic target site and cleaves both DNA strands upon recognition of a target sequence by a guide RNA, but only if the correct protospacer-adjacent motif (PAM) is appropriately oriented at the 3' end of the target sequence.

[070] The terms “functional fragment”, “fragment that is functionally equivalent”, “functionally equivalent fragment”, and the like, are used interchangeably and refer to a portion or subsequence of a parent biological sequence, e.g., a polypeptide that retains the qualitative enzymatic activity of the parent polypeptide, or a polynucleotide that retains the main function of the parent polynucleotide. For example, a functional fragment of a Cas endonuclease retains the ability to create a double-strand break with a guide polynucleotide. It is noted here that a functional fragment may have altered quantitative enzymatic activity as compared to the parent polypeptide. Other examples include a functional fragment of a gene promoter which retains the ability to promote transcription, a functional fragment of an intron which retains the ability to facilitate transcription, and a functional fragment of an enzyme-encoding gene sequence which encodes a functional fragment of an enzyme.

[071] The terms “functional variant”, “variant that is functionally equivalent”, “functionally equivalent variant”, and the like are used interchangeably and refer to a variant of a parent polypeptide that retains the qualitative enzymatic activity of the

parent polypeptide. For example, a functional variant of a Cas endonuclease retains the ability to create a double-strand break with a guide polynucleotide. It is noted here that a functional variant may have altered quantitative enzymatic activity as compared to the parent polypeptide.

5 **[072]** Fragments and variants can be obtained via any convenient method, including site-directed mutagenesis and synthetic construction.

[073] The term “genome” as it applies to fungal cells encompasses not only chromosomal DNA found within the nucleus, but organelle DNA found within subcellular components (e.g., mitochondria) of the cell.

10 **[074]** A “codon-modified gene” or “codon-preferred gene” or “codon-optimized gene” is a gene having its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell. The nucleic acid changes made to codon-optimize a gene are “synonymous”, meaning that they do not alter the amino acid sequence of the encoded polypeptide of the parent gene. However, both native and
15 variant genes can be codon-optimized for a particular host cell, and as such no limitation in this regard is intended.

[075] “Coding sequence” refers to a polynucleotide sequence which codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5’ non-coding sequences), within, or downstream (3’ non-coding
20 sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include, but are not limited to: promoters, translation leader sequences, 5’ untranslated sequences, 3’ untranslated sequences, introns, polyadenylation target sequences, RNA processing sites, effector binding sites, and stem-loop structures.

25 **[076]** “Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. An “enhancer” is a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level
30 or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, and/or comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different

tissues or cell types, or at different stages of development, or in response to different environmental conditions. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity. As is well-known in the art, promoters can be categorized according to their strength and/or the conditions under which they are active, e.g., constitutive promoters, strong promoters, weak promoters, inducible/repressible promoters, tissue-specific/developmentally regulated promoters, cell-cycle dependent promoters, etc.

[077] “RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. “Messenger RNA” or “mRNA” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a DNA that is complementary to, and synthesized from, a mRNA template using the enzyme reverse transcriptase. “Sense” RNA refers to RNA transcript that includes the mRNA and can be translated into protein within a cell or in vitro. “Antisense RNA” refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA, and that, under certain conditions, blocks the expression of a target gene (see, e.g., U.S. Patent No. 5,107,065). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5’ non-coding sequence, 3’ non-coding sequence, introns, or the coding sequence. “Functional RNA” refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated into a polypeptide but yet has an effect on cellular processes. The terms “complement” and “reverse complement” are used interchangeably herein with respect to mRNA transcripts, and are meant to define the antisense RNA of the message.

[078] As used herein, “functionally attached” or “operably linked” means that a regulatory region or functional domain of a polypeptide or polynucleotide sequence having a known or desired activity, such as a promoter, enhancer region, terminator, signal sequence, epitope tag, etc., is attached to or linked to a target (e.g., a gene or polypeptide) in such a manner as to allow the regulatory region or functional domain to control the expression, secretion or function of that target according to its known or desired activity. For example, a promoter is operably linked with a coding sequence when it is capable of regulating the expression of that coding sequence (i.e., the coding sequence is under the transcriptional control of the promoter).

[079] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art.

[080] “PCR” or “polymerase chain reaction” is a technique for the synthesis of specific DNA segments and consists of a series of repetitive denaturation, annealing, and extension cycles and is well known in the art.

[081] The term “recombinant,” when used in reference to a biological component or composition (e.g., a cell, nucleic acid, polypeptide/enzyme, vector, etc.) indicates that the biological component or composition is in a state that is not found in nature. In other words, the biological component or composition has been modified by human intervention from its natural state. For example, a recombinant cell encompass a cell that expresses one or more genes that are not found in its native parent (i.e., non-recombinant) cell, a cell that expresses one or more native genes in an amount that is different than its native parent cell, and/or a cell that expresses one or more native genes under different conditions than its native parent cell. Recombinant nucleic acids may differ from a native sequence by one or more nucleotides, be operably linked to heterologous sequences (e.g., a heterologous promoter, a sequence encoding a non-native or variant signal sequence, etc.), be devoid of intronic sequences, and/or be in an isolated form. Recombinant polypeptides/enzymes may differ from a native sequence by one or more amino acids, may be fused with heterologous sequences, may be truncated or have internal deletions of amino acids, may be expressed in a manner not found in a native cell (e.g., from a recombinant cell that over-expresses the polypeptide due to the presence in the cell of an expression vector encoding the polypeptide), and/or be in an isolated form. It is emphasized that in some embodiments, a recombinant polynucleotide or polypeptide/enzyme has a sequence that is identical to its wild-type counterpart but is in a non-native form (e.g., in an isolated or enriched form).

[082] The term “engineered”, when used in reference to a biological component or composition (e.g., a cell, nucleic acid, polypeptide/enzyme, vector, etc.) indicates that the biological component or composition is designed by human and is at least not completely derived from or completely identical to biological component or composition in nature, as far as the person who designs the “engineered” biological component or composition is aware at the time of designing. An engineered biological component or composition, e.g., an engineered nucleic acid, may be derived from various parts of

different naturally existing biological components or compositions. An engineered biological component or composition may be a recombinant biological component or composition.

[083] The terms “plasmid”, “vector” and “cassette” refer to an extra chromosomal element that carries a polynucleotide sequence of interest, e.g., a gene of interest to be expressed in a cell (an “expression vector” or “expression cassette”). Such elements are generally in the form of double-stranded DNA and may be autonomously replicating sequences, genome integrating sequences, phage, or nucleotide sequences, in linear or circular form, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a polynucleotide of interest into a cell. The polynucleotide sequence of interest may be a gene encoding a polypeptide or functional RNA that is to be expressed in the target cell. Expression cassettes/vectors generally contain a gene with operably linked elements that allow for expression of that gene in a host cell.

[084] The term “expression”, as used herein, refers to the production of a functional end-product (e.g., an mRNA, guide RNA, or a protein) in either precursor or mature form.

[085] “Introduced” in the context of inserting a polynucleotide or polypeptide into a cell (e.g., a recombinant DNA construct/expression construct) refers to any method for performing such a task, and includes any means of “transfection”, “transformation”, “transduction”, physical means, or the like, to achieve introduction of the desired biomolecule.

[086] By “introduced transiently”, “transiently introduced”, “transient introduction”, “transiently express” and the like is meant that a biomolecule is introduced into a host cell (or a population of host cells) in a non-permanent manner. With respect to double stranded DNA, transient introduction includes situations in which the introduced DNA does not integrate into the chromosome of the host cell and thus is not transmitted to all daughter cells during growth as well as situations in which an introduced DNA molecule that may have integrated into the chromosome is removed at a desired time using any convenient method (e.g., employing a cre-lox system, by removing positive selective pressure for an episomal DNA construct, by promoting looping out of all or part of the integrated polynucleotide from the chromosome using a selection media, etc.). No

limitation in this regard is intended. In general, introduction of RNA (e.g., a guide RNA, a messenger RNA, ribozyme, etc.) or a polypeptide (e.g., a Cas polypeptide) into host cells is considered transient in that these biomolecules are not replicated and indefinitely passed down to daughter cells during cell growth. With respect to the

5 Cas/guide RNA complex, transient introduction covers situations when either of the components is introduced transiently, as both biomolecules are needed to exert targeted Cas endonuclease activity. Thus, transient introduction of a Cas/guide RNA complex includes embodiments where either one or both of the Cas endonuclease and the guide RNA are introduced transiently. For example, a host cell having a genome-
10 integrated expression cassette for the Cas endonuclease (and thus not transiently introduced) into which a guide RNA is transiently introduced can be said to have a transiently introduced Cas/guide RNA complex (or system) because the functional complex is present in the host cell in a transient manner. In certain embodiments, the introducing step includes: (i) obtaining a parental fungal cell population that stably
15 expresses the Cas endonuclease, and (ii) transiently introducing the guide RNA into the parental fungal cell population. Conversely, the introducing step can include: (i) obtaining a parental fungal cell population that stably expresses the guide RNA, and (ii) transiently introducing the Cas endonuclease into the parental fungal cell population.

[087] “Mature” protein refers to a post-translationally processed polypeptide (i.e., one
20 from which any pre- or propeptides present in the primary translation product have been removed). “Precursor” protein refers to the primary product of translation of mRNA (i.e., with pre- and propeptides still present). Pre- and propeptides may be but are not limited to intracellular localization signals.

[088] “Stable transformation” refers to the transfer of a nucleic acid fragment into a
25 genome of a host organism, including both nuclear and organellar genomes, resulting in genetically stable inheritance (the resulting host cell is sometimes referred to herein as a “stable transformant”). In contrast, “transient transformation” refers to the transfer of a nucleic acid fragment into the nucleus, or other DNA-containing organelle, of a host organism resulting in gene expression without integration or stable inheritance
30 (sometimes referred to herein as “unstable transformation”, and the resulting host cell sometimes referred to herein as an “unstable transformant”). Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” organisms.

[089] “Fungal cell”, “fungi”, “fungal host cell”, and the like, as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth et al., In, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited
 5 in Hawksworth et al., supra) and all mitosporic fungi (Hawksworth et al., supra). In certain embodiments, the fungal host cell is a yeast cell, where by “yeast” is meant ascosporogenous yeast (Endomycetales), basidiosporogenous yeast, and yeast belonging to the Fungi Imperfecti (Blastomycetes). As such, a yeast host cell includes a Candida, Hansenula, Kluyveromyces, Pichia, Saccharomyces, Schizosaccharomyces,
 10 or Yarrowia cell. Species of yeast include, but are not limited to, the following: Saccharomyces carlsbergensis, Saccharomyces cerevisiae, Saccharomyces diastaticus, Saccharomyces douglasii, Saccharomyces kluyveri, Saccharomyces norbensis, Saccharomyces oviformis, Kluyveromyces lactis, and Yarrowia lipolytica cell.

[090] The term “filamentous fungal cell” includes all filamentous forms of the
 15 subdivision Eumycotina or Pezizomycotina. Suitable cells of filamentous fungal genera include, but are not limited to, cells of Acremonium, Aspergillus, Chrysosporium, Corynascus, Chaetomium, Fusarium, Gibberella, Humicola, Magnaporthe, Myceliophthora, Neurospora, Paecilomyces, Penicillium, Scytalidium, Talaromyces, Thermoascus, Thielavia, Tolypocladium, Hypocrea, and Trichoderma.

[091] Suitable cells of filamentous fungal species include, but are not limited to, cells of
 20 Aspergillus awamori, Aspergillus fumigatus, Aspergillus foetidus, Aspergillus japonicus, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Chrysosporium lucknowense, Hypocrea jecorina, Fusarium bactridioides, Fusarium cerealis, Fusarium crookwellense, Fusarium culmorum, Fusarium graminearum, Fusarium graminum,
 25 Fusarium heterosporum, Fusarium negundi, Fusarium oxysporum, Fusarium reticulatum, Fusarium roseum, Fusarium sambucinum, Fusarium sarcochrom, Fusarium sporotrichioides, Fusarium sulphureum, Fusarium torulosum, Fusarium trichothecioides, Fusarium venenatum, Humicola insolens, Humicola lanuginosa, Myceliophthora thermophila, Neurospora crassa, Neurospora intermedia, Penicillium
 30 purpurogenum, Penicillium canescens, Penicillium solitum, Penicillium funiculosum, Phanerochaete chrysosporium, Talaromyces flavus, Thielavia terrestris, Trichoderma harzianum, Trichoderma koningii, Trichoderma longibrachiatum, Trichoderma reesei, and Trichoderma viride.

[092] The terms “target site”, “target sequence”, “genomic target site”, “genomic target sequence” (and equivalents) are used interchangeably herein and refer to a polynucleotide sequence in the genome of a fungal cell at which a Cas endonuclease cleavage is desired to promote a genome modification, e.g., modification of the DNA sequence at the target site. The context in which this term is used, however, can slightly alter its meaning. For example, the target site for a Cas endonuclease is generally very specific and can often be defined to the exact nucleotide position, whereas in some cases the target site for a desired genome modification can be defined more broadly than merely the site at which DNA cleavage occurs. The target site can be an endogenous site in the fungal cell genome, or alternatively, the target site can be heterologous to the fungal cell and thereby not be naturally occurring in the genome, or the target site can be found in a heterologous genomic location compared to where it occurs in nature.

[093] As used herein, “nucleic acid” means a polynucleotide and includes a single or a double-stranded polymer of deoxyribonucleotide or ribonucleotide bases. Nucleic acids may also include fragments and modified nucleotides. Thus, the terms “polynucleotide”, “nucleic acid sequence”, “nucleotide sequence” and “nucleic acid fragment” are used interchangeably to denote a polymer of RNA and/or DNA that is single- or double-stranded, optionally containing synthetic, non-natural, or altered nucleotide bases. Nucleotides (usually found in their 5'-monophosphate form) are referred to by their single letter designation as follows: “A” for adenosine or deoxyadenosine (for RNA or DNA, respectively), “C” for cytosine or deoxycytosine, “G” for guanosine or deoxyguanosine, “U” for uridine, “T” for deoxythymidine, “R” for purines (A or G), “Y” for pyrimidines (C or T), “K” for G or T, “H” for A or C or T, “I” for inosine, and “N” for any nucleotide.

[094] The term “derived from” encompasses the terms “originated from,” “obtained from,” “obtainable from,” “isolated from,” and “created from,” and generally indicates that one specified material find its origin in another specified material or has features that can be described with reference to the another specified material.

[095] As used herein, the term “hybridization conditions” refers to the conditions under which hybridization reactions are conducted. These conditions are typically classified by degree of “stringency” of the conditions under which hybridization is measured. The degree of stringency can be based, for example, on the melting temperature (T_m) of the

nucleic acid binding complex or probe. For example, “maximum stringency” typically occurs at about $T_m - 5^\circ\text{C}$ (5°C below the T_m of the probe); “high stringency” at about $5-10^\circ\text{C}$ below the T_m ; “intermediate stringency” at about $10-20^\circ\text{C}$ below the T_m of the probe; and “low stringency” at about $20-25^\circ\text{C}$ below the T_m . Alternatively, or in addition, hybridization conditions can be based upon the salt or ionic strength conditions of hybridization, and/or upon one or more stringency washes, e.g.: 6X SSC = very low stringency; 3X SSC = low to medium stringency; 1X SSC = medium stringency; and 0.5X SSC = high stringency. Functionally, maximum stringency conditions may be used to identify nucleic acid sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify nucleic acid sequences having about 80% or more sequence identity with the probe. For applications requiring high selectivity, it is typically desirable to use relatively stringent conditions to form the hybrids (e.g., relatively low salt and/or high temperature conditions are used).

[096] As used herein, the term “hybridization” refers to the process by which a strand of nucleic acid joins with a complementary strand through base pairing, as known in the art. More specifically, “hybridization” refers to the process by which one strand of nucleic acid forms a duplex with, i.e., base pairs with, a complementary strand, as occurs during blot hybridization techniques and PCR techniques. A nucleic acid sequence is considered to be “selectively hybridizable” to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions. Hybridization conditions are based on the melting temperature (T_m) of the nucleic acid binding complex or probe. For example, “maximum stringency” typically occurs at about $T_m - 5^\circ\text{C}$ (5° below the T_m of the probe); “high stringency” at about $5-10^\circ\text{C}$ below the T_m ; “intermediate stringency” at about $10-20^\circ\text{C}$ below the T_m of the probe; and “low stringency” at about $20-25^\circ\text{C}$ below the T_m . Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while intermediate or low stringency hybridization can be used to identify or detect polynucleotide sequence homologs.

[097] Intermediate and high stringency hybridization conditions are well known in the art. For example, intermediate stringency hybridizations may be carried out with an overnight incubation at 37°C in a solution comprising 20% formamide, 5 x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's

solution, 10% dextran sulfate and 20 mg/mL denatured sheared salmon sperm DNA, followed by washing the filters in 1x SSC at about 37 - 50 °C. High stringency hybridization conditions may be hybridization at 65 °C and 0.1X SSC (where 1X SSC = 0.15 M NaCl, 0.015 M Na citrate, pH 7.0). Alternatively, high stringency hybridization conditions can be carried out at about 42°C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 µg/mL denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at 42°C. And very high stringent hybridization conditions may be hybridization at 68 °C and 0.1X SSC. Those of skill in the art know how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

[098] The phrase “substantially similar” or “substantially identical,” in the context of at least two nucleic acids or polypeptides, means that a polynucleotide or polypeptide comprises a sequence that has at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or even at least 99% identical to a parent or reference sequence, or does not include amino acid substitutions, insertions, deletions, or modifications made only to circumvent the present description without adding functionality.

[099] “Sequence identity” or “identity” in the context of nucleic acid or polypeptide sequences refers to the nucleic acid bases or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window.

[0100] The term “percentage of sequence identity” refers to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. Useful examples of percent sequence identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%,

90% or 95%, or any integer percentage from 50% to 100%. These identities can be determined using any of the programs described herein.

[0101] Sequence alignments and percent identity or similarity calculations may be determined using a variety of comparison methods designed to detect homologous sequences including, but not limited to, the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the “default values” of the program referenced, unless otherwise specified. As used herein “default values” will mean any set of values or parameters that originally load with the software when first initialized.

[0102] The “Clustal V method of alignment” corresponds to the alignment method labeled Clustal V (described by Higgins and Sharp, (1989) CABIOS 5:151-153; Higgins et al., (1992) Comput Appl Biosci 8:189-191) and found in the MegAlign™ program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). For multiple alignments, the default values correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method are KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids these parameters are KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. After alignment of the sequences using the Clustal V program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program.

[0103] The “Clustal W method of alignment” corresponds to the alignment method labeled Clustal W (described by Higgins and Sharp, (1989) CABIOS 5:151-153; Higgins et al., (1992) Comput Appl Biosci 8:189-191) and found in the MegAlign™ v6.1 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for multiple alignment (GAP PENALTY=10, GAP LENGTH PENALTY=0.2, Delay Divergen Seqs (%)=30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB). After alignment of the sequences using the Clustal W program, it is possible to obtain a “percent identity” by viewing the “sequence distances” table in the same program.

[0104] Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 (GCG, Accelrys, San Diego, CA) using the following parameters: % identity and % similarity for a nucleotide sequence using a gap creation penalty weight of 50 and a gap length extension penalty weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using a GAP creation penalty weight of 8 and a gap length extension penalty of 2, and the BLOSUM62 scoring matrix (Henikoff and Henikoff, (1989) Proc. Natl. Acad. Sci. USA 89:10915). GAP uses the algorithm of Needleman and Wunsch, (1970) J Mol Biol 48:443-53, to find an alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps, using a gap creation penalty and a gap extension penalty in units of matched bases.

[0105] It is well understood by one skilled in the art that many levels of sequence identity are useful in identifying polypeptides from other species or modified naturally or synthetically wherein such polypeptides have the same or similar function or activity. Useful examples of percent identities include, but are not limited to, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% or 95%, or any integer percentage from 50% to 100%. Indeed, any integer amino acid identity from 50% to 100% may be useful in describing the present disclosure, such as 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%.

[0106] "Gene" includes a nucleic acid fragment that encodes and is capable to express a functional molecule such as, but not limited to, a specific polypeptide (e.g., an enzyme) or a functional RNA molecule (e.g., a guide RNA, an anti-sense RNA, ribozyme, etc.), and includes regulatory sequences preceding (5' non-coding sequences) and/or following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. A recombinant gene refers to a gene that is regulated by a different gene's regulatory sequences which could be from a different organism or the same organism.

[0107] A "mutated gene" is a gene that has been altered through human intervention. Such a "mutated gene" has a sequence that differs from the sequence of the

corresponding non-mutated gene by at least one nucleotide addition, deletion, or substitution. In certain embodiments of the disclosure, the mutated gene comprises an alteration that results from a guide polynucleotide/Cas endonuclease system as disclosed herein. A mutated fungal cell is a fungal cell comprising a mutated gene.

5 **[0108]** As used herein, a “targeted mutation” is a mutation in a native gene that was made by altering a target sequence within the native gene using a method involving a double-strand-break-inducing agent that is capable of inducing a double-strand break in the DNA of the target sequence as disclosed herein or known in the art.

[0109] The term “donor DNA” or “donor nucleic acid sequence” or “donor
10 polynucleotide” refers to a polynucleotide that contains a polynucleotide sequence of interest that is to be inserted at or near a target site or to replace a region at or near a target site, generally in conjunction with the activity of a Cas/guide polynucleotide complex (where the guide polynucleotide defines the target site, as detailed above). As such, the polynucleotide sequence of interest in the donor DNA may include a novel
15 region to be inserted at or near the target site and/or a modified polynucleotide sequence when compared to the nucleotide sequence to be replaced/edited at or near the target site. In certain embodiments, the donor DNA construct further comprises a first and a second region of homology that flank the polynucleotide sequence of interest. The first and second regions of homology of the donor DNA share homology
20 to a first and a second genomic region, respectively, present in or flanking the target site of the fungal cell genome. By “homology” is meant DNA sequences that are similar. For example, a “region of homology to a genomic region” that is found on the donor DNA is a region of DNA that has a similar sequence to a given “genomic region” in the fungal cell genome. A region of homology can be of any length that is sufficient
25 to promote homologous recombination at the cleaved target site. For example, the region of homology can comprise at least 5-10, 5-15, 5-20, 5-25, 5-30, 5-35, 5-40, 5-45, 5- 50, 5-55, 5-60, 5-65, 5- 70, 5-75, 5-80, 5-85, 5-90, 5-95, 5-100, 5-200, 5-300, 5-400, 5-500, 5-600, 5-700, 5-800, 5-900, 5-1000, 5-1100, 5-1200, 5-1300, 5-1400, 5-1500, 5-1600, 5-1700, 5-1800, 5-1900, 5-2000, 5-2100, 5-2200, 5-2300, 5-2400, 5-
30 2500, 5-2600, 5-2700, 5-2800, 5-2900, 5-3000, 5-3100 or more bases in length such that the region of homology has sufficient homology to undergo homologous recombination with the corresponding genomic region. “Sufficient homology” indicates that two polynucleotide sequences have sufficient structural similarity to act as substrates for a homologous recombination reaction. The structural similarity includes

overall length of each polynucleotide fragment, as well as the sequence similarity of the polynucleotides. Sequence similarity can be described by the percent sequence identity over the whole length of the sequences, and/or by conserved regions comprising localized similarities such as contiguous nucleotides having 100% sequence identity, and percent sequence identity over a portion of the length of the sequences.

[0110] The amount of homology or sequence identity shared by a target and a donor polynucleotide can vary and includes total lengths and/or regions having unit integral values in the ranges of about 1-20 bp, 20-50 bp, 50-100 bp, 75-150 bp, 100-250 bp, 150-300 bp, 200-400 bp, 250-500 bp, 300-600 bp, 350-750 bp, 400-800 bp, 450-900 bp, 500-1000 bp, 600-1250 bp, 700-1500 bp, 800-1750 bp, 900-2000 bp, 1-2.5 kb, 1.5-3 kb, 2-4 kb, 2.5-5 kb, 3-6 kb, 3.5-7 kb, 4-8 kb, 5-10 kb, or up to and including the total length of the target site. These ranges include every integer within the range, for example, the range of 1-20 bp includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 and 20 bp. The amount of homology can also be described by percent sequence identity over the full aligned length of the two polynucleotides which includes percent sequence identity of about at least 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%.

Sufficient homology includes any combination of polynucleotide length, global percent sequence identity, and optionally conserved regions of contiguous nucleotides or local percent sequence identity, for example sufficient homology can be described as a region of 75-150 bp having at least 80% sequence identity to a region of the target locus. Sufficient homology can also be described by the predicted ability of two polynucleotides to specifically hybridize under high stringency conditions, see, for example, Sambrook et al., (1989) *Molecular Cloning: A Laboratory Manual*, (Cold Spring Harbor Laboratory Press, NY); *Current Protocols in Molecular Biology*, Ausubel et al., Eds (1994) *Current Protocols*, (Greene Publishing Associates, Inc. and John Wiley & Sons, Inc); and, Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, (Elsevier, New York).

[0111] A "phenotypic marker" is a screenable or selectable marker that includes visual markers and selectable markers whether it is a positive or negative selectable marker. Any phenotypic marker can be used. Specifically, a selectable or screenable marker comprises a DNA segment that allows one to identify, select for, or screen for or against a molecule or a cell that contains it, often under particular conditions. These markers

can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like.

[0112] Examples of selectable markers include, but are not limited to, DNA segments
5 that comprise restriction enzyme sites; DNA segments that encode products which
provide resistance against otherwise toxic compounds and antibiotics, such as,
chlorimuron ethyl, benomyl, Basta, and hygromycin phosphotransferase (HPT); DNA
segments that encode products which are otherwise lacking in the recipient cell (e.g.,
tRNA genes, auxotrophic markers, dominant heterologous marker-*amdS*); DNA
10 segments that encode products which can be readily identified (e.g., phenotypic
markers such as β -galactosidase, GUS; fluorescent proteins such as green fluorescent
protein (GFP), cyan (CFP), yellow (YFP), red (RFP), and cell surface proteins); the
generation of new primer sites for PCR (e.g., the juxtaposition of two DNA sequence
not previously juxtaposed), the inclusion of DNA sequences not acted upon or acted
15 upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.;
and, the inclusion of a DNA sequences required for a specific modification (e.g.,
methylation) that allows its identification.

Methods and compositions for Modifying a Fungal Cell Genome

20 **[0113]** Methods are provided employing a guide RNA /Cas endonuclease system for
modifying the DNA sequence at a target site in the genome of a fungal cell, e.g., a
filamentous fungal cell.

[0114] Aspects of the present disclosure include methods for modifying the DNA
sequence at a target site in the genome of a fungal cell by transiently introducing a Cas
25 endonuclease/guide polynucleotide complex into the cell. The Cas endonuclease/
guide polynucleotide complex is capable of introducing a double-strand break at the
target site in the genome of the fungal cell, and repair of this break can result in
sequence modification (e.g., insertions or deletions).

[0115] Introduction of the Cas endonuclease or guide polynucleotide (or other
30 biomolecule) can be done in any convenient manner, including transfection,
transduction, transformation, electroporation, particle bombardment (biolistic particle
delivery), cell fusion techniques, etc. Each of these components can be introduced
simultaneously or sequentially as desired by the user. For example, a fungal cell can

first be stably transfected with a Cas expression DNA construct followed by introduction of a guide polynucleotide into the stable transfectant (either directly or using a guide polynucleotide expressing DNA construct). This set up may even be advantageous as the user can generate a population of stable Cas transfectant fungal cells into which
5 different guide polynucleotides can be introduced independently (in some cases, more than one guide polynucleotide can be introduced into the same cells should this be desired). In some embodiments, a Cas expressing fungal cell is obtained by the user, and thus the user does not need to introduce a recombinant DNA construct capable of expressing a Cas endonuclease into the cell, but rather only need introduce a guide
10 polynucleotide into the Cas expressing cell.

[0116] In certain embodiments, a guide polynucleotide is introduced into the fungal cell by introducing a recombinant DNA construct that includes an expression cassette (or gene) encoding the guide polynucleotide. In some embodiments, the expression cassette is operably linked to a eukaryotic RNA pol III promoter. These promoters are
15 of particular interest as transcription by RNA pol III does not lead to the addition of a 5' cap structure or polyadenylation that occurs upon transcription by RNA polymerase II from an RNA pol II dependent promoter. In certain embodiments, the RNA pol III promoter is a filamentous fungal cell U6 polymerase III promoter (e.g., SEQ ID NO:11 and functional variants thereof, e.g., SEQ ID NO:12).

[0117] When a double-strand break is induced in the genomic DNA of a host cell (e.g., by the activity of a Cas endonuclease/guide RNA complex at a target site, the complex having double-strand endonuclease activity), the cell's DNA repair mechanism is activated to repair the break which, due to its error-prone nature, can produce mutations at double-strand break sites. The most common repair mechanism to bring
25 the broken ends together is the nonhomologous end-joining (NHEJ) pathway (Bleuyard et al., (2006) DNA Repair 5:1-12). The structural integrity of chromosomes is typically preserved by the repair, but deletions, insertions, or other rearrangements are possible (Siebert and Puchta, (2002) Plant Cell 14:1121-31; Pacher et al., (2007) Genetics 175:21-9).

[0118] Surprisingly, we have found in filamentous fungi that non-homologous insertion of transformed DNA at the double-strand break is highly favored over simple end-joining between the two ends of the chromosomal DNA at a double-strand break. Therefore, in cases where the Cas endonuclease or guide RNA is provided by transformation with an

expression cassette containing DNA construct or constructs, those DNA constructs, or fragments thereof, are inserted at the double-strand break at high frequency. This insertion occurs in the absence of homology between DNA sequences on the Cas endonuclease or guide RNA expression constructs and the sequences around the
5 double-strand break.

[0119] DNA taken up by transformation may integrate in a stable fashion in the genome or it may be transiently maintained. Transient maintenance can be recognized by an unstable phenotype. For example, DNA uptake can be recognized by selection for a marker gene present on the transforming DNA. After transformation and selection, the
10 transformants may be grown under non-selective conditions for several generations before transfer back to selective conditions. A stable transformant will be able to grow after transfer back to selective conditions whereas an unstable transformant will be unable to grow after transfer back to selective conditions due to loss of the transforming DNA. We have demonstrated that it is possible to transiently express Cas
15 endonuclease and/or guide RNA in fungal cells.

[0120] In embodiments where unstable transformants are desired, a plasmid with telomere sequences to encourage autonomous replication can be used. Other types of plasmids that are designed for autonomous replication, such as those with autonomous replication sequences, centromere sequences or other sequences, can also be
20 employed. Surprisingly, in *Trichoderma reesei* we have found that one can use plasmids with no known origin of replication, autonomous replication sequence, centromere or telomere sequences. By screening those transformants that show an unstable phenotype with respect to the selectable marker, efficient target site gene modification without vector DNA insertion is obtained.

[0121] Certain embodiments of the present disclosure include integrating a Cas endonuclease expression cassette and first selectable marker in the genome of a fungus, optionally flanked by repeats to allow subsequent removal (loop-out) of the expression cassette and first selectable marker, to produce a Cas endonuclease expressing host cell. These cells can be employed in numerous ways to obtain a
25 genetic modification of interest, including modification of the DNA sequence at a desired target site.
30

[0122] For example, a Cas endonuclease expressing host cell can be transformed with a DNA construct including a guide RNA expression cassette containing a second

selectable marker. Host cells that are selected for using the second selectable marker will express the guide RNA from this DNA construct, which enables Cas endonuclease activity and targeting to a defined target site of interest in the genome. Screening these host cells for transformants that show an unstable phenotype with respect to the
5 second selectable marker will enable obtaining host cells with a modified site of interest without DNA construct insertion.

[0123] As another example, a Cas endonuclease expressing host cell can be induced to uptake an in vitro synthesized guide RNA to enable Cas endonuclease activity and targeting to a defined site in the genome. In some cases, it will be desirable to induce
10 uptake of both guide RNA and a separate DNA construct bearing a selectable marker gene to allow for selection of those cells that have taken up DNA and, at high frequency, are expected to have simultaneously taken up guide RNA. As above, screening those transformants that show an unstable phenotype with respect to the selectable marker for the genetic modification of interest without vector DNA insertion is
15 obtained.

[0124] As yet another example, a Cas endonuclease expressing host cell can be used to create a "helper strain" that can provide, in trans, the Cas endonuclease to a "target strain". In brief, a heterokaryon can be created between the helper strain and the target strain, e.g., by fusion of protoplasts from each strain or by anastomosis of hyphae
20 depending on the species of filamentous fungus. Maintenance of the heterokaryon will depend on appropriate nutritional or other marker genes or mutations in each parental strain and growth on suitable selective medium such that the parental strains are unable to grow whereas the heterokaryon, due to complementation, is able to grow. Either at the time of heterokaryon formation or subsequently, a guide RNA is introduced
25 by transfection. The guide RNA may be directly introduced or introduced via a DNA construct having a Cas endonuclease expression cassette and a selectable marker gene. Cas endonuclease is expressed from the gene in the helper strain nucleus and is present in the cytoplasm of the heterokaryon. The Cas endonuclease associates with the guide RNA to create an active complex that is targeted to the desired target site(s)
30 in the genome to induce modification of the DNA sequence. Subsequently, spores are recovered from the heterokaryon and subjected to selection or screening to recover the target strain with modification of the DNA sequence at the target site. In cases in which an expression cassette is used to introduce the guide RNA, heterokaryons are chosen in which the guide RNA expression construct is not stably maintained.

[0125] In certain embodiments, the Cas endonuclease is a Cas9 endonuclease (see, e.g., WO 2013141680 entitled “RNA-directed DNA Cleavage by the Cas9-crRNA Complex”). Examples of Cas9 endonucleases include those from *Streptococcus* sp. (e.g., *S. pyogenes*, *S. mutans*, and *S. thermophilus*), *Campylobacter* sp. (e.g., *C. jejuni*), *Neisseria* sp. (e.g., *N. meningitidis*), *Francisella* sp. (e.g., *F. novicida*), and *Pasteurella* sp. (e.g., *P. multocida*) (see, e.g., Cas9 endonucleases described in Fonfara et al., *Nucleic Acids Res.*, 2013, pages 1-14: incorporated herein by reference). In some embodiments, the Cas endonuclease is encoded by an optimized Cas9 endonuclease gene, e.g., optimized for expression in a fungal cell (e.g., Cas9 encoding genes containing SEQ ID NO:8, e.g., SEQ ID NO:9, as described below).

[0126] In certain instances, the Cas endonuclease gene is operably linked to one or more polynucleotides encoding nuclear localization signals such that the Cas endonuclease/guide polynucleotide complex that is expressed in the cell is efficiently transported to the nucleus. Any convenient nuclear localization signal may be used, e.g., a polynucleotide encoding an SV40 nuclear localization signal present upstream of and in-frame with the Cas codon region and a polynucleotide encoding a nuclear localization signal derived from the *T. reesei* blr2 (blue light regulator 2) gene present downstream and in frame with the Cas codon region. Other nuclear localization signals can be employed.

[0127] In certain embodiments of the disclosure, the guide polynucleotide is a guide RNA that includes a crRNA region (or crRNA fragment) and a tracrRNA region (or tracrRNA fragment) of the type II CRISPR/Cas system that can form a complex with a type II Cas endonuclease. As indicated above, the guide RNA/Cas endonuclease complex can direct the Cas endonuclease to a fungal cell genomic target site, enabling the Cas endonuclease to introduce a double strand break into the genomic target site. In some cases, the RNA that guides the RNA/ Cas9 endonuclease complex is a duplex that includes a crRNA and a separate tracrRNA. In other instances, the guide RNA is a single RNA molecule that includes both a crRNA region and a tracrRNA region (sometimes referred to herein as a fused guide RNA). One advantage of using a fused guide RNA versus a duplexed crRNA-tracrRNA is that only one expression cassette needs to be made to express the fused guide RNA.

[0128] Host cells employed in the methods disclosed herein may be any fungal host cells are from the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota

(as defined by Hawksworth et al., In, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in Hawksworth et al., supra) and all mitosporic fungi (Hawksworth et al., supra). In certain embodiments, the fungal host cells are yeast cells, e.g.,

5 Candida, Hansenula, Kluyveromyces, Pichia, Saccharomyces, Schizosaccharomyces, or Yarrowia cell. Species of yeast include, but are not limited to, the following: Saccharomyces carlsbergensis, Saccharomyces cerevisiae, Saccharomyces diastaticus, Saccharomyces douglasii, Saccharomyces kluyveri, Saccharomyces norbensis, Saccharomyces oviformis, Kluyveromyces lactis, and Yarrowia lipolytica cell.

10 In additional embodiments, the fungal cells are filamentous fungal cells including but not limited to species of Trichoderma, Penicillium, Aspergillus, Humicola, Chrysosporium, Fusarium, Neurospora, Myceliophthora, Hypocrea, and Emericella. For example, the filamentous fungi T. reesei and A. niger find use in aspects of the disclosed methods.

15 **[0129]** Virtually any site in a fungal cell genome may be targeted using the disclosed methods, so long as the target site includes the required protospacer adjacent motif, or PAM. In the case of the S. pyogenes Cas9, the PAM has the sequence NGG (5' to 3'; where N is A, G, C or T), and thus does not impose significant restrictions on the selection of a target site in the genome. Other known Cas9 endonucleases have
20 different PAM sites (see, e.g., Cas9 endonuclease PAM sites described in Fonfara et al., Nucleic Acids Res., 2013, pages 1-14: incorporated herein by reference).

[0130] The length of the target site can vary, and includes, for example, target sites that are at least 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides in length. It is further possible that the target site can be palindromic,
25 that is, the sequence on one strand reads the same in the opposite direction on the complementary strand. The cleavage site can be within the target sequence or the cleavage site could be outside of the target sequence. In another variation, the cleavage could occur at nucleotide positions immediately opposite each other to produce a blunt end cut or, in other cases, the incisions could be staggered to produce
30 single-stranded overhangs, also called "sticky ends", which can be either 5' overhangs, or 3' overhangs.

[0131] In some cases, active variant target sequences in the genome of the fungal cell can also be used, meaning that the target site is not 100% identical to the relevant

sequence in the guide polynucleotide (within the crRNA sequence of the guide polynucleotide). Such active variants can comprise at least 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the given target site, wherein the active variant target sequences retain biological activity and hence are capable of being recognized and cleaved by a Cas endonuclease. Assays to measure the double-strand break of a target site by an endonuclease are known in the art and generally measure the overall activity and specificity of the agent on DNA substrates containing recognition sites.

[0132] Target sites of interest include those located within a region of a gene of interest.

Non-limiting examples of regions within a gene of interest include an open reading frame, a promoter, a transcriptional regulatory element, a translational regulatory element, a transcriptional terminator sequence, an mRNA splice site, a protein coding sequence, an intron site, and an intron enhancing motif.

[0133] In certain embodiments, modification of the genome of the fungal cell results in a phenotypic effect that can be detected and, in many instances, is a desired outcome of the user. Non-limiting examples include acquisition of a selectable cell growth phenotype (e.g., resistance to or sensitivity to an antibiotic, gain or loss of an auxotrophic characteristic, increased or decreased rate of growth, etc.), expression of a detectable marker (e.g., fluorescent marker, cell-surface molecule, chromogenic enzyme, etc.), and the secretion of an enzyme whose activity can be detected in culture supernatant.

[0134] In some instances, the genomic modification in the fungal cells is detected directly using any convenient method, including sequencing, PCR, Southern blot, restriction enzyme analysis, and the like, including combinations of such methods.

[0135] In some embodiments, specific genes are targeted for modification using the disclosed methods, including genes encoding enzymes, e.g., acetyl esterases, aminopeptidases, amylases, arabinases, arabinofuranosidases, carboxypeptidases, catalases, cellulases, chitinases, cutinase, deoxyribonucleases, epimerases, esterases, α -galactosidases, β -galactosidases, α -glucanases, glucan lysases, endo- β -glucanases, glucoamylases, glucose oxidases, α -glucosidases, β -glucosidases, glucuronidases, hemicellulases, hexose oxidases, hydrolases, invertases, isomerases, laccases, lipases, lyases, mannosidases, oxidases, oxidoreductases, pectate lyases, pectin acetyl esterases, pectin depolymerases, pectin methyl esterases, pectinolytic

enzymes, peroxidases, phenoloxidases, phytases, polygalacturonases, proteases, rhamno-galacturonases, ribonucleases, transferases, transport proteins, transglutaminases, xylanases, hexose oxidases, and combinations thereof.

[0136] There are numerous variations for implementing the methods described herein.

5 For example, instead of having the Cas expression cassette present as an exogenous sequence in the fungal host cell, this cassette can be integrated into the genome of the fungal host cell. Generating this parental cell line would allow a user to simply introduce a desired guide RNA (e.g., as a guide RNA expression vector) which would then target the genomic site of interest as detailed elsewhere herein. In some of these
10 embodiments, the integrated Cas gene can be designed to include polynucleotide repeats flanking it for subsequent loop-out /removal from the genome if needed.

[0137] Non-limiting examples or embodiments of compositions and methods disclosed herein are as follows:

1. A method for modifying the DNA sequence at a target site in the genome
15 of a filamentous fungal cell, the method comprising:
 - a) introducing into a population of filamentous fungal cells a Cas endonuclease and a guide RNA, wherein the Cas endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cells;
20 and
 - b) identifying at least one fungal cell from the population that has a modification of the DNA sequence at the target site, wherein the Cas endonuclease, the guide RNA, or both are introduced transiently into the population of fungal cells.
- 25 2. The method of embodiment 1, wherein the modification of the DNA sequence at said target site is selected from the group consisting of a deletion of one or more nucleotides, an insertion of one or more nucleotides, a substitution of one or more nucleotides, and any combination thereof.
- 30 3. The method of embodiment 1 or 2, wherein introducing the Cas endonuclease into the population of fungal cells is achieved using a method selected from the group consisting of transfection, transduction, transformation,

electroporation, particle bombardment (biolistic particle delivery), and cell fusion techniques.

4. The method of any preceding embodiment, wherein introducing the guide RNA into the population of fungal cells is achieved using a method selected from the group consisting of transfection, transduction, transformation, electroporation, particle bombardment (biolistic particle delivery), and cell fusion techniques.

5. The method of any preceding embodiment, wherein the identifying step comprises culturing the population of fungal cells from step (a) under conditions to select for or screen for the modification of the DNA sequence at the target site.

6. The method of any preceding embodiment, wherein the identifying step comprises culturing the population of cells from step (a) under conditions to screen for unstable transformants.

7. The method of any preceding embodiment, wherein the Cas endonuclease is a Cas9 endonuclease or variant thereof.

8. The method of embodiment 7, wherein the Cas9 endonuclease or variant thereof comprises a full length Cas9 or a functional fragment thereof from a species selected from the group consisting of: *Streptococcus* sp., *S. pyogenes*, *S. mutans*, *S. thermophilus*, *Campylobacter* sp., *C. jejuni*, *Neisseria* sp., *N. meningitides*, *Francisella* sp., *F. novicida*, and *Pasteurella* sp., *P. multocida*.

9. The method of embodiment 8, wherein the Cas9 endonuclease or variant thereof comprises an amino acid sequence that has at least 70% identity to any one of SEQ ID NOs:1 to 7 or a functional fragment thereof.

10. The method of any preceding embodiment, wherein the introducing step comprises introducing a DNA construct comprising an expression cassette for the Cas endonuclease into the fungal cells.

11. The method of any preceding embodiment, wherein the introducing step comprises introducing a DNA construct comprising an expression cassette for the guide RNA into the fungal cells.

5

12. The method of any one of embodiments 1 to 9 and 11, wherein the introducing step comprises directly introducing the Cas endonuclease into the fungal cells.

10

13. The method of any one of embodiments 1 to 10 and 12, wherein the introducing step comprises directly introducing the guide RNA into the fungal cells.

15

14. The method of embodiment 10, wherein the expression cassette for the Cas endonuclease comprises a Cas coding sequence that is optimized for expression in the fungal cell.

20

15. The method of embodiment 14, wherein the Cas coding sequence is a Cas9 coding sequence comprising a polynucleotide sequence that is at least 70% identical to SEQ ID NO:8 or a functional fragment thereof.

16. The method of any preceding embodiment, wherein the Cas endonuclease is operably linked to a nuclear localization signal.

25

17. The method of embodiment 11, wherein the expression cassette for the guide RNA comprises a RNA polymerase III dependent promoter functional in a Euscomycete or Pezizomycete, and wherein the promoter is operably linked to the DNA encoding the guide RNA.

30

18. The method of embodiment 17, wherein the promoter is derived from a Trichoderma U6 snRNA gene.

19. The method of embodiment 17 or 18, wherein the promoter comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%,

96%, 97%, 98%, or 99% identity to SEQ ID NO: 11 or 12 or a functional fragment thereof.

20. The method of embodiment 19, wherein the promoter comprises the sequence of SEQ ID NO: 11 or 12.

21. The method of any one of embodiments 11 and 17-20, wherein the expression cassette for the guide RNA comprises a guide RNA-encoding DNA with an intron sequence from a *Trichoderma* U6 snRNA gene.

22. The method of embodiment 21, wherein the intron sequence derived from *Trichoderma* U6 snRNA gene comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 90 or a functional fragment thereof.

23. The method of embodiment 22, wherein the intron sequence derived from *Trichoderma* U6 snRNA gene comprises the sequence of SEQ ID NO: 90.

24. The method of any preceding embodiment, wherein the filamentous fungal cell is a *Eumycotina* or *Pezizomycotina* fungal cell.

25. The method of any preceding embodiment, wherein filamentous fungal cell is selected from the group consisting of *Trichoderma*, *Penicillium*, *Aspergillus*, *Humicola*, *Chrysosporium*, *Fusarium*, *Myceliophthora*, *Neurospora*, *Hypocrea*, and *Emericella*.

26. The method of any preceding embodiment, wherein the target site is located within a region of a gene of interest selected from the group consisting of: an open reading frame, a promoter, a regulatory sequence, a terminator sequence, a regulatory element sequence, a splice site, a coding sequence, a polyubiquitination site, an intron site, and an intron enhancing motif.

27. The method of any one of embodiments 1, 2, 4-9, 11, 13, and 16-19, wherein the introducing step comprises: (i) obtaining a parental fungal cell

population that stably expresses the Cas endonuclease, and (ii) transiently introducing the guide RNA into the parental fungal cell population.

28. The method of any one of embodiments 1-3, 5-10, 12, and 14-19, wherein the introducing step comprises: (i) obtaining a parental fungal cell population that stably expresses the guide RNA, and (ii) transiently introducing the Cas endonuclease into the parental fungal cell population.

29. The method of any preceding embodiment, wherein the modification of the DNA sequence at the target site is not caused by a homologous recombination.

30. The method of any preceding embodiment, wherein the method does not involve introducing a donor DNA into the population of fungal cells.

31. A recombinant fungal cell produced by the method of any preceding embodiment.

32. An engineered nucleic acid encoding a Cas endonuclease or variant thereof, wherein the Cas endonuclease or variant thereof comprises an amino acid sequence that has at least 70%, 75%, 80%, 85%, 90%, or 95% identity to any one of SEQ ID NOs:1 to 7 or a functional fragment thereof, and wherein the nucleic acid comprises a polynucleotide sequence that is at least 70% 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO:8 or a functional fragment thereof.

33. The engineered nucleic acid of embodiment 32, wherein the nucleic acid comprises the sequence of SEQ ID NO:8.

34. An engineered nucleic acid encoding a guide RNA which enables a Cas endonuclease to introduce a double-strand break at a target site in the genome of a filamentous fungal cell, wherein the nucleic acid encoding the guide RNA comprises a RNA polymerase III dependent promoter functional in a

Euascomycete or Pezizomycete, and the promoter is derived from a Trichoderma U6 snRNA gene

35. The engineered nucleic acid of embodiment 34, wherein the promoter comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 11 or 12 or a functional fragment thereof.

36. An engineered nucleic acid encoding a guide RNA which enables a Cas endonuclease to introduce a double-strand break at a target site in the genome of a filamentous fungal cell, wherein the nucleic acid encoding the guide RNA comprises a guide RNA-encoding DNA with an intron sequence derived from a Trichoderma U6 snRNA gene.

37. The engineered nucleic acid of embodiment 36, wherein the intron sequence derived from Trichoderma U6 snRNA gene comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 100% identity to SEQ ID NO: 90 or a functional fragment thereof.

38. The engineered nucleic acid of embodiment 34 or 36, wherein the nucleic acid encoding the guide RNA comprises both a promoter derived from a Trichoderma U6 snRNA gene and an intron sequence derived from a Trichoderma U6 snRNA gene, wherein the promoter comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 11 or 12 or a functional fragment thereof, and wherein the intron sequence derived from Trichoderma U6 snRNA gene comprises a nucleotide sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 90 or a functional fragment thereof.

EXAMPLES

[0138] In the following Examples, unless otherwise stated, parts and percentages are by weight and degrees are Celsius. It should be understood that these Examples, while indicating embodiments of the disclosure, are given by way of illustration only. From

the above discussion and these Examples, one skilled in the art can make various changes and modifications of the disclosure to adapt it to various usages and conditions. Such modifications are also intended to fall within the scope of the appended claims.

5

Section A: Introduction of Cas/guide RNA by Expression Vectors

Example 1: Identification of *T. reesei* U6 snRNA gene

[0139] An RNA polymerase III directed promoter is desired for production of guide RNA in *T. reesei* without the addition of a 5' cap structure or polyadenylation that would result from the use of a RNA polymerase II dependent promoter. However, no RNA polymerase III dependent promoter that is functional in *T. reesei* has been described. Known RNA polymerase III dependent promoters from other species were considered to be tested for their ability to function in *T. reesei* including the 5' upstream regions from the *Saccharomyces cerevisiae* *snr52* gene, the human U6 snRNA gene, or the corn U6 snRNA gene.

[0140] More desirable was to identify a native *T. reesei* sequence that would function as an RNA polymerase III dependent promoter. The DNA sequence encoding the human U6 small nuclear RNA (snRNA; GenBank accession number M14486) was used to search the *T. reesei* v2 genome sequence (www.jgi.doe.gov) using the BLAST algorithm. A short region of *T. reesei* DNA sequence was identified with similarity to the human sequence. Examination of the surrounding DNA sequence and comparison with the U6 genes of yeasts, particularly *Schizosaccharomyces pombe* (Marck et al., 2006, *Nucleic Acids Research* 34:1816-1835), allowed a number of features of the *T. reesei* U6 gene to be putatively identified (SEQ ID NO:22, shown below). The start of the transcribed sequence and the terminator were identified as were an upstream TATA box. An intron apparently interrupts the transcribed region and possible A-box and B-box promoter elements can be recognized within the transcribed region, the latter within the intron. (see FIG. 1).

AAAAAACA CTAGTAAGTACTTACTTATGTATTATTA ACTACTTTAGCTAACTTCTGCA
GTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTAT
TTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTATTATTATAATAT
ATATTATATAATAATTTTAAGCCTGGAATAGTAAAGAGGCTTATATAATAATTTATAG
TAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATT

ATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTGGCTATAAGTC
 TGCTGCAAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCT
 ATCGCCTTCGGGCATTTGGTCAATTTATAACGATACAGGTTTCGTTTCGGCTTTTCC
 TCGGAACCCCCAGAGGTCATCAGTTCGAATCGCTAACAGGTCAACAGAGAAGATT
 5 AGCATGGCCCCTGCACTAAGGATGACACGCTCACTCAAAGAGAAGCTAAACATTTT
 TTTTCTCTTCCAAGTCGTGATGGTTATCTTTTTTGCTTAGAGAATCTATTCTTGTGGA
 CGATTAGTATTGGTAAATCCCTGCTGCACATTGCGGCGGATGGTCTCAACGGGCAT
 AATACCCCATTCGTGATGCAGCGGTGATCTTCAATATGTAGTGTAATACGTTGCAT
 ACACCACCAGGTTTCGGTGCCTCCTGTATGTACAGTACTGTAGTTTCGACTCCTCCG
 10 CGCAGGTGGAAACGATTCCCTAGTGGGCAGGTATTTTGGCGGGGTCAAGAA (SEQ
 ID NO:22)

Example 2: sgRNA sequences to target *T. reesei* genes

[0141] It has been shown that a single guide RNA (sgRNA) molecule can interact with
 15 the *Streptococcus pyogenes* Cas9 protein to target this endonuclease in vivo to a
 specific locus in a eukaryote genome (REFS). The sgRNA is a hybrid molecule
 designed as a fusion between the tracrRNA and crRNA observed naturally to be
 components of the *Streptococcus pyogenes* type II CRISPR-Cas system (Gasiunas et
 al. (2012) Proc. Natl. Acad. Sci. USA 109:E2579-86, Jinek et al. (2012) Science
 20 337:816-21, Mali et al. (2013) Science 339:823-26, and Cong et al. (2013) Science
 339:819-23). The first 20 nucleotides of the sgRNA are complementary to the target site
 in the genome. An additional sequence (PAM, protospacer adjacent motif) is also
 required to be present at the target site in the genome adjacent to the sgRNA-
 complementary region. In the case of the *S. pyogenes* Cas9 the PAM has the sequence
 25 NGG (where N is A, G, C or T).

[0142] The sequence of sgRNA used in these experiments is shown below where the
 20 nucleotides designed to be complementary to the target site are shown as N
 residues (SEQ ID NO:23) (N = A, G, C, or U).

NNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGG
 30 CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0143] sgRNAs were designed to target different loci in the *T. reesei* genome. The
 sequence of an sgRNA (called gAd3A TS1) to target the *T. reesei* ad3A gene
 (Phosphoribosylamidoimidazole-succinocarboxamide synthase) at a site designated as

target site 1 (TS1) is shown below (SEQ ID NO:24). The 20 nucleotide region that is complementary to the *T. reesei* genome sequence is shown in lower case.

guccucgagcaaaaggugccGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU
CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

5

[0144] The sequence of an sgRNA (called gTrGA TS2) to target the *T. reesei* gla1 (glucoamylase) gene at a site designated as target site 2 (TS2) is shown below (SEQ ID NO:25). The 20 nucleotide region that is complementary to the *T. reesei* genome sequence is shown in lower case.

10 guucagugcaauaggcgucuGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU
CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

[0145] The sequence of an sgRNA (called gTrGA TS11) to target the *T. reesei* gla1 (glucoamylase) gene at a site designated as target site 11 (TS11) is shown below (SEQ
15 ID NO:26). The 20 nucleotide region that is complementary to the *T. reesei* genome sequence is shown in lower case.

gccaauggcgacggcagcacGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU
CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

20 **[0146]** The sequence of an sgRNA (called gPyr2 TS6) to target the *T. reesei* pyr2 (orotate phosphoribosyltransferase) gene at a site designated as target site 6 (TS6) is shown below (SEQ ID NO:27). The 20 nucleotide region that is complementary to the *T. reesei* genome sequence is shown in lower case.

gcacagcgggaugcccuuguGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU
25 CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

Example 3: Cas9 DNA and protein sequences for expression in T. reesei

[0147] A codon optimized *Streptococcus pyogenes* Cas9-encoding gene, including NLS sequences, was designed, synthesized and tested for expression in *T. reesei* (SEQ ID
30 NO:9). The encoded protein (SEQ ID NO:10) has an N- terminal SV40 nuclear localization signal (NLS; SEQ ID NO:19) and a C-terminal NLS derived from the *T. reesei* blr2 (blue light regulator 2) gene (SEQ ID NO:20; both are underlined in SEQ ID NO:10 below).

SEQ ID NO:9

atggcaccgaagaagaagcgcaaggtgatggacaagaagtacagcatcggcctcgacatcggcaccaactcgggtgg
 gctgggcccgtcatcacggacgaatataaggtcccgtcgaagaagttcaaggtcctcggcaatacagaccgccacagca
 tcaagaaaaacttgatcggcgccctcctgttcgatagcggcgagaccgcggaggcgaccagggtcaagaggaccgcc
 5 aggagacggtacactagggcgcaagaacaggatctgctacctgcaggagatcttcagcaacgagatggcgaaggtgg
 acgactccttctccaccgcctggaggaatcattcctgggtggaggaggacaagaagcatgagcggcacccaatcttcgg
 caacatcgtcgacgaggtggcctaccacgagaagtacccgacaatctaccacctccggaagaaactgggtggacagca
 cagacaaggcggacctccggctcatctaccttgccctcgcgcatatgatcaagttccgcggccacttcctcatcgagggc
 gacctgaacccggacaactccgacgtggacaagctgttcacccagctcgtgcagacgtacaatcaactgttcgaggaga
 10 accccataaacgctagcggcgtggacgccaaggccatcctctcggccaggctctcgaaatcaagaaggctggagAAC
 cttatcgcgcagttgccaggcgaaaagaagaacggcctcttcggcaaccttattgcgctcagcctcggcctgacgccga
 acttcaaatacaacttcgacctcgcgaggacgccaagctccagctctcaaaggacacctacgacgacgacctcgaca
 acctcctggcccagataggagaccagtacgcggacctcttctcgcgcccaagaacctctccgacgctatcctgctcagc
 gacatccttcgggtcaacaccgaaattaccaaggcaccgctgtccgccagcatgattaaacgctacgacgagcaccatc
 15 aggacctcacgctgctcaaggcactcgtccgccagcagctccccgagaagtacaaggagatcttcttcgaccaatcaaa
 aaacggctacgcgggatataatcgacggcgggtgccagccaggaagagttctacaagttcatcaaaccaatcctggagaa
 gatggacggcacccgaggagtgtgtgtcaagctcaacaggaggacctcctcaggaagcagaggaccttcgacaac
 ggctccatcccgcacatcagatccacctggggcgaactgcatgccatcctgcggcgccaggaggacttctacccgttcctgaa
 ggataaccgggagaagatcgagaagatcttgacgttcgcacatccatactacgtgggcccgtggctcgcggaactcc
 20 cggttcgcctggatgacccggaagtcggaggagaccatcacacctggaactttgaggagggtggtcgataagggcgct
 agcgtcagagcttcacgagcgcacatgaccaacttcgataaaaacctgcccaatgaaaaagtcctccccaagcactcgc
 tgctctacgagtacttcaccgtgtacaacgagctaccaagggtcaaatacgtcaccgagggcatgcggaagccggcggtt
 cctgagcggcgagcagaagaaggcgatagtggacctcctcttcaagaccaacaggaagggtgacctgaagcaattaa
 aagaggactacttcaagaaaatagagtgttcgactccgtggagatctcgggcgtggaggatcgggtcaacgcctcactc
 25 ggcacgtatcacgacctcctcaagatcattaaagacaaggacttctcgacaacgaggagaacgaggacatcctcgag
 gacatcgtcctcacctgacctgttcgaggaccgcgaaatgatcgaggagaggctgaagacctacgcgcacctgttcg
 acgacaaggctcatgaaacagctcaagaggcgccgctacactgggtggggaaggctgtcccgcaagctcattaatggca
 tcagggacaagcagagcggcaagaccatcctggacttctcaagtcgcgacgggttcgccaaccgcaacttcacgcagc
 tcattcacgacgactcgtcacgttcaaggaagacatccagaaggcacagggtgagcgggcagggtgactccctccacg
 30 aacacatcgccaacctggccggctcgccggccattaaaaagggcacatcctgcagacgggtcaaggctcgtcgacgagctc
 gtgaagggtgatgggcccggcacaagcccgaataatcgtcatagagatggccagggagaaccagaccacccaaaaa
 gggcagaagaactcgcgcgagcggatgaaacggatcgaggagggcattaaagagctcgggtcccagatcctgaag
 gagcaccctgtgaaaataccagctccagaatgaaaagctctaccttactacctgcagaacggccgcgacatgtac
 gtggaccaggagctggacattaatcggctatcggactacgacgtcgaccacatcgtgccgcagtcgttctcaaggacg

atagcatcgacaacaaggtgctcaccggtcggataaaaatcggggcaagagcgacaacgtgcccagcgaggaggt
 cgtgaagaagatgaaaaactactggcgccagctcctcaacgcgaaactgatcaccagcgcaagttcgacaacctga
 cgaaggcggaacgcggtggcttgagcgaactcgataaggcgggcttcataaaaaggcagctggtcgagacgcgcca
 gatcacgaagcatgtcgcccagatcctggacagccgcatgaatactaagtacgatgaaaacgacaagctgatccggg
 5 aggtgaaggtgatcacgctgaagtccaagctcgtgtcggacttccgcaaggacttccagttctacaaggtccgcgagatc
 aacaactaccaccacgcccacgacgcctacctgaatgcggtggtcgggaccgcccctgatcaagaagtaccggaagct
 ggagtcggagttcgtgtacggcgactacaaggtctacgacgtgcgcaaaatgatcgccaagtccgagcaggagatcgg
 caaggccacggcaaaatacttcttctactcgaacatcatgaacttctcaagaccgagatcacctcgcgaacggcgag
 atccgcaagcgcccgtcatcgaaaccaacggcgagacgggcgagatcgtctgggataagggccgggatttcgcgac
 10 ggtccgcaaggtgctctccatgccgcaagtcaatatcgtgaaaaagacggaggtccagacgggcccgggttcagcaagg
 agtccatcctcccgaagcgcaactccgacaagctcatcgcgaggaagaaggattgggaccgaaaaaatatggcggc
 ttcgacagcccgaccgtcgcatacagcgtcctcgtcgtggcgaaggtggagaagggcaagtcaaagaagctcaagtcc
 gtgaaggagctgctcgggatcacgattatggagcgggtcctcctcgagaagaacccgatcgacttcttagaggccaagg
 gatataaggaggtcaagaaggacctgattattaaactgccgaagtactcgctcttcgagctggaaaacggccgcaaga
 15 ggatgctcgctccgcaggcgagttgcagaagggcaacgagctcgccctcccagcaataacgtcaatttctgtacctc
 gctagccactatgaaaagctcaagggcagcccggaggacaacgagcagaagcagctcttcgtggagcagcacaag
 cattacctggacgagatcatcgagcagatcagcgagttctcgaagcgggtgatcctcgccgacgcgaacctggacaag
 gtgctgtcggcatataacaagcacccgcgacaaaccaatacgcgagcaggccgaaaatatcatccacctcttcacctca
 ccaacctcggcgctccggcagccttcaagtacttcgacaccacgattgaccggaagcgggtacacgagcacgaaggag
 20 gtgctcgatgcgacgctgatccaccagagcatcacagggtctatgaaacacgcatcgacctgagccagctgggcccga
 gacaagaagaagaagctcaagctctag

SEQ ID NO:10

MAPKKKRKVMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 25 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEE
 DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHMIKFRGHFLI
 EGDLPDNDSDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLP
 GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
 FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 30 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
 SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPIYYVGPLARGNSRFAWMTRK
 SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
 KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR
 FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV

MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
 EDIQKAQVSGQGDSLHEHIANLAGSPAIAKKGILQTVKVDELVKVMGRHKPENIVIEMA
 RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
 MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
 5 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
 MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
 IKKYPKLESEFVYGDYKVYDVRKMIKSEGEIGKATAKYFFYSNIMNFFKTEITLANGEI
 RKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRN
 SDKLIARKKDWDPKKYGGFDSPTVAYSVLVAKVEKGKSKKLKSVKELLGITIMERS
 10 EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
 NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL
 YNKHARDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
 SITGLYETRIDLSQLGGDKKKKLKL

15 Example 4: Construction of Cas9 expression vectors

[0148] The synthetic DNA sequence encoding Cas9 shown above was inserted into pENTR/D-TOPO so that it would be between flanking attL1 and attL2 sites to enable transfer by Gateway cloning (Invitrogen) into suitable expression vectors. A Gateway compatible expression vector, pTrex2gHyg, was available that comprises the following
 20 features; the promoter region from the *T. reesei* pki1 (pyruvate kinase) gene and terminator region from the *T. reesei* cbh1 (cellobiohydrolase I) gene separated by Gateway cloning sites, a bacterial hygromycin phosphotransferase gene functionally linked to the *Neurospora crassa* cpc1 (cross pathway control 1) promoter region and the *Aspergillus nidulans* trpC (trifunctional protein with glutamine amido transferase,
 25 indoleglycerolphosphate synthase and phosphoribosylanthranilate isomerase activity) terminator region, and bacterial vector sequences for selection and maintenance in *E. coli*. The cas9 gene was cloned into pTrex2gHyg using the Gateway cloning procedure (Invitrogen) to give pTrex2gHyg MoCas (see FIG. 2).

30 Example 5: Construction of sgRNA expression vectors

[0149] Synthetic DNA sequences were obtained that encode the gAd3A TS1 sgRNA flanked by different putative RNA polymerase III dependent promoters and terminators. Each of these synthetic DNA sequences also had restriction enzyme recognition sites (EcoRI and BamHI) at either end.

[0150] The following sequence encodes the gAd3A TS1 sgRNA (underlined) with the *Saccharomyces cerevisiae* snr52 promoter and *S. cerevisiae* sup4 terminator (denoted gAd3A TS1-1; SEQ ID NO:28):

gaattcggatccTCTTTGAAAAGATAATGTATGATTATGCTTTCACTCATATTTATACAGA
 5 AACTTGATGTTTTCTTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACA
 ACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTTCAC
 ACCCTACAATGTTCTGTTCAAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTG
 GTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCggtcctcg
agcaaaagggtgccGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
 10 AACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTTGTTTTTTATGTCTgaattcg
 gatcc

[0151] The following sequence encodes the gAd3A TS1 sgRNA (underlined) with the *T. reesei* U6 promoter and terminator (denoted gAd3A TS1-2; SEQ ID NO:29):

gaattcggatccAAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTA
 15 ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA
 TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTT
 ATTATAATATATATTATATAATAATTTTAAAGCCTGGAATAGTAAAGAGGCTTATATAA
 TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG
 20 AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG
 GCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTG
 ATGGTAGTCTATCggtcctcgagcaaaagggtgccGTTTTAGAGCTAGAAATAGCAAGTTAAA
TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTT
 TTCTCTTgaattcggatcc

25

[0152] The following sequence encodes the gAd3A TS1 sgRNA (underlined) with the *T. reesei* U6 promoter, terminator and an intron (in italics) (denoted gAd3A TS1-3; SEQ ID NO:30):

gaattcggatccAAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTA
 30 ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA
 TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTT
 ATTATAATATATATTATATAATAATTTTAAAGCCTGGAATAGTAAAGAGGCTTATATAA
 TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG

AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG
 GCTATAAGTCTGCTGCAAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTG
 ATGGTAGTCTATCgctcctcgagcaaaagggtgccGTTTTAGAGCTAGAGTTCGTTTCGGCTTT
TCCTCGGAACCCCCAGAGGTCATCAGTTCGAATCGCTAACAGAATAGCAAGTTAAA
 5 ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTT
 TTTCTCTTgaattcgatcc

[0153] Plasmid p219M (FIG. 3) is an E. coli vector containing the T. reesei pyr4
 (orotidine monophosphate decarboxylase) gene including its native promoter and
 10 terminator. This vector was digested with EcoRI and BamHI and the ends were
 dephosphorylated. Each of the above synthetic DNA molecules was digested with
 EcoRI and BamHI and ligated with the cut p219M to create a series of vectors
 containing an sgRNA expression cassette and the pyr4 gene. Each vector was
 designated by the name of the sgRNA that it encoded (for example, p219M gAd3A TS1-
 15 1 incorporates the gAd3A expression cassette with the S. cerevisiae snr52 promoter
 and sup4 terminator).

[0154] Guide RNA expression cassettes with a shorter T. reesei U6 promoter region
 were obtained as synthetic DNA. An example is provided here that includes the
 sequence for an sgRNA targeting the T. reesei gla1 gene at TS11 (SEQ ID NO:31;
 20 intron sequence is underlined).

AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCA
 CTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAAACTACCCCCAACCTCG
 TAGGTATATAAGTACTGTTTGATGGTAGTCTATCgccaatggcgacggcagcacGTTTTAGA
 GCTAGAGTTCGTTTCGGCTTTTCCTCGGAACCCCCAGAGGTCATCAGTTCGAATC
 25 GCTAACAGAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
 CACCGAGTCGGTGGTGCTTTTTTTTTCTCTT

[0155] The above gRNA expression cassette was amplified by PCR using primers
 gRNA fwd aflIII (5'- cgtcagcttaagaattcctaaagAAACAGCATGAAATGG; SEQ ID NO:32)
 30 and gRNA rev sfil (5'-
 cgtcagggccacgtgggcccAAGAGAAAAAAAAGCACCCGACTCGG; SEQ ID NO:33).
 These primers add an aflIII to the 5' end and an sfil site to the 3' end of the guide RNA
 expression cassette. The PCR product was purified using a Qiagen PCR Purification Kit
 according to the manufacturer's directions. The PCR product was then digested with

SfiI and AflII and cleaned again on a Qiagen PCR Purification Kit. Plasmid pTrex2g/Hyg MoCas was digested with SfiI and AflII and dephosphorylated using the Roche Rapid alkaline phosphatase kit (Roche Diagnostics Corp., IN). The digested plasmid and PCR product were finally ligated using the Roche Rapid DNA ligase kit to create

5 pTrex2g/Hyg MoCas gTrGA TS11B. Other sgRNA expression cassettes were inserted into pTrex2g/Hyg MoCas in a similar manner.

Example 6: Cas9-mediated gene inactivation in Trichoderma reesei

[0156] A series of experiments are described below in which a *Trichoderma reesei* strain is either co-transformed with two separate expression vectors, one for production of Cas9 and one for production of gRNA, or is transformed with a single vector for expression of both Cas9 and gRNA. These experiments demonstrate that the 5' upstream region from the *T. reesei* U6 gene promotes gRNA transcription only when the U6 intron is also present within the gRNA transcribed region. The experiments also demonstrate that targeted gene inactivation can occur with high efficiency in *T. reesei* transformants.

Inactivation of the ad3A gene

[0157] A strain of *Trichoderma reesei* derived from the publicly available strain RL-P37 in which the genes (*cbh1*, *cbh2*, *egl1*, and *egl2*) encoding the four major secreted cellulases were deleted was used. This strain also lacked a functional *pyr4* gene. Biolistic transformation (as described in US20060003408A1) was used to co-transform with a mixture of equal amounts of pTrex2gHyg MoCas and either p219M gAd3ATS1-1, p219M gAd3ATS1-2 or p219M gAd3ATS1-3. Transformants were selected on agar plates with Vogel's minimal medium containing 2% glucose, 100 mg/L hygromycin B and 200 mg/L adenine. After selection on the first plates transformant colonies were picked to fresh plates of the same selective medium. During growth on the second plate it was possible to distinguish between stable and unstable hygromycin-resistant transformants. Stable transformants grew more rapidly, the colonies had a smooth outline and the mycelium was more dense. Unstable transformants grew slower, had less dense mycelium and colonies had a ragged irregular outline. After growth on the second plate transformants were transferred to Vogel's medium with glucose, without hygromycin and with 14mg/L adenine to screen for those which exhibited a red/brown color indicating that they were adenine auxotrophs. Five stable and 23 unstable

transformants were obtained with p219M gAd3ATS1-1 and all were adenine prototrophs. Eleven stable and 38 unstable transformants were obtained with p219M gAd3ATS1-2 and all 11 stable and 29 of the unstable transformants were adenine prototrophs. Nineteen stable and 2 unstable transformants were obtained with p219M gAd3ATS1-3 and all were adenine auxotrophs. Clearly, adenine auxotrophs were only obtained with gAd3ATS1-3 that utilizes the T. reesei U6 promoter, intron and terminator to control transcription of sgAd3A TS1. Adenine auxotrophy indicates targeted Cas9 cleavage at the native T. reesei ad3A locus. It can be concluded that Cas9-mediated gene inactivation is efficient because all transformants with gAd3ATS1-3 that were tested were adenine auxotrophs.

[0158] In order to determine the mutations at the ad3A locus in co-transformants with pTrex2gHyg MoCas and p219M gAd3ATS1-3 genomic DNA was extracted from 10 stable adenine auxotrophic transformants. This DNA was used as template for PCR using several different primer pairs designed to generate products that spanned the Cas9 target site or were upstream or downstream of the target site. PfuUltra II Fusion HS DNA polymerase (Agilent Technologies) was used for the PCR according to the manufacturer's directions. In each case, the extension time was that suggested by the manufacturer for the expected size of the PCR product as described below. The sizes of the PCR products were evaluated by agarose gel electrophoresis.

[0159] A PCR product of the expected size (872 bp) was obtained in all transformants using Ad3 5' fwd + Ad3 5' rev primers (5'- tgaacacagccaccgacatcagc [SEQ ID NO:34] and 5'- gctggtgaggggttgctattg [SEQ ID NO:35] respectively) that amplify a region on the 5' side of the TS1 target site.

[0160] A PCR product of the expected size (1214 bp) was obtained in all transformants using Ad3 5' fwd + Ad3a 5005 rev primers (5'- tgaacacagccaccgacatcagc [SEQ ID NO:34] and 5'- gattgcttgggaggaggacat [SEQ ID NO:36] respectively) that amplify a region on the 5' side of the TS1 target site.

[0161] A PCR product of the expected size (904 bp) was obtained in all transformants using Ad3 3' fwd + Ad3 3' rev primers (5'- cgaggccactgatgaagttgttc [SEQ ID NO:37] and 5'- cagttttccaaggctgccaacgc [SEQ ID NO:38] respectively) that amplify a region on the 3' side of the TS1 target site.

[0162] A PCR product of the expected size (757 bp) was obtained in all transformants using Ad3a 5003 fwd + Ad3mid rev primers (5'- ctgatcttgccacctggaaatc [SEQ ID NO:39]

and 5'- ctctctatcatttgccaccctcc [SEQ ID NO:40] respectively) that amplify a region on the 3' side of the TS1 target site.

[0163] The above PCR results demonstrated that the genomic DNA preparations were of a quality sufficient to obtain PCR products from either upstream or downstream of the Cas9 target site.

[0164] No PCR product could be obtained for any transformants using Adfrag fwd + Adfrag rev primers (5'- ctccattcaccctcaattctcc [SEQ ID NO:41] and 5'- gttcccttggcgggtgcttgatc [SEQ ID NO:42] respectively) spanning the TS1 target site in ad3A. The expected size for this PCR product presuming no large size change caused by Cas9 activity was approximately 764 bp.

[0165] No PCR product could be obtained for any transformants using Adfrag fwd + Ad3 3' rev primers (5'- ctccattcaccctcaattctcc [SEQ ID NO:41] and 5'- cagttttccaaggctgccaacgc [SEQ ID NO:38] respectively) spanning the TS1 target site in ad3A. The expected size for this PCR product presuming no large size change caused by Cas9 activity was approximately 2504 bp.

[0166] No PCR product could be obtained for any transformants using Ad3a 2k fwd + Ad3a 2k rev primers (5'- caatagcacaaccctcaccagc [SEQ ID NO:43] and 5'- gaacaacttcacagtggcctcg [SEQ ID NO:44] respectively) spanning the TS1 target site in ad3A. The expected size for this PCR product presuming no large size change caused by Cas9 activity was approximately 1813 bp.

[0167] Five of the transformants also gave no PCR product using Adfrag fwd + Ad3 mid rev primers (5'- ctccattcaccctcaattctcc [SEQ ID NO:41] and 5'- ctctctatcatttgccaccctcc [SEQ ID NO:40] respectively) spanning the TS1 target site. The expected size for this PCR product presuming no large size change caused by Cas9 activity was approximately 1438 bp.

[0168] Based on published data, Cas9-mediated inactivation of genes typically involves error-prone repair of a double-strand break in the DNA at the target site. The end result is small deletions or insertions (indels) at the target site. The above results from PCR analysis were surprising in that it was not possible to obtain a PCR product of the expected size that spanned the target site suggesting that inactivation of ad3A was not due to small insertions or deletions (indels) at the target site. Instead, these data are consistent with the possibilities that inactivation of ad3A was caused by a chromosomal rearrangement or large insertion at the target site.

Inactivation of the glucoamylase (GA) gene

[0169] A strain of *Trichoderma reesei* derived from the publicly available strain RL-P37 in which the genes (*cbh1*, *cbh2*, *egl1*, and *egl2*) encoding the four major secreted cellulases were deleted was used. This strain also lacked a functional *pyr4* gene. This strain was co-transformed using the biolistic method with a mixture of equal amounts of pTrex2gHyg MoCas and p219M gTrGA TS2. Transformants were selected on agar plates with Vogel's minimal medium containing 1% glucose, 100 ug/ml hygromycin B and 2 mg/ml uridine. After selection on the first plates transformant colonies were picked to fresh plates of the same selective medium. During growth on the second plate it was possible to distinguish between stable and unstable hygromycin-resistant transformants. Seventeen stable and 4 unstable transformants were obtained. These transformants were transferred to Vogel's agar plates without glucose and with 1% insoluble starch to screen for presence or absence of secreted glucoamylase. Colonies able to secrete glucoamylase grow well and sporulate. Colonies unable to secrete glucoamylase grow with very sparse mycelium and are clearly distinguishable. Fourteen of the 17 stable transformants were unable to secrete glucoamylase and all 4 of the unstable transformants did not secrete glucoamylase.

[0170] In order to determine the mutations at the *gla1* (glucoamylase) locus in co-transformants with pTrex2gHyg MoCas and p219M gTrGA TS2 genomic DNA was extracted from 5 stable glucoamylase non-producing transformants. This DNA was used as template for PCR using different primer pairs designed to generate products that spanned the Cas9 target site or were upstream or downstream of the target site. PfuUltra II Fusion HS DNA polymerase (Agilent Technologies) was used for the PCR according to the manufacturer's directions. In each case, the extension time was that suggested by the manufacturer for the expected size of the PCR product as described below. The sizes of the PCR products were evaluated by agarose gel electrophoresis.

[0171] No PCR product could be obtained for any transformants using *glaA* + *glaB* primers (5'- ccgtagtggaagatccttgccg [SEQ ID NO:45] and 5'- gtcgaggatttgcttcatacctc [SEQ ID NO:46] respectively) spanning the TS2 target site in *gla1*. The expected size for this PCR product presuming no large size change caused by Cas9 activity was approximately 1371 bp.

[0172] A band of the expected size (364 bp) was obtained in all transformants using glaA + glaJ primers (5'- ccgtagttgaagatccttgccg [SEQ ID NO:45] and 5'- tgccgactttgtccagtgattcg [SEQ ID NO:47] respectively) that amplify a region on the 5' side of the TS2 target site.

5 **[0173]** A band of the expected size (520 bp) was obtained in 4 of the transformants using glaK + glaB primers (5'- ttacatgtggacgcgagatagcg [SEQ ID NO:48] and 5'- gtcgaggatttgcttcatacctc [SEQ ID NO:46] respectively) that amplify a region on the 3' side of the TS2 target site. One of the transformants gave no PCR product with this primer pair.

10 **[0174]** A separate experiment intended to demonstrate inactivation of the gla1 gene by targeted Cas9 action was performed using a strain of *T. reesei* derived from RL-P37 and having an inactive pyr4 gene. Protoplasts of this strain were transformed with pTrex2gHyg MoCas gTrGA TS11 using a polyethylene glycol-mediated procedure (as described below). Transformants were selected on agar plates of Vogel's minimal
15 medium with 2% glucose, 2 mg/ml uridine, 1.1M sorbitol and 100 ug/ml hygromycin B. After selection on the first plates transformant colonies were picked to fresh plates of the same selective medium without sorbitol. During growth on the second plate it was possible to distinguish between stable and unstable hygromycin-resistant transformants. Transformants were transferred to Vogel's agar plates without glucose
20 and with 1% insoluble starch to screen for presence or absence of secreted glucoamylase. Five stable transformants, designated B#1, B#2, B#4, B#5 and B#6, which did not secrete glucoamylase were selected for further analysis. Genomic DNA was extracted from each of these transformants.

[0175] PCR was performed using genomic DNA as template and primers gla1repF and
25 gla1repR (5'- gtgtgtctaatagcctccaccac [SEQ ID NO:49] and 5'- gatcgtgctagcgctgctgttg [SEQ ID NO:50] respectively) that generate a product of 983 bp from the wild-type gla1 locus spanning the TS11 target site. The PCR conditions included gradually reducing the primer annealing temperature with each PCR cycle and a long extension time to determine if there had been a large insertion at the target site. The specific PCR
30 conditions were as follows.

Step 1: 94C for 1 minute

Step 2: 94C for 25 seconds

Step 3: 63C for 30 seconds (temperature reduced by 0.2C per cycle)

Step 4: 70C for 8 minutes

Steps 2-4 repeated 24 more times

Step 5: Hold at 4C

5 **[0176]** A clear PCR product of greater than 12 kb was obtained from two of the transformants (B#1 and B#6) suggesting an increase of greater than 11 kb in the DNA region spanning the target site. The other three transformants gave only non-specific PCR products that appeared as low intensity bands on agarose gel electrophoresis. Sequence analysis of the >12 kb PCR product from B#6 demonstrated that DNA
10 derived from plasmid pTrex2gHyg MoCas gTrGA TS11 was inserted at the TS11 target site.

[0177] PCR was performed using genomic DNA samples B#2, B#4, and B#5 and primer pair 1553R and 1555F (5'- CCGTGATGGAGCCCGTCTTCT [SEQ ID NO:51] and 5'- CGCGGTGAGTTCAGGCTTTTTC [SEQ ID NO:52] respectively). Primer 1553R binds
15 to the gla1 gene on the 3' side of target site 11. Primer 1555F binds near the start codon of the hygromycin phosphotransferase (hygB) gene on the plasmid pTrex2gHyg MoCas gTrGA TS11. The same PCR conditions were used as above. PCR products of 4.5 kb and 6.5 were obtained for transformants B#4 and B#5 respectively. PCR products should only be obtained if the plasmid with the hygB gene had inserted into
20 the gla1 gene. Presumably, the inserted plasmid DNA in transformants B#4, and B#5 was so large that it was not possible to obtain a PCR product using primers gla1repF and gla1repR.

[0178] Taken together, the PCR data demonstrated that stable hygromycin-resistant transformants with glucoamylase inactivation have arisen through insertion of large
25 segments of the Cas9 and guide RNA expression vector at the target site in the gla1 gene.

Inactivation of the pyr2 gene

[0179] Transformants of *T. reesei* strains QM6a or RL-P37 were generated by PEG-mediated transformation of protoplasts with derivatives of plasmid pTrex2gHyg MoCas
30 that included guide RNA expression cassettes targeting different positions within the *T. reesei* pyr2 gene. Inactivation of this gene confers uridine auxotrophy and resistance to 5-fluoroorotic acid (FOA). Transformants were initially selected on medium containing

hygromycin B. Upon transfer to fresh agar plates containing hygromycin B they were scored as stable or unstable. Transformants were then transferred to agar plates of Vogel's minimal medium with 2 mg/ml uridine and 1.2 mg/ml FOA. The ability to grow in the presence of FOA is indicative of uridine auxotrophy due to Cas9-mediated

5 inactivation of the pyr2 gene.

[0180] Genomic DNA was extracted from some of the FOA resistant hygromycin stable and unstable transformants for PCR analysis. The primers used for this analysis were pyr2F (5'-gtataagagcaggaggaggag [SEQ ID NO:53]) and pyr2R (5'-gaacgcctcaatcagtcagtcg [SEQ ID NO:54]) designed to amplify a region of the pyr2
10 locus spanning the target sites and approximately 0.8kb in length.

[0181] Among the QM6a transformants shown to be FOA resistant 18 stable and 5 unstable hygromycin resistant transformants were tested using the PCR protocol with an extension time sufficient to amplify the region of the pyr2 locus presuming the size to be similar to that in a wild-type strain. None of the stable transformants gave a PCR
15 product with this short extension time whereas 2 of the unstable transformants did give a PCR product. DNA sequence analysis of these two PCR products showed that one had a single nucleotide deletion and the other had a 111 nucleotide deletion at the expected target site.

[0182] Among the RL-P37 transformants shown to be FOA resistant 4 stable and 2
20 unstable hygromycin resistant transformants were tested using the PCR protocol with a short extension time. None of the stable transformants gave a PCR product with this short extension time whereas both of the unstable transformants did give a PCR product. DNA sequence analysis of these two PCR products showed that one had a single nucleotide deletion and the other had an insertion of 134 nucleotides at the
25 expected target site. This insertion consisted of two small fragments of the pTrex2gHyg vector.

[0183] A different 6 stable hygromycin resistant RL-P37 transformants were analyzed using the PCR protocol described earlier designed to enable amplification of the region of the pyr2 locus presuming a large DNA fragment was inserted at the target site in the
30 pyr2 locus. All 6 transformants gave a large PCR product (between approximately 5 kb and >12 kb depending on the transformant) with this long extension time protocol. DNA sequence analysis of 5 of these PCR products showed that pTrex2gHyg vector DNA, or fragments thereof, was integrated in all cases.

[0184] Taken together, these data show that repair of a double strand break caused by Cas9 predominantly involves integration of large vector fragments in stable transformants. This can be a very efficient method of gene inactivation. This also demonstrates that a DNA fragment or vector bearing a functional gene and having no sequence homology with the target site can integrate in a site-specific manner at the target site following Cas9 cleavage and double strand break formation. In contrast, small deletions or insertions (indels) are associated with inactivation of a gene by Cas9 in unstable transformants. This is the method of choice for gene inactivation if vector integration is undesirable.

Example 7: Expression of cas9 and sgRNA using expression vector with telomeres

[0185] A version of the Cas9 and guide RNA expression vector pTrex2gHyg MoCAS gPyr2 TS6 was constructed that contained *Trichoderma reesei* telomere sequences (shown in FIG. 6). The DNA sequence shown below (SEQ ID NO:55) was inserted into the vector. The underlined regions contain the repeated telomere sequences, each reading in towards center of this fragment. The central portion is a bacterial kanamycin resistance gene with promoter and terminator that enables selection in *E. coli* to ensure maintenance of the telomere repeats. In *Trichoderma*, a vector with telomeres is expected to linearize with the telomere sequences at each end and should be maintained autonomously at low copy number although occasional integration into the chromosomal DNA can also occur.

tcaggaaatagctttaagtagcttattaagtattaaaattatatatatttttaataataactatatttctttaataaataaggtattttaag
ctttatatataaataataataaaaataatattatatagctttttattaataaataaaaatagctaaaaatataaaaaaaatag
ctttaaaatacttatttttaattagaattttatatatttttaatatataagatcttttactttttataagcttcctacctaataaattttta
cttttttactattttactatatcttaataaaggcttttaaaaaatataaaaaaaatcttcttatattataagctataaggattatat
atatatttttttaatttttaagtaagtagtaaaagctagaattaaagtttaatttttaaggctttatttaaaaaaaggcagtaata
gctataaaaagaaatttcttttctttataactaaaagtagcttttttttaataagggttaggggttaggggttaggggttaggggttaaggggttaacaaagccacgtt
caaccacatcttaggggttaggggttaggggttaggggttaggggttaggggttaggggttaggggttaaggggttaacaaagccacgtt
gtgtctcaaaatctctgatgttacattgcacaagataaaaatatcatcatgaacaataaaaactgtctgcttacataaacag
taatacaaggggtgttatgagccatattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctga
tttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcg
ccagagttgttctgaaacatggcaaaggtagcggttgccaatgatgttacagatgagatgggtcagactaaactggctgacg
gaatttatgcctcttccgaccatcaagcattttatccgtactcctgatgatgcattggttactcaccactgcgatccccgggaaa

20 **[0186]** This vector was inserted into *T. reesei* strain RL-P37 by PEG-mediated transformation of protoplasts. Transformants were selected for hygromycin resistance and transferred to fresh agar plates with hygromycin. The majority of transformants showed an unstable hygromycin resistance phenotype. Individual transformed colonies were transferred to minimal medium agar plates containing 2 mg/ml uridine and 1.2
25 mg/ml 5-fluoroorotic acid to select for those that were able to grow and thus had a Pyr-minus phenotype. Eight out of 142 (6%) of the unstable transformants were Pyr-minus. Analysis by PCR of the *pyr2* locus and sequencing of three of these transformants showed that two had small deletions at the target site (1 bp and 27 bp respectively) and one had a 1 bp deletion combined with an insertion of 68 bp derived from the bacterial
30 vector portion of pTrex2gHyg MoCAS gPyr2 TS6. The other 5 transformants did not give a PCR product despite using PCR conditions designed to amplify large DNA fragments [PCR conditions: Step 1: 94°C for 1 minute; Step 2: 94°C for 25 seconds; Step 3: 63C for 30 seconds (temperature reduced by 0.2C per cycle); Step 4: 70°C for

8 minutes; Steps 2-4 repeated 24 more times; Step 5: Hold at 4 °C. Polymerase: PfuUltra II Fusion HS DNA polymerase (Agilent Technologies)].

[0187] These results demonstrate that expression of Cas9 and guide RNA from an autonomously replicating vector enables Cas9 targeting to a specific locus (pyr2 in this case). The resulting gene inactivation can occur without insertion of vector DNA at the target site.

Section B: Direct Introduction of Cas and/or guide RNA

Example 8: Heterologous expression of CRISPR SpyCas9 in E.coli

[0188] E. coli codon-optimized Streptococcus pyogenes Cas9 (SpyCas9) gene was synthesized and inserted into the expression vector pET30a at NcoI and HindIII sites by Generay (Shanghai, China), resulting in the plasmid pET30a-SpyCas9 (FIG. 7). As indicated in the plasmid map in FIG. 8A, the full coding sequence of the expression cassette contains, in 5' to 3' orientation, a sequence encoding an N-terminal His6 tag / thrombin / S•Tag™ / enterokinase region (SEQ ID NO:13; includes a start codon methionine), a sequence encoding an SV40 nuclear localization signal (SEQ ID NO:14), a sequence encoding the SpyCas9 (SEQ ID NO:15), and a sequence encoding the BLR nuclear localization signal (SEQ ID NO:16) all in operable linkage. This entire coding sequence is shown in SEQ ID NO:17. The amino acid sequence of the N-terminal His6 tag / thrombin / S•Tag™ / enterokinase region encoded by SEQ ID NO:13 is shown in SEQ ID NO:18 (including the methionine at position 1), the amino acid sequence of the SV40 nuclear localization signal encoded by SEQ ID NO:14 is shown in SEQ ID NO:19, the amino acid sequence of the SpyCas9 encoded by SEQ ID NO:15 is shown in SEQ ID NO:1, and the amino acid sequence of the BLR nuclear localization signal encoded by SEQ ID NO:16 is shown in SEQ ID NO:20. The amino acid sequence encoded by SEQ ID NO:17 is shown in SEQ ID NO:21.

[0189] The pET30a-SpyCas9 plasmid was transformed into Rosetta2 (De3)plysS E. coli strain (Novagen®, EMD Biosciences, Inc., Merck KGaA, Darmstadt, Germany) and the transformation products were spread on Luria Agar plates supplemented with 34ppm Chloramphenicol and 50ppm Kanamycin. Colonies were picked and cultivated for 24 hours in a 250ml shake flask with 25 ml of the Invitrogen MagicMedia™ E.coli Expression Medium (Thermo Fisher Scientific Inc., Grand Island, NY).

Example 9: Purification of SpyCas9

[0190] For purification of SpyCas9, a combination of affinity, hydrophobic interaction and size exclusion chromatographic steps were applied. Briefly, SpyCas9 expressing E. coli cells (Rosetta2 (De3)plysS, as described above) were cultured in a 250ml shake flask with 25 ml MagicMedia™ for 24 hours and harvested by centrifugation. Cells (approximately 40 grams) were pelleted and resuspended in 400 ml lysis buffer (20mM HEPES, pH7.5, 500mM NaCl, 0.1% Triton X-100, 1mM DTT and 1 mM TCEP, protease inhibitor cocktail purchased from Roche) and lysed via ultra-sonicator (35% power, 20 min, 2s on/3s off) (SCIENT2-II D, Ningbo Scientz Biotechnology Co., LTD). The lysate was cleared by centrifugation at 20000g for 40 min.

[0191] Approximately 400 ml of clarified lysate was incubated with 5 ml Ni-NTA resin (GE Healthcare) overnight at 4° C with shaking at 30 rpm/min using a Rolling Incubator (Kylin-Bell Lab. Instruments Co., Ltd. Haimen, China). After centrifugation, the resin was transferred to a XK26/20 column (GE Healthcare) and connected to AKTA Explorer system (GE Healthcare). After being washed extensively with equilibration buffer (20 mM HEPES, pH 7.5, 300 mM NaCl, 0.1% Triton X-100) followed by wash buffer (25 mM imidazole in equilibration buffer), the target protein was eluted with 250 mM imidazole in equilibration buffer.

[0192] To the active fraction collected from the affinity step, ammonium sulfate was added to a final concentration of 0.8 M and loaded onto a 20 ml phenyl-Sepharose HP column (GE Healthcare). The column was eluted with a gradient of 0.8 M to 0.0 M ammonium sulfate in 50 mM HEPES buffer pH 7.5 and the flow through was collected.

[0193] Finally, the protein was further purified by size exclusion chromatography on a Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES pH7.5, 150 mM KCl and 10% glycerol. The fraction with the highest purity were pooled and concentrated via Amicon 30 KDa membrane filter (Millipore). The final protein sample was stored at -20°C freezer in the 40% glycerol until use.

Example 10: Guide RNA Design and Expression Vector Cloning

[0194] We used the Cas9 Target Finder to identify viable target sites. Target sequences with an appropriate PAM site were identified on the sense or antisense strand of the

xyr1 gene of *Trichoderma reesei* (Transcription factor Xylanase regulator 1 involved in Xylan degradation (Protein ID 122208)) as well as the pyr4 gene of *Trichoderma reesei* (orotidine-5'-monophosphate decarboxylase (Protein ID 74020)). Using this program, we identified all 20-nucleotide long target sequences followed by a 3-nucleotide PAM sequence (NGG) that matches the sequence pattern GGN18NGG or GN19NGG. Basic local alignment search tool (BLAST) was performed using the *Trichoderma reesei* genome sequence database (genome.jgi-psf.org/Trire2/Trire2.home) to check for uniqueness of the 20-nt sequence and to avoid off target effects. The following sequences were used to generate in vitro guide RNA expression constructs in the pSM1guide plasmid (shown in FIG. 8A) for two xyr1 specific target sites (xyr1 Ta and xyr1 Tc) and for one pyr4 specific target site (pyr4 TS2). The target sequences with the associated PAM sites as well as the oligos used for annealing and cloning into the pSM1guide plasmid at the BSA1 restriction sites are shown:

Xyr1 Ta

(1) Target sequence (5'-3', PAM bold underlined):

GCAGCACCTCGCACAGCATG**CGG** (SEQ ID NO:56)

(2) oligo 1: TAGGCAGCACCTCGCACAGCATG (SEQ ID NO:57)

(3) oligo 2: AAACCATGCTGTGCGAGGTGCT (SEQ ID NO:58)

Xyr1 Tc

(1) Target sequence (5'-3', PAM bold underlined):

GCTGCCAGGAAGAATTCAAC**GGG** (SEQ ID NO:59)

(2) oligo 1: TAGGCTGCCAGGAAGAATTCAAC (SEQ ID NO:60)

(3) oligo 2: AAACGTTGAATTCTTCCTGGCA (SEQ ID NO:61)

Pyr4 TS2

(1) Target sequence (5'-3', PAM bold underlined):

GCTCAAGACGCACTACGACAT**TGG** (SEQ ID NO:62)

(2) oligo 1: TAGGCTCAAGACGCACTACGACA (SEQ ID NO:63)

(3) oligo 2: AAAGTGTGCTAGTGCGTCTTGAGC (SEQ ID NO:64)

[0195] The sequences below show the template sequence derived from the respective pSM1guide plasmid constructs for transcription of each of the three guide RNAs (i.e., for the xyr1 Ta, xyr1 Tc and pyr4 TS2 target sites above). Each sequence below shows

the T7 promoter (bold), the VT domain (shown in uppercase), the CER domain (shown in lowercase), and a transcriptional terminator (bold underline).

Xyr1 Ta (SEQ ID NO:65)

taatacgactcactataggGCAGCACCTCGCACAGCATGgttttagagctagaaatagcaagtt
5 aaaataaggctagtagccgttatcaacttgaaaaagtggcaccgagtcggtgct**ttttt**acg

Xyr1 Tc (SEQ ID NO:66)

taatacgactcactataggGCTGCCAGGAAGAATTCAACgtttttagagctagaaatagcaagtta
10 aaataaggctagtagccgttatcaacttgaaaaagtggcaccgagtcggtgct**ttttt**acg

Pyr4 TS2 (SEQ ID NO:67)

taatacgactcactataggGCTCAAGACGCACTACGACAgtttttagagctagaaatagcaagtta
aaataaggctagtagccgttatcaacttgaaaaagtggcaccgagtcggtgct**ttttt**acg

15 Example 11: In vitro DNA Cleavage assay

[0196] Guide RNAs were produced in vitro from templates for xyr1 Ta and xyr1 Tc using the MEGAshortscript™ T7 transcription kit from Thermo Fisher according to the manufacturer's instructions. In vitro transcription was carried out at 37°C for at least 5 hours. Transcribed guide RNAs were purified using MEGAclean™ Transcription Clean-Up kit from Thermo Fisher. The RNA concentration was measured with NanoDrop™
20 (Thermo Fisher). Denaturing urea- PAGE gel (10%) was used to confirm the quality of the guide RNA produced (data not shown).

[0197] Purified Cas9 protein (200ng) was incubated with: (1) 300 ng substrate DNA alone (substrate DNA is plasmid pXA3 [shown in FIG. 9A] linearized with NdeI ; pXA3
25 contains the xyr1 gene [SEQ ID NO:89] that has the 20bp target sequence and appropriately spaced PAM site for both of the xyr1 guide RNAs)(2) 300ng substrate DNA in the presence of 100ng in vitro synthesized xyr1 Ta guide RNA; and (3) 300ng substrate DNA in the presence of 100ng in vitro synthesized xyr1 Tc guide RNA. The reactions were carried out in NEB buffer 3 in a reaction volume of 20ul for 1 h at 37 °C.
30 (1X NEB3 Buffer Components consists of 100mM NaCl, 50mM Tris-HCl, 10mM MgCl₂ 10mM MgCl₂, 1mM DTT, pH 7.9 at 25 °C.)

[0198] As shown in FIG. 9B, each of the xyr1 specific guide RNA with purified SpyCas9 can successfully cut substrate DNA into the expected fragments(Lanes 3 and 4),

confirming the function of the synthesized guide RNA/Cas9 complex. Lane 1 shows molecular weight markers; Lane 2 shows NdeI-linearized plasmid pXA3 substrate in the absence of Cas9 and guide RNA; Lane 3 shows cleavage of linearized plasmid pXA3 substrate in the presence of Cas9 and a guide RNA with the xyr1 Ta VT domain; Lane 4 shows cleavage of the linearized plasmid pXA3 substrate in the presence of Cas9 and a guide RNA with the xyr1 Tc VT domain. Positions of the linearized plasmid pXA3 substrate and products are indicated at the right.

Example 12: guide RNA introduction into Cas9-expressing fungal cells

Methods

(i) Protoplast preparation

[0199] For protoplast preparation, 5×10^8 spores of the desired *T. reesei* strain are inoculated into 50 ml germination medium (recipe described in US Patent No. 8,679,815) in a 250 ml shake flask with 4 baffles and incubated at 27°C for 17 hours at 170 rpm. The mycelia are recovered by transferring the liquid volume into 50 ml conical tubes and spinning at 3000 rpm for 10 minutes. The supernatant is decanted and the mycelial pellets are washed twice using 1.2 M MgSO₄ -10 mM Na-phosphate buffer and resuspended in 15 ml lysing enzyme buffer (lysing Enzyme from *Trichoderma harzianum* (Sigma catalog #L1412)) dissolved in 1.2 M MgSO₄ – 10 mM Na-phosphate buffer (pH 5.8), 50 mg/ml). The cell suspension is transferred into a 250 ml shake flask with 4 baffles and shaken at room temperature for at least 2 hours at 200 rpm. The protoplasts are harvested by filtration through Miracloth (Calbiochem Art. No. 475855) folded in a glass funnel into a Greiner tube. 0.6 M Sorbitol - 0.1 M Tris-HCl buffer is added carefully on top of the filtered protoplasts. The protoplasts are collected by centrifugation for 15 minutes at 4000 rpm. The middle phase containing the protoplasts is transferred into a new tube and added at least an equal volume of 1.2 M Sorbitol - 10 mM Tris-HCl buffer. The protoplasts are collected by centrifugation for 5 minutes at 4000 rpm, and washed two times with 1.2M sorbitol-10mM Tris-HCl buffer. The pellet is resuspended into at least 1ml 1.2 M Sorbitol - 10 mM Tris-HCl pH 7.5 - 10 mM CaCl₂ buffer and the number of protoplasts counted under a microscope. The protoplast suspension is diluted using 4 parts of 1.2 M Sorbitol – 10 mM Tris-HCl – 10 mM CaCl₂ and 1 part of 25% PEG6000 – 50 mM CaCl₂ – 10mM Tris-HCl until 5×10^8 per ml for use in subsequent transformation.

(ii) Transformation

[0200] The desired cargo (e.g., a DNA construct, guide RNA, Cas9/guide RNA complex, etc.) is added to 200 μ L protoplast ($\sim 1 \times 10^8$) and kept on ice for 30 min. After incubation, protoplasts are added to cooled molten sorbitol/Vogel agar (1.1 M sorbitol of minimal Vogel agar) to be as the top layer of the minimal Vogel plate (Davis et al., (1970) Methods in Enzymology 17A, pp. 79-143 and Davis, Rowland, NEUROSPORA, CONTRIBUTIONS OF A MODEL ORGANISM, Oxford University Press, (2000)). The plates are incubated at 30° C for a week. The detailed steps are described in US Patent No. 8,679,815 (incorporated herein by reference).

Experimental

[0201] Protoplasts of a *Trichoderma reesei* strain having an inactivated *pyr2* gene (encoding orotate phosphoribosyl transferase, Protein ID 21435) (strain T4 mpg1 Δ *pyr2*) was transformed as described above with a DNA construct containing an expression cassette for Cas9 under the control of the pyruvate kinase (*pki*) promoter and an expression cassette for the *pyr2* gene from *T. reesei* under the control of the its native promoter. A transformant with the Cas9-*pyr2* cassette integrated into the genome and constitutively expressing the Cas9 gene was identified by selecting for cells having a functional *pyr2* gene (growth without uridine supplementation on Vogels media).

[0202] Twenty (20) μ g of in vitro synthesized *Pyr4* TS2 guide RNA as described above (with target site 5'GCTCAAGACGCACTACGACA3', SEQ ID NO:92) was introduced into the Cas9 expressing *T. reesei* cells by the protoplast transformation method described above. Analysis of the *pyr4* gene from isolated strains that are resistant to FOA and require uridine for growth by sequencing and alignment showed the presence of changes to the DNA sequence at the *pyr4* gene target site. Sequence changes included insertions of a few nucleotides (1-2 nucleotides; clones T4 4-3, T4 4-11, T4 4-18, T4 4-19, T4 4-4, and T4 4-7) as well as larger insertions (68 nucleotides, clone T4 4-20) (FIG. 10). This demonstrates that direct, transient introduction of guide RNA into a Cas-expressing fungal host cell can be used to modify the DNA sequence at a desired target site in the genome of the cell.

Example 13: In vivo SpyCas9/guide RNA uptake experiment

[0203] To form the Cas9/guide RNA complex in vitro, purified Cas9 protein 20 µg was mixed with Pyr4 TS2 guide RNA 20 µg in 20mM Hepes, 100mM NaCl, 5mM MgCl₂, 0.1 mM EDTA pH6.5 (final volume is 40 µL), and incubated at room temperature from 15-30 minutes to allow for complex formation. The Cas9/guide RNA complex was transformed into *T. reesei* protoplasts as described above and grown on Vogel's Uridine FOA plates.

[0204] PCR analysis of the isolated strains from this transformation is shown in FIGS.

11A and B. FIG. 11A shows agarose gel analysis of pyr4 specific PCR products (encompassing the target site) of two isolated strains (P37 2.2. and P37 4.1; both resistant to FOA and that require uridine for growth). Strain P37 2.2 (Lane 2) showed a PCR product that is of lower molecular weight than the T4 4.1 clone (Lane 3; which is equivalent to the control, shown in FIG. 11B, Lane 2), indicating a large deletion in the pyr4 gene. FIG. 11B shows similar PCR/agarose gel analysis as in FIG. 11A, and includes analysis of P37 strains 4.1, 4.2, 4.3, and 4.4 (all of which are resistant to FOA and require uridine for growth). Strain 4.3 (Lane 5) showed PCR product of the pyr4 gene that is of lower molecular weight than the control (C+; Lane 2), indicating a large deletion in the pyr4 gene.

[0205] Sequence analysis of the pyr4 genes derived from clones T4 2.2 (shown in FIG. 11A) and T4 2.4 (not shown in FIG. 11A or 11B) is shown in FIG. 12. Note that the wild type pyr4 sequence is the first sequence (top) in the alignments. This analysis shows that the T4 2.2 clone (top alignment) has a deletion of 611 base pairs at the target site of the introduced Cas9/guide RNA complex. The sequence corresponding to the VT domain sequence of the guide RNA is boxed and the PAM site is circled. The bottom alignment shows a 1 base pair insertion in the pyr4 gene at the target site of the isolated T4 2.4 strain (a "G" residue). The sequence corresponding to the VT domain sequence of the guide RNA is indicated with a line over the alignment and the PAM site is circled.

[0206] FIG. 13 shows sequence analysis of the pyr4 genes derived from clones P37 4.1 and 4.2 (top alignment), 4.3 (bottom alignment) and 4.4 (middle alignment) (which were shown in FIG. 11B). The wild type pyr4 sequence is the first sequence (top) in all alignments and a consensus is shown on the bottom of all alignments. The top alignment shows that the P37 4.1 clone (third sequence in the alignment) has an insertion of a T nucleotide while the P37 4.2 clone (second sequence in the alignment)

has an insertion of a G nucleotide at the target site in the pyr4 gene. The middle alignment shows that the P37 4.4 clone (second sequence in the alignment) has a deletion of an A nucleotide at the target site in the pyr4 gene. The bottom alignment shows that the pyr4 gene sequence in the P37 4.3 clone (second sequence in the alignment) diverges abruptly at the target site. Further alignment analysis (not shown) confirmed that the P37 4.3 clone has a deletion of 988 base pairs at the target site of the introduced Cas9/guide RNA complex.

[0207] This demonstrates that direct, transient introduction of a Cas9/guide RNA complex into a fungal host cell can be used to modify the DNA sequence at a desired target site in the genome of the cell.

[0208] Although the foregoing compositions and methods have been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings herein that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

[0209] Accordingly, the preceding merely illustrates the principles of the present compositions and methods. It will be appreciated that those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the present compositions and methods and are included within its spirit and scope. Furthermore, all examples and conditional language recited herein are principally intended to aid the reader in understanding the principles of the present compositions and methods and the concepts contributed by the inventors to furthering the art, and are to be construed as being without limitation to such specifically recited examples and conditions. Moreover, all statements herein reciting principles, aspects, and embodiments of the present compositions and methods as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents and equivalents developed in the future, i.e., any elements developed that perform the same function, regardless of structure. The scope of the present compositions and methods, therefore, is not intended to be limited to the exemplary embodiments shown and described herein.

Sequences:

SEQ ID NO:1

Streptococcus pyogenes Cas9, no NLS (encoded by SEQ ID NO:8 and SEQ ID NO:15)

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA
 5 EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPI
 FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN
 SDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLF
 GNLIALLSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS
 DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQQLPEKYKEIFFDQSKN
 10 GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL
 GELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITP
 WNFEEVVDKGASASQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG
 MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFRNASLG
 TYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTFEDREMIEERLKTYAHLFDDKVMKQLKR
 15 RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFQMQLIHDDSLTFKEDIQKAQ
 VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDDELVKVMGRHKPENIVIAMARENQTTQ
 KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
 INRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKKNYWRQLL
 NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDEN
 20 DKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG TALIKKYPKLES
 EFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN
 GETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK
 DWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERSSSFENPIDFLE
 AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY
 25 EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVLSAYNKHHRDKPI
 REQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ SITGLYETRIDL SQ
 LGGD

SEQ ID NO:2

30 *Streptococcus thermophilus* LMD-9 Cas9
 MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITA
 EGRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSF LVPDDKRDSKYPI
 FGNLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSK
 NNDIQKNFQDFLD TYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLPGEKNSGIF

SEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGYIGDDYSDVFLKAKKLY
 DAILLSGFLTVDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISKTYNEVFKDDTKN
 GYAGYIDGKTNQEDFYVYLKKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHL
 QEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKITPW
 5 NFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRD
 YQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLL
 NIINDKEFLDDSSNEAIIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRHYTGWG
 KLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGN
 KEVVKSLPGSPAIIKKGILQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQ
 10 RLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYLQNGKDMYTGDDLDIDRL
 SNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDVPSLEVVKRKRKTFWYQLLKSCLIS
 QRKFDNLTKAERGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRVR
 TVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYG
 DYPKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKE
 15 SDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLSSKPKPNSNENLVGAK
 EYLDPKKYGGYAGISNSFTVLVKGTEIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLE
 KGYKDIELIIELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAK
 RISNTINENHRKYVENHKKEFEELFYIILEFNENYVGAKKNGKLLNSAFQSWQNHSIDE
 LCSSFIGPTGSEKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGL
 20 YETRIDLAKLGEG

SEQ ID NO:3

Streptococcus mutans UA159 Cas9

MKKPYSIGLDIGTNSVGWAVVTDDYKVPAAKKMKVLGNTDKSHIEKNLLGALLFDSGNT
 25 AEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFHRLSDSFLVTEDEKRGERRH
 PIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTR
 NNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAAKDRVLKLPNEKSN
 GRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEEELEVLLAQIGDNYAELFLSAK
 KLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSD
 30 VSKDGYAGYIDGKTNQEAIFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPH
 QIHLQEMRAIRRQAEFYFPLADNQDRIEKLTLFRIPYYVGPLARGKSDFAWLSRKSAD
 KITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYK
 TEQGKTAFDANMKQEIFDGVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKV
 FNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRENYSDLLTKEQ

VKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKE
 EIAKAQVIGETDNLNQVVSDIAGSPAIAKKGILQSLKIVDELVKIMGHQPENIVVEMAREN
 QFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQQLQNDRLFLYYLQNGRDMYT
 GEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKS
 5 SKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTET
 DENNKKIRQVKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYP
 QLEPEFVYGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHIS
 NIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPRKTCKFYWDTKKYGGFDSP
 IVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKL
 10 PKYSLFKLENGRKRLLASARELQKGNEIVLPNHLGTLTYHAKNIHKVDEPKHLDYVDKH
 KDEFKELLDVVSNFSSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTAIGAPA
 TFKFFDKNIDRKRYTSTTEILNATLIHQSTGLYETRIDLNLKGGD

SEQ ID NO:4

15 *Campylobacter jejuni* Cas9

MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLAR
 RKARLNHLKHLIANEFKLNIEDYQSFDESLAKAYKGSLISPYELRFRALNELLSKQDFA
 RVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENS
 KEFTNVRNKKESYERCIASFLKDELKLIFKKQREFGFSFSKKFEEVLSVAFYKRALK
 20 DFSHLVGNCSSFTDEKRAPKNSPLAFMFVALTRIINLLNNLNKNTGILYTKDDLNALLNE
 VLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAK
 DITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPMLLEGKKYDEAC
 NELNLKVAINEDKKDFLPANETYYKDEVTPVVLRAIKEYRKVLNALLKKYGKVHKINI
 ELAREVGKNHSQRAKIEKEQNENYKAKKDAELEECEKLGLKINSKNILKLRLFKEQKEFC
 25 AYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMKNKVLVFTKQNQEKLNQTPFEAFGN
 DSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVNLYTKDYL
 DFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVI
 IAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEI
 FVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNGDMF
 30 RVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCSLYK
 DSLILIQTKDMQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKS
 IGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK

SEQ ID NO:5

Neisseria meningitides Cas9

MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRFERAEVPKTGDSLAM
 ARRLARSVRRLTRRRRAHRLLRTRRLKREGVLQAANFDENGLIKSLPNTPWQLRAAAL
 5 DRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRT
 PAELALNKFEKESGHIRNQSRSDYSHTFSRKDLQAEILLFEKQKEFGNPHVSGGLKEGI
 ETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTCLNNLRILEQGS
 ERPLTDTERATLMDEPYRKSCLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEM
 KAYHAISRALKKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEAL
 10 LKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEI
 RNPVVLRLALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDR
 EKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHKGKCLYSGKEINLGRLENEKGYVEIDHA
 LPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSFRP
 RSKKQRILLQKFDEEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQI
 15 TNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTID
 KETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSR
 PEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLDLEK
 MVNREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKAVRVEQVQKT
 GVWVRNHNGIADNATMVRVDVFEKGDYLLVPIYSWQVAKGILPDRAVVQKGDEED
 20 WQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRG TGNNIRIHDLDHKIGKNGIL
 EGIGVKTALS FQKYQIDELGKEIRPCRLKKRPPVR

SEQ ID NO:6

Francisella tularensis subsp. novicida Cas9

MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGKVVYELSKDSYTLLMNNRTARR
 HQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLFNRRGFSFITDGYSPEYLN
 VPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLMQKILEFKLMKLCTDIKD
 DKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYYHHDKYNIQE
 FLKRHATINDRILDTLLTDDLDIWNFNFEKFDKNEEKLQNQEDKDHIQAHLHHFVFA
 30 VNLIKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVKNLV
 NLIGNLSNLELKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGEWRVGVKDQDKKD
 GAKYSYKDLCLNELKQKVTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQSLILNPKFL
 DNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQQIASGQ
 RDYKDL DARILQFIFDRV KASDELL NEIYFQA KKLKQKASSELEKLESSKKLDEVANS

QLSQILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHKYNNTG
 RFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFISKWLVEHI
 RGFKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNYKHGL
 AYELGVLLFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNA
 5 HRMQQIKITEPVEDNKKDKIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQ
 VLSAKHQLHIPITESNAFEFEPALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFA
 KGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKNKGNRIF
 CLRDLADNYKLKQFETTDDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFRH
 ALFLADENPIKQAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDYF
 10 GIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGS
 IGLEIDKNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKKAIEGFNTHRQMTRD
 GIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFVDKPISIDI
 QISTLEELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEFLRSLA
 YRSERVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYREWQNTTIKDDYEFLKSFF
 15 NVKSITKLHKKVRKDFSLPISTNEGKFLVKRKTWDNNFIYQILNDSDSRADGTPFIPAF
 DISKNEIVEAIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDIGIATIQ
 YKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQSTIIEFESSGFNK
 TIKEMLGMKLAGIYNETSNN

20 SEQ ID NO:7

Pasteurella multocida Cas9

MQTTNLSYILGLDLGASVGWAVVEINENEDPIGLIDVGVRIFERAIEVPKTGESLALSRR
 LARSTRRLIRRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGLERRLSAIEW
 GAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQSDDYRTPAELALKKF
 25 AKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTELLMWQ
 KPALSGEAILKMLGKCTHEKNEFKAACKHTYSAERFVWLTKLNNLRILEDGAERALNEEE
 RQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKHLRYSKENAESATFMELKAWHAIRKAL
 ENQGLKDTWQDLAKKPDLLDEIGTAFSLYKTDEDIQQYLTKNVPNSVINALLVSLNFDK
 FIELSLKSLRKILPLMEQGKRYDQACREIYGHYGEANQKTSQLLPAIPAQEIRNPVVL
 30 TLSQARKVINAIIRQYGSPARVHIETGRELGKSFKERREIQKQQEDNRTKRESAVQKFK
 ELFSDFSSEPCKSKDILKFRLYEQQHKGKCLYSGKEINIHRLNEKGYVEIDHALPFSRTWD
 DSFNNKVLVLASENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAAKKQRLLT
 QVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSRWGLIK
 ARENNNRHHALDAIVVACATPSMQQKITRFIRFKEVHPYKIENRYEMVDQESGEIISPH

FPEPWAYFRQEVNIRVFDNHPDTVLKEMPLDRPQANHQFVQPLFVSRAPTRKMSGQ
 GHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREPALYAGLKARLAEFNQDPA
 KAFATPFYKQGGQQVKAIRVEQVQKSGVLVRENNGVADNASIVRTDVFIKNNKFFLVPI
 YTWQVAKGILPNKAIVAHKNEDEWEEMDEGAKFKFSLFPNDLVELKTKKEYFFGYIIG
 5 LDRATGNISLKEHDGEISKGKDG VYRVGVKLALSFEKYQVDELGKNRQICRPQQRQ
 PVR

SEQ ID NO:8

Filamentous fungal cell codon optimized Streptococcus pyogenes Cas9-encoding gene;
 10 no NLS

atggacaagaagtacagcatcggcctcgacatcggcaccaactcgggtgggctgggcccgtcatcacggacgaatataa
 ggtcccgtcgaagaagttcaaggtcctcggcaatacagaccgccacagcatcaagaaaaactgatcggcgccctcct
 gttcgatagcggcgagaccgcggaggcgaccaggctcaagaggaccgccaggagacgggtacactaggcgcaaga
 acaggatctgctacctgcaggagatcttcagcaacgagatggcgaagggtggacgactccttctccaccgcctggagga
 15 atcattcctggtggaggaggacaagaagcatgagcggcacccaatcttcggcaacatcgtcgacgaggtggcctacca
 cgagaagtacccgacaatctaccacctccggaagaaactggtggacagcacagacaaggcggacctccgggtcatct
 acctgccctcgcgcatatgatcaagttccgcggccacttctcatcgagggcgacctgaaccgcggacaactccgacgtg
 gacaagctgttcatccagctcgtgcagacgtacaatcaactgttcgaggagaaccccataaacgctagcggcgtggacg
 ccaaggccatcctctcggccaggctctcgaaatcaagaaggctggagaaccttatcgcgacgttgccaggcgaaaaga
 20 agaacggcctcttcggcaaccttattgcgctcagcctcggcctgacgccgaacttcaaatcaaacttcgacctcgcgag
 gacgccaaagctccagctctcaaaggacacctacgacgacacctcgacaacctcctggcccagataggagaccagta
 cgcggaacctcttctcgccgccaagaacctctccgacgctatcctgctcagcgacatccttcgggtcaacaccgaaattac
 caaggcaccgctgtccgccagcatgattaaacgctacgacgagcaccatcaggacctcacgctgctcaaggcactcgt
 ccgccagcagctccccgagaagtacaaggagatcttcttcgaccaatcaaaaaacggctacgcgggatatatcgacgg
 25 cgggtgccagccaggaagagttctacaagttcatcaaaccaatcctggagaagatggacggcaccgaggagttgctggt
 caagctcaacagggaggacctcctcaggaagcagaggaccttcgacaacgggtccatcccgcacatcagatccacctgg
 gcgaactgcatgccatcctgcggcgccaggaggacttctacccgttcctgaaggataaccggggagaagatcgagaag
 atcttgacgttccgcatcccatactacgtgggcccgtggctcgcggaactcccgggttcgcttgatgacctcggaagtcg
 gaggagaccatcacacctggaactttgaggaggtggtcgataagggcgctagcgtcagagcttcatcgagcgcacg
 30 accaacttcgataaaaacctgcccattgaaaaagtcctccccaagcactcgtgctctacgagtacttcacctgttaca
 cgagctaccaaggtcaaatacgtcaccgagggcatgcggaagccggcggttcctgagcggcgagcagaagaaggc
 gatagtggaacctcctcttcaagaccaacaggaagggtgacctgaagcaattaaaagaggactacttcaagaaaataga
 gtgcttcgactccgtggagatctcgggcgtggaggatcggttcaacgcctcactcggcacgtatcacgacctcctcaagat
 cattaaagacaaggacttctcgacaacgaggagaacgaggacatcctcgaggacatcgtcctcacctgacctgttc

gaggaccgcgaaatgatcgaggagaggctgaagacctacgcgcacctgttcgacgacaaggatcatgaaacagctca
 agaggcgccgctacactggttggggaaggctgtcccgcaagctcattaatggcatcagggacaagcagagcggcaag
 accatcctggacttcctcaagtccgacgggttcgccaaccgcaacttcagctcattcacgacgactcgctcacgttc
 aaggaagacatccagaaggcacagggtgagcgggcagggtgactccctccacgaacacatcgccaacctggccggct
 5 cgccggccattaaaaagggcacctgcagacgggtcaaggctcgtcagcagctcgtgaaggatgggcccggcacaag
 cccgaaaatatcgtcatagagatggccaggggagaaccagaccacccaaaaagggcagaagaactcgcgcgagcg
 gatgaaacggatcgaggagggcattaaagagctcgggtcccagatcctgaaggagcaccctggaaaataccag
 ctccagaatgaaaagctctacctctactacctgcagaacggccgcgacatgtacgtggaccaggagctggacattaatc
 ggctatcggactacgacgtcgaccacatcgtgccgcagtcgttcctcaaggacgatagcatcgacaacaagggtgctcac
 10 ccggctcggataaaaaatcggggcaagagcgacaacgtgccagcgaggaggtcgtgaagaagatgaaaaactactg
 gcgccagctcctcaacgcgaaactgatcaccagcgcaagttcgacaacctgacgaaggcggaacgcgggtggcttga
 gcgaactcgataaggcgggcttcataaaaaggcagctggtcgagacgcgccagatcacgaagcatgtcgcccagatc
 ctggacagccgcatgaataactaagtacgatgaaaacgacaagctgatccgggagggtgaaggatcacgctgaagtcc
 aagctcgtgtcggacttcgcaaggacttcagttctacaagggtccgcgagatcaacaactaccaccacgcccacgacg
 15 cctacctgaatgcgggtggtcgggaccgccctgatcaagaagtacccgaagctggagtcggagttcgtgtacggcgacta
 caaggctctacgacgtgcgcaaaatgatcgccaagtccgagcaggagatcggcaaggccacggcaaaatacttctcta
 ctgaacatcatgaacttctcaagaccgagatcacctcgcgaacggcgagatccgcaagcgcccgtcatcgaaac
 caacggcgagacgggcgagatcgtctgggataagggccgggatttcgcgacgggtccgcaagggtgctctccatgccgca
 agtcaatatcgtgaaaaagacggagggtccagacgggcccgggttcagcaaggagtcctcctcccgaagcgcaactccg
 20 acaagctcatcgcgaggaagaaggattgggacccgaaaaaatatggcgggttcgacagcccagaccgtcgcatacag
 cgtcctcgtcgtggcgaagggtggagaagggaaggtcaaagaagctcaagtcctgaaggagctgctcgggatcacgat
 tatggagcgggtcctccttcgagaagaacccgatcgacttcctagaggccaagggatataaggagggtcaagaaggacct
 gattattaaactgccgaagtactcgctcttcgagctggaaaacggccgcaagaggatgctcgctccgcaggcgagttgc
 agaagggaacgagctcgccctcccagcaaatatcgtcaatttctgtacctcgctagccactatgaaaagctcaaggg
 25 cagcccggaggacaacgagcagaagcagctcttcgtggagcagcacaagcattacctggacgagatcatcgagcag
 atcagcgagttctgaagcgggtgatcctcgccgacgcgaacctggacaagggtgctgtcggcatataacaagcaccgc
 gacaaaccaatacgcgagcaggccgaaaatatcatccacctcttcacctcaccaacctcggcgctccggcagccttca
 agtacttcgacaccacgattgaccggaagcgggtacacgagcacgaaggagggtgctcgatgcgacgctgatccaccag
 agcatcacagggtctatgaaacacgcatcgacctgagccagctgggcccggagac

30

SEQ ID NO:9

Filamentous fungal cell codon optimized *Streptococcus pyogenes* Cas9-encoding gene;
 with N- and C-terminal NLS sequences

atggcaccgaagaagaagcgcaaggtgatggacaagaagtacagcatcggcctcgacatcggcaccaactcgggtgg
 gctgggcccgtcatcacggacgaatataaggtcccgtcgaagaagttcaaggtcctcggcaatacagaccgccacagca
 tcaagaaaaacttgatcggcgccctcctgttcgatagcggcgagaccgcggaggcgaccagggtcaagaggaccgcc
 aggagacgggtacactaggcgcaagaacaggatctgctacctgcaggagatcttcagcaacgagatggcgaaggtgg
 5 acgactccttctccaccgcctggaggaatcattcctggtggaggaggacaagaagcatgagcggcacccaatcttcgg
 caacatcgtcgcgaggtggcctaccacgagaagtacccgacaatctaccacctccggaagaaactggtggacagca
 cagacaaggcggacctccggctcatctacctgccctcgcgcataatgatcaagttccgcggccacttcctcatcgagggc
 gacctgaacccggacaactccgacgtggacaagctgttcacccagctcgtgcagacgtacaatcaactgttcgaggaga
 accccataaacgctagcggcgtggacgccaaggccatcctctcggccagggtctcgaaatcaagaaggctggagAAC
 10 cttatcgcgcagttgccaggcgaaaagaagaacggcctcttcggcaaccttattgcgctcagcctcggcctgacgccga
 acttcaaatacaacttcgacctcgcggaggacgccaagctccagctctcaaaggacacctacgacgacgacctcgaca
 acctcctggcccagataggagaccagtacgcggacctcttcctcgcggccaagaacctctccgacgctatcctgctcagc
 gacatccttcgggtcaacaccgaaattaccaaggcaccgctgtccgccagcatgattaaacgctacgacgagcaccatc
 aggacctcacgctgctcaaggcactcgtccgccagcagctccccgagaagtacaaggagatcttcttcgaccaatcaaa
 15 aaacggctacgcgggatataatcgacggcggtgccagccaggaagagttctacaagttcatcaaaccaatcctggagaa
 gatggacggcacccgaggagttgctggtcaagctcaacagggaggacctcctcaggaagcagaggaccttcgacaac
 ggctccatcccgcacatcagatccacctgggccaactgcatgccatcctcgcggcgccaggaggacttctaccggttcctgaa
 ggataaccggggagaagatcgagaagatcttgacgttcgcacatccatactacgtgggcccgtggctcgcggcaactcc
 cggttcgcctggatgacctggaagtcggaggagaccatcacacctggaactttgaggaggtggtcgataagggcgct
 20 agcgctcagagcttcatcgagcgcacatgaccaacttcgataaaaacctgcccaatgaaaaagtcctccccaagcactcgc
 tgctctacgagtacttcacctgttacaacgagctcaccaagggtcaaatacgtcacccgagggcatgcggaagccggcggt
 cctgagcggcgagcagaagaaggcgatagtgacacctcctctcaagaccaacaggaagggtgacctgaagcaattaa
 aagaggactacttcaagaaaatagagtgttcgactccgtggagatctcgggctggaggatcggttcaacgcctcactc
 ggcacgtatcacgacctcctcaagatcattaaagacaaggacttctcgacaacgaggagaacgaggacatcctcgag
 25 gacatcgtcctcacctgacctgttcgaggaccgcgaaatgatcgaggagagggtgaagacctacgcgcacctgttcg
 acgacaaggctcatgaaacagctcaagaggcgccgctacactggttggggaaggctgtcccgaagctcattaatggca
 tcagggacaagcagagcggcaagaccatcctggacttctcaagtccgacgggttcgccaaccgcaacttcatgcagc
 tcattcacgacgactcgtcacgttcaaggaagacatccagaaggcacagggtgagcgggcagggtgactccctccacg
 aacacatcgccaacctggccggctcgcgggccattaaaaagggcacatcctgcagacgggtcaaggctcgtcgacgagctc
 30 gtgaagggtgatgggcccggcacaagcccgaataatcgtcatagagatggccagggagaaccagaccacccaaaaa
 gggcagaagaactcgcgcgagcggatgaaacggatcgaggagggcattaaagagctcgggtcccagatcctgaag
 gagcacccttggaataatccagctccagaatgaaaagctctaccttactacctgcagaacggccgcgacatgtac
 gtggaccaggagctggacattaatcggctatcggactacgacgtcgaccacatcgtgccgcagtcgttctcaaggacg
 atagcatcgacaacaagggtgctcacccggctcggaataaaatcggggcaagagcgcacaacgtgccagcggaggagt

cgtgaagaagatgaaaaactactggcgccagctcctcaacgcgaaactgatcaccagcgcaagttcgacaacctga
 cgaaggcggaacgcggtggcttgagcgaactcgataaggcgggcttcataaaaaggcagctggctcgagacgcgcca
 gatcacgaagcatgtcgcccagatcctggacagccgcatgaatactaagtacgatgaaaacgacaagctgatccggg
 aggtgaaggatgatcacgctgaagtccaagctcgtgtcggacttccgcaaggacttccagttctacaagggtccgagatc
 5 aacaactaccaccacgcccacgacgcctacctaagtgcggtggctcgggaccgcccctgatcaagaagtacccgaagct
 ggagtcggagttcgtgtacggcgactacaagggtctacgacgtgcgcaaaatgatcgccaagtccgagcaggagatcgg
 caaggccacggcaaaatacttcttctactcgaacatcatgaacttctcaagaccgagatcacctcgcgaacggcgag
 atccgcaagcgcccgcctcatcgaaaccaacggcgagacggggcgagatcgtctgggataagggccgggatttcgacgac
 ggtccgcaagggtgctctccatgccgcaagtcaatatcgtgaaaaagacggaggtccagacggggcggttcagcaagg
 10 agtccatcctcccgaagcgcaactccgacaagctcatcgcgaggaagaaggattgggaccgaaaaaatatggcggc
 ttcgacagcccagccgtcgcatcacgctcctcgtcgtggcgaagggtggagaagggcaagtcaaagaagctcaagtcc
 gtgaaggagctgctcgggatcacgattatggagcggctcctcctcgagaagaacccgatcgacttctagaggccaagg
 gatataaggaggtcaagaaggacctaattataaactgccgaagtactcgctcttcgagctggaaaacggccgcaaga
 ggatgctcgctccgcaggcgagttgcagaagggcaacgagctcgccctcccgagcaatacgtcaatttctgtacctc
 15 gctagccactatgaaaagctcaagggcagcccggaggacaacgagcagaagcagctcttcgtggagcagcacaag
 cattacctggacgagatcatcgagcagatcagcgagttctcgaagcgggtgatcctcgccgacgcgaacctggacaag
 gtgctgtcggcatataacaagcaccgcgacaaaccaatacgcgagcaggccgaaaatatcatccacctcttcacctca
 ccaacctcggcgctccggcagccttcaagtacttcgacaccacgattgaccggaagcgggtacacgagcacgaaggag
 gtgctcgatgcgacgctgatccaccagagcatcacagggctctatgaaacacgcatcgacctgagccagctgggcgga
 20 gacaagaagaagaagctcaagctctag

SEQ ID NO:10

Streptococcus pyogenes Cas9 with N- and C-terminal NLS sequences (encoded by
SEQ ID NO:9)

25 MAPKKKRKVMDDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA
 LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEE
 DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRILIYLAHAMIKFRGHFLI
 EGDLPDNDSDVDKLFQILVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLP
 GEKKNGLFGNLIASLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADL
 30 FLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI
 FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG
 SIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRK
 SEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKV
 KYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDR

FNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTTLTLFEDREMIEERLKTIAHLFDDKV
 MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK
 EDIQKAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVDELVKVMGRHKPENIVIEMA
 RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
 5 MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
 KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
 MNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTA
 IKKYPKLESEFVYGDYKVYDVRKMIKSESEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
 RKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRN
 10 SDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEKGKSKKLKSVKELLGITIMERS
 SF EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV
 NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIIEQISEFSKRVLADANLDKVL
 SA YNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQ
 SITG LYETRIDLSQLGGDKKKKKLKL

15

SEQ ID NO:11

Full U6 gene promoter sequence (not including transcription start site)

AAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTAACTTCTGCA
 GTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTAT
 20 TTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTATTATAATAT
 ATATTATATAATAATTTTAAAGCCTGGAATAGTAAAGAGGCTTATATAATAATTTATAG
 TAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATGAAATGGTATT
 ATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTGGCTATAAGTC
 TGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCT
 25 ATC

SEQ ID NO:12

Truncated/shorter U6 gene promoter sequence (not including transcription start site)

AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCA
 30 CTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCG
 TAGGTATATAAGTACTGTTTGATGGTAGTCTATC

SEQ ID NO:13

N-terminal His6 tag / thrombin / S•Tag™ / enterokinase region polynucleotide sequence (with start codon); encodes SEQ ID NO:18

5 atgcaccatcatcatcatcattcttctggtctggtgccacgcggttctggtatgaaagaaaccgctgctgctaaattcgaacg
ccagcacatggacagcccagatctgggtaccgacgacgacgacaaggccatggcc

SEQ ID NO:14

SV40 NLS coding sequence (encodes SEQ ID NO:19)

10 ccaaaaaagaaacgcaagggtt

SEQ ID NO:15

E. coli codon-optimized Cas9 gene (no stop codon)

atggataaaaaatacagcattggtctggatatcggaaccaacagcgttgggtgggcagtaataacagatgaatacaaa
15 gtgccgtcaaaaaaatttaagggtctggggaatacagatcgccacagcataaaaaagaatctgattggggcattgctgttt
gattcgggtgagacagctgaggccacgcgtctgaaacgtacagcaagaagacgttacacacgtcgtaaaaatcgtattt
gctacttacaggaaatTTTTTctaacgaaatggccaaggtagatgatagtttctccatcgtctcgaagaatctttctggtgag
gaagataaaaaacacgaacgtcacctatctttggcaatatcgtggatgaagtggcctatcatgaaaaataccctacgatt
tatcatcttcgcaagaagttggtgatagtagcgacaaagcggatctgcgtttaatctatcttgcgttagcgcacatgatcaa
20 atttcgtggtcatttcttaattgaaggtagatctgaatcctgataactctgatgtggacaaattgtttatacaattagtgcaaacct
taatcagctgttcgaggaaaacccattaatgcctctggagttgatgccaaagcgattttaagcgcgagactttctaagtcc
cggcgtctggagaatctgatcgcccagttaccaggggaaaagaaaaatgggtctgtttggtaatctgattgccctcagctctgg
ggcttaccgccgaacttcaaattccaattttgacctgggtgaggacgcaaagctgcagctgagcaaagatacttatgatgatg
acctcgacaatctgctcgcccagattggtgaccaatatgcggatctgttctggcagcgaagaatcttcggatgctatcttgc
25 tgtcggatattctgcgtgttaataccgaaatcaccaaagcgcctctgtctgcaagtagatcaagagatacgacgagcacc
accaggacctgactcttctaaggcactggtacgccaacagcttccggagaaatacaaaagaatattcttcgaccagtcc
aagaatgggttacgcgggctacatcgatggtggtgcatcacaggaagagttctataaatttataaaccaatccttgagaaa
atggatggcacggaagagttacttgtaaacttaaccgcgaagacttgcttagaaagcaacgtacattcgacaacggctc
catccacaccagattcatttaggtgaacttcacgccatcttgcgcagacaagaagattctatcccttcttaaagacaatc
30 gggagaaaatcgagaagatcctgacgttccgcattccctattatgtcgggtcccctggcacgtggtaattctcggtttgcctgg
atgacgcgcaaaagtgaggaaaccatcaccccttgaactttgaagaagtcgtggataaagggtgctagcgcgcagctctt
tatagaaagaatgacgaacttcgataaaaacttgcccaacgaaaaagtcctgcccaagcactctcttttatatgagtacttt
actgtgtacaacgaactgactaaagtgaatacgttacggaaggtagcgcaaacctgccttcttagtggcgagcagaa
aaaagcaattgtcgatcttctttaaacgaatcgcaaggtaactgtaaaacagctgaaggaagattatttcaaaaagat

cgaatgctttgattctgtcgagatctcgggtgtcgaagatcgtttcaacgcttccttagggacctatcatgatttgctgaagata
 ataaaagacaaagactttctcgacaatgaagaaaatgaagatatctggaggatattgtttgaccttgaccttattcgaag
 atagagagatgatcgaggagcgcttaaaaacctatgccacctgtttgatgacaaagtcatgaagcaattaaagcgccg
 cagatatacgggggtggggccgcttgagccgcaagttgattaacggtattagagacaagcagagcggaaaaactatcct
 5 ggatttcctcaaactctgacggatttgcaaccgcaattttatgcagcttatacatgatgattcgcttacattcaaagaggatatt
 cagaaggctcaggtgtctgggcaaggtgattcactccacgaacatatagcaaatttggccggctctcctgcgattaagaa
 ggggatcctgcaaacagttaaagttgtggatgaacttgtaaaagtaatgggcccgcacaagccggagaatatcgtgata
 gaaatggcgcgcgagaatcaaacgacacaaaaagggtcaaaagaactcaagagagagaatgaagcgcattgagga
 ggggataaaggaacttgatctcaaattctgaaagaacatccagttgaaaacactcagctgcaaaatgaaaaattgtac
 10 ctgtactacctgcagaatggaagagacatgtacgtggatcaggaattggatatcaatagactctcggactatgacgtagat
 cacattgtccctcagagcttcctcaaggatgattctatagataataaagtacttacgagatcggacaaaaatcgcggtaaat
 cggataacgtcccatcggaggaagtcgttaaaaagatgaaaaactattggcgtcaactgctgaacgccaagctgatcac
 acagcgtaagtttgataatctgactaaagccgaacgcggtggtcttagtgaactcgataaagcaggattataaaacggc
 agttagtagaaacgcgccaaattacgaaacacgtggctcagatcctcgattctagaatgaatacaaagtacgatgaaaa
 15 cgataaactgatccgtgaagtaaaagtcattaccttaaaatctaaacttggtccgatttccgcaaagattttcagttttacaa
 ggtccgggaaatcaataactatcacatgcacatgatgcataatttaaagtcgggtttaggcacggcccttattaagaaatac
 cctaaactcgaaagtgagtttgttatggggattataaagtgtatgacgttcgcaaaatgatcgcgaaatcagaacaggaa
 atcggtaaggctaccgctaaatacttttttattccaacattatgaattttttaagaccgaaataactctcgcgaaatgggtgaaat
 ccgtaaacggcctcttatagaaaccaatggtgaaacgggagaaatcgttgggataaaggctcgtgactttgccaccgttcg
 20 taaagtcctctcaatgccgcaagttaacattgtcaagaagacggaagttcaaacagggggattctccaaagaatctatcct
 gccgaagcgtaacagtataaacttattgccagaaaaaaagattgggatccaaaaaaatacggaggccttgattccct
 accgtcgcgtatagtgtgctggtggttgctaaagtcgagaaagggaaaagcaagaaattgaaatcagttaaagaactgc
 tgggtattacaattatggaaagatcgtcctttgagaaaaatccgatcgacttttagaggccaaggggtataaggaagtga
 aaaaagatctcatcatcaaattaccgaagtatagtccttttgagctggaaaacggcagaaaaagaatgctggcctccgcg
 25 ggcgagttacagaaggggaaatgagctggcgctgccttccaaatatgttaattttctgtaccttgccagtcattatgagaaact
 gaagggcagccccgaagataacgaacagaaacaattattcgtggaacagcataagcactatttagatgaaattataga
 gcaaattagtgaaattttctaagcgcgttatcctcgcggtatgctaatttagacaaagtactgtcagcttataataaacatcggg
 ataagccgattagagaacaggccgaaaatatcattcatttgtttaccttaaccaaccttgagaccagctgccttcaaata
 tttcgataccacaattgatcgtaaacgggtatacaagtacaaaagaagtcttggacgcaaccctcattcatcaatcttactg
 30 gattatatgagacacgcattgatctttcacagctgggaggagac

SEQ ID NO:16

BLR2 nuclear localization signal coding sequence (encodes SEQ ID NO:20)

aagaagaaaaaactgaaactg

SEQ ID NO:17

The nucleotide sequence of the SpyCas9 synthetic gene in plasmid pET30a- SpyCas9. The oligonucleotides encoding the N-terminal His6 tag, the SV40 nuclear localization signal, and the BLR nuclear localization signal are shown in bold underline, italic underline, and underlined, respectively.

atg**caccatcatcatcatcat**tcttctggtctggtgccacgcggttctggtatgaaagaaaccgctgctgctaaattcgaac
 gccagcacatggacagcccagatctgggtaccgacgacgacgacaaggccatggccccaaaaaagaaacgcaag
gttatggataaaaaatacagcattggtctggatatcggaaccaacagcggtgggtgggcagtaataacagatgaataca
 10 aagtgccgtcaaaaaaatttaagggtctggggaatacagatcgccacagcataaaaaagaatctgattggggcattgctg
 ttgattcgggtgagacagctgaggccacgcgtctgaaacgtacagcaagaagacgttacacacgctcgtaaaaatcgtat
 ttgctacttacaggaaatTTTTCTAACGAAATGGCCAAGGTAGATGATAGTTTCTTCATCGTCTCGAAGAATCTTTCTGGTTGA
 ggaagataaaaaacacgaacgtcacctatctttggcaatatcgtaggaagtgccctatcatgaaaaataccctacga
 ttatcatcttcgcaagaagtgggtgatagtagcgacaagcggtatctgcgtttaatctatcttgcgttagcgccacatgatca
 15 aatttcgtggtcatttcttaattgaagggtgatctgaatcctgataactctgatgtggacaaattgtttatacaattagtgcacacct
 ataatcagctgttcgaggaaaacccattaatgcctctggagttgatgccaaagcgattttaagcgcgagactttctaagtc
 ccggcgtctggagaatctgatcgccagttaccaggggaaaagaaaaatggctctgttggtaatctgattgccctcagctctg
 gggcttaccgccgaacttcaaatccaattttgacctggctgaggacgcaaagctgcagctgagcaaagatacttatgatgat
 gacctcgacaatctgctcgcccagattggtgaccaatatgcggatctgtttctggcagcgaagaatctttcggatgctatctt
 20 gctgtcggatattctgcgtgttaataccgaaatcaccaaagcgctctgtctgcaagtatgatcaagagatacgcgagca
 ccaccaggacctgactcttcttaaggcactgggtacgccaacagcttccggagaaatacaaaagaatattcttcgaccagt
 ccaagaatggttacgcgggctacatcgatgggtggtcatcacaggaagagttctataaattattaaaccaatccttgaga
 aaatggatggcacggaagagttacttgtaaacttaaccgcgaagacttgcttagaaagcaacgtacattcgacaacgg
 ctccatcccacaccagattcatttaggtgaacttcacgccatcttgcgcagacaagaagatttctatcccttcttaaagaca
 25 atcgggagaaaaatcgagaagatcctgacgttccgcattccctattatgtcgggtccctggcacgtggtaattctcggtttgcct
 ggatgacgcgcaaaagtgaggaaaccatcaccccttggaaactttgaagaagtcgtggataaagggtgctagcgcgcagt
 ctttatagaaagaatgacgaacttcgataaaaacttgcccaacgaaaaagtcctgccaagcactctcttttatatgagta
 ctttactgtgtacaacgaactgactaaagtgaatacgttacggaagggtatgcgcaaacctgcctttcttagtggcgagcag
 aaaaaagcaattgtcgatcttctctttaaacgaatcgcaaggtaactgtaaaacagctgaaggaagattatttcaaaaag
 30 atcgaatgctttgattctgtcgagatctcgggtgtcgaagatcgtttcaacgcttccttagggacctatcatgatttgctgaagat
 aataaaagacaaagactttctcgacaatgaagaaaatgaagatattctggaggatattgtttgaccttgaccttattcgaa
 gatagagagatgatcgaggagcgcttaaaaacctatgcccacctgtttgatgacaaagtcataagcaattaaagcgcc
 gcagatatacggggtggggccgcttgagccgcaagttgattaacggtattagagacaagcagagcggaaaaactatcc
 tggatttcctcaaatctgacggatttgcaaccgcaattttatgcagcttatacatgatgattcgcttacattcaaagaggatat

tcagaaggctcaggtgtctgggcaaggtgattcactccacgaacatatagcaaattggccggctctcctgcgattaagaa
 ggggatcctgcaaacagttaaagttgtggatgaacttgtaaaagtaatgggcccacaaagccggagaatatcgtgata
 gaaatggcgcgcgagaatcaaacgacacaaaaagggtcaaaagaactcaagagagagaatgaagcgcattgagga
 ggggataaaggaacttggatctcaaattctgaaagaacatccagttgaaaacactcagctgcaaaatgaaaaattgtac
 5 ctgtactacctgcagaatggaagagacatgtacgtggatcaggaattggatatcaatagactctcggactatgacgtagat
 cacattgtccctcagagcttcctcaaggatgattctatagataataaagtacttacgagatcggacaaaaatcgcggtaaat
 cggataacgtcccacgcggaggaagtcgttaaaaagatgaaaaactattggcgtcaactgctgaacgccaagctgatcac
 acagcgtaagtttgataatctgactaaagccgaacgcggtgggtcttagtgaactcgataaagcaggatttataaacggc
 agttagtagaaacgcgccaattacgaaacacgtggctcagatcctcgattctagaatgaatacaaagtacgatgaaaa
 10 cgataaactgatccgtgaagtaaaagtcattaccttaaaatctaaacttgtgtccgatttccgcaaagattttcagttttacaa
 ggtccgggaaatcaataactatcacatgcacatgatgcataatttaaatgcggtttaggcacggcccttattaagaaatac
 cctaaactcgaaagtgagtttgtttatggggattataaagtgtatgacgttcgcaaaatgatcgcgaaatcagaacaggaa
 atcggtaaggctaccgctaaatacttttttattccaacattatgaattttttaagaccgaaataactctcgcgaaatgggtgaaat
 ccgtaaacggcctcttatagaaaccaatgggtgaaacggggagaaatcgtttgggataaaggctcgtgactttgccaccgttcg
 15 taaagtcctctcaatgccgcaagttaacattgtcaagaagacggaagttcaaacagggggatttccaaagaatctatcct
 gccgaagcgtaacagtataaacttattgccagaaaaaaagattgggatccaaaaaaatacggaggctttgattccct
 accgtcgctatagtgtgctgggtggtgctaaagtcgagaaagggaaaagcaagaaattgaaatcagttaaagaactgc
 tgggtattacaattatggaaagatcgtcctttgagaaaaatccgatcgacttttagaggccaaggggtataaggaagtga
 aaaaagatctcatcatcaaattaccgaagtatagtcttttgagctggaaaacggcagaaaaagaatgctggcctccgcg
 20 ggcgagttacagaaggggaaatgagctggcgctgccttccaaatatgttaattttctgtaccttgccagtcattatgagaaact
 gaagggcagccccgaagataacgaacagaaacaattattcgtggaacagcataagcactatttagatgaaattataga
 gcaaattagtgaattttctaagcgcgttatcctcgcggtatgctaatttagacaaagtactgtcagcttataataaacatcggg
 ataagccgattagagaacaggccgaaaatatcattcatttgtttaccttaaccaaccttgagcaccagctgccttcaaata
 tttcgataccacaattgatcgtaaacgggtatacaagtacaaaagaagtcttggacgcaaccctcattcatcaatcttactg
 25 gattatatgagacacgcattgatctttcacagctgggcccggagacaagaagaaaaaaactgaaactg

SEQ ID NO:18

N-terminal His6 tag / thrombin / S•Tag™ / enterokinase region amino acid sequence
 (with start methionine)

30 Mhhhhhhssglvprgsgmketaaakferqhmdspdlgtddddkama

SEQ ID NO:19

SV40 NLS

PKKKRKV

SEQ NO:20

T. reesei blr2 (blue light regulator 2) gene NLS

KKKKLKL

5

SEQ ID NO:21

The amino acid sequence of the SpyCas9 protein expressed from plasmid pET30a-SpyCas9. The N-terminal His6 tag, the SV40 nuclear localization signal, and the BLR nuclear localization signal are shown in bold underline, italic underline, and underlined, respectively.

10

15

20

25

mssglvprgsgmketaaakferqhmdspdlgtdddddkama

dkkysigldigtnsvgwavit

deykvpskkfkvlgntrhsikknligallfdsgetaeatrkrtrrrytrrknricylqeifsnemakvddsffhrleesflveed

kkherhpifgnivdevayhekyptiyhlrkklvdstkdalrliylalahmikfrghfliegdlnpdnsdvdklfiqlvtynqlfe

enpinasgvdakailsarlsksrrlenliaqlpgekknglfgnlialslgltpnfksnfdlaedakqlskdydddldnllaqigd

qyadlflaaknlsdaillsdilrvnteitkapsasmikrydehhqdltlkalvrqqpkykeiffdqskngyagyidggasqe

efykfikpilekmdgteellvklredllrkqrtfdngsiphqihlgelhailrrqedfypflkdnrekiekiltfripyyvgplargnsr

fawmtrkseetitpwnfeevvdkgasaqsfierntnfdknlpnekvlpkhsllieyftvyneltkvkyvtegmrkpaflsge

qkkaivdlfktnrkvtkqlkedyfkkiecfsveisgvedrfnaslgyhdllkiikdkdfldneenediledivltltlfedremi

eerlkyahlfddkvmkqlkrrrytgwgrlsrklingirdkqsgktildflksdganrnmqlihddsltfkediqaqvsgqgd

slhehianlagspaikkgilqtkvkvdelvkvmgrhkpeniviemarenqttqkgqknsrermkrieegikelgsqilkehp

ventqlqneklylylqngrdmyvdqeldinrlsdydvdhivpqsfkddsidnkvltrsdknrgksdnpseevvkkmkkn

ywrqllnaklitqrkfdnltkaergglseidkagfikrqlvetrqitkhvaqildsrmntkydendklirevkvitlksklvsdfrkdf

qfykvreinnyhahdaylnavvgtalikkypklesefvygdykvdyvrkmiakseqeigkatakyffysnimnffkteitla

ngeirkrpietngetgeivwdkgrdfatvrkvlsmppqvnivkktevqgtggfskesilpkrnsdkliarkkdwdpkkyggfds

ptvaysvlvvakvekgkskklksvkellgitimerssfeknpidfleakgykevkkdliiklpkyslfelengrkrmlasagelq

kgnelalpskyvnflylashyeklkgspedneqkqlfveqhkhyldeiieqisefskrviladanldkvlsaynkhrdkpireq

aeniihlftltnlgapaafkyfdttidrkrytstkevldatlihqsitglyetridlsqlggd

SEQ ID NO:22

30

Putative *T. reesei* U6 gene

AAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTAACTTCTGCA

GTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTATTATTTTTAT

TTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTATTATAATAT

ATATTATATAATAATTTTAAGCCTGGAATAGTAAAGAGGCTTATATAATAATTTATAG

TAATAAAAGCTTAGCAGCTGTAATATAATTCTAAAGAAACAGCATGAAATGGTATT
 ATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTGGCTATAAGTC
 TGCTGCAAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTGATGGTAGTCT
 ATCGCCTTCGGGCATTTGGTCAATTTATAACGATACAGGTTCGTTTCGGCTTTTCC
 5 TCGGAACCCCCAGAGGTCATCAGTTCGAATCGCTAACAGGTCAACAGAGAAGATT
 AGCATGGCCCCTGCACTAAGGATGACACGCTCACTCAAAGAGAAGCTAAACATTTT
 TTTTCTCTTCCAAGTCGTGATGGTTATCTTTTTTGCTTAGAGAATCTATTCTTGTGGA
 CGATTAGTATTGGTAAATCCCTGCTGCACATTGCGGCGGATGGTCTCAACGGGCAT
 AATACCCCATTCGTGATGCAGCGGTGATCTTCAATATGTAGTGTAATACGTTGCAT
 10 ACACCACCAGGTTCGGTGCCTCCTGTATGTACAGTACTGTAGTTCGACTCCTCCG
 CGCAGGTGGAAACGATTCCCTAGTGGGCAGGTATTTTGGCGGGGTCAAGAA

SEQ ID NO:23

sequence of sgRNA (N is sequence complementary to target site)

15 NNNNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGG
 CUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

SEQ ID NO:24

sgRNA: gAd3A TS1

20 guccucgagcaaaaggugccGUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGU
 CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

SEQ ID NO:25

sgRNA: gTrGA TS2

25 guucagugcaauaggcgucuGUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGU
 CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

SEQ ID NO:26

sgRNA: gTrGA TS11

30 gccaauggcgacggcagcacGUUUUAGAGCUAGAAAUAGCAAGUUAAAUAAGGCUAGU
 CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

SEQ ID NO:27

sgRNA: gPyr2 TS6

gcacagcgggaugcccuuguGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGU
5 CCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGGUGC

SEQ ID NO:28

Synthetic DNA: gAd3A TS1-1 (gAd3A TS1 sgRNA (SEQ ID NO:3) with *Saccharomyces cerevisiae* snr52 promoter and *S. cerevisiae* sup4 terminator)

10 gaattcggatccTCTTTGAAAAGATAATGTATGATTATGCTTTCACATATTTATACAGA
AACTTGATGTTTTCTTTCGAGTATATACAAGGTGATTACATGTACGTTTGAAGTACA
ACTCTAGATTTTGTAGTGCCCTCTTGGGCTAGCGGTAAAGGTGCGCATTTTTTTCAC
ACCCTACAATGTTCTGTTCAAAGATTTTGGTCAAACGCTGTAGAAGTGAAAGTTG
GTGCGCATGTTTCGGCGTTCGAAACTTCTCCGCAGTGAAAGATAAATGATCggtcctcg
15 agcaaaaggtgccGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTTTTGTGTTTTTATGTCTgaattcg
gatcc

SEQ ID NO:29

20 Synthetic DNA: gAd3A TS1-2 (gAd3A TS1 sgRNA (SEQ ID NO:3) with *T. reesei* U6 promoter and terminator)

gaattcggatccAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTA
ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA
TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTT
25 ATTATAATATATATTATATAATAATTTTAAGCCTGGAATAGTAAAGAGGCTTATATAA
TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG
AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG
GCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTG
ATGGTAGTCTATCggtcctcgagcaaaaggtgccGTTTTAGAGCTAGAAATAGCAAGTTAAA
30 TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTT
TTCTCTTgaattcggatcc

SEQ ID NO:30

Synthetic DNA: gAd3A TS1-3 (gAd3A TS1 sgRNA (SEQ ID NO:3) with T. reesei U6 promoter, terminator and intron)

5 gaattcggatccAAAAACACTAGTAAGTACTTACTTATGTATTATTAAGTACTTTAGCTA
 ACTTCTGCAGTACTACCTAAGAGGCTAGGGGTAGTTTTATAGCAGACTTATAGCTA
 TTATTTTTATTTAGTAAAGTGCTTTTAAAGTAAGGTCTTTTTTATAGCACTTTTTATTT
 ATTATAATATATATTATATAATAATTTTAAGCCTGGAATAGTAAAGAGGCTTATATAA
 TAATTTATAGTAATAAAAGCTTAGCAGCTGTAATATAATTCCTAAAGAAACAGCATG
 10 AAATGGTATTATGTAAGAGCTATAGTCTAAAGGCACTCTGCTGGATAAAAATAGTG
 GCTATAAGTCTGCTGCAAACTACCCCCAACCTCGTAGGTATATAAGTACTGTTTG
 ATGGTAGTCTATCgtcctcgagcaaaaggtgccGTTTTAGAGCTAGAGTTTCGTTTCGGCTTT
 TCCTCGGAACCCCCAGAGGTCATCAGTTCGAATCGCTAACAGAATAGCAAGTTAAA
 ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGGTGCTTTTTT
 15 TTTCTCTTgaattcggatcc

SEQ ID NO:31

Guide RNA expression cassettes with a shorter T. reesei U6 promoter region were obtained as synthetic DNA. An example is provided here that includes the sequence for an sgRNA targeting the T. reesei gla1 gene at TS11.

20 AATTCCTAAAGAAACAGCATGAAATGGTATTATGTAAGAGCTATAGTCTAAAGGCA
 CTCTGCTGGATAAAAATAGTGGCTATAAGTCTGCTGCAAACTACCCCCAACCTCG
 TAGGTATATAAGTACTGTTTGATGGTAGTCTATCgccaatggcgacggcagcacGTTTTAGA
 GCTAGAGTTTCGTTTCGGCTTTTCCTCGGAACCCCCAGAGGTCATCAGTTCGAATC
 25 GCTAACAGAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGG
 CACCGAGTCGGTGGTGCTTTTTTTTTCTCTT

SEQ ID NO:32

Primer: gRNA fwd afIII

30 cgtcagcttaagAATTCCTAAAGAAACAGCATGAAATGG

SEQ ID NO:33

Primer: gRNA rev sfil

cgtcagggccacgtgggcccAAGAGAAAAAAAAGCACCCGACTCGG

SEQ ID NO:34

Primer: Ad3 5' fwd

tgaacacagccaccgacatcagc

5

SEQ ID NO:35

Primer: Ad3 5' rev

gctgggtgagggtttgctattg

10 SEQ ID NO:36

Primer: Ad3a 5005 rev

gattgcttgggaggaggacat

SEQ ID NO:37

15 Primer: Ad3 3' fwd

cgaggccactgatgaagttgttc

SEQ ID NO:38

Primer: Ad3 3' rev

20 Cagttttccaaggctgccaacgc

SEQ ID NO:39

Primer: Ad3a 5003 fwd

ctgatcttgaccctggaaatc

25

SEQ ID NO:40

Ad3mid rev

ctctctatcattgccaccctcc

30 SEQ ID NO:41

Primer: Adfrag fwd

ctccattcacccctcaattctcc

SEQ ID NO:42

Primer: Adfrag rev

gttccttggcgggtgcttgatc

5

SEQ ID NO:43

Primer: Ad3a 2k fwd

caatagcacaaccctcaccagc

10 SEQ ID NO:44

Ad3a 2k rev

gaacaactcatcagtggcctcg

SEQ ID NO:45

15 Primer: glaA

ccgtagttgaagatccttgccg

SEQ ID NO:46

Primer: glaB

20 gtcgaggattgcttcatacctc

SEQ ID NO:47

Primer: glaJ

tgccgactttgtccagtgattcg

25

SEQ ID NO:48

Primer: glaK

ttacatgtggacgcgagatagcg

30 SEQ ID NO:49

Primer: gla1repF

gtgtgtctaatacctccaccac

gatcgtgctagcgctgctgttg

5

CCGTGATGGAGCCCGTCTTCT

CGCGGTGAGTTCAGGCTTTTTC

gtataagagcaggaggaggag

20 gaacgcctcaatcagtcagtcg

Bacterial kanamycin resistance gene (with promoter and terminator) between
Trichoderma reesei telomere sequences

25 tcaggaaatagctttaagtagcttattaagtattaaaattatatatattttaataataactatatcttctaataaataaggtattttaag
ctttatatataaataataataaaaataatatattatatagctttttattaataaataaaaatagctaaaaatataaaaaaaaaatag
ctttaaataacttatttttaattagaattttatatatttttaatatataagatcttttactttttataagcttcctacctaaattaaatttta
cttttttactattttactatatcttaaataaaggctttaaaaatataaaaaaaaaatcttcttatattataagctataaggattatat
atatattttttttaatttttaagtaagtattaaagctagaattaaagtttaatttttaaggcttatttaaaaaaaggcagtaata
30 gcttataaaagaaatttcttttcttttataactaaaagtagctttttttaataagggttaggggttaggggttagctcacaccgaccatcc
caaccacatcttaggggttaggggttaggggttaggggttaggggttaggggttaggggttaggggttaaacaaagccacgtt
gtgtctcaaaatctctgatgttacattgcacaagataaaaatataatcatcatgaacaataaaaactgtctgcttacataaacag
taatacaaggggtgttatgagccatattcaacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctga
tttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagccccgatgcg

ccagagttgtttctgaaacatggcaaaggtagcggtgccaatgatgttacagatgagatggtcagactaaactggctgacg
gaatttatgcctcttcgaccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccgggaaa
acagcattccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtggtcctgcgccggttgca
ttcgattcctgtttgtaattgtccttttaacagcgatcgcgatttctgtctcgctcaggcgcaatcacgaatgaataacggtttgg
5 tgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagctttgccattct
caccggattcagtcgtcactcatgggtgatttctcacttgataaccttattttgacgaggggaaattaatagggtgtattgatgtg
gacgagtcggaatcgagaccgataaccaggatcttgccatcctatggaactgcctcgggtgagtttctccttcattacagaa
acggctttttcaaaaatatgggtattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagttttctaatacaga
attggtaattgggtgtaacactggcagagcattacgctgacttgacgggacggcggctttgttgaataaatcgaacttttgc
10 gagttgaaggatcagatcacgcacttcccgacaacgcagaccgttccgtggcaaagcaaaagttcaaaatcaccaact
ggtccacctacaacaaagctctcatcaaccgtggctccctcactttctggctggatgatggggcgattcaggcctgggtatga
gtcagcaacaccttcttcacgaggcagacctcagcggtttaaacctaaccctaaccctaaccctaaccctaaccctaacc
ctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaaccctaacc
gaggggtctatagactaatctacaggccgtacatgggtgtgattgcagatgcgacgggcaagggttacagtggtccagaagg
15 aggagagcggcataggattgtaatagaccagctttacataataatcgccgtgttgctactgactgatgaccttctccctaac
cagtttcctaattaccactgcagtgaggataaccctaactcgctctgggggttattattatactgattagcaggtggcttatatagt
gctgaagtactataagagtttctgcgggaggaggtggaaggactataaactggacacagttagggatagagtgatgaca
agacctgaatgttatcctccgggtgtggtatagcgaattggctgaccttgacagatggtaatggtttaggcagggttttgcagag
ggggacgagaacgcgttctgcgatttaacggctgctgccgccaagctttacgggttcttaatgggcggccgc

20

SEQ ID NO:56

Xyr1 Ta Target sequence (5'-3', PAM bold underlined):GCAGCACCTCGCACAGCATG**CGG**

25 SEQ ID NO:57

Xyr1 Ta (2) oligo 1

TAGGCAGCACCTCGCACAGCATG

SEQ ID NO:58

30 Xyr1 Ta oligo 2

AAACCATGCTGTGCGAGGTGCT

SEQ ID NO:59

Xyr1 Tc Target sequence (5'-3', PAM bold underlined):

GCTGCCAGGAAGAATTCAAC**GGG**

SEQ ID NO:60

Xyr1 Tc oligo 1

5 TAGGCTGCCAGGAAGAATTCAAC

SEQ ID NO:61

Xyr1 Tc oligo 2

AAACGTTGAATTCTTCCTGGCA

10

SEQ ID NO:62

Pyr4 TS2 Target sequence (5'-3', PAM bold underlined)

GCTCAAGACGCACTACGACA**TGG**

15 SEQ ID NO:63

Pyr4 TS2 oligo 1

TAGGCTCAAGACGCACTACGACA

SEQ ID NO:64

20 Pyr4 TS2 oligo 2

AAACTGTCGTAGTGCGTCTTGAGC

SEQ ID NO:65

Xyr1 Ta

25 **taatacgaactcactatagg**GCAGCACCTCGCACAGCATGgttttagagctagaaatagcaagttaaata
aggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc**tttt**acg

SEQ ID NO:66

Xyr1 Tc

30 **taatacgaactcactatagg**GCTGCCAGGAAGAATTCAACggttttagagctagaaatagcaagttaaataa
ggctagtccgttatcaactgaaaaagtggcaccgagtcggtgc**tttt**acg

SEQ ID NO:67

Pyr4 TS2

taatacgactcactataggGCTCAAGACGCACTACGACAgtttttagagctagaaatagcaagttaaata
aggctagtcggttatcaactgaaaaagtggcaccgagtcggtgctttttacg

SEQ ID NO:68

5 K21 control T4

tggcccgtcgattgtcgtgctcaagacgcactacgacatggtctcgg

SEQ ID NO:69

T4 4-3

10 tggcccgtcgattgtcgtgctcaagacgcactacgCGacatggtctcgg

SEQ ID NO:70

T4 4-13

tggcccgtcgattgtcgtgctcaagacgcactacgacatggtctcgg

15

SEQ ID NO:71

T4 4-11

tggcccgtcgattgtcgtgctcaagacgcactacgGacatggtctcgg

20 SEQ ID NO:72

T4 4-12

tggcccgtcgattgtcgtgctcaagacgcactacgacatggtctcgg

SEQ ID NO:73

25 T4 4-18

tggcccgtcgattgtcgtgctcaagacgcactacgGacatggtctcgg

SEQ ID NO:74

T4 4-20

30 tggcccgtcgattgtcgtgctcaagacgcactacgAGCCGACAGGGCGCCTGGCTAAATCCAAGGT
CAAGACAGGCTGGTGGTTGTTTAGTGCGAGTCCTCTGacatggtctcgg

SEQ ID NO:75

T4 4-19

tggcccgctcgattgtcgtgctcaagacgcactacgGacatgggtctcgg

SEQ ID NO:76

T4 4-4

5 tggcccgctcgaatgttggtcaaggcgcccttcgGacatgggtctcgg

SEQ ID NO:77

T4 4-7

tggcccgctcgattgtcgtgctcaagacgcactacgGacatgggtctcgg

10

SEQ ID NO:78

9-96

CCGCTGACGGCTTACCTGTTCAAGCTCATGGACCTCAAGGCGTCCAACCTGTGCC
TGAGCGCCGACGTGCCGACAGCGCGCGAGCTGCTGTACCTGGCCGACAAGATTG
15 GCCCGTCGATTGTCGTGCTCAAGACGCACTACGCAGGCCTGCGTCGAGGCCGCC
CGGGAGCACAAGGACTTTGTCATG

SEQ ID NO:79

Pyr4 Tr

20 CCGCTGACGGCTTACCTGTTCAAGCTCATGGACCTCAAGGCGTCCAACCTGTGCC
TGAGCGCCGACGTGCCGACAGCGCGCGAGCTGCTGTACCTGGCCGACAAGATTG
GCCCGTCGATTGTCGTGCTCAAGACGCACTACGACATGGTCTCGGGCTGGGACTT
CCACCCGGAGACGGGGACGGGAGCCCAGCTGGCGTCGCTGGCGCGCAAGCACG
GCTTCCTCATCTTCGAGGACCGCAAGTTTGGCGACATTGGCCACACCGTCGAGCT
25 GCAGTACACGGGCGGGTCTGGCGCGCATCATCGACTGGGCGCACATTGTCAACGT
CAACATGGTGCCCGGCAAGGCGTCGGTGGCCTCGCTGGCCCAGGGCGCCAAGC
GCTGGCTCGAGCGCTACCCCTGCGAGGTCAAGACGTCCGTCACCGTCGGCACGC
CCACCATGGACTCGTTTGACGACGACGCCGACTCCAGGGACGCCGAGCCCGCCG
GCGCCGTCAACGGCATGGGCTCCATTGGCGTCCTGGACAAGCCCATCTACTCGA
30 ACCGGTCCGGCGACGGCCGCAAGGGCAGCATCGTCTCCATCACCACCGTCACCC
AGCAGTACGAGTCCGTCTCCTCGCCCCGGTTAACAAGGCCATCGCCGAGGGCG
ACGAGTCGCTCTTCCCGGGCATCGAGGAGGCGCCGCTGAGCCGCGGCCTCCTGA
TCCTCGCCCAAATGTCCAGCCAGGGCAACTTCATGAACAAGGAGTACACGCAGGC
CTGCGTCGAGGCCGCCCGGGAGCACAAGGACTTTGTCATG

SEQ ID NO:80

Query

ctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgacatggtctc

5

SEQ ID NO:81

Subject

ctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgGacatggtctc

10 SEQ ID NO:82

Pyr4 Tr

gacagcgcgcgagctgctgtacctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgacatggtct
cgggctgggacttccacccgg

15 SEQ ID NO:83

P37 #13 4.2 rc

gacagcgcgcgagctgctgtacctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgGacatggt
ctcgggctgggacttccacccgg

20 SEQ ID NO:84

P37 4.1 #12 rc

gacagcgcgcgagctgctgtacctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgTacatggt
ctcgggctgggacttccacccgg

25 SEQ ID NO:85

P37 #15 4.4 rc

gacagcgcgcgagctgctgtacctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgcatggtctc
gggctgggacttccacccgg

30 SEQ ID NO:86

P37 #14 4.3

Gacagcgcgcgagctgctgtacctggccgacaagattggcccgctcgattgtcgtgctcaagacgcactacgacatggtc
tcgggctgggacttccacccgg

SEQ ID NO:87

Consensus (deletion alignment)

Gacagcgcgagctgctgtacctggccgacaagattggcccgtcgattgtcgtgctcaagacgcannangnnnnngg
nnnnnggnngggannncnancngg

5

SEQ ID NO:88

Wild type pyr4 full coding sequence

Atggcaccacacccgacgctcaaggccaccttcgcgccaggagcgagacggcgacgcacccgctgacggcttacc
tggtcaagctcatggacctcaaggcgtccaacctgtgctgagcgccgacgtgccgacagcgcgcgagctgctgtacctg
10 gccgacaagattggcccgtcgattgtcgtgctcaagacgcactacgacatgggtctcgggctgggacttccacccggaga
cgggcacgggagcccagctggcgtcgtggcgcgcaagcacggcttcctcatcttcgaggaccgcaagtttggcgaca
ttggccacaccgtcgagctgcagtacacgggcgggtcggcgcgcatcatcgactgggcgcacattgtcaacgtcaacat
ggtgcccggcaaggcgtcgggtggcctcgtcgtggcccagggcgccaagcgctggctcgagcgctacccctgcgaggtca
agacgtccgtcacctcggcacgcccaccatggactcgtttgacgacgacgccgactccagggacgccgagcccgcc
15 ggcgccgtcaacggcatgggctccattggcgtcctggacaagcccattactcgaaccgggtccggcgacggccgcaag
ggcagcatcgtctccatcaccaccgtcacccagcagtagcagtcctcctcgccccgggtaacaaaggccatcgccg
agggcgacgagtcgtcttcccgggcatcgaggaggcgccgctgagccgcgccctcctgatcctcgcccaaagtcca
gccagggcaacttcatgaacaaggagtagcacgcaggcctgcgtcgaggccgcccgggagcacaaggactttgtcatg
ggcttcatctcgcaggagacgctcaacaccgagcccgacgatgcctttatccacatgacgcccggctgccagctgcccc
20 ccgaagacgaggaccagcagaccaacggatcggtcgggtggagacggccagggccagcagtagacaacacgccgcac
aagctgattggcatcgccggcagcgacattgccattgtgggcccggggcatcctcaaggcctcagaccccgtagaggag
gcagagcgggtaccgatcagcagcgtggaaagcctacaccgagaggctgctgcgatag

SEQ ID NO:89

25 Xyr1 gene coding sequence

atgtgtccaatcctctccgtcgtattctgcctaccccgacatctcctcggcgtcatttgacccgaactaccatgggtcacagt
cgcatctccactcgatcaacgtcaacacattcggcaacagccacccctatcccatgcagcacctcgcacagcatgcgga
gctttcgagttcacgcatgataagggccagtcgggtgcagccaaagcagcgccagggctctcttattgctgccaggaaga
attcaacGGTactgctgggcccattcggcggaggatcagtcgcgcttgtagaccagtgcaaccagcttcgtaccaagtg
30 cgatggcttacacccatgtgcccattgtataggtatgtcccttttctctacacagtgatgctgcgtcaagcacatgtactgat
cgatcttgtttagaattcggccttggatgcgaatatgtccgagagagaaagaagcgtggcaaagcttcgcgcaaggatatt
gctgcccagcaagccgcgggcggctgcagcacaacactccggccaggtccaggatgggtccagaggatcaacatcgca
aactctcacgccagcaaagcgaatcttcgcgtggcagcgctgagcttgcccagcctgccacgacccgcctcatggcca
cattgagggtctgtcagctccttcagcgacaatggcctttcccagcatgctgccatgggcccagcatggatggcctggaaga

tcaccatggccacgtcggagttgatcctgccctgggcccgaactcagctggaagcgtcatcagcaatgggcctgggcgca
 tacggtgaagtccaccccggtatgagagccccggcatgaatggccatgtgatggtgcccccgctgatggcgcgcaga
 ccaccatggccgggtattccggtatctcgatgctgcgcaagccccgagtcgggtacgtatagcagcgacggtaacttc
 gactcaccgggtcacatccatgattacccgctggcaaatgggagctcgccctcatggggagttctcgctggcctcgccctcga
 5 accagttccagcttcagctctcgcagcccatcttcaagcaaagcgatttgcgatatcctgtgcttgagcctctgctgcctcac
 ctgggaaacatcctccccgtgtctttggcgtgcgatctgattgacctgtacttctcctcgctcttcacacagatgcacca
 atgtcccatacgttctgggcttcgtcttcgggaagcgctccttctgcaccccacgaaccacgaagggtgccagccccg
 ctgcttgcgagcatgctgtgggtggcggcacagactagcgaagcgctccttctgacgagcctgccgtcggcgaggagca
 aggtctgccagaagctgctcgagctgaccgttgggcttcttcagccccgatccacaccggcaccaacagcccgtctccc
 10 aagactagccccgctcgtcgggtgctgctgccctgggagttcttgggggtggccatgccgggctcgtgaacatggattcactg
 gccggcgaaacgggtgcttttggggccatagggagccttgacgacgtcatcacctatgtgcacctcgccacgggtcgtctc
 ggccagcgagtacaagggcgccagcctgcggtgggtggggtgcggcatggtctctgccagagagctcaagcttggccg
 tgagctgccgcctggcaatccacctgccaacaggaggagcgggcgagggccttagcgaagacgtggatgagcacgact
 tgaacagaaacaacactcgcttcgtgacggaagaggagcgcggaagagcgacggcgagcatggtggctcgtttacatc
 15 gtcgacaggcacctggcgctctgctacaaccgccccctgtttctctggacagcgagtgacgagcacttgaccacccgatg
 gacgacatcaagtggcaggcaggcaaatttcgcagccacgatgcagggaactccagcatcaacatcgatagctccatg
 acggacgagtttggcgatagtccccgggcggtcgcggcgccacactacgagtgccgcggctcgtagcattttggctacttc
 ttgtccttgatgacaatcctgggcgagattgtcgatgtccaccatgctaaaagccacccccgggtcggcggttgattccgctc
 cgcgcggttggtggacgagcaggttgcgtgaaatcacccgacacctggacatgtatgaggagagcctcaagagggtcgt
 20 ggccaagcatctgccattgtcctcaaaggacaaggagcagcatgagatgcacgacagtgagcggttaacagacatgc
 aatctccactctcggtgcggaccaacgcgtccagccgcatgacggagagcgagatccaggccagcatcgtgggtggctt
 acagcaccatgtgatgcatgtcctccacatcctccttgcggataagtgggatcccatcaaccttctagacgacgacgact
 gtggatctcgtcgggaaggattcgtgacggcgacgagccacgcggtatcggctgccgaagctattagccagattctcgagt
 ttgacctggcctggagtttatgccattcttctacggcgctctatctcctgcagggttccttcctcctcctgctcatcgccgacaag
 25 ctgcaggccgaagcgtctccaagcgtcatcaaggcttgcgagaccattgttagggcacacgaagcttgcgttgtagcgt
 gagcacagagtatcaggtaagccctatcagttcaaacgtctatcttgcgtgtaataaagactgacttgacatcagcgca
 actttagcaagggttatgcgaagcgcgctggctctgattcggggccgtgtgccggaagatttagctgagcagcagcagcga
 cgacgcgagcttctgcactataccgatggactggtaacggaaccgggtctggccctctaa

30 SEQ ID NO:90

U6 intron

GTTCGTTTCGGCTTTTCCTCGGAACCCCCAGAGGTCATCAGTTCGAATCGCTAACA
 G

SEQ ID NO:91

U6 gene transcriptional terminator sequence

TTTTTTTTCTCTT

5 SEQ ID NO:92

Target Sequence for Pyr4 TS2 guide RNA

GCTCAAGACGCACTACGACA

THAT WHICH IS CLAIMED:

1. A method for modifying the DNA sequence at a target site in the genome of a filamentous fungal cell, the method comprising:
 - a) introducing into a population of filamentous fungal cells a Cas
5 endonuclease and a guide RNA, wherein the Cas endonuclease and guide RNA are capable of forming a complex that enables the Cas endonuclease to introduce a double-strand break at a target site in the genome of the fungal cells; and
 - b) identifying at least one fungal cell from the population that has a
10 modification of the DNA sequence at the target site,
wherein the Cas endonuclease, the guide RNA, or both are introduced transiently into the population of fungal cells.
2. The method of claim 1, wherein the modification of the DNA sequence at said target site is selected from the group consisting of a deletion of one or more
15 nucleotides, an insertion of one or more nucleotides, a substitution of one or more nucleotides, and any combination thereof.
3. The method of claim 1 or 2, wherein introducing the Cas endonuclease into the population of fungal cells is achieved using a method selected from the group
20 consisting of transfection, transduction, transformation, electroporation, particle bombardment (biolistic particle delivery), and cell fusion techniques.
4. The method of any preceding claim, wherein introducing the guide RNA into the population of fungal cells is achieved using a method selected from the group
25 consisting of transfection, transduction, transformation, electroporation, particle bombardment (biolistic particle delivery), and cell fusion techniques.
5. The method of any preceding claim, wherein the identifying step comprises culturing the population of fungal cells from step (a) under conditions to select for or
30 screen for the modification of the DNA sequence at the target site.
6. The method of any preceding claim, wherein the identifying step comprises culturing the population of cells from step (a) under conditions to screen for unstable transformants.

7. The method of any preceding claim, wherein the Cas endonuclease is a Cas9 endonuclease or variant thereof.

8. The method of claim 7, wherein the Cas9 endonuclease or variant thereof comprises a full length Cas9 or a functional fragment thereof from a species selected from the group consisting of: *Streptococcus sp.*, *S. pyogenes*, *S. mutans*, *S. thermophilus*, *Campylobacter sp.*, *C. jejuni*, *Neisseria sp.*, *N. meningitides*, *Francisella sp.*, *F. novicida*, and *Pasteurella sp.*, *P. multocida*.

9. The method of claim 8, wherein the Cas9 endonuclease or variant thereof comprises an amino acid sequence that has at least 70% identity to any one of SEQ ID NOs:1 to 7.

10. The method of any preceding claim, wherein the introducing step comprises introducing a DNA construct comprising an expression cassette for the Cas endonuclease into the fungal cells.

11. The method of any preceding claim, wherein the introducing step comprises introducing a DNA construct comprising an expression cassette for the guide RNA into the fungal cells.

12. The method of any one of claims 1 to 9 and 11, wherein the introducing step comprises directly introducing the Cas endonuclease into the fungal cells.

13. The method of any one of claims 1 to 10 and 12, wherein the introducing step comprises directly introducing the guide RNA into the fungal cells.

14. The method of claim 10, wherein the expression cassette for the Cas endonuclease comprises a Cas coding sequence that is optimized for expression in the fungal cell.

15. The method of claim 14, wherein the Cas coding sequence is a Cas9 coding sequence comprising a polynucleotide sequence that is at least 70% identical to SEQ ID NO:8.

5 16. The method of any preceding claim, wherein the Cas endonuclease is operably linked to a nuclear localization signal.

10 17. The method of claim 11, wherein the expression cassette for the guide RNA comprises a RNA polymerase III dependent promoter functional in a Euscomycete or Pezizomycete, and wherein the promoter is operably linked to the DNA encoding the guide RNA.

18. The method of claim 17, wherein the promoter is derived from a Trichoderma U6 snRNA gene.

15

19. The method of claim 17 or 18, wherein the promoter comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 11 or 12.

20 20. The method of any one of claims 11 and 17-19, wherein the expression cassette for the guide RNA comprises a guide RNA-encoding DNA with an intron sequence from a Trichoderma U6 snRNA gene.

25 21. The method of claim 20, wherein the intron sequence derived from Trichoderma U6 snRNA gene comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 90.

22. The method of any preceding claim, wherein the filamentous fungal cell is a Eumycotina or Pezizomycotina fungal cell.

30 23. The method of any preceding claim, wherein filamentous fungal cell is selected from the group consisting of Trichoderma, Penicillium, Aspergillus, Humicola, Chrysosporium, Fusarium, Myceliophthora, Neurospora, Hypocrea, and Emericella.

24. The method of any preceding claim, wherein the target site is located within a region of a gene of interest selected from the group consisting of: an open reading frame, a promoter, a regulatory sequence, a terminator sequence, a regulatory element sequence, a splice site, a coding sequence, a polyubiquitination site, an intron site, and an intron enhancing motif.

25. The method of any one of claims 1, 2, 4-9, 11, 13, and 16-19, wherein the introducing step comprises: (i) obtaining a parental fungal cell population that stably expresses the Cas endonuclease, and (ii) transiently introducing the guide RNA into the parental fungal cell population.

26. The method of any one of claims 1-3, 5-10, 12, and 14-19, wherein the introducing step comprises: (i) obtaining a parental fungal cell population that stably expresses the guide RNA, and (ii) transiently introducing the Cas endonuclease into the parental fungal cell population.

27. The method of any preceding claim, wherein the modification of the DNA sequence at the target site is not caused by a homologous recombination.

28. The method of any preceding claim, wherein the method does not involve introducing a donor DNA into the population of fungal cells.

29. A recombinant fungal cell produced by the method of any preceding claim.

30. An engineered nucleic acid encoding a Cas endonuclease or variant thereof, wherein the Cas endonuclease or variant thereof comprises an amino acid sequence that has at least 70% identity to any one of SEQ ID NOs:1 to 7, and wherein the nucleic acid comprises a polynucleotide sequence that is at least 70% identical to SEQ ID NO:8.

31. An engineered nucleic acid encoding a guide RNA which enables a Cas endonuclease to introduce a double-strand break at a target site in the genome of a filamentous fungal cell, wherein the nucleic acid encoding the guide RNA comprises a RNA polymerase III dependent promoter functional in a Euscomycete or Pezizomycete, and the promoter is derived from a Trichoderma U6 snRNA gene

32. The engineered nucleic acid of claim 31, wherein the promoter comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 11 or 12.

5 33. An engineered nucleic acid encoding a guide RNA which enables a Cas endonuclease to introduce a double-strand break at a target site in the genome of a filamentous fungal cell, wherein the nucleic acid encoding the guide RNA comprises a guide RNA-encoding DNA with an intron sequence derived from a Trichoderma U6 snRNA gene.

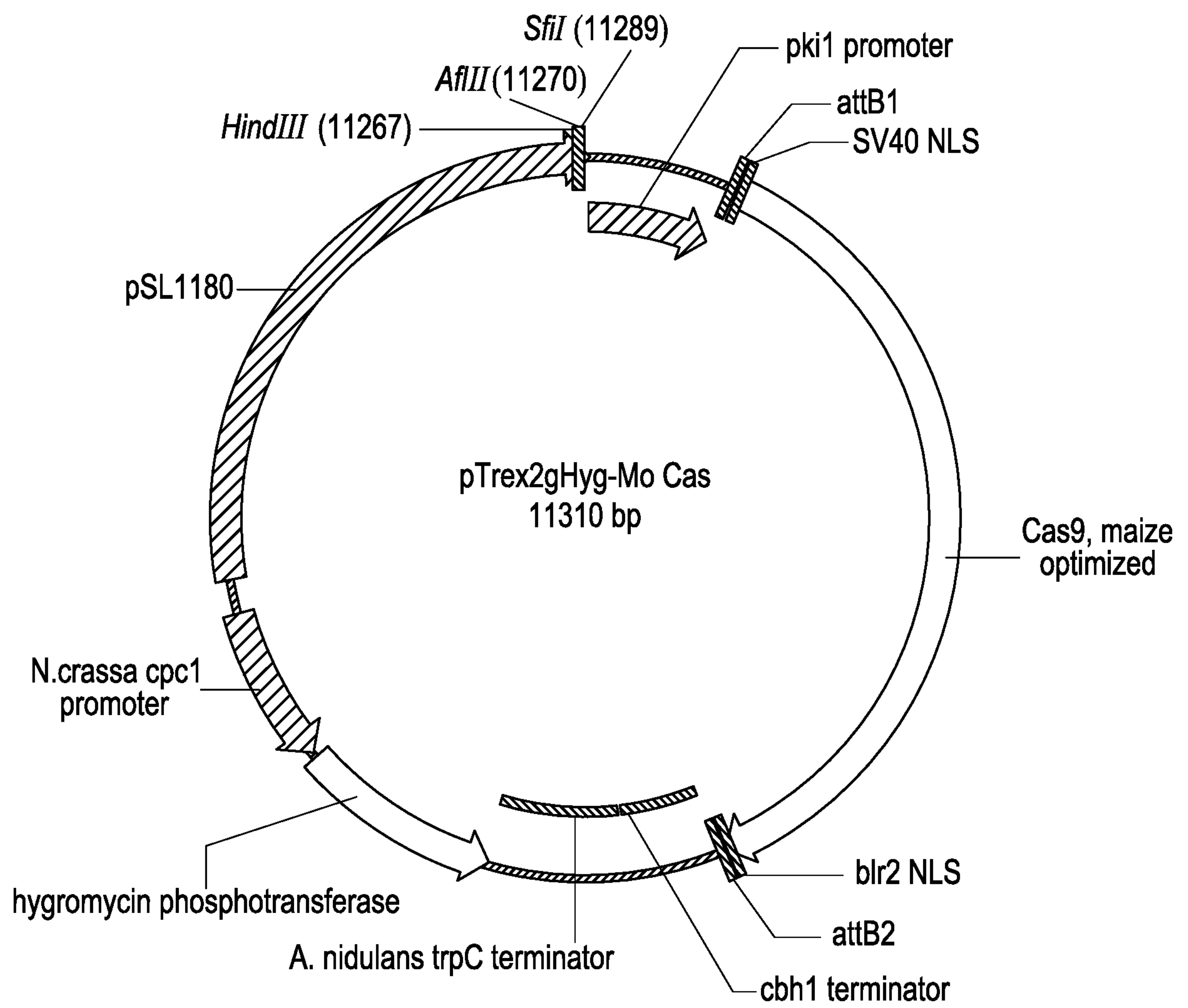
10

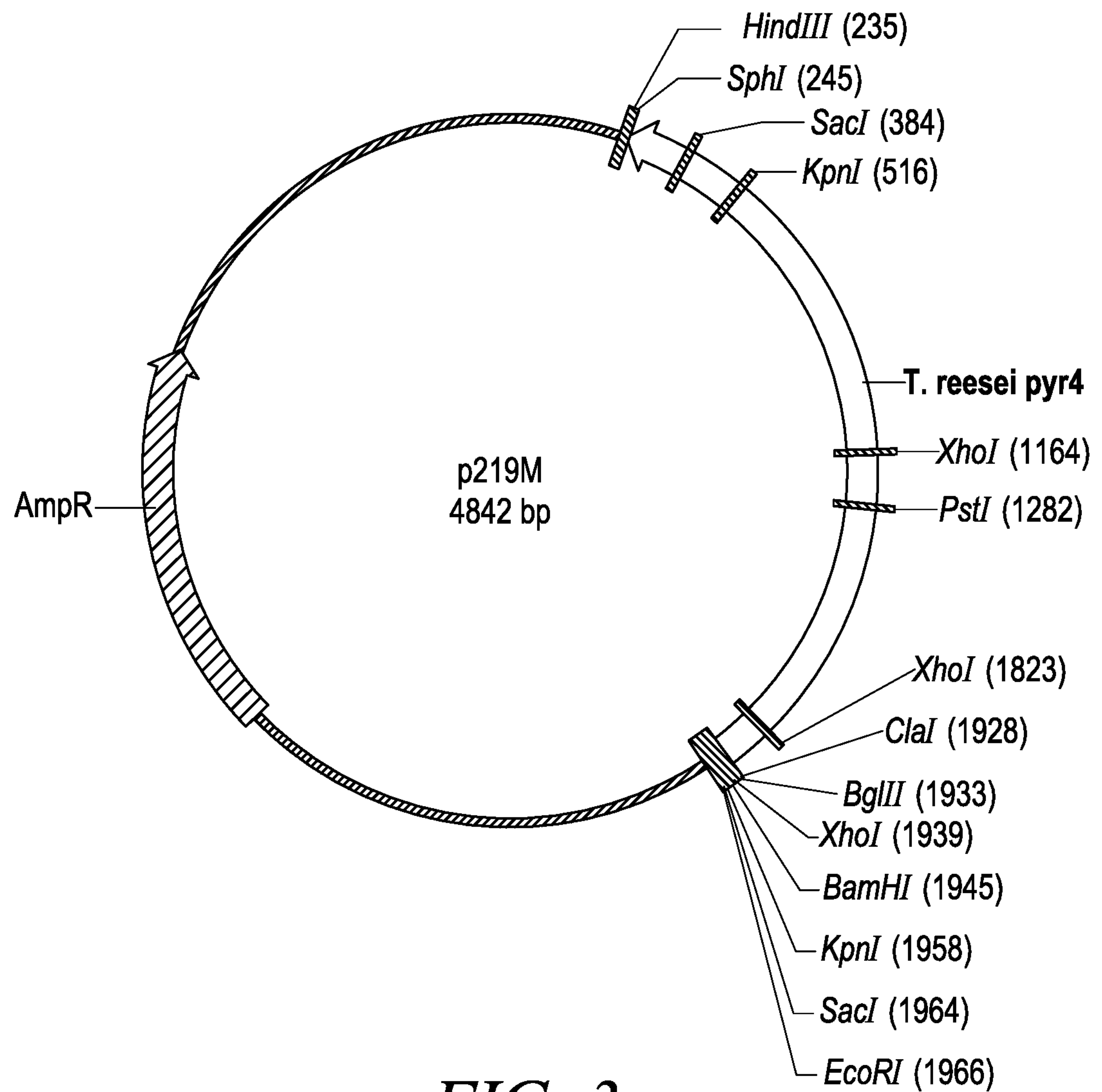
34. The engineered nucleic acid of claim 33, wherein the intron sequence derived from Trichoderma U6 snRNA gene comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 90.

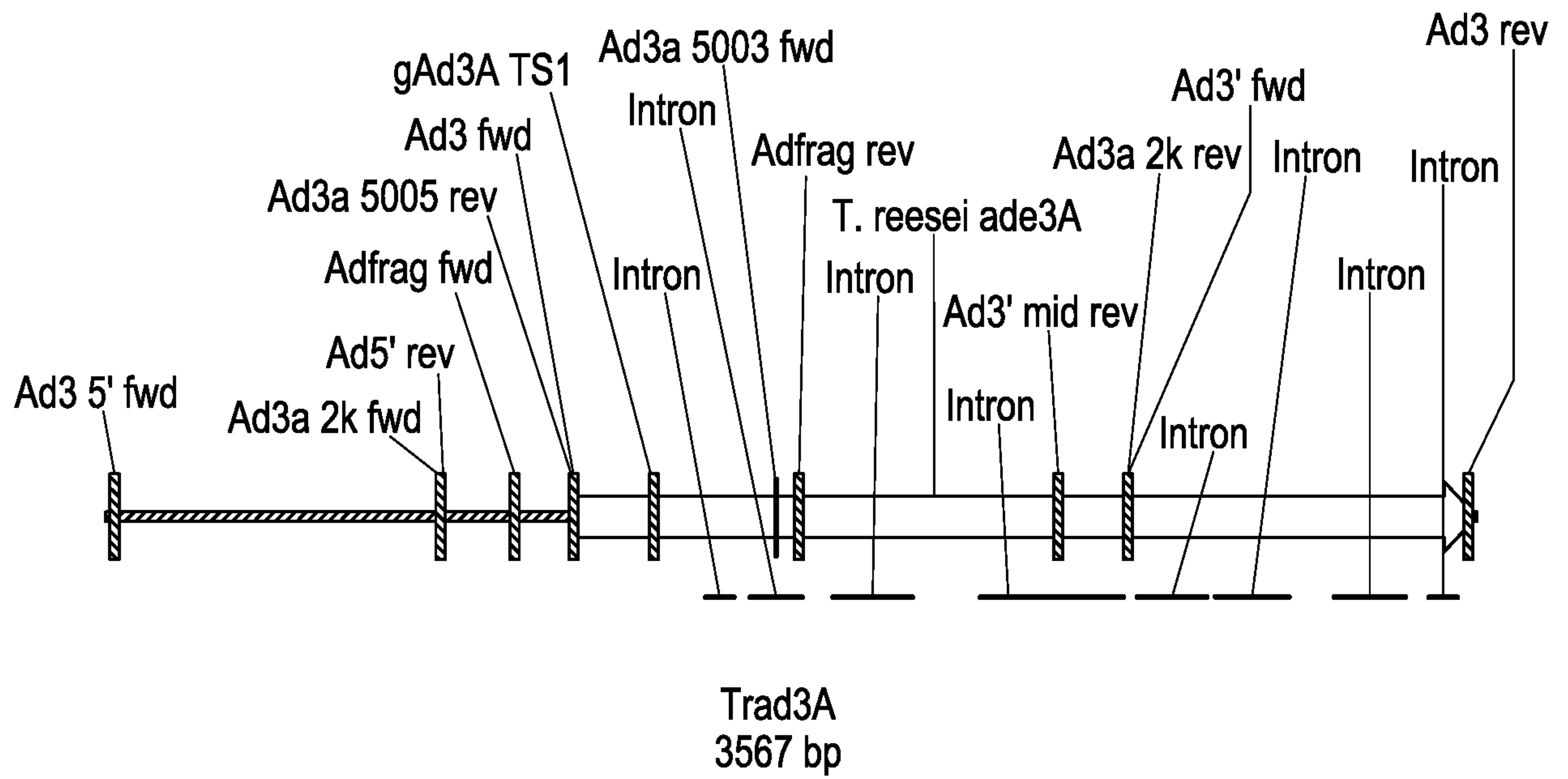
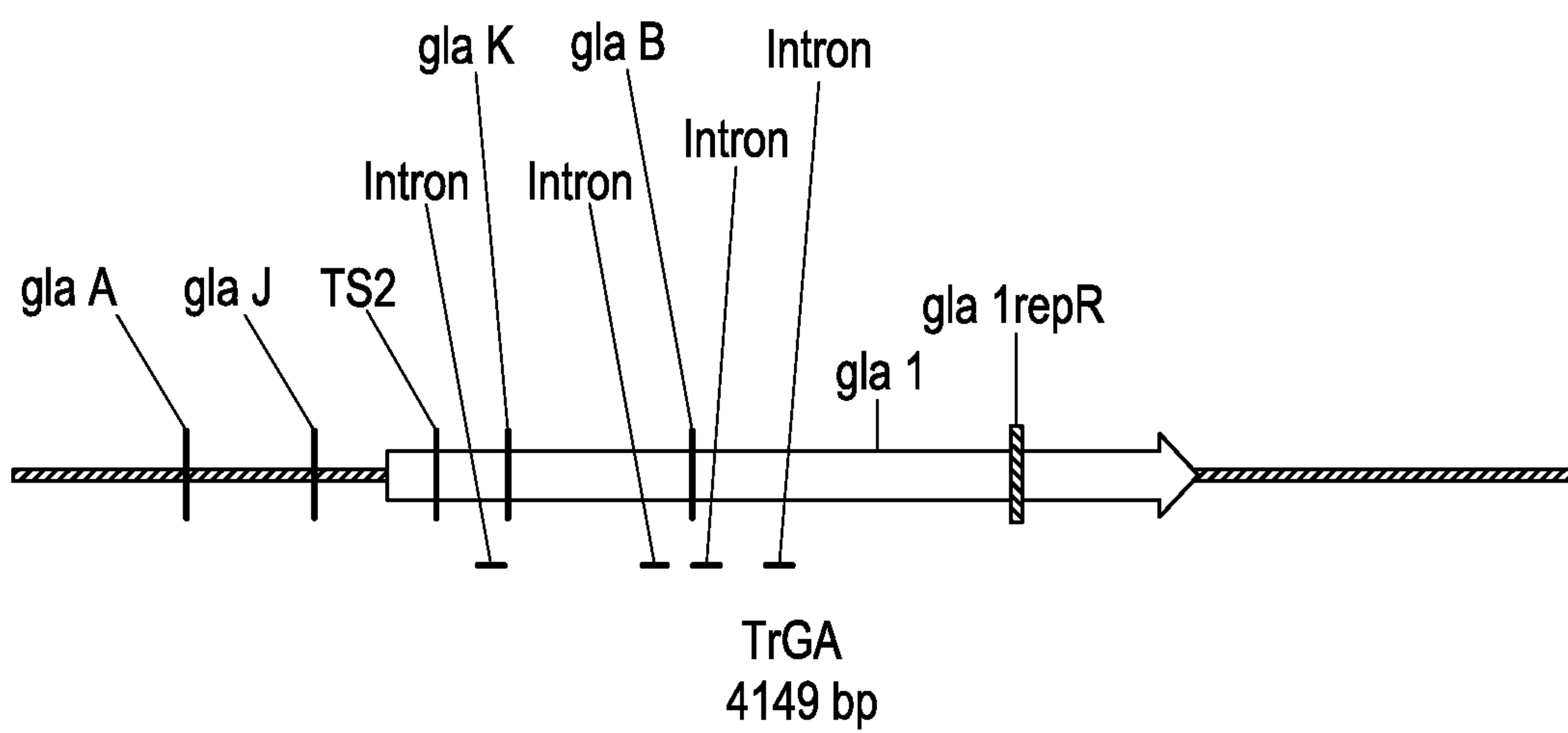
15 35. The engineered nucleic acid of claim 31 or 33, wherein the nucleic acid encoding the guide RNA comprises both a promoter derived from a Trichoderma U6 snRNA gene and an intron sequence derived from a Trichoderma U6 snRNA gene, wherein the promoter comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 11 or 12, and wherein the intron sequence derived from Trichoderma U6 snRNA gene
20 comprises a nucleotide sequence with at least 70% identity to SEQ ID NO: 90.

	Terminator	
AAGGATGACACGC	TCACTCAAAGAGAACTAAACA	<u>TTTTTTTTCCTT</u> CCAAGTCGTGATGGTTA
TCTTTTTGCTTAGAGAA	TCTATTCTTGTGGACGATTAGTATTGGTAAAT	CCCTGCTGCACATTGCGGC
GGATGCTCAACGGCATA	AATACCCATTCTGATGCAGCGTGATCTTCAATATG	TAGTGTAAATACG
TTGCATACACCACAGGT	TCGGTGCCTCCTGTATGTACAGTACTGTAGTTCGACTCCTCCGCGCAGG	
TGGAACGATTCCCTAGTGGCAGGTATTTGGCGGGTCAAGAA		

FIG. 1

2/12**FIG. 2**

3/12**FIG. 3**

4/12**FIG. 4****FIG. 5**

5/12

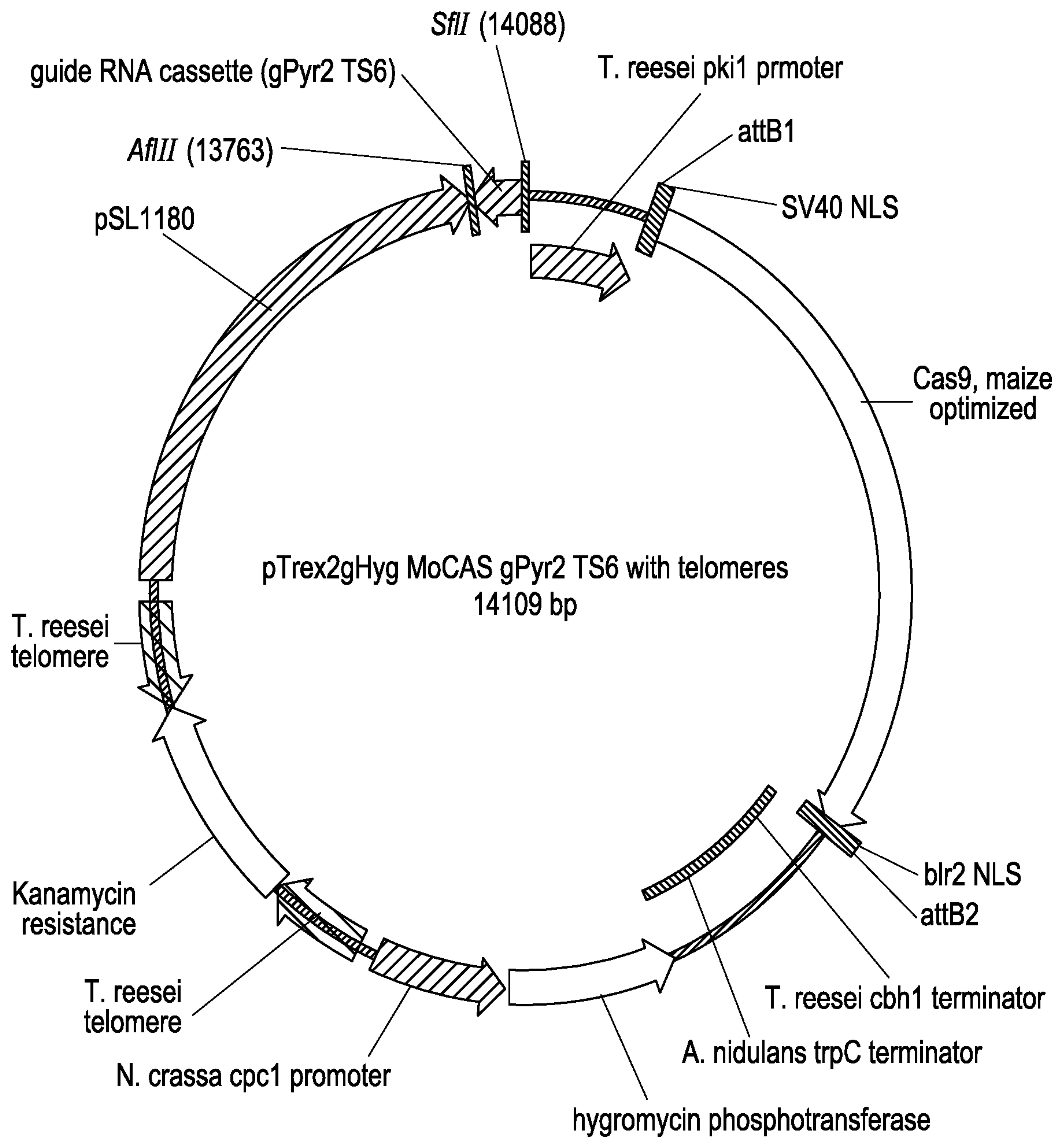
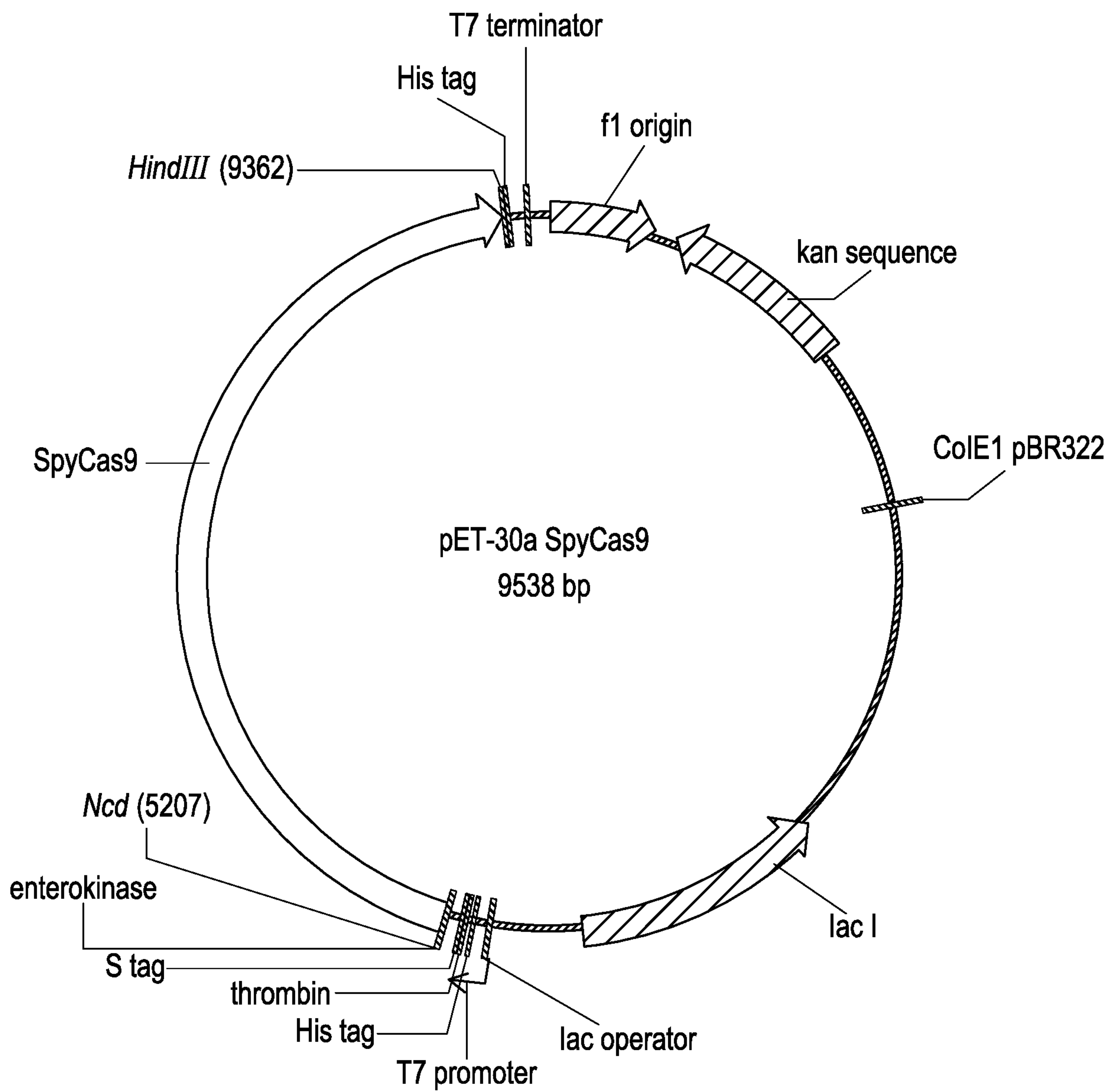


FIG. 6

6/12**FIG. 7**

7/12

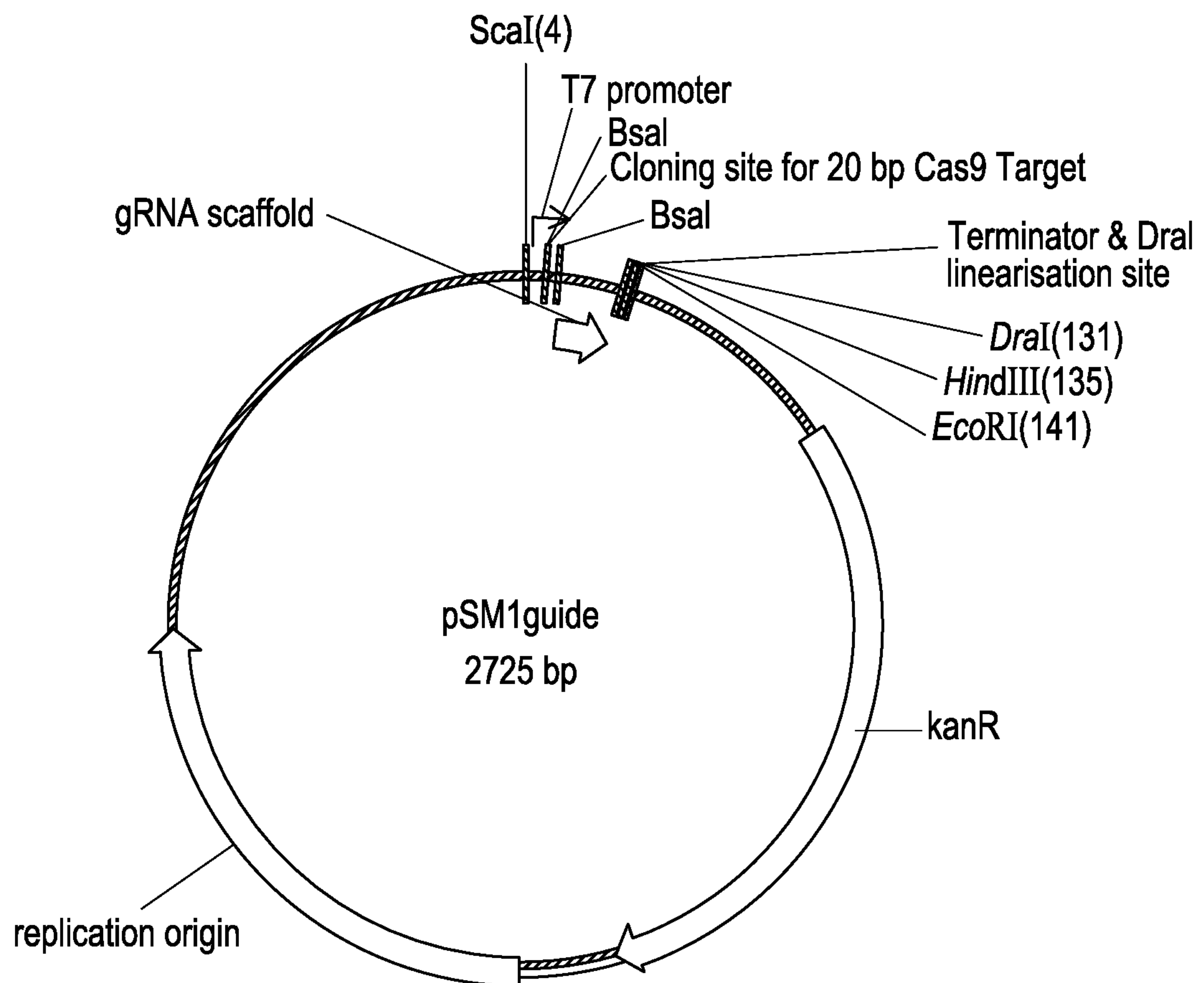


FIG. 8A

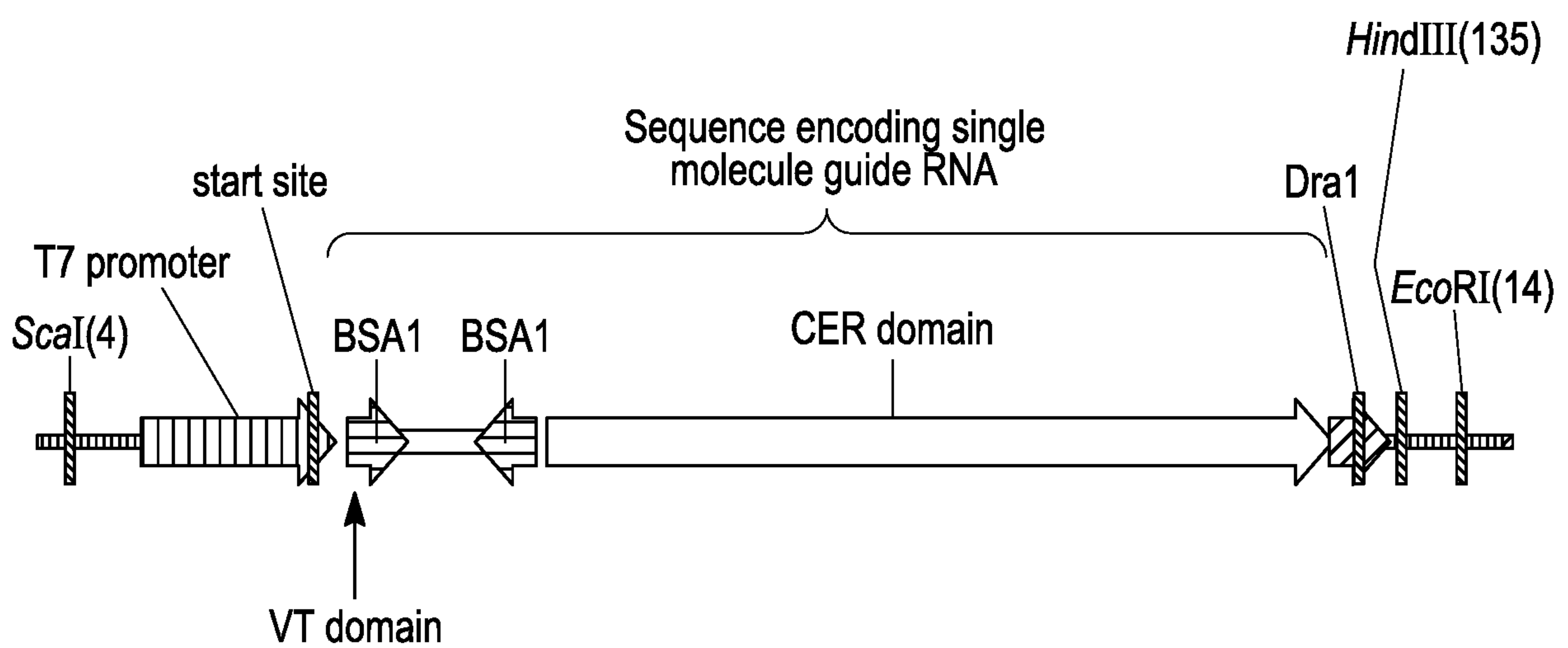


FIG. 8B

8/12

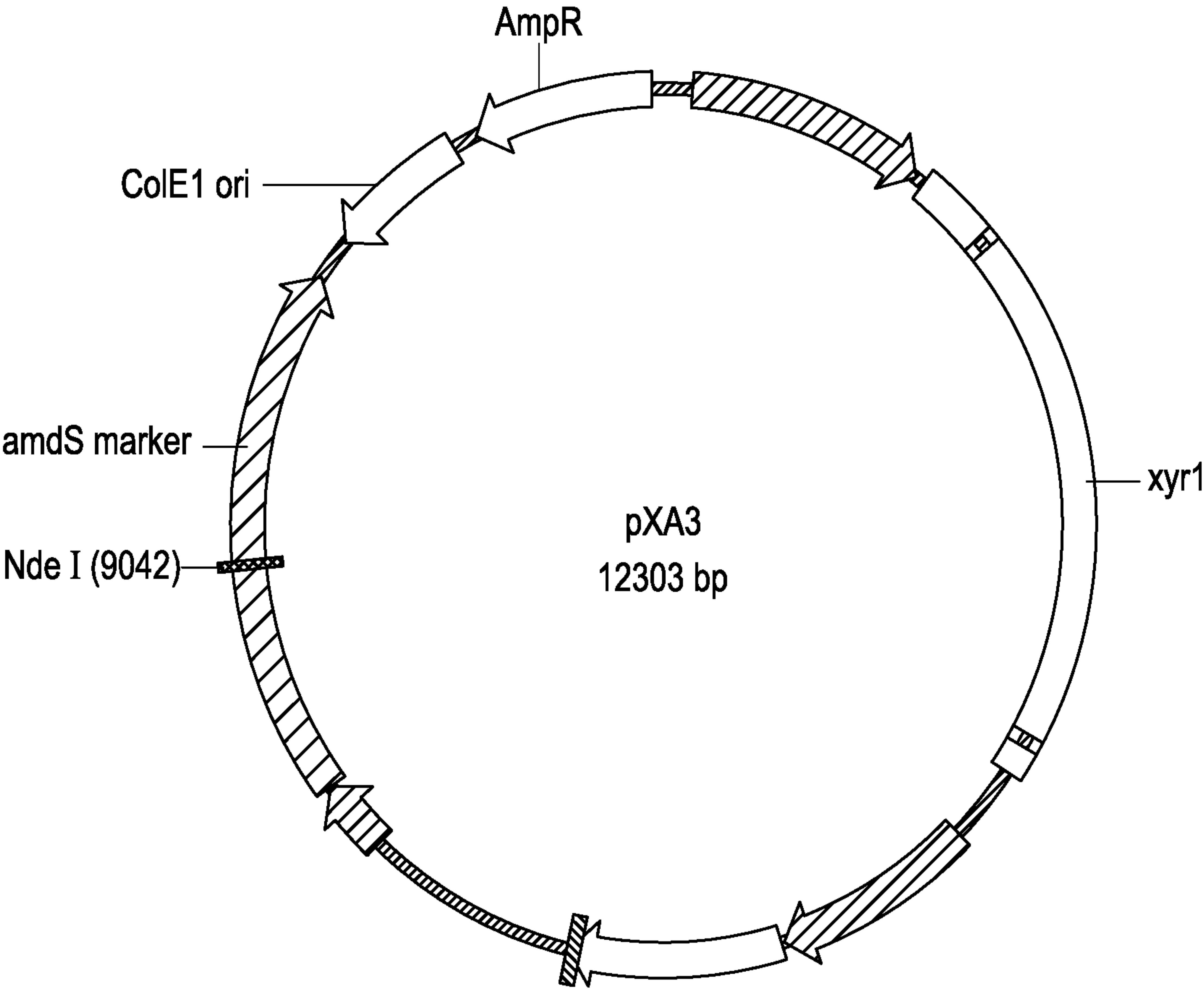


FIG. 9A

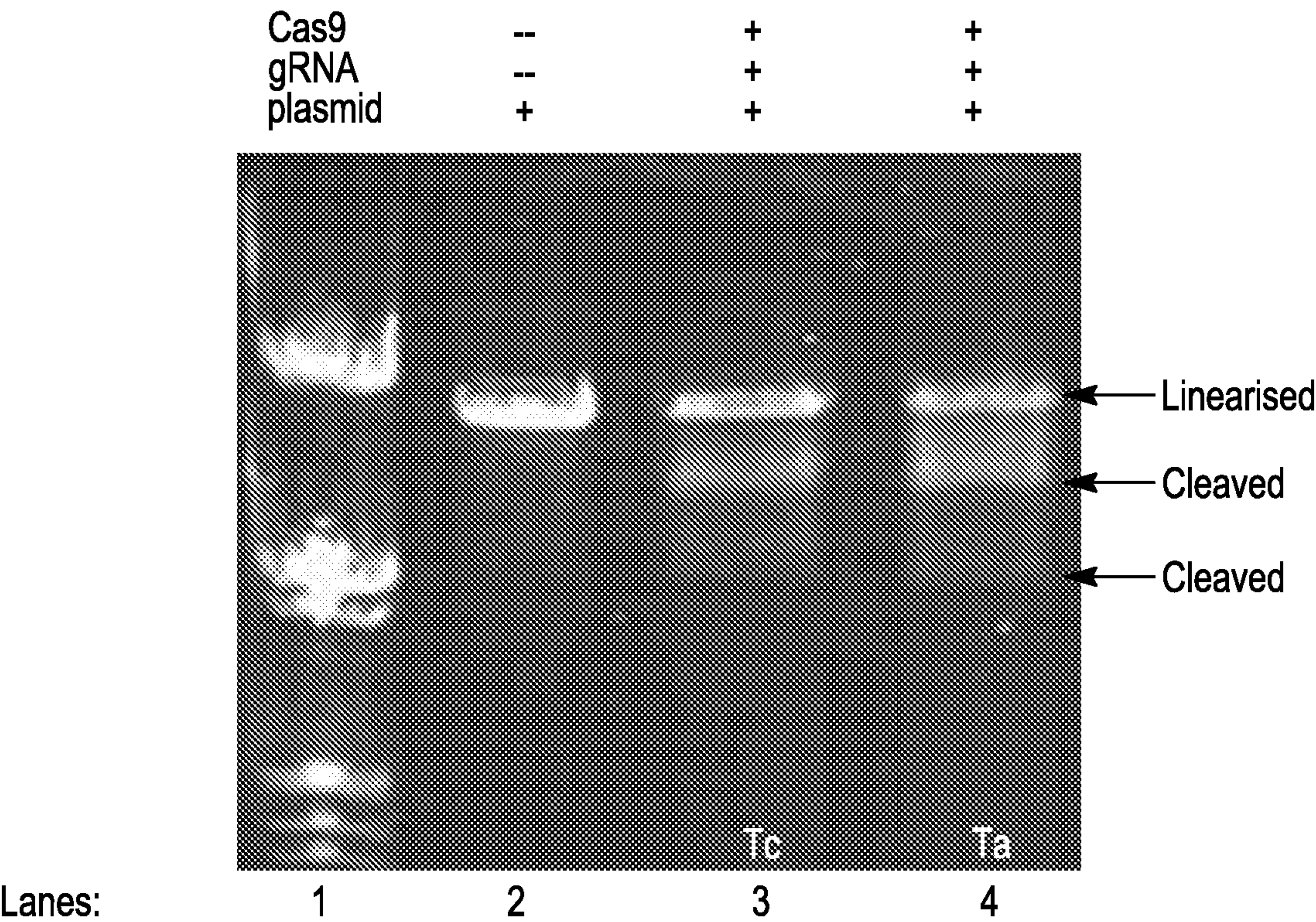
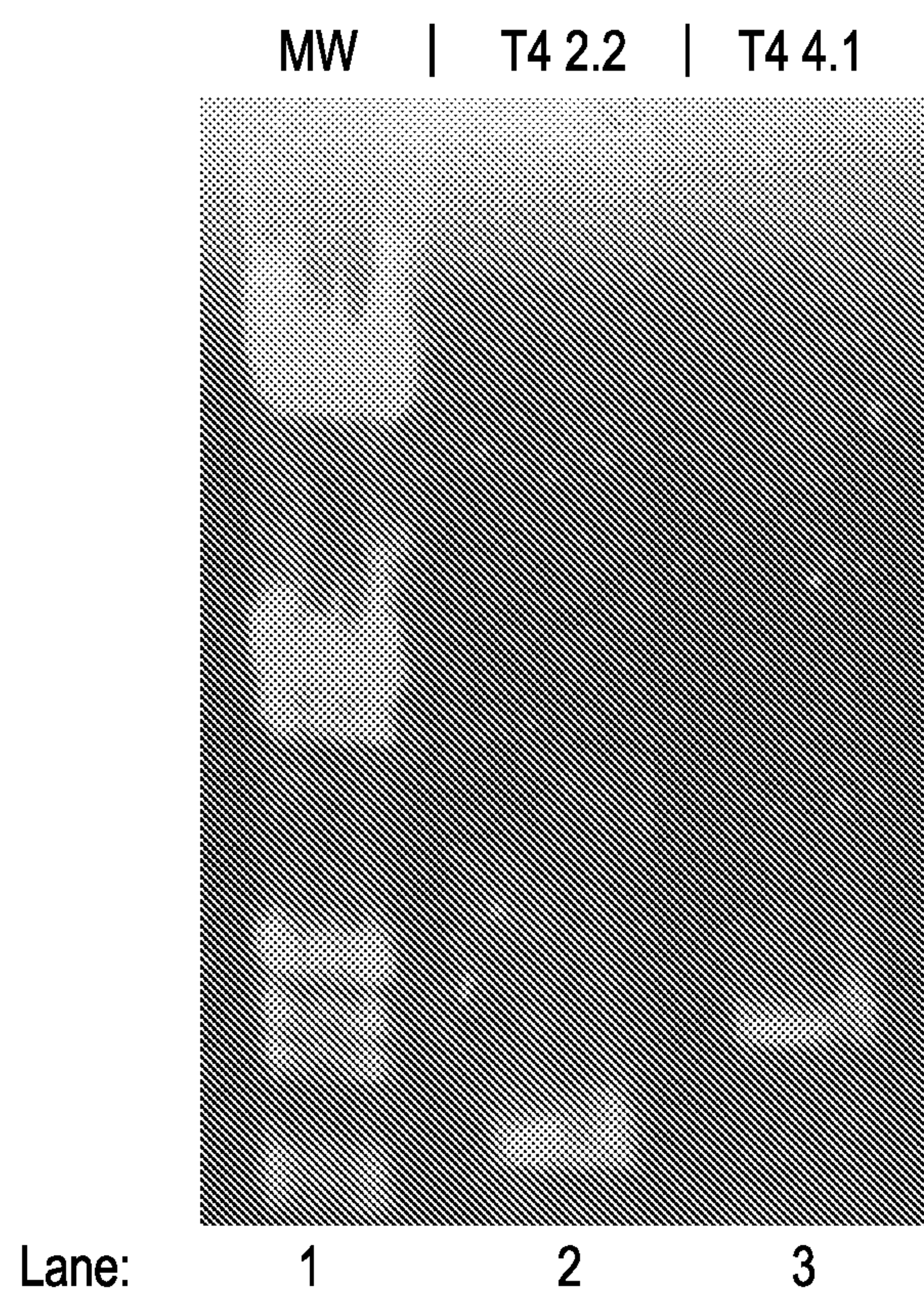
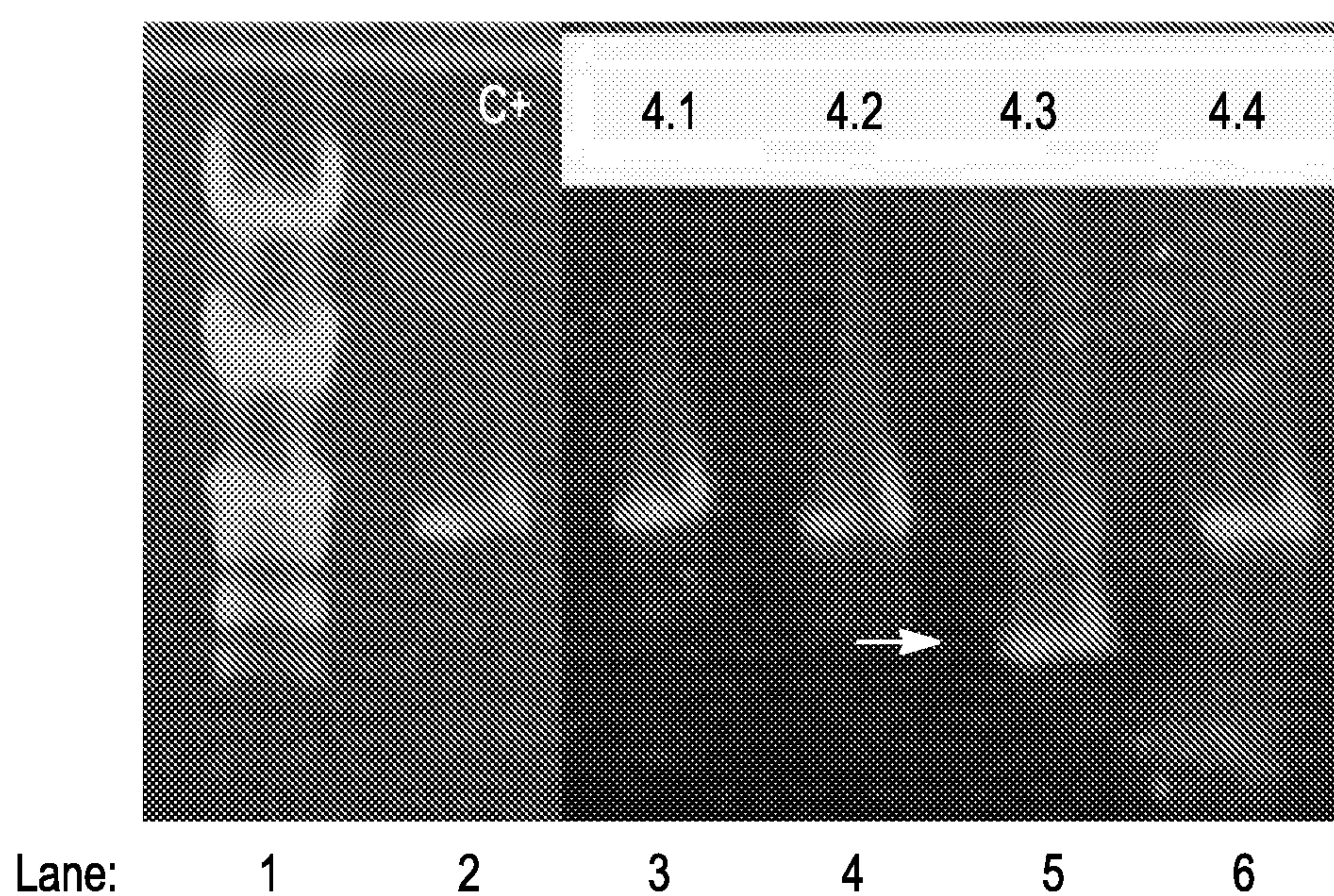


FIG. 9B

	581	590	600	610	620
K21 control T4	---	---	---	---	ACA TGGTCTCGG
T4 4-3	---	---	---	---	ACA TGGTCTCGG
T4 4-13	---	---	---	---	ACA TGGTCTCGG
T4 4-11	---	---	---	---	ACA TGGTCTCGG
T4 4-12	---	---	---	---	ACA TGGTCTCGG
T4 4-18	---	---	G	---	ACA TGGTCTCGG
T4 4-20	ACAGGCTGGTGTGTTAGTGCAGTCCTCTGACA	TGGTCTCGG			
T4 4-19	---	---	---	---	ACA TGGTCTCGG
T4 4-4	---	---	---	---	ACA TGGTCTCGG
T4 4-7	---	---	---	---	ACA TGGTCTCGG

FIG. 10

10/12**FIG. 11A****FIG. 11B**

11/12

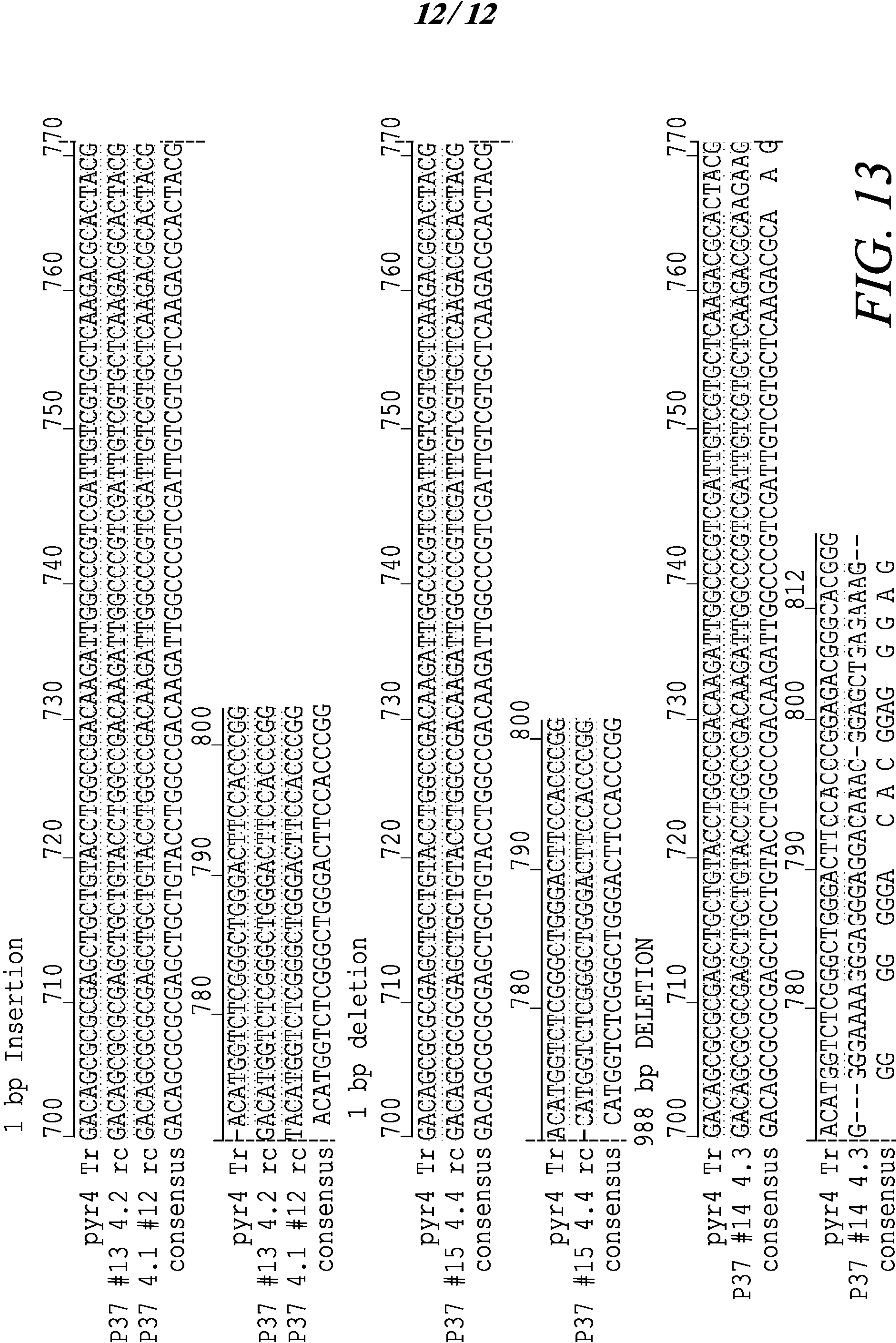
Deletion of 611 nucleotides in pyr4 gene

							Section 12
	(628)	628	640	650	660	670	684
9-96	(345)	CCGCTGACGGCTTACCTGTTCAAGCTCATGGACCTCAAGGCGTCCAACCTGTGCCTG					
pyr4 Tr	(628)	CCGCTGACGGCTTACCTGTTCAAGCTCATGGACCTCAAGGCGTCCAACCTGTGCCTG					
							Section 13
	(685)	685	690	700	710	720	730
9-96	(402)	AGCGCCGACGTGCCGACGGCGCGGAGCTGCTGTACCTGGCCGACAAGATTGGCCCG					
pyr4 Tr	(685)	AGCGCCGACGTGCCGACAGCGCGCGGAGCTGCTGTACCTGGCCGACAAGATTGGCCCG					
							Section 14
	(742)	742	750	760	770	780	798
9-96	(459)	TCGATTGTCGTGCTCAAGACGCACTACG					
pyr4 Tr	(742)	TCGATTGTCGTGCTCAAGACGCACTACGACATGGTCTCGGGCTGGGACTTCCACCCG					
							Section 15
	(799)	799	810	820	830	840	855
9-96	(487)	-----					
pyr4 Tr	(799)	GAGACGGGCACGGGAGCCCAGCTGGCGTCGCTGGCGCGCAAGCACGGCTTCCTCATC					
							Section 16
	(856)	856	870	880	890	900	912
9-96	(487)	-----					
pyr4 Tr	(856)	TTCGAGGACCGCAAGTTTGGCGACATTGGCCACACCGTCGAGCTGCAGTACACGGGC					
	(1084)	1084	1090	1100	1110	1120	1130
9-96	(487)	-----					
pyr4 Tr	(1084)	GACGCCGACTCCAGGGACGCCGAGCCCGCCGGCGCCGTCAACGGCATGGGCTCCATT					
							Section 21
	(1141)	1141	1150	1160	1170	1180	1197
9-96	(487)	-----					
pyr4 Tr	(1141)	GGCGTCCTGGACAAGCCCATCTACTCGAACCGGTCCGGCGACGGCCGCAAGGGCAGC					
							Section 22
	(1198)	1198	1210	1220	1230	1240	1254
9-96	(487)	-----					
pyr4 Tr	(1198)	ATCGTCTCCATCACCACCGTCACCCAGCAGTACGAGTCCGTCTCCTCGCCCCGGTTA					
							Section 23
	(1255)	1255	1260	1270	1280	1290	1300
9-96	(487)	-----					
pyr4 Tr	(1255)	ACAAAGGCCATCGCCGAGGGCGACGAGTCGCTCTTCCCGGGCATCGAGGAGGCGCCG					
							Section 24
	(1312)	1312	1320	1330	1340	1350	1368
9-96	(487)	-----					
pyr4 Tr	(1312)	CTGAGCCGCGGCCTCCTGATCCTCGCCCAAATGTCCAGCCAGGGCAACTTCATGAAC					
	(1369)	1369	1380	1390	1400	1410	1425
9-96	(487)	-----					
pyr4 Tr	(1369)	AAGGAGTACACGCAGGCCTGCGTCGAGGCCGCGCCGGGAGCACAAGGACTTTGTTCATG					

P37 (2.4) #10MH94 1 bp insertion

Query 521 CTGGCCGACAAGATTGGCCCGTCGATTGTCGTGCTCAAGACGCACTACG-ACATGGTCTC 579
 |||||
 Sbjct 523 CTGGCCGACAAGATTGGCCCGTCGATTGTCGTGCTCAAGACGCACTACGGACATGGTCTC 464

FIG. 12



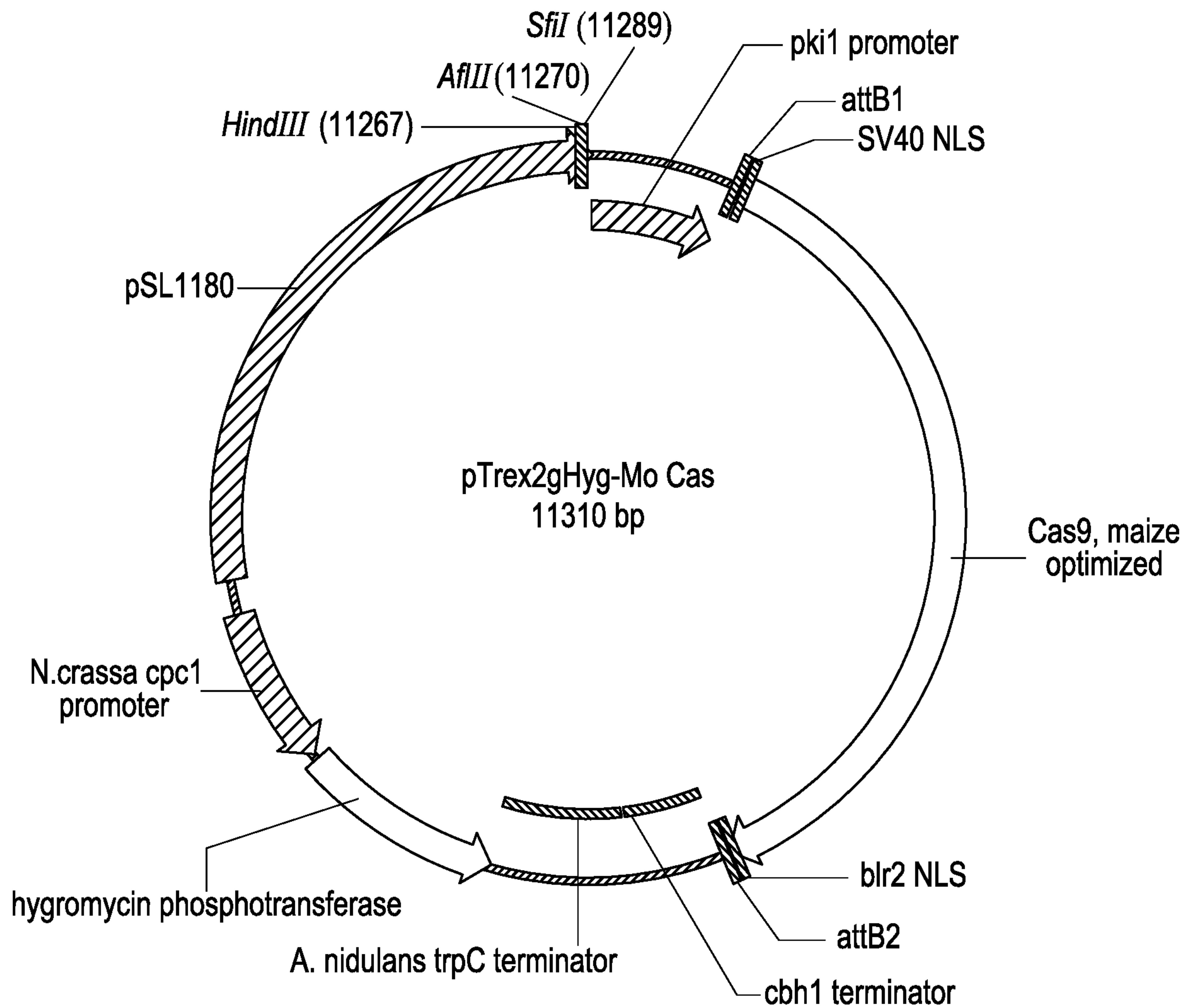


FIG. 2