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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2016 (08.12.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/196361 A1

(51) International Patent Classification:
C12Q 1/02 (2006.01) *C12N 1/21* (2006.01)
C12Q 1/68 (2006.01) *C12N 15/63* (2006.01)

(21) International Application Number:
PCT/US2016/034812

(22) International Filing Date:
27 May 2016 (27.05.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/168,355 29 May 2015 (29.05.2015) US
62/296,853 18 February 2016 (18.02.2016) US

(71) Applicant: **NORTH CAROLINA STATE UNIVERSITY** [US/US]; 1021 Main Campus Drive, 2nd Floor, Raleigh, NC 27606 (US).

(72) Inventors: **BARRANGOU, Rodolphe**; 1021 Main Campus Drive, 2nd Floor, Raleigh, NC 27606 (US). **SELLE, Kurt, M.**; 1021 Main Campus Drive, 2nd Floor, Raleigh, NC 27606 (US).

(74) Agent: **BONNEN, Alice, M.**; Myers Bigel & Sibley, P.A., P.O. Box 37428, Raleigh, NC 27627 (US).

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

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

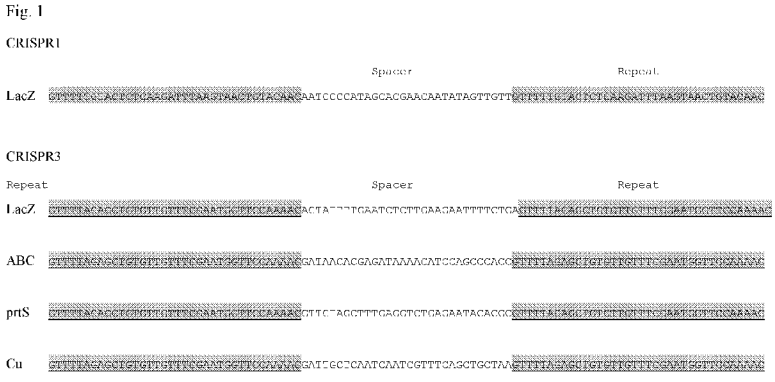
Published:

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

WO 2016/196361 A1

(54) Title: METHODS FOR SCREENING BACTERIA, ARCHAEA, ALGAE, AND YEAST USING CRISPR NUCLEIC ACIDS

Fig. 1



(57) Abstract: This invention relates to the use of CRISPR nucleic acids to screen for essential and non-essential genes and expendable genomic islands in bacteria, archaea, algae and/or yeast, to kill bacteria, archaea, algae and/or yeast, to identify the phenotype of a gene or genes, and/or to screen for reduced genome size and/or a gene deletion in bacteria, archaea, algae and/or yeast.

METHODS FOR SCREENING BACTERIA, ARCHAEA, ALGAE, AND YEAST USING CRISPR NUCLEIC ACIDS

STATEMENT OF PRIORITY

This application claims the benefit, under 35 U.S.C. § 119 (e), of U.S. Provisional Application No. 62/168,355 filed on May 29, 2015, and U.S. Provisional Application No. 62/296,853, filed on February 18, 2016, the entire contents of each of which is incorporated by reference herein.

FIELD OF THE INVENTION

The invention relates to the use of CRISPR nucleic acids to screen for essential and non-essential genes and expendable genomic islands in bacteria, archaea, algae and/or yeast, to kill bacteria, archaea, algae and/or yeast, to identify the phenotype of a gene or genes, and/or to screen for reduced genome size and/or a gene deletion in bacteria, archaea, algae and/or yeast.

BACKGROUND OF THE INVENTION

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), in combination with associated sequences (*cas*), constitute the CRISPR-Cas system, which confers adaptive immunity in many bacteria. CRISPR-mediated immunization occurs through the uptake of DNA from invasive genetic elements such as plasmids and phages, as novel "spacers."

CRISPR-Cas systems consist of arrays of short DNA repeats interspaced by hypervariable sequences, flanked by *cas* genes, that provide adaptive immunity against invasive genetic elements such as phage and plasmids, through sequence-specific targeting and interference (Barrangou et al. 2007. *Science*. 315:1709-1712; Brouns et al. 2008. *Science* 321:960-4; Horvath and Barrangou. 2010. *Science*. 327:167-70; Marraffini and Sontheimer. 2008. *Science*. 322:1843-1845; Bhaya et al. 2011. *Annu. Rev. Genet.* 45:273-297; Terns and Terns. 2011. *Curr. Opin. Microbiol.* 14:321-327; Westra et al. 2012. *Annu. Rev. Genet.* 46:311-339; Barrangou R. 2013. *RNA*. 4:267-278). Typically, invasive DNA sequences are acquired as novel "spacers" (Barrangou et al. 2007. *Science*. 315:1709-1712), each paired with a CRISPR repeat and inserted as a novel repeat-spacer unit in the CRISPR locus. Subsequently, the repeat-spacer array is transcribed as a long pre-CRISPR RNA (pre-crRNA) (Brouns et al. 2008. *Science* 321:960-4), which is processed into small interfering CRISPR

RNAs (crRNAs) that drive sequence-specific recognition. Specifically, crRNAs guide nucleases towards complementary targets for sequence-specific nucleic acid cleavage mediated by Cas endonucleases (Garneau et al. 2010. *Nature*. 468:67-71; Haurwitz et al. 2010. *Science*. 329:1355-1358; Sapranauskas et al. 2011. *Nucleic Acid Res*. 39:9275-9282; Jinek et al. 2012. *Science*. 337:816-821; Gasiunas et al. 2012. *Proc. Natl. Acad. Sci.* 109:E2579-E2586; Magadan et al. 2012. *PLoS One*. 7:e40913; Karvelis et al. 2013. *RNA* Biol. 10:841-851). These widespread systems occur in nearly half of bacteria (~46%) and the large majority of archaea (~90%).

In general terms, there are two main classes (Makarova et al. *Nat Rev Microbiol.* **13**:722–736 (2015)) of CRISPR-Cas systems, which encompass five major types and 16 different subtypes based on cas gene content, cas operon architecture, Cas protein sequences, and process steps (Makarova et al. *Biol Direct*. 6:38 (2011); Makarova and Koonin *Methods Mol Biol*. 1311:47–75 (2015); Barrangou, R. *Genome Biology* 16:247 (2015)). In types I and III, the specialized Cas endonucleases process the pre-crRNAs, which then assemble into a large multi-Cas protein complex capable of recognizing and cleaving nucleic acids complementary to the crRNA. Type I systems are the most frequent and widespread systems, which target DNA in a Cascade-driven and PAM-dependent manner, destroying target nucleic acids by using the signature protein Cas3. A different process is involved in Type II CRISPR-Cas systems. Here, the pre-CRNAs are processed by a mechanism in which a trans-activating crRNA (tracrRNA) hybridizes to repeat regions of the crRNA. The hybridized crRNA-tracrRNA are cleaved by RNase III and following a second event that removes the 5' end of each spacer, mature crRNAs are produced that remain associated with the both the tracrRNA and Cas9. The mature complex then locates a target dsDNA sequence ('protospacer' sequence) that is complementary to the spacer sequence in the complex and cuts both strands. Target recognition and cleavage by the complex in the type II system not only requires a sequence that is complementary between the spacer sequence on the crRNA-tracrRNA complex and the target 'protospacer' sequence but also requires a protospacer adjacent motif (PAM) sequence located at the 3' end of the protospacer sequence.

SUMMARY OF THE INVENTION

One aspect of the invention provides a method of screening a population of bacterial cells for essential genes, non-essential genes, and/or expendable genomic islands, comprising: introducing into said population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence

or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby producing a population of transformed bacterial cells; determining the presence or absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells, wherein the presence of a deletion in the population of transformed bacterial, archaeal or yeast cells means that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells means that the target region is comprised within an essential gene. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A second aspect of the invention provides a method of screening a population of bacterial, archaeal, algal or yeast cells for essential genes, non-essential genes, and/or expendable genomic islands, comprising: introducing into the population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells; and determining the presence or absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells, wherein the presence of a deletion in the population of transformed bacterial, archaeal or yeast cells means that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells means that the target region is comprised within an essential gene.

A third aspect of the invention provides a method of killing one or more bacterial cells within a population of bacterial cells, comprising: introducing into the population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-

spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby killing one or more bacterial cells that comprise the target region within the population of bacterial cells. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A fourth aspect of the invention provides a method of killing one or more cells within a population of bacterial, archaeal, algal or yeast cells, comprising: introducing into the population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing one or more cells within a population of bacterial, archaeal, algal or yeast cells that comprise the target region in their genome.

A fifth aspect of the invention provides a method of identifying a phenotype associated with a bacterial gene, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, wherein the target region comprises at least a portion of an open reading frame encoding a polypeptide or functional nucleic acid, thereby killing the cells comprising the target region and producing a population of transformed bacterial cells without the target region; and analyzing the phenotype of the population. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A sixth aspect of the invention provides a method of identifying a phenotype of a bacterial, archaeal, algal or yeast gene, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-

spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing the bacterial, archaeal or yeast cells comprising the target region and producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and analyzing the phenotype of the population of transformed bacterial, archaeal, algal or yeast cells, and/or (i) growing individual bacterial, archaeal, algal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells and (ii) analyzing the phenotype of the individual colonies.

A seventh aspect of the invention provides a method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the cells comprising the target region are killed, thereby selecting one or more bacterial cells without the target region and having a reduced genome size from the population of bacterial cells. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

An eighth aspect of the invention provides a method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells, comprising: introducing into a population of bacterial cells (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat

sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby selecting one or more bacterial cells without the target region and having a reduced genome size from the population of transgenic bacterial cells. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A ninth aspect of the invention provides a method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced the genome size from a population of bacterial, archaeal, algal or yeast cells, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and the at least one repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of bacterial, archaeal, algal or yeast cells.

A tenth aspect of the invention provides a method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced the genome size from a population of bacterial, archaeal or yeast cells, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells:(a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b)(i) a

heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and the at least one repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of one or more bacterial, archaeal, algal or yeast cells of said population, and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of transgenic bacterial, archaeal, algal or yeast cells.

An eleventh aspect of the invention provides a method of identifying in a population of bacteria at least one isolate having a deletion in its genome, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein cells comprising the target region are killed, thereby producing a population of transformed bacterial cells without the target region; and growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of transformed bacteria having a deletion in its genome. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A twelfth aspect of the invention provides a method of identifying in a population of bacteria at least one isolate having a deletion in its genome, comprising: introducing into the population of bacterial cells (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into

the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby producing a population of transformed bacterial cells without the target region; and growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of bacteria having a deletion in its genome. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

A thirteenth aspect of the invention provides a method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells: (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population of transformed bacterial, archaeal, algal or yeast cells having a deletion in its genome.

A fourteenth aspect of the invention provides a method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome, comprising: introducing into the population of bacterial, archaeal, algal or yeast cells (a)(i)

one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b)(i) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of one or more bacterial, archaeal, algal or yeast cells of said population, and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population having a deletion in its genome.

Further provided herein are expression cassettes, cells and kits comprising the nucleic acid constructs, nucleic acid arrays, nucleic acid molecules and/or nucleotide sequences of the invention.

These and other aspects of the invention are set forth in more detail in the description of the invention below.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the sequence of SthCRISPR1 and SthCRISPR 3 arrays for targeting each composite transposon.

Fig. 2 shows a schematic of the splicing overlap extension (SOE) method for construction of targeting plasmids.

Figs. 3A-3E show a map of essential genes, insertion sequences and genomic islands. (**A**) The location and distribution of putative essential ORFs (top row), insertion sequences (2^{nd} row) and putative genomic islands (3^{rd} row). Potential targets for CRISPR-Cas (4^{th} row) mediated deletion were identified by mapping transposable elements of various families within the genome of *Streptococcus thermophilus* LMD-9. Genetic organization of putative genomic islands and the protospacer/PAM combinations corresponding to each. (**B**) Genomic island 1, encoding an oligopeptide transport system. (**C**) Genomic island 2, containing the cell-envelope proteinase PrtS. (**D**) Genomic island 3, encoding an ATPase copper-efflux protein. (**E**) The genomic island encoding selected ORFs including the Lac operon.

Fig. 4 provides a dendrogram of transposon coding sequences distributed throughout the genome of *S. thermophilus* LMD-9. The alignment was created using the Geneious® Software. Family designations were assigned using www-is.biotoul.fr. The letters correspond to alignments in **Fig. 5** for each family.

Fig. 5A-5D show the alignments of transposon coding sequences (using Geneious® software) for each major IS family found in the *S. thermophilus* LMD-9 genome. Different families exhibited varying levels of conservation of length and nucleotide identity. (**A**) Sth6 transposons were highly polymorphic and apparently degenerate due to internal deletions in some of the copies. In contrast, IS1167 (**B**) and IS1191 (**C**) had fewer copies but maintained high fidelity in length and identity. IS1193 (**D**) had high fidelity copies but exhibited the greatest intra-family diversity in length.

Fig. 6A-6C show the structural basis for and apparent cytotoxicity of DNA targeting by CRISPR-Cas9. Spacer sequences for SthCRISPR1 (**A**) and SthCRISPR3 (**B**) for targeting *lacZ*. Cas9 interrogates DNA and binds reversibly to PAM sequences with stabilization of Cas9 at the target occurring via formation of the tracrRNA::crRNA duplex. Activation of the Cas9 causes simultaneous cleavage of each strand by the RuvC and HNH domains, as denoted by black wedges. Transformants recovered following electroporation of control and self-targeting plasmids (**C**). Average clones ± SD screened across independent transformation experiments ($n = 4$) for each of the plasmids tested.

Figs. 7A-7D shows genome sequencing and phenotypic analysis of Lac$^-$ clones. Sequence data revealed an absence of the chromosomal segment encoding *lacZ* in two mutants independently created by targeting the (**A**) 5' end and (**B**) cation-binding residue coding sequences of *lacZ* using the CRISPR3 system. The size of the deletions ranged from 101,865-102,146 bp in length, constituting approximately 5.5% of the genome of *S. thermophilus*. (**C**) Growth of large deletion strains compared to wild-type in semi-synthetic Elliker medium represented as mean ± SD OD 600 nm of three independent biological replicates (**D**) Acidification capacity of *S. thermophilus* strains in skim milk.

Fig. 8A-8E provides a depiction of recombination events between insertion sequences (IS). (**A**) Gel electrophoresis image of large deletion amplicons yielded by PCR analysis of gDNA recovered from transformants. Screening was performed using primers flanking the IS1193 elements upstream and downstream of the putative deletion site. Lanes denoted with Δ were amplified from gDNA of Lac$^-$ clones recovered following CRISPR-Cas mediated targeting of *lacZ,* whereas WT is from wild-type. (**B**) Sequences of predicted recombination sites were determined by mapping single nucleotide polymorphisms corresponding to either upstream (gray) or downstream (black) IS elements. The three sites are predicted based on sequences conserved in both IS elements (light gray). The sites depicted represent genotypes from independent clones and are representative of the Lac$^-$ phenomenon observed at nine different recombination sites. Chimeric IS element footprints were similarly found in each genomic island locus at the deletion junction. (**C**) Schematic of IS's predicted to recombine during chromosomal deletion of the island encoding *lacZ*. (**D**) Amplicons generated from primers flanking genomic islands 1, 2, and 3 to confirm deletions. (**E**) Amplicons generated from internal primers to confirm the absence of wild-type sequences in each CRISPR-induced deletion culture. Lanes denoted with Δ were amplified from gDNA of clones recovered following CRISPR-Cas mediated targeting, and WT is wild-type.

Fig. 9 provides targets of lethality and shows use of defined genetic loci for assessing type II CRISPR-Cas system-based lethality via targeting the genome of *Streptococcus thermophilus* LMD-9. Both orthogonal type II systems (CRISPR1 and CRISPR3) were tested; CRISPR1 targets in dark grey, CRISPR3 targets in light grey. Specific genetic features were selected to test (i) intergenic regions (INT), (ii) mobile genetic elements (ISSth7, *oppC*-GEI1, *prtS*-GEI2, *copA*-GEI3, *cI*, *lacZ*-GEI4, *epsU*), (iii) essential genes (*dltA*,

ltaS), (iv) poles of the replichore (OriC, *xerS*), and forward vs. reverse strands of DNA (outer targets vs. inner targets).

Fig. 10 shows CRISPR-based lethality achieved by targeting the regions defined in **Fig. 9**. Log reduction in CFU (cell forming units) was calculated with regard to transformation of a non-targeting plasmid control; pORI28. Lethality ranged from 2-3 log reduction for all targets tested, regardless of chromosomal location, coding sequence, or essentiality. ISSth7-insertion sequence element, ltaS-lipoteichoic acid synthase; prtS-genomic island 2; INT-intergenic region; dltA-D-alanine ligase; rheB -chi site deficient locus; oppC-genomic island 1; comS-chi site dense locus; xerS-terminus of replication; copA-genomic island 3; cI-prophage remnant; OriC-origin of replication; Cas9-CRISPR3 Cas9 coding sequence; epsU-exopolysaccharide cassette.

Fig. 11 shows transcriptional profiles of CRISPR-mediated genomic island deletion strains.

Fig 12 shows $\log_2$ transformed RNA-sequencing read coverage of genomic island deletion strains,GEI1, GEI2, GEI3, and GEI4.

Fig 13 shows XY plots of genomic island deletion strain expression values (X-axes) verses wild-type expression values (Y-axes). For each of the genomic island deletion strains (GEI1-GEI4), the expression of genes encoded on each of the target islands (black) was minimal. Genes encoded in GEI1 are shown in the upper left panel, genes encoded in GEI2 are shown in the upper right panel, genes encoded in GEI3 are shown in the lower left panel, and genes encoded in GEI4 are shown in the lower right panel.

Fig 14A-14B shows introduction of an exogenous phage, plasmid or phagemid encoding CRISPR arrays (Type II system) to co-opt endogenous systems for programmed cell death in *Streptococcus thermophiles*.

Fig 15. The Type II guides of *Lactobacillus casei*. The top structure is the predicted guide. The figure on the bottom left is the correct dual guide crRNA:tracrRNA as confirmed by RNA Sequencing. The figure on the bottom right is an example of a predicted artificial single guide.

Fig 16 provides exemplary Type II guides of *Lactobacillus gasseri*. The top structure is the predicted guide. The figure on the bottom left is the correct dual guide crRNA:tracrRNA as confirmed by RNA Sequencing. The figure on the bottom right is an example of a predicted artificial single guide.

Fig 17 provides exemplary Type II guides of *Lactobacillus pentosus*. The top structure is the predicted guide. The figure on the bottom left is the correct dual guide crRNA:tracrRNA as confirmed by RNA Sequencing. The figure on the bottom right is an example of a predicted artificial single guide.

Fig 18 provides exemplary Type II guides of *Lactobacillus jensenii*. The left panel is the correct dual guide crRNA:tracrRNA as confirmed by RNA Sequencing. The right panel provides an example of a predicted artificial single guide.

Fig 19 shows the results of transformation of plasmids containing a protospacer that matches the most highly transcribe crRNA in the native *L. gasseri* Type II CRISPSR array. From left to right, four different plasmids were transformed into *L gasseri*: an empty pTRK563 vector, a construct with the correct protospacer but an incorrect PAM, the correct PAM but a protospacer that is not in the array, and the correct protospacer with the PAM that demonstrated the most interference targeting and cell death. The reported values represent the mean ± SEM of three independent replicates.

Fig 20 shows transformation of plasmids containing a protospacer that matches the most highly transcribe crRNA in the native *L. pentosus* Type II CRISPSR array. From left to right, four different plasmids were transformed into *L. pentosus*: a construct with the correct protospacer but an incorrect PAM (Lpe4 ctGttt), the correct PAM but a protospacer that is not in the array (Lpe8 noSPCR), an empty pTRK563 vector (pTRK563), and a plasmid with the correct protospacer and correct PAM (Lpe1 gttaat). The reported values represent the mean ± SEM of three independent replicates.

Fig 21 provides an exemplary Type I CRISPR-Cas guide of *Lactobacillus casei*. The sequence provided is the native Type I leader and repeat that is found in *Lactobacillus casei*

NCK 125. This artificial array contains a spacer that targets the 16s rDNA gene in the host genome.

Fig 22 shows transformation of plasmids containing a protospacer that matches the most highly transcribe crRNA in the native *L. jensenii* Type II CRISPSR array. From left to right, four different plasmids were transformed into *L jensenii*: an empty pTRK563 vector, a construct with the correct protospacer but an incorrect PAM, the correct PAM but a protospacer that is not in the array, and the correct protospacer with the PAM that demonstrated the most interference targeting and cell death.

Fig 23 shows targeted self-killing using the native Type I system in *Lactobacillus casei* NCK 125. Two targets were designed in the 16s rDNA gene. The PAM 5'-YAA-3' was predicted using the native spacer sequences in the organism. An artificial array containing the native Type I leader, repeats and the selected spacers was cloned into pTRK870. The constructs introduced included an empty vector (pTRK563) and two different artificial arrays: one containing a single spacer targeting the + strand in the 16s gene (1-2 alt) and the other array containing the original spacer targeting the + strand but containing an additional spacer targeting the – strand in the 16s gene (1, 2-3). The reported values represent the mean ± SEM of three independent replicates.

DETAILED DESCRIPTION

The present invention now will be described hereinafter with reference to the accompanying drawings and examples, in which embodiments of the invention are shown. This description is not intended to be a detailed catalog of all the different ways in which the invention may be implemented, or all the features that may be added to the instant invention. For example, features illustrated with respect to one embodiment may be incorporated into other embodiments, and features illustrated with respect to a particular embodiment may be deleted from that embodiment. Thus, the invention contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted. In addition, numerous variations and additions to the various embodiments suggested herein will be apparent to those skilled in the art in light of the instant disclosure, which do not depart from the instant invention. Hence, the following descriptions are intended to illustrate some particular embodiments of the invention, and not to exhaustively

specify all permutations, combinations and variations thereof.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The terminology used in the description of the invention herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention.

All publications, patent applications, patents and other references cited herein are incorporated by reference in their entireties for the teachings relevant to the sentence and/or paragraph in which the reference is presented.

Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination. Moreover, the present invention also contemplates that in some embodiments of the invention, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a composition comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination.

As used in the description of the invention and the appended claims, the singular forms "a," "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise.

Also as used herein, "and/or" refers to and encompasses any and all possible combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative ("or").

The term "about," as used herein when referring to a measurable value such as a dosage or time period and the like refers to variations of ± 20%, ± 10%, ± 5%, ± 1%, ± 0.5%, or even ± 0.1% of the specified amount.

As used herein, phrases such as "between X and Y" and "between about X and Y" should be interpreted to include X and Y. As used herein, phrases such as "between about X and Y" mean "between about X and about Y" and phrases such as "from about X to Y" mean "from about X to about Y."

The term "comprise," "comprises" and "comprising" as used herein, specify the presence of the stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

As used herein, the transitional phrase "consisting essentially of" means that the scope of a claim is to be interpreted to encompass the specified materials or steps recited in the claim and those that do not materially affect the basic and novel characteristic(s) of the claimed invention. Thus, the term "consisting essentially of" when used in a claim of this invention is not intended to be interpreted to be equivalent to "comprising."

"Cas9 nuclease" refers to a large group of endonucleases that catalyze the double stranded DNA cleavage in the CRISPR Cas system. These polypeptides are well known in the art and many of their structures (sequences) are characterized (See, e.g., WO2013/176772; WO/2013/188638). The domains for catalyzing the cleavage of the double stranded DNA are the RuvC domain and the HNH domain. The RuvC domain is responsible for nicking the (−) strand and the HNH domain is responsible for nicking the (+) strand (See, e.g., Gasiunas et al. *PNAS* 109(36):E2579-E2586 (September 4, 2012)).

As used herein, "chimeric" refers to a nucleic acid molecule or a polypeptide in which at least two components are derived from different sources (e.g., different organisms, different coding regions).

"Complement" as used herein can mean 100% complementarity or identity with the comparator nucleotide sequence or it can mean less than 100% complementarity (e.g., about 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%, and the like, complementarity).

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A." Complementarity between two single-stranded molecules may be "partial," in which only some of the nucleotides bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands.

As used herein, "contact," contacting," "contacted," and grammatical variations thereof, refers to placing the components of a desired reaction together under conditions suitable for carrying out the desired reaction (e.g., integration, transformation, screening, selecting, killing, identifying, amplifying, and the like). The methods and conditions for carrying out such reactions are well known in the art (See, e.g., Gasiunas et al. (2012) *Proc. Natl. Acad. Sci.* 109:E2579-E2586; M.R. Green and J. Sambrook (2012) Molecular Cloning:

A Laboratory Manual. 4th Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

A "deletion" as used herein can comprise the loss or deletion of genetic material including but not limited to a deletion of a portion of a chromosome or a plasmid, a deletion of a gene or a portion of a gene from a chromosome or a plasmid. In some embodiments, a deletion can comprise one gene or more than one gene. In some embodiments, a deletion may also comprise the loss of non-protein-coding regions that may encode small non-coding RNAs. In some embodiments, a deletion can comprise the loss of an entire plasmid or of an entire mobile genetic element. In some embodiments, the loss of a mobile genetic element may be defined as, for example, an inability to replicate or persist.

In some embodiments, a a phasmid of the invention may comprise a CRISPR array from a Type I CRISPR-Cas system, a Type II CRISPR-Cas system, a Type III CRISPR-Cas system, a Type IV CRISPR-Cas system, and/or a Type V CRISPR-Cas system (see, Makarova et al. *Nature Reviews Biotechnology* 13:722736 (2015)).

Thus, in some embodiments, in addition to a Type I crRNA, a phasmid of the invention may comprise Type I polypeptides and/or Type I Cascade polypeptides (i.e., a Type I CRISPR-Cas system).

As used herein, "Type I polypeptide" refers to any of a Cas3 polypeptide, Cas3' polypeptide, a Cas3" polypeptide, fusion variants thereof, and any one or more of the Type I Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated complex for antiviral defense ("Cascade") polypeptides. Thus, the term "Type I polypeptide" refers to the polypeptides that make up a Type I-A CRISPR-Cas system, a Type I-B CRISPR-Cas system, a Type I-C CRISPR-Cas system, a Type I-D CRISPR-Cas system, a Type I-E CRISPR-Cas system, a Type I-F CRISPR-Cas system, and/or a Type I-U CRISPR-Cas system. Each Type-I CRISPR-Cas system comprises at least one Cas3 polypeptide. Cas3 polypeptides generally comprise both a helicase domain and an HD domain. However, in some Type I CRISPR-Cas systems, the helicase and HD domain are found in separate polypeptides, Cas3' and Cas3". In particular, Cas3' encodes the helicase domain whereas Cas3" encodes the HD domain. Consequently, because both domains are required for Cas3 function, Type I subtypes either encode Cas3 (I-C, I-D, I-E, I-F, I-U) or Cas3' and Cas3" (I-A, I-B).

As used herein, "Type I Cascade polypeptides" refers to a complex of polypeptides involved in processing of pre-crRNAs and subsequent binding to the target DNA in Type I

CRISPR-Cas systems. These polypeptides include, but are not limited to, the Cascade polypeptides of Type I subtypes I-A, I-B, I-C, I-D, I-E and I-F. Non-limiting examples of Type I-A Cascade polypeptides include Cas7 (Csa2), Cas8a1 (Csx13), Cas8a2 (Csx9), Cas5, Csa5, Cas6a, Cas3' and/or a Cas3". Non-limiting examples of Type I-B Cascade polypeptides include Cas6b, Cas8b (Csh1), Cas7 (Csh2) and/or Cas5. Non-limiting examples of Type I-C Cascade polypeptides include Cas5d, Cas8c (Csd1), and/or Cas7 (Csd2). Non-limiting examples of Type I-D Cascade polypeptides include Cas10d (Csc3), Csc2, Csc1, and/or Cas6d. Non-limiting examples of Type I-E Cascade polypeptides include Cse1 (CasA), Cse2 (CasB), Cas7 (CasC), Cas5 (CasD) and/or Cas6e (CasE). Non-limiting examples of Type I-F Cascade polypeptides include Cys1, Cys2, Cas7 (Cys3) and/or Cas6f (Csy4). Non-limiting examples of Type I-U Cascade polypeptides include Cas8c, Cas7, Cas5, Cas6 and/or Cas4.

In some embodiments, a phasmid of the invention may comprise may comprise a Type II CRISPR-Cas system in addition to a Type II crRNA. Type II CRISPR-Cas systems comprise three subtypes: Type II-A, Type II-B and Type II-C, each of which comprise the multidomain protein, Cas9, in addition to the adaptation polypeptides, Cas1, Cas2 and optionally, Csn2 and/or Cas4. Most Type II loci also encode a tracrRNA. Organisms comprising exemplary Type II CRISPR-Cas systems include Legionella pneumophila str. Paris, Streptococcus thermophilus CNRZ1066 and Neisseria lactamica 020-06.

In additional embodiments, a phasmid of the invention may comprise may comprise a Type III CRISPR-Cas system in addition to a Type III crRNA. Similar to Type I CRISPR-Cas systems, in Type III systems processing and interference is mediated by multiprotein CRISPR RNA (crRNA)-effector complexes (Makarova et al. *Nature Reviews Biotechnology* 13:722736 (2015)) – "CASCADE" in Type I and "Csm" or "Cmr" in Type III. Thus, in some embodiments, a Type III CRISPR-Cas system can comprise a Csm complex (e.g., Type III-A Csm) and/or a Cmr complex (e.g., Type III-B Cmr), and optionally a Cas6 polypeptide. In representative embodiments, a Csm complex may comprise Cas10 (or Csm1), Csm2, Csm3, Csm4, Csm5, and Csm6 polypeptides and a Cmr complex may comprise Cmr1, Cas10 (or Cmr2), Cmr3, Cmr4, Cmr5, and Cmr6 polypeptides. In addition to the Csm complex or Cmr complex, a Type III CRISPR-Cas system may further comprise a Cas7 polypeptide. Four subtypes of a Type III CRISPR-Cas system have been characterized, III-A, III-B, III-C, III-D. In some embodiments, a Type III-A CRISPR-Cas system comprises Cas6, Cas10, Csm2, Cas7 (Csm3), Cas5 (Csm4), Cas7 (Csm5), and Csm6 polypeptides. In some embodiments, a Type III-B CRISPR-Cas system comprises Cas7 (Cmr1), Cas10, Cas5

(Cmr3), Cas7 (Cmr4), Cmr5, Cas6, and Cas7 (Cmr6) polypeptides. In some embodiments, a Type III-C CRISPR-Cas system comprises Cas7 (Cmr1), Cas7 (Cmr6), Cas10, Cas7 (Cmr4), Cmr5 and Cas5 (Cmr3), polypeptides. In some embodiments, a Type III-D CRISPR-Cas system comprises Cas10, Cas7 (Csm3), Cas5 (Cs×10), Csm2, Cas7 (Csm3), and all1473 polypeptides.

In some embodiments, a phasmid of the invention may comprise a Type IV CRISPR-Cas system, in addition to a Type IV crRNA. Type IV CRISPR-Cas systems can comprise a Csf4 polypeptide (dinG) and/or a Csf1, Cas7 (Csf2) and/or Cas5 (csf3) polypeptide. (Makarova et al. *Nature Reviews Microbiology* 13:722-736 (2015)).

In some embodiments, a phasmid of the invention further comprises a Type V CRISPR-Cas system, in addition to a Type V crRNA. Type V CRISPR-Cas systems can comprise a Cpf1 polypeptide and/or a Cas1, Cas2 and/or Cas4 polypeptide. (Makarova et al. *Nature Reviews Microbiology* 13:722-736 (2015)).

A "fragment" or "portion" of a nucleotide sequence of the invention will be understood to mean a nucleotide sequence of reduced length relative (e.g., reduced by 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides) to a reference nucleic acid or nucleotide sequence and comprising, consisting essentially of and/or consisting of a nucleotide sequence of contiguous nucleotides identical or substantially identical (e.g., 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% identical) to the reference nucleic acid or nucleotide sequence. Such a nucleic acid fragment or portion according to the invention may be, where appropriate, included in a larger polynucleotide of which it is a constituent. Thus, hybridizing to (or hybridizes to, and other grammatical variations thereof), for example, at least a portion of a target DNA (e.g., target region in the genome), refers to hybridization to a nucleotide sequence that is identical or substantially identical to a length of contiguous nucleotides of the target DNA. In some embodiments, a repeat of a repeat spacer sequence or a repeat-spacer-repeat sequence can comprise a fragment of a repeat sequence of a wild-type CRISPR locus or a repeat sequence of a synthetic CRISPR array, wherein the fragment of the repeat retains the function of a repeat in a CRISPR array of hybridizing with the tracr nucleic acid.

In some embodiments, the invention may comprise a functional fragment of a Cas9 , Cas3, Cas3', Cas3", or Cpf1 nuclease. A Cas9 functional fragment retains one or more of the activities of a native Cas9 nuclease including, but not limited to, HNH nuclease activity, RuvC nuclease activity, DNA, RNA and/or PAM recognition and binding activities. A

functional fragment of a Cas9 nuclease may be encoded by a fragment of a Cas9 polynucleotide. A Cas3, Cas3' or Cas3" functional fragment retains one or more of the activities of a native Cas9 nuclease including, but not limited to, nickase activity, exonuclease activity, DNA-binding, and/or RNA binding. A functional fragment of a Cas3, Cas3' or Cas3" nuclease may be encoded by a fragment of a Cas3, Cas3' or Cas3" polynucleotide, respectively.

As used herein, the term "gene" refers to a nucleic acid molecule capable of being used to produce mRNA, antisense RNA, RNAi (miRNA, siRNA, shRNA), anti-microRNA antisense oligodeoxyribonucleotide (AMO), and the like. Genes may or may not be capable of being used to produce a functional protein or gene product. Genes can include both coding and non-coding regions (e.g., introns, regulatory elements, promoters, enhancers, termination sequences and/or 5' and 3' untranslated regions). A gene may be "isolated" by which is meant a nucleic acid that is substantially or essentially free from components normally found in association with the nucleic acid in its natural state. Such components include other cellular material, culture medium from recombinant production, and/or various chemicals used in chemically synthesizing the nucleic acid.

The term "genome" as used herein includes an organism's chromosomal/nuclear genome as well as any mitochondrial, and/or plasmid genome.

A "hairpin sequence" as used herein, is a nucleotide sequence comprising hairpins (e.g., that forms one or more hairpin structures). A hairpin (e.g., stem-loop, fold-back) refers to a nucleic acid molecule having a secondary structure that includes a region of complementary nucleotides that form a double strand that are further flanked on either side by single stranded-regions. Such structures are well known in the art. As known in the art, the double stranded region can comprise some mismatches in base pairing or can be perfectly complementary. In some embodiments of the present disclosure, a hairpin sequence of a nucleic acid construct can be located at the 3'end of a tracr nucleic acid.

A "heterologous" or a "recombinant" nucleotide sequence is a nucleotide sequence not naturally associated with a host cell into which it is introduced, including non- naturally occurring multiple copies of a naturally occurring nucleotide sequence.

Different nucleic acids or proteins having homology are referred to herein as "homologues." The term homologue includes homologous sequences from the same and other species and orthologous sequences from the same and other species. "Homology" refers to the level of similarity between two or more nucleic acid and/or amino acid sequences in terms of percent of positional identity (*i.e.*, sequence similarity or identity).

Homology also refers to the concept of similar functional properties among different nucleic acids or proteins. Thus, the compositions and methods of the invention further comprise homologues to the nucleotide sequences and polypeptide sequences of this invention. "Orthologous," as used herein, refers to homologous nucleotide sequences and/ or amino acid sequences in different species that arose from a common ancestral gene during speciation. A homologue of a nucleotide sequence of this invention has a substantial sequence identity (e.g., at least about 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%, and/or 100%) to said nucleotide sequence of the invention. Thus, for example, a homologue of a Type I, Type II, Type III, Type IV, or Type V polynucleotide or polypeptide can be about 70% homologous or more to any one of any known or later identified Type I, Type II, Type III, Type IV, or Type V polynucleotide or polypeptide.

As used herein, hybridization, hybridize, hybridizing, and grammatical variations thereof, refer to the binding of two fully complementary nucleotide sequences or substantially complementary sequences in which some mismatched base pairs may be present. The conditions for hybridization are well known in the art and vary based on the length of the nucleotide sequences and the degree of complementarity between the nucleotide sequences. In some embodiments, the conditions of hybridization can be high stringency, or they can be medium stringency or low stringency depending on the amount of complementarity and the length of the sequences to be hybridized. The conditions that constitute low, medium and high stringency for purposes of hybridization between nucleotide sequences are well known in the art (See, e.g., Gasiunas et al. (2012) *Proc. Natl. Acad. Sci.* 109:E2579-E2586; M.R. Green and J. Sambrook (2012) Molecular Cloning: A Laboratory Manual. 4th Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

As used herein, the terms "increase," "increasing," "increased," "enhance," "enhanced," "enhancing," and "enhancement" (and grammatical variations thereof) describe an elevation of at least about 25%, 50%, 75%, 100%, 150%, 200%, 300%, 400%, 500% or more as compared to a control.

A "native" or "wild type" nucleic acid, nucleotide sequence, polypeptide or amino acid sequence refers to a naturally occurring or endogenous nucleic acid, nucleotide sequence, polypeptide or amino acid sequence. Thus, for example, a "wild type mRNA" is a mRNA that is naturally occurring in or endogenous to the organism. A "homologous" nucleic acid sequence is a nucleotide sequence naturally associated with a host cell into which it is introduced.

Also as used herein, the terms "nucleic acid," "nucleic acid molecule," "nucleic acid construct," "nucleotide sequence" and "polynucleotide" refer to RNA or DNA that is linear or branched, single or double stranded, or a hybrid thereof. The term also encompasses RNA/DNA hybrids. When dsRNA is produced synthetically, less common bases, such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others can also be used for antisense, dsRNA, and ribozyme pairing. For example, polynucleotides that contain C-5 propyne analogues of uridine and cytidine have been shown to bind RNA with high affinity and to be potent antisense inhibitors of gene expression. Other modifications, such as modification to the phosphodiester backbone, or the 2'-hydroxy in the ribose sugar group of the RNA can also be made. The nucleic acid constructs of the present disclosure can be DNA or RNA, but are preferably DNA. Thus, although the nucleic acid constructs of this invention may be described and used in the form of DNA, depending on the intended use, they may also be described and used in the form of RNA.

As used herein, the term "nucleotide sequence" refers to a heteropolymer of nucleotides or the sequence of these nucleotides from the 5' to 3' end of a nucleic acid molecule and includes DNA or RNA molecules, including cDNA, a DNA fragment or portion, genomic DNA, synthetic (*e.g.*, chemically synthesized) DNA, plasmid DNA, mRNA, and anti-sense RNA, any of which can be single stranded or double stranded. The terms "nucleotide sequence" "nucleic acid," "nucleic acid molecule," "oligonucleotide" and "polynucleotide" are also used interchangeably herein to refer to a heteropolymer of nucleotides. Except as otherwise indicated, nucleic acid molecules and/or nucleotide sequences provided herein are presented herein in the 5' to 3' direction, from left to right and are represented using the standard code for representing the nucleotide characters as set forth in the U.S. sequence rules, 37 CFR §§1.821 - 1.825 and the World Intellectual Property Organization (WIPO) Standard ST.25.

As used herein, the term "percent sequence identity" or "percent identity" refers to the percentage of identical nucleotides in a linear polynucleotide sequence of a reference ("query") polynucleotide molecule (or its complementary strand) as compared to a test ("subject") polynucleotide molecule (or its complementary strand) when the two sequences are optimally aligned. In some embodiments, "percent identity" can refer to the percentage of identical amino acids in an amino acid sequence.

A "protospacer sequence" refers to the target double stranded DNA and specifically to the portion of the target DNA (e.g., or target region in the genome) that is fully or substantially complementary (and hybridizes) to the spacer sequence of the CRISPR repeat-

spacer sequences, CRISPR repeat-spacer-repeat sequences, and/or CRISPR arrays.

As used herein, the terms "reduce," "reduced," "reducing," "reduction," "diminish," "suppress," and "decrease" (and grammatical variations thereof), describe, for example, a decrease of at least about 5%, 10%, 15%, 20%, 25%, 35%, 50%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% as compared to a control. In particular embodiments, the reduction can result in no or essentially no (*i.e.*, an insignificant amount, *e.g.*, less than about 10% or even 5%) detectable activity or amount of the component being measured (e.g., the population of cells or a genome size). Thus, for example, a reduced genome size can mean a reduction in the size of a genome of at least about 5%, 10%, 15%, 20%, 25%, 35%, 50%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% as compared to a control.

A control as used herein may be, for example, a population of bacterial, archaeal, algal or yeast cells that has not been transformed with a heterologous nucleic acid construct of this invention. In some embodiments, a control may be a wild-type population of bacterial, archaeal, algal or yeast cells, or it may be a population of bacterial, archaeal or yeast cells transformed with a heterologous construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is not complementary to a target region in the genome of the bacterial, archaeal or yeast cells of said population (i.e., non-self targeting/"scrambled spacer"). In additional aspects, a control may be, for example, a wild-type population of bacterial, archaeal, algal or yeast cells, or a population of bacterial, archaeal, algal or yeast cells transformed with a heterologous construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population that is not located adjacent to a protospacer adjacent motif (PAM).

A "repeat sequence" as used herein refers, for example, to any repeat sequence of a wild-type CRISPR locus or a repeat sequence of a synthetic CRISPR array that are separated by "spacer sequences" (e.g., a repeat-spacer sequence or a repeat-spacer-repeat sequence of the invention). A repeat sequence useful with this invention can be any known or later identified repeat sequence of a CRISPR locus. Accordingly, in some embodiments, a repeat-spacer sequence or a repeat-spacer-repeat comprises a repeat that is substantially identical (e.g. at least about 70% identical (e.g., at least about 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 more)) to a repeat from a wild-type Type II

CRISPR array. In some embodiments, a repeat sequence is 100% identical to a repeat from a wild type Type I CRISPR array, a wild type Type II CRISPR array, wild type Type III CRISPR array, wild type Type IV CRISPR array, or wild type Type V CRISPR array. In additional embodiments, a repeat sequence useful with this invention can comprise a nucleotide sequence comprising a partial repeat that is a fragment or portion of a consecutive nucleotides of a repeat sequence of a CRISPR locus or synthetic CRISPR array of any of a Type I crRNA, Type II crRNA, Type III crRNA, Type IV crRNA, or Type V crRNA.

As used herein, "CRISPR array"of a Type I, Type II, Type III, Type IV, or Type V CRISPR-Cas system refers to a nucleic acid construct that comprises from 5' to 3' a repeat-spacer-repeat sequence or comprises from 5' to 3' at least one repeat-spacer sequence (e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 repeat-spacer sequences, and any range or value therein). When more than one repeat-spacer is comprised in a CRISPR array, the spacer of the prior (5' to 3') repeat-spacer sequence can be linked to the repeat of the following repeat-spacer (e.g., the spacer of a first repeat-spacer sequence is linked to the repeat of a second repeat-spacer sequence). In some embodiments, a CRISPR array can comprise two repeats (or two partial repeats) separated by a spacer (e.g., a repeat-spacer-repeat sequence).

As used herein "sequence identity" refers to the extent to which two optimally aligned polynucleotide or peptide sequences are invariant throughout a window of alignment of components, *e.g.*, nucleotides or amino acids. "Identity" can be readily calculated by known methods including, but not limited to, those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heinje, G., ed.) Academic Press (1987); and *Sequence Analysis Primer* (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991).

A "spacer sequence" as used herein is a nucleotide sequence that is complementary to a target DNA (i.e., target region in the genome or the "protospacer sequence", which is adjacent to a protospacer adjacent motif (PAM) sequence). The spacer sequence can be fully complementary or substantially complementary (e.g., at least about 70% complementary (e.g., about 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 more)) to a target DNA. In representative embodiments, the spacer sequence has 100%

complementarity to the target DNA. In additional embodiments, the complementarity of the 3' region of the spacer sequence to the target DNA is 100% but is less than 100% in the 5' region of the spacer and therefore the overall complementarity of the spacer sequence to the target DNA is less than 100%. Thus, for example, the first 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, and the like, nucleotides in the 3' region of a 20 nucleotide spacer sequence (seed sequence) can be 100% complementary to the target DNA, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target DNA. In some embodiments, the first 7 to 12 nucleotides of the spacer sequence can be 100% complementary to the target DNA, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target DNA. In other embodiments, the first 7 to 10 nucleotides of the spacer sequence can be 100% complementary to the target DNA, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target DNA. In representative embodiments, the first 7 nucleotides (within the seed) of the spacer sequence can be 100% complementary to the target DNA, while the remaining nucleotides in the 5' region of the spacer sequence are substantially complementary (e.g., at least about 70% complementary) to the target DNA.

As used herein, a "target DNA," "target region" or a "target region in the genome" refers to a region of an organism's genome that is fully complementary or substantially complementary (e.g., at least 70% complementary (e.g., 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 more)) to a spacer sequence in a repeat-spacer sequence or repeat-spacer-repeat sequence. In some embodiments, a target region may be about 10 to about 40 consecutive nucleotides in length located immediately adjacent to a PAM sequence (PAM sequence located immediately 3' of the target region) in the genome of the organism (e.g., Type I CRISPR-Cas systems and Type II CRISPR-Cas systems). In the some embodiments, e.g., Type I systems, the PAM is on the alternate side of the protospacer (the 5' end). There is no known PAM for Type III systems. Makarova et al. describes the nomenclature for all the classes, types and subtypes of CRISPR systems (*Nature Reviews Microbiology* **13**:722–736 (2015)). Guide structures and PAMs are described in by R. Barrangou (*Genome Biol.* 16:247 (2015)).

In some embodiments, a target region useful with this invention is located within an essential gene or a non-essential gene.

In representative embodiments, a target region can be randomly selected or can be specifically selected. In some embodiments, a randomly selected target region may be selected from any at least 10 consecutive nucleotides (e.g., 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, and the like, and any range or value therein) located immediately adjacent to a PAM sequence in a bacterial, archaeal, algal or yeast genome. In some embodiments, the target region can be about 10 to about 20 consecutive nucleotides, about 10 to about 30 consecutive nucleotides, and/or about 10 to about 40 consecutive nucleotides and the like, or any range or value therein, located immediately adjacent to a protospacer adjacent motif (PAM) sequence in a bacterial, archaeal, algal or yeast genome. In some embodiments, specifically selecting a target region can comprise selecting two or more target regions that are located about every 100 nucleotides to about every 1000 nucleotides, about every 100 nucleotides to about every 2000, about every 100 nucleotides to about every 3000, about every 100 nucleotides to about every 4000, and/or about every 100 nucleotides to about every 5000 nucleotides, and the like, from one another in the genome of the one or more bacteria, archaea, algal or yeast cells. In particular embodiments, specifically selecting a target region comprises specifically selecting a target region from a gene, open reading frame, a putative open reading frame or an intergenic region comprising at least about 10 to about 40 consecutive nucleotides immediately adjacent to a PAM sequence in a bacterial, archaeal, algal or yeast genome.

A "trans-activating CRISPR (tracr) nucleic acid" or "tracr nucleic acid" as used herein refers to any tracr RNA (or its encoding DNA). A tracr nucleic acid comprises from 5' to 3' a lower stem, an upper stem, a bulge, a nexus hairpin and terminal hairpins (*See,* Briner et al. (2014) *Molecular Cell.* 56(2):333-339). A trans-activating CRISPR (tracr) nucleic acid functions in hybridizing to the repeat portion of mature or immature crRNAs, recruits Cas9 protein to the target site, and may facilitate the catalytic activity of Cas9 by inducting structural rearrangement. The functional composition of tracrRNA molecules is listed above. Sequences for tracrRNAs are specific to the CRISPR-Cas system and can be variable. Any tracr nucleic acid, known or later identified, can be used with this invention.

As used herein, the phrase "substantially identical," or "substantial identity" in the context of two nucleic acid molecules, nucleotide sequences or protein sequences, refers to two or more sequences or subsequences that have at least about 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%, and/or 100% nucleotide or amino acid

residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. In some embodiments of the invention, the substantial identity exists over a region of the sequences that is at least about 50 residues to about 150 residues in length. Thus, in some embodiments of the invention, the substantial identity exists over a region of the sequences that is at least about 3 to about 15 (e.g., 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 residues in length and the like or any value or any range therein), at least about 5 to about 30 , at least about 10 to about 30, at least about 16 to about 30, at least about 18 to at least about 25, at least about 18, at least about 22, at least about 25, at least about 30, at least about 40, at least about 50, about 60, about 70, about 80, about 90, about 100, about 110, about 120, about 130, about 140, about 150, or more residues in length, and any range therein. In representative embodiments, the sequences can be substantially identical over at least about 22 nucleotides. In some particular embodiments, the sequences are substantially identical over at least about 150 residues. In some embodiments, sequences of the invention can be about 70% to about 100% identical over at least about 16 nucleotides to about 25 nucleotides. In some embodiments, sequences of the invention can be about 75% to about 100% identical over at least about 16 nucleotides to about 25 nucleotides. In further embodiments, sequences of the invention can be about 80% to about 100% identical over at least about 16 nucleotides to about 25 nucleotides. In further embodiments, sequences of the invention can be about 80% to about 100% identical over at least about 7 nucleotides to about 25 nucleotides. In some embodiments, sequences of the invention can be about 70% identical over at least about 18 nucleotides. In other embodiments, the sequences can be about 85% identical over about 22 nucleotides. In still other embodiments, the sequences can be 100% identical over about 16 nucleotides. In a further embodiment, the sequences are substantially identical over the entire length of a coding region. Furthermore, in exemplary embodiments, substantially identical nucleotide or polypeptide sequences perform substantially the same function (e.g., the function or activity of a crRNA, tracr nucleic acid, repeat sequence, Cas9 nuclease (nickase, DNA, RNA and/or PAM recognition and binding), Cas3, Cas3', Cas3" or any other CRISPR-Cas polynucleotide or polypeptide).

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence

comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for aligning a comparison window are well known to those skilled in the art and may be conducted by tools such as the local homology algorithm of Smith and Waterman, the homology alignment algorithm of Needleman and Wunsch, the search for similarity method of Pearson and Lipman, and optionally by computerized implementations of these algorithms such as GAP, BESTFIT, FASTA, and TFASTA available as part of the GCG® Wisconsin Package® (Accelrys Inc., San Diego, CA). An "identity fraction" for aligned segments of a test sequence and a reference sequence is the number of identical components which are shared by the two aligned sequences divided by the total number of components in the reference sequence segment, *i.e.*, the entire reference sequence or a smaller defined part of the reference sequence. Percent sequence identity is represented as the identity fraction multiplied by 100. The comparison of one or more polynucleotide sequences may be to a full-length polynucleotide sequence or a portion thereof, or to a longer polynucleotide sequence. For purposes of this invention "percent identity" may also be determined using BLASTX version 2.0 for translated nucleotide sequences and BLASTN version 2.0 for polynucleotide sequences.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an

expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix *(see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences *(see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleotide sequence to the reference nucleotide sequence is less than about 0.1 to less than about 0.001. Thus, in some embodiments of the invention, the smallest sum probability in a comparison of the test nucleotide sequence to the reference nucleotide sequence is less than about 0.001.

Two nucleotide sequences can also be considered to be substantially complementary when the two sequences hybridize to each other under stringent conditions. In some representative embodiments, two nucleotide sequences considered to be substantially complementary hybridize to each other under highly stringent conditions.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes* part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York (1993). Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleotide sequences which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of

highly stringent wash conditions is 0.1 5M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example of a medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example of a low stringency wash for a duplex of, *e.g.,* more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes (*e.g.,* about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleotide sequences that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This can occur, for example, when a copy of a nucleotide sequence is created using the maximum codon degeneracy permitted by the genetic code.

The following are examples of sets of hybridization/wash conditions that may be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the invention. In one embodiment, a reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C. In another embodiment, the reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C or in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C. In still further embodiments, the reference nucleotide sequence hybridizes to the "test" nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, or in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

Any nucleotide sequence and/or heterologous nucleic acid construct of this invention can be codon optimized for expression in any species of interest. Codon optimization is well known in the art and involves modification of a nucleotide sequence for codon usage bias using species specific codon usage tables. The codon usage tables are generated based on a sequence analysis of the most highly expressed genes for the species of interest. When the

nucleotide sequences are to be expressed in the nucleus, the codon usage tables are generated based on a sequence analysis of highly expressed nuclear genes for the species of interest. The modifications of the nucleotide sequences are determined by comparing the species specific codon usage table with the codons present in the native polynucleotide sequences. As is understood in the art, codon optimization of a nucleotide sequence results in a nucleotide sequence having less than 100% identity (e.g., 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%, and the like) to the native nucleotide sequence but which still encodes a polypeptide having the same function as that encoded by the original, native nucleotide sequence. Thus, in representative embodiments of the invention, the nucleotide sequence and/or heterologous nucleic acid construct of this invention can be codon optimized for expression in the particular species of interest.

In some embodiments, the heterologous or recombinant nucleic acids molecules, nucleotide sequences and/or polypeptides of the invention are "isolated." An "isolated" nucleic acid molecule, an "isolated" nucleotide sequence or an "isolated" polypeptide is a nucleic acid molecule, nucleotide sequence or polypeptide that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule, nucleotide sequence or polypeptide may exist in a purified form that is at least partially separated from at least some of the other components of the naturally occurring organism or virus, for example, the cell or viral structural components or other polypeptides or nucleic acids commonly found associated with the polynucleotide. In representative embodiments, the isolated nucleic acid molecule, the isolated nucleotide sequence and/or the isolated polypeptide is at least about 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more pure.

In other embodiments, an isolated nucleic acid molecule, nucleotide sequence or polypeptide may exist in a non-native environment such as, for example, a recombinant host cell. Thus, for example, with respect to nucleotide sequences, the term "isolated" means that it is separated from the chromosome and/or cell in which it naturally occurs. A polynucleotide is also isolated if it is separated from the chromosome and/or cell in which it naturally occurs in and is then inserted into a genetic context, a chromosome and/or a cell in which it does not naturally occur (e.g., a different host cell, different regulatory sequences, and/or different position in the genome than as found in nature). Accordingly, the heterologous nucleic acid constructs, nucleotide sequences and their encoded polypeptides are "isolated" in that, by the hand of man, they exist apart from their native environment and

therefore are not products of nature, however, in some embodiments, they can be introduced into and exist in a recombinant host cell.

In some embodiments, the heterologous or recombinant nucleic acid constructs of the invention are "synthetic." A "synthetic" nucleic acid molecule, a "synthetic" nucleotide sequence or a "synthetic" polypeptide is a nucleic acid molecule, nucleotide sequence or polypeptide that is not found in nature but is created by the hand of man and is therefore not a product of nature.

In any of the embodiments described herein, the nucleotide sequences and/or heterologous nucleic acid constructs of the invention can be operatively associated with a variety of promoters and other regulatory elements for expression in various organisms cells. Thus, in representative embodiments, a nucleic acid construct of this invention can further comprise one or more promoters operably linked to one or more nucleotide sequences.

By "operably linked" or "operably associated" as used herein, it is meant that the indicated elements are functionally related to each other, and are also generally physically related. Thus, the term "operably linked" or "operably associated" as used herein, refers to nucleotide sequences on a single nucleic acid molecule that are functionally associated. Thus, a first nucleotide sequence that is operably linked to a second nucleotide sequence, means a situation when the first nucleotide sequence is placed in a functional relationship with the second nucleotide sequence. For instance, a promoter is operably associated with a nucleotide sequence if the promoter effects the transcription or expression of said nucleotide sequence. Those skilled in the art will appreciate that the control sequences (*e.g.*, promoter) need not be contiguous with the nucleotide sequence to which it is operably associated, as long as the control sequences function to direct the expression thereof. Thus, for example, intervening untranslated, yet transcribed, sequences can be present between a promoter and a nucleotide sequence, and the promoter can still be considered "operably linked" to the nucleotide sequence.

A "promoter" is a nucleotide sequence that controls or regulates the transcription of a nucleotide sequence (*i.e.*, a coding sequence) that is operably associated with the promoter. The coding sequence may encode a polypeptide and/or a functional RNA. Typically, a "promoter" refers to a nucleotide sequence that contains a binding site for RNA polymerase II and directs the initiation of transcription. In general, promoters are found 5', or upstream, relative to the start of the coding region of the corresponding coding sequence. The promoter region may comprise other elements that act as regulators of gene expression. These include a TATA box consensus sequence, and often a CAAT box consensus sequence (Breathnach and

Chambon, (1981) *Annu. Rev. Biochem.* 50:349). In plants, the CAAT box may be substituted by the AGGA box (Messing *et al.*, (1983) in Genetic Engineering of Plants, T. Kosuge, C. Meredith and A. Hollaender (eds.), Plenum Press, pp. 211-227).

Promoters can include, for example, constitutive, inducible, temporally regulated, developmentally regulated, chemically regulated, tissue-preferred and/or tissue-specific promoters for use in the preparation of heterologous nucleic acid constructs, i.e., "chimeric genes" or "chimeric polynucleotides." These various types of promoters are known in the art.

The choice of promoter will vary depending on the temporal and spatial requirements for expression, and also depending on the host cell to be transformed. Promoters for many different organisms are well known in the art. Based on the extensive knowledge present in the art, the appropriate promoter can be selected for the particular host organism of interest. Thus, for example, much is known about promoters upstream of highly constitutively expressed genes in model organisms and such knowledge can be readily accessed and implemented in other systems as appropriate.

In some embodiments, a nucleic acid construct of the invention can be an "expression cassette" or can be comprised within an expression cassette. As used herein, "expression cassette" means a heterologous nucleic acid construct comprising a nucleotide sequence of interest (e.g., the nucleic acid constructs of the invention (e.g., a synthetic tracr nucleic acid construct, a synthetic CRISPR nucleic acid construct, a synthetic CRISPR array, a chimeric nucleic acid construct; a nucleotide sequence encoding a polypeptide of interest, a Type I polypeptide, Type II polypeptide, Type III polypeptide, Type IV polypeptide, and/orType V polypeptide)), wherein said nucleotide sequence is operably associated with at least a control sequence (e.g., a promoter). Thus, some aspects of the invention provide expression cassettes designed to express the nucleotides sequences of the invention.

An expression cassette comprising a nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. An expression cassette may also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression.

An expression cassette also can optionally include a transcriptional and/or translational termination region (*i.e.*, termination region) that is functional in the selected host cell. A variety of transcriptional terminators are available for use in expression cassettes and are responsible for the termination of transcription beyond the heterologous nucleotide sequence of interest and correct mRNA polyadenylation. The termination region may be native to the transcriptional initiation region, may be native to the operably linked nucleotide

sequence of interest, may be native to the host cell, or may be derived from another source (*i.e.*, foreign or heterologous to the promoter, to the nucleotide sequence of interest, to the host, or any combination thereof).

An expression cassette also can include a nucleotide sequence for a selectable marker, which can be used to select a transformed host cell. As used herein, "selectable marker" means a nucleotide sequence that when expressed imparts a distinct phenotype to the host cell expressing the marker and thus allows such transformed cells to be distinguished from those that do not have the marker. Such a nucleotide sequence may encode either a selectable or screenable marker, depending on whether the marker confers a trait that can be selected for by chemical means, such as by using a selective agent (*e.g.*, an antibiotic and the like), or on whether the marker is simply a trait that one can identify through observation or testing, such as by screening (*e.g.*, fluorescence). Of course, many examples of suitable selectable markers are known in the art and can be used in the expression cassettes described herein.

In addition to expression cassettes, the nucleic acid molecules and nucleotide sequences described herein can be used in connection with vectors. The term "vector" refers to a composition for transferring, delivering or introducing a nucleic acid (or nucleic acids) into a cell. A vector comprises a nucleic acid molecule comprising the nucleotide sequence(s) to be transferred, delivered or introduced. Vectors for use in transformation of host organisms are well known in the art. Non-limiting examples of general classes of vectors include but are not limited to a viral vector, a plasmid vector, a phage vector, a phagemid vector, a cosmid vector, a fosmid vector, a bacteriophage, an artificial chromosome, or an Agrobacterium binary vector in double or single stranded linear or circular form which may or may not be self transmissible or mobilizable. A vector as defined herein can transform prokaryotic or eukaryotic host either by integration into the cellular genome or exist extrachromosomally (e.g. autonomous replicating plasmid with an origin of replication). Additionally included are shuttle vectors by which is meant a DNA vehicle capable, naturally or by design, of replication in two different host organisms, which may be selected from actinomycetes and related species, bacteria and eukaryotic (e.g. higher plant, mammalian, yeast or fungal cells). In some representative embodiments, the nucleic acid in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell. The vector may be a bi-functional expression vector which functions in multiple hosts. In the case of genomic DNA, this may contain its own promoter or other regulatory elements and in the case of cDNA this may be under the control of an appropriate promoter or other regulatory elements for expression in

the host cell. Accordingly, the nucleic acid molecules of this invention and/or expression cassettes can be comprised in vectors as described herein and as known in the art.

"Introducing," "introduce," "introduced" (and grammatical variations thereof) in the context of a polynucleotide of interest means presenting the polynucleotide of interest to the host organism or cell of said organism (e.g., host cell) in such a manner that the polynucleotide gains access to the interior of a cell. Where more than one polynucleotide is to be introduced these polynucleotides can be assembled as part of a single polynucleotide or nucleic acid construct, or as separate polynucleotide or nucleic acid constructs, and can be located on the same or different expression constructs or transformation vectors. Accordingly, these polynucleotides can be introduced into cells in a single transformation event, in separate transformation/transfection events, or, for example, they can be incorporated into an organism by conventional breeding protocols. Thus, in some aspects, one or more nucleic acid constructs of this invention can be introduced singly or in combination into a host organism or a cell of said host organism. In the context of a population of cells, "introducing" means contacting the population with the heterologous nucleic acid constructs of the invention under conditions where the heterologous nucleic acid constructs of the invention gain access to the interior of one or more cells of the population, thereby transforming the one or more cells of the population.

The term "transformation" or "transfection" as used herein refers to the introduction of a heterologous nucleic acid into a cell. Transformation of a cell may be stable or transient. Thus, in some embodiments, a host cell or host organism is stably transformed with a nucleic acid construct of the invention. In other embodiments, a host cell or host organism is transiently transformed with a nucleic acid construct of the invention. Thus, in representative embodiments, a heterologous nucleic acid construct of the invention can be stably and/or transiently introduced into a cell.

"Transient transformation" in the context of a polynucleotide means that a polynucleotide is introduced into the cell and does not integrate into the genome of the cell.

By "stably introducing" or "stably introduced," in the context of a polynucleotide, means that the introduced polynucleotide is stably incorporated into the genome of the cell, and thus the cell is stably transformed with the polynucleotide.

"Stable transformation" or "stably transformed" as used herein means that a nucleic acid construct is introduced into a cell and integrates into the genome of the cell. As such, the integrated nucleic acid construct is capable of being inherited by the progeny thereof, more particularly, by the progeny of multiple successive generations. "Genome" as used

herein includes the nuclear, mitochondrial and plasmid genome, and therefore may include integration of a nucleic acid construct into, for example, the plasmid or mitochondrial genome. Stable transformation as used herein may also refer to a transgene that is maintained extrachromasomally, for example, as a minichromosome or a plasmid.

Transient transformation may be detected by, for example, an enzyme-linked immunosorbent assay (ELISA) or Western blot, which can detect the presence of a peptide or polypeptide encoded by one or more transgene introduced into an organism. Stable transformation of a cell can be detected by, for example, a Southern blot hybridization assay of genomic DNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into an organism (e.g., a bacterium, an archaea, a yeast, an algae, and the like). Stable transformation of a cell can be detected by, for example, a Northern blot hybridization assay of RNA of the cell with nucleic acid sequences which specifically hybridize with a nucleotide sequence of a transgene introduced into a plant or other organism. Stable transformation of a cell can also be detected by, *e.g.*, a polymerase chain reaction (PCR) or other amplification reactions as are well known in the art, employing specific primer sequences that hybridize with target sequence(s) of a transgene, resulting in amplification of the transgene sequence, which can be detected according to standard methods. Transformation can also be detected by direct sequencing and/or hybridization protocols well known in the art.

Accordingly, in some embodiments, the nucleotide sequences, constructs, expression cassettes can be expressed transiently and/or they can be stably incorporated into the genome of the host organism.

A heterologous nucleic acid construct of the invention can be introduced into a cell by any method known to those of skill in the art. In some embodiments of the invention, transformation of a cell comprises nuclear transformation. In still further embodiments, the heterologous nucleic acid construct (s) of the invention can be introduced into a cell via conventional breeding techniques.

Procedures for transforming both eukaryotic and prokaryotic organisms are well known and routine in the art and are described throughout the literature (*See*, for example, Jiang et al. 2013. *Nat. Biotechnol.* 31:233-239; Ran et al. *Nature Protocols* 8:2281–2308 (2013))

A nucleotide sequence therefore can be introduced into a host organism or its cell in any number of ways that are well known in the art. The methods of the invention do not depend on a particular method for introducing one or more nucleotide sequences into the

organism, only that they gain access to the interior of at least one cell of the organism. Where more than one nucleotide sequence is to be introduced, they can be assembled as part of a single nucleic acid construct, or as separate nucleic acid constructs, and can be located on the same or different nucleic acid constructs. Accordingly, the nucleotide sequences can be introduced into the cell of interest in a single transformation event, or in separate transformation events, or, alternatively, where relevant, a nucleotide sequence can be incorporated into an organism as part of a breeding protocol.

Mobile genetic elements (MGEs) present bacteria with continuous challenges to genomic stability, promoting evolution through horizontal gene transfer. The term MGE encompasses plasmids, bacteriophages, transposable elements, genomic islands, and many other specialized genetic elements (1). MGEs encompass genes conferring high rates of dissemination, adaptive advantages to the host, and genomic stability, leading to their near universal presence in bacterial genomes. To cope with the permanent threat of predatory bacteriophages and selfish genetic elements, bacteria have evolved both innate and adaptive immune systems targeting exogenous genetic elements. Innate immunity includes cell-wall modification, restriction/modification systems, and abortive phage infection (2). Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated genes (Cas) are an adaptive immune system targeted against invasive genetic elements in bacteria (3). CRISPR-Cas mediated immunity relies on distinct molecular processes, categorized as *acquisition*, *expression*, and *interference* (3). *Acquisition* occurs via molecular 'sampling' of foreign genetic elements, from which short sequences, termed spacers, are integrated in a polarized fashion into the CRISPR array (4). *Expression* of CRISPR arrays is constitutive and inducible by promoter elements within the preceding leader sequence (5-6). *Interference* results from a corresponding transcript that is processed selectively at each repeat sequence, forming CRISPR RNAs (crRNAs) that guide Cas proteins for sequence-specific recognition and cleavage of target DNA complementary to the spacer (7). CRISPR-Cas technology has applications in strain typing and detection (8-10), exploitation of natural/engineered immunity against mobile genetic elements (11), programmable genome editing in diverse backgrounds (12), transcriptional control (13-14), and manipulation of microbial populations in defined consortia (15).

The various CRISPR systems are known in the art. For example, see Makarova et al., which describes the nomenclature for all the classes, types and subtypes of CRISPR systems (*Nature Reviews Microbiology* **13**:722–736 (2015)); see also, R. Barrangou (*Genome Biol.* 16:247 (2015)).

Although sequence features corresponding to CRISPR arrays were described previously in multiple organisms (16-17), *Streptococcus thermophilus* was the first microbe where the roles of specific *cas* genes and CRISPR-array components were elucidated (4). *S. thermophilus* is a non-pathogenic thermophilic Gram-positive bacterium used as a starter culture that catabolizes lactose to lactic acid in the syntrophic production of yogurt and various cheeses (18). *S. thermophilus* encodes up to four CRISPR-Cas systems, two of them (SthCRISPR1 and SthCRISPR3) are classified as Type II-A systems that are innately active in both acquisition and interference (4, 19). Accordingly, genomic analysis of *S. thermophilus* and its bacteriophages established a likely mechanism of CRISPR-Cas systems for phage/DNA protection. Investigation of CRISPR-Cas systems in *S. thermophilus* led to bioinformatic analysis of spacer origin (4, 20), discovery of the proto-spacer adjacent motif (PAM) sequences (19; 21), understanding of phage-host dynamics (22-23), demonstration of Cas9 endonuclease activity (7, 24-25), and recently, determination of the tracrRNA structural motifs governing function and orthogonality of Type II systems (26). Genomic analysis of *S. thermophilus* revealed evolutionary adaptation to milk through loss of carbohydrate catabolism and virulence genes found in pathogenic streptococci (18). *S. thermophilus* also underwent significant acquisition of niche-related genes, such as those encoding including cold-shock proteins, copper resistance proteins, proteinases, bacteriocins, and lactose catabolism proteins (18). Insertion sequences (ISs) are highly prevalent in *S. thermophilus* genomes and contribute to genetic heterogeneity between strains by facilitating dissemination of islands associated with dairy adaptation genes (18). The concomitant presence of MGEs and functional CRISPR-Cas systems in *S. thermophilus* suggests that genome homeostasis is governed at least in part by the interplay of these dynamic forces. Thus, *S. thermophilus* constitutes an ideal host for investigating the genetic outcomes of CRISPR-Cas targeting of genomic islands.

CRISPR-Cas systems have recently been the subject of intense research in genome editing applications (12), but the evolutionary roles of most endogenous microbial systems remain unknown (27). Even less is known concerning evolutionary outcomes of housing active CRISPR-Cas systems beyond the prevention of foreign DNA uptake (7), spacer acquisition events (4), and mutation caused by chromosomal self-targeting (28-32). Thus, the present inventors sought to determine the outcomes of targeting integrated MGEs with endogenous Type II CRISPR-Cas systems. Four islands were identified in *S. thermophilus* LMD-9, with lengths ranging from 8 to 102 kbp and totaling approximately 132 kbp, or 7% of the genome. In order to target genomic islands, plasmid-based expression of engineered

CRISPR arrays with self-targeting spacers were transformed into *S. thermophilus* LMD-9. Collectively, our results elucidate fundamental genetic outcomes of self-targeting events and show that CRISPR-Cas systems can direct genome evolution at the bacterial population level.

Utilizing these discoveries, the present inventors have developed novel methods for screening populations of bacterial, archaeal, algal or yeast cells for essential genes, non-essential genes, and/or expendible genomic islands; for killing one or more cells within a population of bacterial, archaeal, algal or yeast cells; for identifying a phenotype of a bacterial, archaeal, algal or yeast gene; for selecting one or more bacterial, archaeal, algal or yeast cells having a reducing the genome size from a population of bacterial, archaeal or yeast cells; and/or for identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a mutation (e.g., deletion) in its genome.

Thus in one aspect, the present inventors, have developed methods for identifying genetic variants in a population that have altered genetic content that provides them the ability to escape targeting. Here, the target sequence has been modified, and one looks for survivors that have that modification. In some aspects, the modification (i.e., mutation) is a deletion. Further, if the target sequence has been modified, then the wild type genotype is not essential.

Accordingly, in one aspect of the invention a method of screening a population of bacterial cells for essential genes, non-essential genes, and/or expendable genomic islands is provided, comprising: introducing into said population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby producing a population of transformed bacterial cells; and determining the presence or absence of a deletion in the population of transformed bacterial cells, wherein the presence of a deletion in the population of transformed bacterial cells indicates that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population means that the target region is comprised within an essential gene. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In additional aspects, the invention provides a method of screening a population of bacterial, archaeal, algal or yeast cells for essential genes, non-essential genes, and/or expendable genomic islands, comprising: introducing into the population of bacterial,

archaeal, algal or yeast cells: (a) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells; and determining the presence or absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells, wherein the presence of a deletion in the population of transformed bacterial, archaeal or yeast cells means that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells means that the target region is comprised within an essential gene.

In other aspects, a method of killing one or more bacterial cells within a population of bacterial cells is provided, comprising: introducing into the population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby killing one or more bacterial cells that comprise the target region within the population. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In an additional aspect, a method of killing one or more cells within a population of bacterial, archaeal, algal or yeast cells is provided, the method comprising: introducing into the population of bacterial, archaeal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing one

or more cells that comprise the target region in their genome within the population of bacterial, archaeal, algal or yeast cells.

Transformation of bacterial genome-targeting CRISPR RNAs can be used to selectively kill bacterial cells on a sequence-specific basis to subtract genetically distinct subpopulations, thereby enriching bacterial populations lacking the target sequence. This distinction can occur on the basis of the heterogeneous distribution of orthogonal CRISPR-Cas systems within genetically similar populations. Thus, in some embodiments, an CRISPR array that is introduced into a population of cells can be compatible (i.e., functional) with a CRISPR-Cas system in the one or more bacterial cells to be killed but is not compatible (i.e., not functional) with the CRISPR Cas system of at least one or more bacterial cells in the population. For instance, *Escherichia coli* and *Klebsiella pneumoniae* can exhibit either Type I-E or Type I-F CRISPR-Cas systems; *Clostridium difficile* encodes Type I-B systems, and different strains of *S. thermophilus* exhibit both Type II-A and Type I-E systems or just Type II-A systems. Depending on the specific CRISPR RNA transformed into a mixture of bacteria, it can specifically target that subset of the population based on its functional compatibility with its cognate system. This can be applied to diverse species containing endogenous CRISPR-Cas systems such as, but not limited to: *Pseudomonas* spp. (such as: *P. aeruginosa*), *Escherichia* spp. (such as: *E. coli*), *Enterobacter* spp. (such as: *E. cloacae*), *Staphylococcus* spp. (such as: *S. aureus*), *Enterococcus* spp. (such as: E. *faecalis, E. faecium*), *Streptomyces* spp. (such as: *S. somaliensis*), *Streptococcus* spp. (such as: *S. pyo*genes), *Vibrio* spp. (such as: *V. cholerae*), *Yersinia* spp. (such as: *Y. pestis*), *Francisella* spp. (such as: *F. tularensis, F. novicida*), *Bacillus* spp. (such as: *B. anthracis, B. cereus*), *Lactobacillus* spp. (such as: L. *casei, L. reuteri, L. acidophilus, L. rhamnosis)*, *Burkholderia* spp. (such as: *B. mallei, B. pseudomallei*), *Klebsiella* spp. (such as: *K. pneumoniae*), *Shigella* spp. (such as: *S. dysenteriae, S. sonnei*), *Salmonella* spp. (such as: *S. enterica*), *Borrelia* spp. (such as: *B. burgdorfieri*), *Neisseria* spp. (such as: *N. meningitidis*), *Fusobacterium* spp. (such as: *F. nucleatum*), *Helicobacter* spp. (such as: *H. pylori*), *Chlamydia* spp. (such as: *C. trachomatis*), *Bacteroides* spp. (such as: *B. fragilis*), *Bartonella* spp. (such as: *B. quintana*), *Bordetella* spp. (such as: *B. pertussis*), *Brucella* spp. (such as: *B. abortus*), *Campylobacter* spp. (such as: *C. jejuni*), *Clostridium* spp. (such as: *C. difficile*), *Bifidobacterium* spp. (such as: *B. infantis*), *Haemophilus* spp. (such as: *H. influenzae*), *Listeria* spp. (such as: *L. monocytogenes*), *Legionella* spp. (such as: *L. pneumophila*), *Mycobacterium* spp. (such as: *M. tuberculosis*), *Mycoplasma* spp. (such as: *M. pneumoniae*), *Rickettsia* spp. (such as: *R. rickettsii*), *Acinetobacter* spp. (such as: *A. calcoaceticus, A. baumanii*), *Rumincoccus* spp.

(such as: *R. albus*), *Propionibacterium* spp. (such as: *P. freudenreichii*), *Corynebacterium* spp. (such as: *C. diphtheriae*), *Propionibacterium* spp. (such as: *P. acnes*), *Brevibacterium* spp. (such as: *B. iodinum*), *Micrococcus* spp. (such as: *M. luteus*), and/or *Prevotella* spp. (such as: *P. histicola*).

CRISPR targeting can remove specific bacterial subsets on the basis of the distinct genetic content in mixed populations. Support for this claim is presented in examples 4, 5 where Lac^- bacteria are selected for while Lac^+ are removed from the population. The genetic distinction between the Lac^+ and Lac^- strains is presented in examples 8 and 10, where sequencing of the surviving clones revealed up to 5.5% difference in genetic content compared to the reference wild-type *S. thermophilus* strain. CRISPR-targeting spacers can thus be tuned to various levels of bacterial relatedness by targeting conserved or divergent genetic sequences. Thus, in some embodiments, the bacterial cells in the population can comprise the same CRISPR Cas system and the introduced CRISPR array thus may be functional in the bacterial population as a whole but the genetic content of the different strains or species that make up the bacterial population is sufficiently distinct such that the target region for the introduced CRISPR array is found only in the one or more bacterial species of the population that is to be killed. This can be applied to diverse species containing endogenous CRISPR-Cas systems such as, but not limited to: *Pseudomonas* spp. (such as: *P. aeruginosa*), *Escherichia* spp. (such as: *E. coli*), *Enterobacter* spp. (such as: *E. cloacae*), *Staphylococcus* spp. (such as: *S. aureus*), *Enterococcus* spp. (such as: E. *faecalis, E. faecium*), *Streptomyces* spp. (such as: *S. somaliensis*), *Streptococcus* spp. (such as: *S. pyo*genes), *Vibrio* spp. (such as: *V. cholerae*), *Yersinia* spp. (such as: *Y. pestis*), *Francisella* spp. (such as: *F. tularensis, F. novicida*), *Bacillus* spp. (such as: *B. anthracis, B. cereus*), *Lactobacillus* spp. (such as: L. *casei, L. reuteri, L. acidophilus, L. rhamnosis)*, *Burkholderia* spp. (such as: *B. mallei, B. pseudomallei*), *Klebsiella* spp. (such as: *K. pneumoniae*), *Shigella* spp. (such as: *S. dysenteriae, S. sonnei*), *Salmonella* spp. (such as: *S. enterica*), *Borrelia* spp. (such as: *B. burgdorfieri*), *Neisseria* spp. (such as: *N. meningitidis*), *Fusobacterium* spp. (such as: *F. nucleatum*), *Helicobacter* spp. (such as: *H. pylori*), *Chlamydia* spp. (such as: *C. trachomatis*), *Bacteroides* spp. (such as: *B. fragilis*), *Bartonella* spp. (such as: *B. quintana*), *Bordetella* spp. (such as: *B. pertussis*), *Brucella* spp. (such as: *B. abortus*), *Campylobacter* spp. (such as: *C. jejuni*), *Clostridium* spp. (such as: *C. difficile*), *Bifidobacterium* spp. (such as: *B. infantis*), *Haemophilus* spp. (such as: *H. influenzae*), *Listeria* spp. (such as: *L. monocytogenes*), *Legionella* spp. (such as: *L. pneumophila*), *Mycobacterium* spp. (such as: *M. tuberculosis*), *Mycoplasma* spp. (such as: *M. pneumoniae*), *Rickettsia* spp. (such as: *R.*

rickettsii), *Acinetobacter* spp. (such as: *A. calcoaceticus, A. baumanii*), *Rumincoccus* spp. (such as: *R. albus*), *Propionibacterium* spp. (such as: *P. freudenreichii*), *Corynebacterium* spp. (such as: *C. diphtheriae*), *Propionibacterium* spp. (such as: *P. acnes*), *Brevibacterium* spp. (such as: *B. iodinum*), *Micrococcus* spp. (such as: *M. luteus*), and/or *Prevotella* spp. (such as: *P. histicola*).

The extent of killing within a population using the methods of this invention may be affected by the amenability of the particular population to transformation in addition to whether the target region is comprised in a non-essential gene, an essential gene or an expendable island. The extent of killing in a population of bacterial, archaeal or yeast cells can vary, for example, by organism, by genus and species. Accordingly, as used herein "killing" means eliminating 2 logs or more of the cells in a population (1% survival or less). Less than 1 log of killing would be a small reduction in the population; whereas 2-3 logs of killing results in a significant reduction of the population; and more than 3 logs of killing indicates that the population has been substantially eradicated.

In another aspect, a method of identifying a phenotype associated with a bacterial gene is provided, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, wherein the target region comprises at least a portion of an open reading frame encoding a polypeptide or functional nucleic acid, thereby killing the cells comprising the target region and producing a population of transformed bacterial cells without the target region (i.e., surviving cells do not comprise the target region); and (i) analyzing the phenotype of the population of cells, or (ii) growing individual bacterial colonies from the population of transformed bacterial cells and analyzing the phenotype of the individual colonies. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In another aspect, a method of identifying the phenotype of a bacterial, archaeal, algal, or yeast gene is provided, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is

substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing the bacterial, archaeal, algal or yeast cells comprising the target region and producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and (i) analyzing the phenotype of the population of cells, and/or (ii) growing individual bacterial, archaeal, or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells; and analyzing the phenotype of the individual colonies.

In some embodiments, the analysis comprises PCR, optical genome mapping, genome sequencing, restriction mapping and/or restriction analysis to identify and characterize the mutation, and complementation analysis and/or phenotypic assays to analyze the phenotype.

In some embodiments of the invention determining the extent of killing or a reduction in a population can comprise any method for determining population number, including, but not limited to, (1) plating the cells and counting the colonies, (2) optical density, (3) microscope counting, (4) most probable number, and/or (5) methylene blue reduction. In some embodiments, 16S rDNA sequencing can be used to profile a composition of mixed populations. This can be done, for example, by purifying DNA from the sample as a whole, and performing either whole-genome shotgun sequencing using high-throughput technologies or, for example, by PCR amplifying the 16S gene and sequencing the products in the same manner. The sequences can then be computationally assigned to certain bacterial taxa. In other embodiments, quantitative PCR methods may also be used to quantify bacterial levels. Such techniques are well known in the art. For example, primers for qPCR can be designed to amplify specifically from a strain species, genus, or group of organisms that share the sequence. Thus, a threshold number (ct) may be used to quantify said organism or group of organisms. In additional embodiments, any bacterial activity (phenotype) specific to the target population may also be used as a metric to determine depletion of a population.

In further embodiments, a method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells is provided, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the cells comprising the target region are killed, thereby selecting one or more bacterial cells without the target region and having a reduced genome size from

the population of bacterial cells. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In some embodiments, a method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells, comprising: introducing into a population of bacterial cells: (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome, and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, wherein cells comprising the target region are killed and cells not comprising the target region survive, thereby selecting one or more bacterial cells without the target region and having a reduced genome size from the population of transgenic bacterial cells. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

As is well known in the art, transposons can be created via, for example, PCR amplification or through designed DNA synthesis, and may be introduced via any method of transformation.

In some embodiments, the invention provides a method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced genome size from a population of bacterial, archaeal, algal or yeast cells, comprising: introducing into a population of bacterial, archaeal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and the at least

one repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of bacterial, archaeal, algal or yeast cells.

In other embodiments, a method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced genome size from a population of bacterial, archaeal, algal or yeast cells is provided, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells: (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b)(i) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence or the at least one repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial, archaeal, algal or yeast cells of said population, and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome, and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, wherein cells comprising the target region are killed and cells not comprising the target region survive, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of transgenic bacterial, archaeal, algal or yeast cells.

In some aspects, the reduced genome size may be reduced as compared to a control. In some aspects, a control may be a wild-type population of bacterial, archaeal, algal or yeast

cells, or a population of bacterial, archaeal, algal or yeast cells transformed with a heterologous construct comprising a CRISPR array (e.g., a Type I, Type II, Type III, Type IV or Type V CRISPR array) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is not complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population (i.e., non-self targeting/"scrambled spacer"). In additional aspects, a control may be a population of bacterial, archaeal, algal or yeast cells transformed with a heterologous construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial, archaeal, algal or yeast cells of said population but lacks a protospacer adjacent motif (PAM).

In some embodiments, a method of identifying in a population of bacteria at least one isolate having a deletion in its genome (e.g., a chromosomal and/or plasmid deletion) is provided, comprising: introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population and cells comprising the target region are killed, thereby producing a population of transformed bacterial cells without the target region; and growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of transformed bacteria having a deletion in its genome. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In additional embodiments, the invention provides a method of identifying in a population of bacteria at least one isolate having a deletion in its genome, comprising: introducing into the population of bacterial cells: (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into

the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed [and cells not comprising the target region survive], thereby producing a population of transformed bacterial cells without the target region; and growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of bacteria having a deletion in its genome. A CRISPR array useful with this invention may be Type I, Type II, Type III, Type IV or Type V CRISPR array.

In further embodiments, a method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome is provided, comprising: introducing into a population of bacterial, archaeal, algal or yeast cells: (a) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid; (b) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (e.g., chromosomal, mitochondrial and/or plasmid genome) of the bacterial, archaeal, algal or yeast cells of said population; and (c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population of transformed bacterial, archaeal, algal or yeast cells having a deletion in its genome.

In still further embodiments, the invention provides a method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome, comprising: introducing into the population of bacterial, archaeal, algal or yeast cells: (a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide

sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and (b)(i) a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial, archaeal, algal or yeast cells of said population; and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed and cells not comprising the target region survive, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population having a deletion in its genome.

In some embodiments, fitness/growth rate can be increased by reducing genome size or by deleting select genes (encoding polypeptides or functional nucleic acids (e.g., transcriptional regulators)) that require high energy input for transcription and translation. Thus, in some embodiments, a method of increasing the fitness or growth rate of a population of bacterial, archaeal, algal or yeast cells is provided, comprising: selecting for a reduced genome size (e.g., selecting for the absence of a portion of the genome) and/or deletion in the genomes of the bacterial, archaeal, algal or yeast cells of the populations as described herein. In some embodiments, the deletion may comprise one gene or more than one gene. Therefore, through reducing the genome size or deleting a particular gene or genes, the cells of the population no longer expend energy on the transcription/translation of the portion of the genome that is absent or the deleted gene or genes, thereby having reduced energy needs

and increased fitness as compared to a control population still comprising said portion of the genome and/or said gene or genes.

In other embodiments, a method of increasing the amount of a product produced from a population of bacterial, archaeal, algal or yeast cells is provided, comprising increasing the fitness or growth rate of the cell by selecting for a deletion in the genomes of the bacterial, archaeal, algal or yeast cells as described herein. In some embodiments, the products can include, but are not limited to, antibiotics, secondary metabolites, vitamins, proteins, enzymes, acids, and pharmaceuticals.

In some embodiments, a CRISPR array (crRNA, crDNA) useful with this invention may be an array from any Type I CRISPR-Cas system, Type II CRISPR-Cas system, Type III CRISPR-Cas system, Type IV CRISPR-Cas system, or a Type V CRISPR-Cas system.

With regard to the preceding embodiments, a heterologous nucleic acid construct comprising a tracr nucleic acid and a heterologous nucleic acid construct comprising a CRISPR array may be comprised in and introduced in the same construct (e.g., expression cassette or vector) or in different constructs. In particular embodiments, a heterologous nucleic acid construct comprising a tracr nucleic acid and a heterologous nucleic acid construct comprising a CRISPR array may be comprised in single construct (e.g., expression cassette and/or vector) that may optionally further comprise a polynucleotide encoding Cas9 polypeptide. In some embodiments, the heterologous nucleic acid construct comprising a tracr nucleic acid and the heterologous nucleic acid construct comprising a CRISPR array may be operably linked to a single promoter and/or to separate promoters.

In some embodiments, a heterologous nucleic acid construct comprising a trans-activating CRISPR (tracr) nucleic acid and a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) may be comprised in a CRISPR guide (gRNA, gDNA). In some embodiments, a CRISPR guide may be operably linked to a promoter.

In some embodiments, a Cas9 polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cas9 nuclease. Exemplary Cas9 nucleases useful with this invention can be any Cas9 nuclease known to catalyze DNA cleavage in a CRISPR-Cas system. As known in the art, such Cas9 nucleases comprise a HNH motif and a RuvC motif (See, e.g., WO2013/176772; WO/2013/188638). In some embodiments, a functional fragment of a Cas9 nuclease can be used with this invention.

CRISPR-Cas systems and groupings of Cas9 nucleases are well known in the art and include, for example, a *Streptococcus thermophilus* CRISPR 1 (Sth CR1) group of Cas9 nucleases, a *Streptococcus thermophilus* CRISPR 3 (Sth CR3) group of Cas9 nucleases, a *Lactobacillus buchneri* CD034 (Lb) group of Cas9 nucleases, and a *Lactobacillus rhamnosus* GG (Lrh) group of Cas9 nucleases. Additional Cas9 nucleases include, but are not limited to, those of *Lactobacillus curvatus* CRL 705. Still further Cas9 nucleases useful with this invention include, but are not limited to, a Cas9 from *Lactobacillus animalis* KCTC 3501, and *Lactobacillus farciminis* WP 010018949.1.

Furthermore, in particular embodiments, the Cas9 nuclease can be encoded by a nucleotide sequence that is codon optimized for the organism comprising the target DNA. In still other embodiments, the Cas9 nuclease can comprise at least one nuclear localization sequence.

In some embodiments, a Type I polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cas3, Cas3' nuclease, a Cas3" nuclease, fusion variants thereof. In some embodiments, a Type I Cascade polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cas7 (Csa2), Cas8a1 (Csx13), Cas8a2 (Csx9), Cas5, Csa5, Cas6a, Cas6b, Cas8b (Csh1), Cas7 (Csh2), Cas5, Cas5d, Cas8c (Csd1), Cas7 (Csd2), Cas10d (Csc3), Csc2, Csc1, Cas6d, Cse1 (CasA), Cse2 (CasB), Cas7 (CasC), Cas5 (CasD), Cas6e (CasE), Cys1, Cys2, Cas7 (Cys3), Cas6f (Csy4), Cas6 and/or Cas4

Type I CRISPR-Cas systems are well known in the art and include, for example, *Archaeoglobus fulgidus* comprises an exemplary Type I-A CRISPR-Cas system, *Clostridium kluyveri* DSM 555 comprises an exemplary Type I-B CRISPR-Cas system, *Bacillus halodurans* C-125 comprises an exemplary Type I-C CRISPR-Cas system, *Cyanothece sp.* PCC 802 comprises an exemplary Type I-D CRISPR-Cas system, *Escherichia coli* K-12 comprises an exemplary Type I-E CRISPR-Cas system, *Geobacter sulfurreducens* comprises an exemplary Type I-U CRISPR-Cas system and *Yersinia pseudotuberculosis* YPIII comprises an exemplary Type I-F CRISPR-Cas system.

In some embodiments, a Type II polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cas9. Type II CRISPR-Cas systems well known in the art and include, for example, *Legionella pneumophila* str. Paris, *Streptococcus thermophilus* CNRZ1066 and *Neisseria lactamica* 020-06.

In some embodiments, a Type III polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cas6, Cas10 (or Csm1), Csm2, Csm3, Csm4, Csm5, and Csm6, Cmr1, Cas10 (or Cmr2), Cmr3, Cmr4, Cmr5, and Cmr6, Cas7, Cas10, Cas7 (Csm3), Cas5 (Csm4), Cas7 (Csm5), Csm6, Cas7 (Cmr1), Cas5 (Cmr3), Cas7 (Cmr4), Cas7 (Cmr6), Cas7 (Cmr6), Cmr5, Cas5 (Cmr3), Cas5 (Cs×10), Csm2, Cas7 (Csm3), and all1473. Type III CRISPR-Cas systems are well known in the art and include, for example, *Staphylococcus epidermidis* RP62A, which comprises an exemplary Type III-A CRISPR-Cas system, *Pyrococcus furiosus* DSM 3638, which comprises an exemplary Type III-B CRISPR-Cas system, *Methanothermobacter thermautotrophicus* str. Delta H, which comprises an exemplary Type III-C CRISPR-Cas system, and *Roseiflexis* sp. Rs-1, which comprises an exemplary Type III-D CRISPR-Cas system.

In some embodiments, a Type IV polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Csf4 (dinG), Csf1, Cas7 (Csf2) and/or Cas5 (csf3). Type IV CRISPR-Cas systems are well known in the art, for example, *Acidithiobacillus ferrooxidans* ATCC 23270 comprises an exemplary Type IV CRISPR-Cas system.

In some embodiments, a Type V polypeptide useful with this invention comprises at least 70% identity (e.g., about 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%, and the like) to an amino acid sequence of a Cpf1, Cas1, Cas2,or Cas4. Type V CRISPR-Cas systems are well known in the art and include, for example, *Francisella cf. novicida* Fx1 comprises an exemplary Type V CRISPR-Cas system.

Additionally provided herein are expression cassettes and vectors comprising the nucleic acid constructs, the nucleic acid arrays, nucleic acid molecules and/or the nucleotide sequences of this invention, which can be used with the methods of this disclosure.

In further aspects, the nucleic acid constructs, nucleic acid arrays, nucleic acid molecules, and/or nucleotide sequences of this invention can be introduced into a cell of a host organism. Any cell/host organism for which this invention is useful with can be used. Exemplary host organisms include bacteria, archaea, algae and fungi (e.g., yeast).

The invention will now be described with reference to the following examples. It should be appreciated that these examples are not intended to limit the scope of the claims to the invention, but are rather intended to be exemplary of certain embodiments. Any variations in the exemplified methods that occur to the skilled artisan are intended to fall within the scope of the invention.

EXAMPLES

Example 1. Bacterial Strains

All bacterial strains are listed in **Table 1**. Bacterial cultures were cryopreserved in an appropriate growth medium with 25% glycerol (vol/vol) and stored at -80°C. *S. thermophilus* was propagated in Elliker media (Difco) supplemented with 1% beef extract (wt/vol) and 1.9% (wt/vol) β-glycerolphosphate (Sigma) broth under static aerobic conditions at 37°C, or on solid medium with 1.5% (wt/vol) agar (Difco), incubated anaerobically at 37°C for 48 hours. Concentrations of 2 μg/mL of erythromycin (Em) and 5 μg/mL of chloramphenicol (Cm) (Sigma) were used for plasmid selection in *S. thermophilus*, when appropriate. *E. coli* EC1000 was propagated aerobically in Luria-Bertani (Difco) broth at 37°C, or on brain-heart infusion (BHI) (Difco) solid medium supplemented with 1.5 % agar. Antibiotic selection of *E. coli* was maintained with 40 μg/mL kanamycin (Kn) and 150 μg/mL of Em for recombinant *E. coli*, when appropriate. Screening of *S. thermophilus* derivatives for β-galactosidase activity was assessed qualitatively by supplementing a synthetic Elliker medium with 1% lactose, 1.5% agar, and 0.04% bromo-cresol purple as a pH indicator.

Example 2. DNA Isolation and Cloning

All kits, enzymes, and reagents were used according to the manufacturers' instructions. DNA purification and cloning were performed as described previously (41). Briefly, purification of genomic DNA from *S. thermophilus* employed a ZR Fungal/Bacterial MiniPrep kit (Zymo). Plasmid DNA was isolated from *E. coli* using Qiagen Spin miniprep kit (Qiagen). High fidelity PCR amplification of DNA was performed with PFU HS II DNA polymerase (Stratagene). Routine PCRs were conducted with Choice-*Taq* Blue polymerase (Denville). Primers for PCR amplification were purchased from Integrated DNA

Technologies (Coralville, IA). DNA extraction from agarose gels was performed with a Zymoclean DNA gel recovery kit (Zymo). Restriction endonucleases were acquired from Roche Molecular Biochemicals. Ligations were performed with New England Biolabs quick T4 ligase. Sequencing was performed by Davis Sequencing Inc. (Davis, CA). Cryopreserved rubidium chloride competent *E. coli* cells were prepared as previously described (41). Plasmids with *lacZ* targeting arrays were constructed with each consisting sequentially of the (1) native leader sequence specific to SthCRISPR1 or SthCRISPR3 (2) native repeats specific to CRISPR 1 or CRISPR 3 (3) spacer sequence specific to the 5' end of *lacZ* (4) another native repeat (**Fig. 1**). In order to engineer each plasmid, the sequence features listed above were ordered as extended oligomers (**Table 2**), combined using splicing by overlap extension PCR (42) and cloned into pORI28 (**Fig. 2**).

Example 3. Selection and design of CRISPR spacers

The programmable specificity of chromosomal cleavage hinges upon selection of a desired spacer sequence unique to the target allele. Specificity is further compounded by the requisite PAM, a short conserved sequence that must be proximate to the proto-spacer in the target sequence (21, 43). Thus, strict criteria for selection and design of spacers were the location of consensus PAM sequences and incidental sequence identity to extraneous genomic loci. Putative protospacers were constrained by first defining the location of all putative PAM sequences in the sense and antisense strands of *lacZ*. Within the 3,081 nt gene, there were 22 CRISPR1 (AGAAW) and 39 CRISPR3 (GGNG) PAM sites that were identical to their bioinformatically derived consensus sequences (21). After potential spacers were identified, the complete proto-spacer, seed, and PAM sequence were subjected to BLAST analysis against the genome of *S. thermophilus* LMD-9 to prevent additional targeting of non-specific loci. The spacers for CRISPR1 and CRISPR3 were disparate in sequence and corresponding PAM sites, but were designed to target the 5' end of *lacZ*, resulting in predicted cleavage sites residing 6 nt apart. Therefore, the leader sequences, repeats, and spacers on each plasmid represented orthogonal features unique to CRISPR1 or CRISPR3, respectively. To assess target locus-dependent mutations, an additional CRISPR3 plasmid was created with a spacer to the metal cation-binding residue essential for β-galactosidase activity. A CRISPR1 array plasmid containing a non-self-spacer was used as a control to quantify lethality of self-targeting.

Example 4. Transformation

Plasmids were electroporated into competent *S. thermophilus* containing the temperature-sensitive helper plasmid, pTRK669, according to methods described previously (44). Briefly, an overnight culture of *S. thermophilus* was inoculated at 1% (vol/vol) into 50 mL of Elliker medium supplemented with 1% beef extract, 1.9% β-glycerophosphate and Cm selection. When the culture achieved an OD_{600} nm of 0.3, penicillin G was added to achieve a final concentration of 10 μg/mL, in order to promote electroporation efficiency (45). Cells were harvested by centrifugation and washed 3x in 10 mL cold electroporation buffer (1 M sucrose and 3.5 mM $MgCl_2$). The cells were concentrated 100-fold in electroporation buffer and 40 μL of the suspension was aliquoted into 0.1 mm electroporation cuvettes. Each suspension was combined with 700 ng of plasmid. Electroporation conditions were set at 2,500 V, 25 μFd capacitance, and 200 Ohms resistance. Time constants were recorded and ranged from 4.4 to 4.6 ms. The suspensions were immediately combined with 950 μL of recovery medium and incubated for 8 hours at 37°C. Cell suspensions were plated on selective medium and electroporation cuvettes were washed with medium to ensure recovery of cells.

Example 5. β-galactosidase phenotype confirmation

Transformants generated from both CRISPR1 and CRISPR3 were initially screened for the β-galactosidase deficient phenotype by restreaking colonies on semi-synthetic Elliker medium supplemented with 1% lactose as the sold carbohydrate source. Loss of β-galactosidase activity was confirmed by performing Miller assays (*o*-nitrophenyl-β-D-galactoside (ONPG) (46). Briefly, cultures were propagated to late-log phase (OD_{600} nm of 1.2) in 5 mL of medium and harvested by centrifugation (4,000 x *g* for 10 min). Cells were washed and resuspended in 0.5 mL phosphate-buffered saline (Gibco-Invitrogen). Each suspension was combined with 100 uL of 0.1 mm glass beads (Biospec) and then subjected to five 60 s cycles of homogenization in a Mini-Beadbeater (Biospec). Samples were then centrifuged (15,000 x *g* for 5 min) to remove debris and intact cells. Cell lysates (10 μL aliquots) were combined with 600 μL of substrate solution (60 mM Na_2HPO_4; 40 mM NaH_2PO_4; 1 mg/mL ONPG; 2.7 μL/mL β-mercaptoethanol) and incubated for 5 min at room temperature, at which point 700 μL stop solution was added (1 M $NaCO_3$). The absorbance at 420 nm was recorded and activity of β-galactosidase was reported as Miller units, calculated as previously described (46).

Example 6. Growth and activity assessment

Cultures were preconditioned for growth assays by subculturing for 12 generations in a semi-synthethic Elliker medium deficient in lactose. Fresh medium was inoculated with an overnight culture at 1% (vol/vol) and incubated at 37°C statically. OD_{600} monitored hourly until the cultures achieved stationary phase. Acidification of milk was assessed by inoculating skim milk with an overnight culture to a level of 10^8 cfu/mL and incubating at 42°C. The pH was subsequently monitored using a Mettler Toledo Seven Easy pH meter and Accumet probe. Skim milk was acquired from the NCSU Dairy plant and Pasteurized for 30 min at 80°C.

Example 7. Identification of expendable genomic regions

In silico prediction of mobile and expendable loci for CRISPR-Cas targeting was performed on the basis of i) location, orientation, and nucleotide identity of IS elements, and ii) location of essential ORFs. In *Bacillus subtilis,* 271 essential ORFs were identified by determining the lethality of genome-wide gene knockouts (33). The *S. thermophilus* genome was queried for homologues to each essential gene from *B. subtilis* using the BLASTp search tool under the default scoring matrix for amino acid sequences. Homologues to about 239 essential ORFs were identified in *S. thermophilus*, all of which were chromosomally encoded (**Table 4**). Proteins involved in conserved cellular processes including DNA replication/homeostasis, translation machinery, and core metabolic pathways were readily identified. No homologues corresponding to cytochrome biosynthesis/respiration were observed, in accordance with the metabolic profile of fermentative bacteria. Each putative essential ORF was mapped to the reference genome using SnapGene software, facilitating visualization of their location and distribution in *S. thermophilus* LMD-9 (**Fig. 3A**).

IS elements within the *S. thermophilus* genome were grouped by aligning transposon coding sequences using Geneious® software (**Fig. 4**). Family designations were determined according to BLAST analysis within the IS element database (www-is.biotoul.fr//). To predict the potential for recombination-mediated excision of chromosomal segments, the relative location of related IS elements were mapped to the *S. thermophilus* genome (**Fig. 3A**). The IS1193 and Sth6 families of IS elements appeared most frequently in the genome and are commonly found in *Streptococcus pneumoniae* and *Streptococcus mutans* (34). Despite the prevalence of IS1193 elements, many of these loci were shown to be small fragments that exhibited some polymorphism and degeneracy, but there were also several copies present with a high level of sequence identity(**Fig. 5A**). In contrast, the Sth6 family exhibited considerable polymorphism and high degeneracy, with some copies harboring

significant internal deletions (**Fig. 5B**). IS1167 and IS1191 elements were less frequent but exhibited near perfect fidelity between the copies identified in the genome (**Figs. 5C** and **5D**). Based on the conservation of length and sequence of the IS1167 and IS1191 elements of *S. thermophilus,* and their relative proximity to milk adaptation genes, we postulate that these conserved/high fidelity transposons were recently acquired in the genome.

By combining the location of predicted essential ORFs and IS elements, expendable islands flanked by IS elements of high fidelity were identified (**Fig. 3A**) (**Table 3**). The first island contained an operon unique to *S. thermophilus* LMD-9, encoding a putative ATP-dependent oligonucleotide transport system with unknown specificity (**Fig. 3B**) (35). The second harbors the cell-envelope proteinase PrtS which contributes to the fast-acidification phenotype of *S. thermophilus* (**Fig. 3C**)(36). Notably, while *prtS* is not ubiquitous in *S. thermophilus* genomes, it has been demonstrated that the genomic island encoding *prtS* is transferable between strains using natural competence (36). The third island contains a putative ATP-dependent copper efflux protein and is present in every sequenced *S. thermophilus* strain (**Fig. 3D**). The fourth island is the largest by far in terms of length at 102 kbp, and gene content, with 102 predicted ORFs including the *lac* operon (**Fig. 3E**). This island is found in all strains of *S. thermophilus*, but the specific gene content and length varies among strains. In order to determine the outcome of targeting a large genomic island with both endogenous Type II systems, repeat-spacer arrays were generated for the *lacZ* coding sequence (**Fig. 3E**) and cloned into pORI28 (**Fig. 2**). The fourth island was selected for CRISPR-Cas targeting due to its size, ubiquity in *S. thermophilus* strains, and the ability to screen for *lacZ* mutations on the basis of a β-galactosidase negative phenotype.

Example 8. CRISPR-Cas targeting of *lacZ* selects for large deletion events

In Type II systems, Cas9 interrogates DNA and binds reversibly to PAM sequences with activation of Cas9 at the target occurring via formation of the tracrRNA::crRNA duplex (37), ultimately resulting in dsDNA cleavage (**Fig. 6A** and **6B**)(25). **Figs. 14A** and **14B** are schematics showing the general approach for co-opting endogenous CRISPR systems for targeted killing. In particular, these **Figs. 14A** and **14B** show the approach for co-opting endogenous type II systems in *Streptococcus thermophilus* for targeted killing. Thus, in *S. thermophilus*, programmed cell death was achieved using the CRISPR-Sth1 (A) or CRISPR-Sth3 (B) Type II system, by designing a genome targeting spacer sequence flanked by native repeats, whose expression was driven by a native or synthetic promoter. The transcribed repeat-spacer array is processed via host encoded RNAase III and Cas9 to yield mature

crRNAs, which recruit Cas9 to the genome to elicit double-stranded DNA cleavage resulting in cell death.

Transformation with plasmids eliciting chromosomal self-targeting by CRISPR-Cas systems appeared cytotoxic as measured by the relative reduction in surviving transformants compared to non-self-targeting plasmids (15, 29). Targeting the *lacZ* gene in *S. thermophilus* resulted in about a 2.5-log reduction in recovered transformants (**Fig. 6C**), approaching the limits of transformation efficiency. Double-stranded DNA breaks (DSBs) constitute a significant threat to the survival of organisms. The corresponding repair pathways often require end resection to repair blunt-ended DNA. Cas9-effected endonucleolysis further exacerbates the pressure for mutations caused by DSBs to occur, as restoration of the target locus to the wild-type does not circumvent subsequent CRISPR targeting. Identification of spacer origins within lactic acid bacteria revealed that 22% of spacers exhibit complementarity to self and that the corresponding genomic loci were altered, likely facilitating survival of naturally occurring self-targeting events (28).

To determine if the target locus was mutated in response to Cas9-induced cleavage, transformants were first screened for loss of β-galactosidase activity. Clones deficient in activity were genotyped at the *lacZ* locus. No mutations due to classical or alternative end joining, nor any spontaneous single nucleotide polymorphisms were observed in any of the clones sequenced. The absence of single nucleotide polymorphisms may be attributed to a low transformation efficiency compounded by low incidence of point mutations, and the absence of Ku and LigaseIV homologs correlated with an absence of non-homologous end joining (38). PCR screening indicated that the wild-type *lacZ* was not present, but the PCR amplicons did not correspond to the native *lacZ* locus; rather, an IS element-flanked sequence at another genomic locus was amplified. To investigate the genotype responsible for the loss of β-galactosidase activity, Single Molecule Real Time sequencing was performed on two clones; one generated from CRISPR3 targeting the 5' end of *lacZ*, and one generated from CRISPR3 targeting the sequence encoding the ion-binding pocket necessary for β-galactosidase catalysis (**Fig. 7A** and **7B**). This sequencing strategy was employed for its long read length to circumvent difficulty in reliably mapping reads to the proper locus, due to the high number of IS elements in the genome (35). Reads were mapped to the reference genome sequence using Geneious software, and revealed the absence of a large segment (about 102 kbp) encoding the *lacZ* open reading frame (**Fig. 7A** and **7B**). Both sequenced strains confirmed the reproducibility of the large deletion boundaries, and showed that the deletion occurred independently of the *lacZ* spacer sequence or CRISPR-Cas system used for

targeting. However, the sequencing data did not reliably display the precise junctions of the deletion.

The 102 kbp segments deleted constitute approximately 5.5% of the 1.86 Mbp genome of *S. thermophilus*. The region contained 102 putative ORFs (STER_1278-1379), encoding ABC transporters, two-component regulatory systems, bacteriocin synthesis genes, phage related proteins, lactose catabolism genes, and several cryptic genes with no annotated function (35). The effect of the deletion on growth phenotype was assessed in broth culture by measuring OD 600 nm over time (**Fig. 7C**). The deletion clones appeared to have a longer lag phase, lower final OD ($p<0.01$) and exhibited a significantly longer generation time during log phase with an average of 103 min, compared to 62 min for the wild-type ($p<0.001$). Although the deletion derivatives have 5.5% less of the genome to replicate per generation, and expend no resources in transcription or translation of the eliminated ORFs, no apparent increase in fitness was observed relative to the wild-type. β-galactosidase activity is a hallmark feature for industrial application of lactic acid bacteria and is essential for preservation of food systems through acidification. The capacity of *lacZ* deficient *S. thermophilus* strains to acidify milk was therefore assessed by monitoring pH (**Fig. 7D**). Predictably, the deletion strain failed to acidify milk over the course of the experiment, in sharp contrast to the rapid acidification phenotype observed in the wild-type.

Example 9. Genomic deletions occur through recombination between homologous IS elements

In order to investigate the mechanism of deletion, the nucleotide sequences flanking the segment were determined. The only homologous sequences observed at the junctions were two truncated IS1193 insertion sequences exhibiting 91% nucleotide sequence identity globally over 727 bp. Accordingly, a primer pair flanking the two IS elements was designed to amplify genomic DNA of surviving clones exhibiting the deletion. Each of the deletion strains exhibited a strong band of the predicted size (about 1.2 kb), and confirmed the large genomic deletion event (**Fig. 8A**). Interestingly, a faint amplicon corresponding to the chromosomal deletion was observed in the wild-type, indicating that this region may naturally excise from the genome at a low rate within wild-type populations. Sequencing of the junction amplicon was performed for 20 clones generated by chromosomal self-targeting by CRISPR3. Genotyping of the locus revealed the presence of one chimeric IS element in each clone and, furthermore, revealed the transition from the upstream element to the downstream sequence within the chimera for each clone (**Fig. 8B**). The size of deletions

observed ranged from 101,865-102,146 bp. The exact locus of transition was variable, but non-random within the clones, implying the potential bias of the deletion mechanism. *S. thermophilus* harbors typical recombination machinery encoded as RecA (STER_0077), AddAB homologs functioning as dual ATP-dependent DNA exonucleases (STER_ 1681 and STER_ 1682), and a helicase (STER_1742) of the RecD family. The high nucleotide identity between the flanking IS elements and the capacity for *S. thermophilus* to carry out site-specific recombination (4) confirms the potential for RecA-mediated recombination to mediate excision of the genomic segment (**Fig. 8C**).

Next, CRISPR-Cas targeting was evaluated for the ability to facilitate isolation of deletions for each locus with the same genetic architecture. For this purpose, three CRISPR3 repeat-spacer arrays, one targeting the oligonucleotide transporter in the first locus, *prtS* from the second locus, and the ATPase copper efflux gene from the third locus were generated and cloned into pORI28 (**Fig. 5**). In order to screen for deletions, primers flanking the IS elements at each locus were designed to amplify each deletion junction (**Fig. 8D**). The absence of wild-type loci was also confirmed in each case by designing internal primers for each genomic island (**Fig. 8E**) Following transformations with the targeting plasmids, deletions at each locus were isolated and the absence of wild-type confirmed. Sequencing of the deletion junction amplicons confirmed that a single chimeric IS element footprint remained, indicating a common mechanism for deletion at each locus. Interestingly, primers flanking the IS elements also amplified from wild-type gDNA, further suggesting that population heterogeneity naturally occurred at each locus was due to spontaneous genomic deletions. These results imply that sequence-specific Cas9 cleavage selects for the variants lacking protospacer and PAM combinations necessary for targeting. Thus, spontaneous genomic deletions can be isolated using CRISPR-Cas targeting as a strong selection for microbial variants that have already lost those genomic islands.

Example 10. Population screening

In this study, native Type IIA systems harbored in *S. thermophilus* were repurposed for defining spontaneous deletions of large genomic islands. By independently targeting four islands in *S. thermophilus*, stable mutants collectively lacking a total of 7% of the genome were generated. Characterization of the deletion junctions suggested that an IS-dependent recombination mechanism contributes to population heterogeneity and revealed deletion events ranging from 8 to 102 kbp. Precise mapping of the chimeric IS elements indicated that

natural recombination events are likely responsible for the large chromosomal deletions in *S. thermophilus* and could potentially be exploited for targeted genome editing.

Our results demonstrate that wild-type clones were removed from the population while mutants without CRISPR-Cas targeted features survived. Thus, adaptive islands were identified and validated, showing that precise targeting by an endogenous Cas9 can be exploited for isolating large deletion variants in mixed populations.

Genome evolution of bacteria occurs through horizontal gene transfer, intrinsic mutation, and genome restructuring. Genome sequencing and comparative analysis of *S. thermophilus* strains has revealed significant genome decay, but also indicates that adaptation to nutrient-rich food environments occurred through niche-specific gene acquisition (18; 35). The presence of MGEs including integrative and conjugative elements, prophages, and IS elements in *S. thermophilus* genomes is indicative of rapid evolution to a dairy environment (38-39). Mobile genetic features facilitate gene acquisition and conversely, inactivation or loss of non-essential sequences. Consequently, MGEs confer genomic plasticity as a means of increasing fitness or changing ecological lifestyles. Our results strongly indicate that CRISPR-Cas targeting of these elements may influence chromosomal rearrangements and homeostasis. This is in contrast to experiments targeting essential features, which resulted in selection of variants with inactivated CRISPR-Cas machinery (Jiang 2013). Mutation of essential ORFs is not a viable avenue for circumvention of CRISPR-Cas targeting, and thus only those clones with inactivated CRISPR-Cas systems remain. By design, targeting genetic elements predicted to be hypervariable and expendable demonstrated that variants with altered loci were viable, maintaining active CRISPR-Cas systems during self-targeting events.

Despite the near ubiquitous distribution of IS elements in bacterial genomes they remain an enigmatic genetic entity, largely due to their diversity and plasticity in function (34). Our results suggest it is possible to predict recombination between related IS elements by analyzing their location, orientation, and sequence conservation (**Fig. 4** and **Fig. 5**). CRISPR-Cas targeting can then be employed to empirically validate population heterogeneity at each predicted locus, and simultaneously increase the recovery of low incidence mutants. The high prevalence of MGEs in lactic acid bacteria, and especially *S. thermophilus,* is in accordance with their role in speciation of these hyper-adapted bacteria through genome evolution (39-40). Moreover, recovery of genomic deletion mutants using CRISPR-Cas targeting could facilitate phenotypic characterization of genes with unknown function. Mutants exhibiting the deletion of the 102 kb island encoding the *lac* operon had significantly

increased generation times relative to the wild-type and achieved a lower final OD. With 102 predicted ORFs therein, it is likely that additional phenotypes are affected and many of the genes do not have annotated functions. Considering the industrial relevance of niche-specific genes such as *prtS,* this method allows for direct assessment of how island-encoded genes contribute to adaption to grow in milk. Moreover, it is in the natural genomic and ecological context of these horizontally acquired traits, since they were likely acquired as discrete islands. These results establish new avenues for the application of self-targeting CRISPR-Cas9 systems in bacteria for investigation of transposition, DNA repair mechanisms, and genome plasticity.

CRISPR-Cas systems generally limit genetic diversity through interference with genetic elements, but acquired MGEs can also provide adaptive advantages to host bacteria. Thus, the benefit of maintaining genomically integrated MGEs despite CRISPR-Cas targeting is an important driver of genome homeostasis. Collectively, our results establish that i*n silico* prediction of GEIs can be coupled with CRISPR-Cas targeting to isolate clones exhibiting large genomic deletions. Chimeric insertion sequence footprints at each deletion junction indicated a common mechanism of deletion for all four islands. The high prevalence of self-targeting spacers exhibiting identity to genomic loci, combined with experimental demonstrations of genomic alterations, suggest that CRISPR-Cas self-targeting may contribute significantly to genome evolution of bacteria (28;30). Collectively, studies on CRISPR-Cas induced large deletions substantiate this approach as a rapid and effective means to assess the essentiality and functionality of gene clusters devoid of annotation, and define minimal bacterial genomes based on chromosomal deletions occurring through transposable elements.

Fig. 9 shows defined genetic loci for assessing type II CRISPR-Cas system-based lethality via targeting the genome of *Streptococcus thermophilus* LMD-9. The methods to carry out this analysis are known in the art. *See*, Selle and Barrangou *PNAS.* 112(26):8076–8081 (2015).

Both orthogonal type II systems (CRISPR1 and CRISPR3) were tested; CRISPR1 targets in dark gey, CRISPR3 targets in light grey. Specific genetic features were selected to test (i) intergenic regions (INT), (ii) mobile genetic elements (IS, GEI1-GEI3, PRO, lacZ, EPS), (iii) essential genes (dltA, LTA),(iv) poles of the replichore (ORI, TER), and forward versus reverse strands of DNA (outer targets versus inner targets).

Fig. 10 shows CRISPR-based lethality achieved by targeting the regions defined in **Fig. 9**. Log reduction in CFU was calculated with regard to transformation of a non-targeting

plasmid control; pORI28. Lethality ranged from 2-3 log reduction for all targets tested, regardless of chromosomal location, coding sequence, or essentiality.

Fig. 11 shows transcriptional profiles of CRISPR-mediated genomic island deletion strains. Recovery and genotyping of cells surviving CRISPR targeting of the genomic islands 1-4 resulted in identification of stable independent mutants lacking the genomic island targeted in each experiment. Subsequently, the cells were propagated and their total RNA was isolated and sequenced. Using this approach, transcriptional profiles were generated by mapping sequencing reads to the reference genome. In each case, the absence of sequencing reads to the predicted genomic island loci further suggested the loss of the target genetic entity, while having minimal impact on the expression of core genes throughout the rest of the genome.

Furthermore, RNA sequencing data supports the boundaries of the deletions by using read coverage mapping and transcriptional value comparisons and additionally supports the discernment of phenotype using comparative transcriptomics generated using the same data set. Specifically, the lack of transcriptional activity present at the expected deletion regions using high-throughput RNA-sequencing is confirmed as shown in **Fig. 12**. **Fig. 12** shows $\log_2$ transformed RNA-sequencing read coverage of genomic island deletion strains and for each genomic island strain (GEI1-GEI4), the absence of sequencing reads to the predicted genomic island loci further suggested the loss of the target genetic entity, while having minimal impact on the expression of core genes throughout the rest of the genome.

Fig. 13 further confirms lack of transcriptional activity for the deleted genes. For each of the genomic island deletion strains (GEI1-GEI4), the expression of genes encoded on each of the target islands (black) was minimal. Genes encoded in GEI1 are shown in the upper left panel, genes encoded in GEI2 are shown in the upper right panel, genes encoded in GEI3 are shown in the lower left panel, and genes encoded in GEI4 are shown in the lower right panel. In general, genomic island deletions 1 and 2 had minimal impact on the transcription of other genes (gray), whereas genomic island 3 and 4 appeared to affect the transcription of other genes not encoded on the islands.

In addition, RNA-sequencing data was used to compare the transcriptional levels of genes not encoded on the deleted island (GEI4), i.e., other genes still present in the chromosome, and identifying phenotypes associated with genomic deletions. Genes that are differentially transcribed in the deletion strain suggest that cellular processes were impacted by the genes that were lost or that there is compensation for the loss of the activity of these genes. Thus, inferences can be made about the pathways to which these genes or genomic

regions are relevant. **Table 5** provides a list of differentially expressed genes identified in deletion strain GEI4. Many of the genes observed to be differentially expressed relate to the biosynthetic capacity of *Streptococcus thermophilus*, including aromatic amino acid and purine biosynthesis.

Example 11. Targeted killing of *Lactobacillus casei* using a Type II CRISPR-Cas system

Exemplary CRISPR-Cas Type II guides of *L. Casei* are provided in **Fig. 15**. The top panel provides a predicted guide, while the bottom left panel shows an exemplary dual guide structure and the bottom right panel shows an exemplary single guide structure.

Example 12. Targeted killing of *Lactobacillus gasseri* using a Type II CRISPR-Cas system

Exemplary Type II guides for targeted killing of *L. gasseri* are provided in **Fig. 16**. The upper panel provides the predicted guide, while the bottom left panel provides the correct dual guide crRNA:tracrRNA (confirmed by RNA sequencing) and the bottom right panel provides an exemplary predicted single guide.

Plasmids were transformed into *L. gasseri* each carrying different constructs as follows: an empty pTRK563 vector, a construct with the correct protospacer but an incorrect PAM, the correct PAM but a protospacer that is not in the array, and the correct protospacer with the PAM. The results are shown in **Fig. 19**. The plasmid having the correct protospacer and correct PAM showed significantly more interference targeting and cell death.

Example 13. Targeted killing of *Lactobacillus pentosus* using a Type II CRISPR-Cas system

Exemplary Type II guides for targeted killing of *L. pentosus* are provided in **Fig. 17**. The top panel shows the predicted guide. The panel on the bottom left is the correct dual guide crRNA:tracrRNA (confirmed by RNA Sequencing) and the bottom right panel is an exemplary predicted artificial single guide.

Plasmids were transformed into *L. pentosus,* each plasmid carrying different constructs as follows: a construct with the correct protospacer but an incorrect PAM, a construct with a correct PAM but a protospacer that is not in the array, an empty pTRK563 vector, and a correct protospacer with a correct PAM. The results are shown in **Fig. 20**. The plasmid having a correct protospacer and correct PAM (Lpe1 gttaat) showed significantly more interference targeting and cell death.

Example 14. Targeted killing of *Lactobacillus jensenii* using a Type II CRISPR-Cas system

Fig. 18 provides exemplary CRISPR-Cas Type II guides. The panel on the left is the correct dual guide crRNA:tracrRNA as confirmed by RNA sequencing and the panel on the right is an exemplary predicted artificial single guide.

Plasmids were transformed into *L. jensenii* each carrying different constructs as follows: a construct comprising an empty pTRK563 vector, a construct with the correct protospacer but an incorrect PAM, a construct with a correct PAM but a protospacer that is not in the array, and a construct having a correct protospacer with a correct PAM. The results are shown in **Fig. 22**. The plasmid having a correct protospacer and correct PAM showed substantially more interference targeting and cell death.

Example 15. Targeted killing of *Lactobacillus casei* NCK 125 using a Type I CRISPR-Cas system

Fig. 21 provides an exemplary Type I CRISPR- Cas guide for *L. casei,* which comprises the sequence of the native Type I leader and repeat found in *L. casei* NCK 125. PAM 5'-YAA-3' was predicted using the native spacer sequences in the organism. The artificial array contains a spacer that targets the 16s rDNA gene in the host genome. The results are provided in **Fig. 22**, which shows a significant reduction between the empty vector and two different artificial arrays: one of which contains a single spacer targeting the + strand in the 16s gene (1-2 alt) and the other containing the original spacer targeting the + strand but containing an additional spacer targeting the – strand in the 16s gene (1, 2-3).

CRISPR-Cas systems as described herein may be used for, for example, (i) targeted reduction of pathogens in the case of either medical intervention (e.g., pathogens including but not limited to, fungi, nematodes, protozoa (e.g., malaria), cestodes, coccidia (microsporidia), trematodes, pentastomids, acanthocephalans, arthropods, and the like); (ii) for protection of consumables (food systems, animals, crops); (iii) for control and/or removal of undesirable organisms from industrial fermentative processes (raw materials, processing equipment, starter cultures) and (iv) for control of environmental microbial consortia to impact ecosystems and/or chemical cycles as well as for remediation.

Table 1. Bacterial Strains and Plasmids

Strain designation	**Description**	**Original Reference**
E. coli EC1000	Host for pORI plasmids, chromosomal $repA^+$ (pWVO1), Km^R Host for pTRK935	47
S. thermophilus LMD-9 *S thermophilus* LMD-9 with pTRK669	Wild-type Wild-type, $RepA^+$ and Cm^R conferred by pTRK669	40 This study
Plasmids		
pORI28	Broad range non-replicative vector, Em^R	47
pTRK669	Ts-helper plasmid $repA^+$, Cm^R	44
pCRISPR1::*lacZ*	pORI28::CRISPR1-Leader-RSR-*lacZ* N-terminus spacer	This study
pCRISPR3::*lacZ*	pORI28::CRISPR1-Leader-RSR-*lacZ* active site spacer	This study
pCRISPR3::ABC	pORI28::CRISPR1-Leader-RSR-ABC spacer	This study
pCRISPR3::Cu	pORI28::CRISPR1-Leader-RSR-Cu efflux spacer	This study
pCRISPR3::*prts*	pORI28::CRISPR1-Leader-RSR-*prts* spacer	This study
pCRISPR3::*lacZ* pCRISPR1::Non-self	pORI28::CRISPR3-Leader-RSR-*lacZ* N-terminus spacer pORI28::CRISPR1-Leader-RSR-non self spacer	This study This study

Table 2. Primers

Primer Name	Sequence	Function
C1_N-term_F	CAAGAACAGTTATTGATTTTATAATCACTATGTGGGTATGAAAATCTCAAAAATCATTTGAGGTTTT TGTACTCTCAAGATTTAAGTAACTGTACAACATTAAGAGATTGTCTTAACTT	Template for SOE-PCR
C3_N-term	AGCGGATAACAATTTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACTCAGAAAATTCTT CAAGAGATTCAAAATACTGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACCTCGTAGGATATC TTTTCTAC	Template for SOE-PCR
C1_N-term_R	AGCGGATAACAATTTCACGTTGTACAGTTACTTAAATCTTGAGAGTACAAAAACAGGGGAGATGAAG TTAAGACAATCTCTTAATGT	Template for SOE-PCR
C3_A-site	AGCGGATAACAATTTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACGAAGTCTTGGTCT TCCAACCAGCTTGCTGTAGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACCTCGTAGGATATC TTTTCTAC	Template for SOE-PCR
C3_ABC	AGCGGATAACAATTTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACGATAACACGAGATAA AACATCCAGCCCACCGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACctcgtaggatatcttttc tac	
C3_prtS	AGCGGATAACAATTTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACGTTGTAGCTTTGAGG TCTGAGAATACACGCGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACctcgtaggatatcttttc tac	
C3_Cu	AGCGGATAACAATTTCACGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACGATTGCTCAATCA ATCGTTTCAGCTGCTAAGTTTTGGAACCATTCGAAACAACACAGCTCTAAAACctcgtaggatatct tttctac	
C3LF	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	Amplify Sth3 Leader from gDNA
C3LR	CTCGTAGGATATCTTTTCTAC	Amplify Sth3 Leader from gDNA
C1F	AGCAGGGATCCCAAGAACAGTTATTGATTTTATAATC	CRISPR1 SOE-PCR Forward
C3F	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	CRISPR3 SOE-PCR Foward
C1C3R	TGCTGGAGCTCGTGAAATTGTTATCCGCT	CRISPR1 and CRISPR3 SOE-PCR Reverse
1193F	TTGAACACTAGGAACCTCATA	Deletion junction amplification
1193R	CGTAAGGTTTTGATGACTCAAG	Deletion junction amplification
pORI28F	TTGGTTGATAATGAACTGTGCTG	Sequencing MCS of pORI28
pORI28R	TTGTTGTTTTTATGATTACAAAGTGA	Sequencing MCS of pORI28

Table 3. Putative expendable genomic islands and islets.

Genomic island	ORF region	Length kbp	GC content %	Notable genes	IS family
1	STER_139-STER_148	7.81	37.1	Oligopeptide transporters	IS6
2	STER_840-STER_848	10.29	39.9	Proteinase PrtS	ISSth16/IS1167
3	STER_881-STER_888	8.71	39.3	Copper efflux	IS1191
4	STER_1277-STER_1380	101.76	37.2	Lactose catabolism, 2-component reg., bacteriocin synthesis, ABC transporters	IS1193

Table 4. Homologues to about 239 essential ORFs identified in *S. thermophilus.*

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1	101	1465	+	chromosomal replication initiation prot. DnaA	98%	4.00E-124	44%	Chromosomal replication initiator protein DnaA
CSTER	2	1620	2756	+	DNA polymerase III subunit beta	99%	1.00E-88	39%	DNA polymerase III subunit beta dnan
CSTER	6	5818	6387	+	Peptidyl-tRNA hydrolase	97%	1.00E-61	51%	
CSTER	9	10331	10702	+	cell-cycle prot.	94%	1.00E-05	22%	Cell division protein DivIC
CSTER	12	12122	13387	+	tRNA(Ile)-lysidine synthetase, MesJ	92%	3.00E-40	32%	tRNA(Ile)-lysidine synthase tils
CSTER	13	13469	14011	+	Hypoxanthine-guanine phosphoribosyltransferase	99%	2.00E-82	63%	Hypoxanthine-guanine phosphoribosyltransferase hprt
CSTER	40	25595	26425	+	Cell shape-determining protein MreC	94%	1.00E-22	28%	Cell shape-determining protein MreC
CSTER	43	28617	29582	+	ribose-phosphate pyrophosphokinase	97%	3.00E-145	65%	Ribose-phosphate pyrophosphokinase prs
CSTER	47	32842	33846	+	put. glycerol-3-phosphate acyltransferase	94%	9.00E-124	54%	Phosphate acyltransferase PlsX
CSTER	48	33846	34091	+	acyl carrier prot.	76%	1.00E-08	40%	Acyl carrier protein
CSTER	65	53036	54727	-	Arginyl-tRNA synthetase	98%	6.00E-52	28%	
CSTER	95	72798	77192	+	DNA polymerase III PolC	99%	0.00E+00	51%	DNA polymerase III PolC-type
CSTER	105	82956	83723	+	30S ribosomal prot. S2	96%	1.00E-118	69%	
CSTER	106	83841	84881	+	Translation elongation factor Ts	97%	8.00E-72	44%	
CSTER	117	97363	98706	+	Cysteinyl-tRNA synthetase	99%	1.00E-178	54%	
CSTER	127	104406	104852	+	50S ribosomal prot. L13	99%	9.00E-57	57%	
CSTER	128	104880	105272	+	30S ribosomal prot. S9	100%	3.00E-59	68%	
CSTER	193	159363	160967	+	CTP synthetase	99%	0.00E+00	68%	CTP synthase pyrg
CSTER	199	166439	166897	-	conserv. hyp. prot.	81%	2.00E-07	28%	Protein NrdI
CSTER	208	176203	176472	+	30S ribosomal prot. S15	100%	6.00E-38	63%	
CSTER	217	183028	184185	+	undecaprenyl pyrophosphate phosphatase	98%	2.00E-68	38%	Probable undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase tagO
CSTER	218	184346	185116	+	ABC transporter ATPase	95%	8.00E-130	71%	Vegetative protein 296 sufC
CSTER	219	185153	186415	+	hyp. prot.	97%	4.00E-111	42%	FeS cluster assembly protein SufD
CSTER	220	186469	187701	+	put. aminotransferase (class V)	98%	0.00E+00	60%	Cysteine desulfurase SufS

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	221	187688	188122	+	NifU fam. prot.	95%	3.00E-48	51%	Zinc-dependent sulfurtransferase SufU
CSTER	245	208199	208993	+	phosphatidate cytidylyltransferase	98%	3.00E-67	41%	Phosphatidate cytidylyltransferase
CSTER	247	210343	212205	+	Prolyl-tRNA synthetase	98%	0.00E+00	50%	
CSTER	252	215735	216022	+	Co-chaperonin GroES (HSP10)	97%	1.00E-18	43%	
CSTER	253	216071	217690	+	Chaperonin GroEL (HSP60 family)	98%	0.00E+00	75%	
CSTER	261	224055	224204	+	50S ribosomal prot. L33	97%	9.00E-12	50%	
CSTER	262	224216	224392	+	Preprotein translocase subunit SecE	81%	1.40E+00	21%	Protein translocase subunit SecE
CSTER	268	227616	230117	+	Leucyl-tRNA synthetase	99%	0.00E+00	70%	
CSTER	273	232342	233796	+	nicotinate phosphoribosyltransferase	97%	0.00E+00	63%	Nicotinate phosphoribosyltransferase pncb
CSTER	274	233808	234629	+	NAD synthetase	100%	9.00E-109	59%	NH(3)-dependent NAD(+) synthetase nade
CSTER	286	248252	249580	+	UDP-N-acetylmuramate-alanine ligase	99%	3.00E-170	55%	UDP-N-acetylmuramate--L-alanine ligase murC
CSTER	302	260849	261664	+	Glutamate racemase	94%	7.00E-83	48%	Glutamate racemase 1 racE
CSTER	307	264328	265041	+	segregation and condensation prot. A	93%	5.00E-37	39%	Segregation and condensation protein A
CSTER	308	265034	265615	+	segregation and condensation prot. B	94%	9.00E-36	40%	Segregation and condensation protein B
CSTER	313	269788	270297	+	rRNA methytransferase	96%	3.00E-58	51%	rRNA methyltransferase
CSTER	349	303983	305959	+	Transketolase	98%	0.00E+00	58%	Transketolase tkt
CSTER	357	312857	314032	+	chromosome replication initiation/membrane attachment protein DnaB	63%	3.00E-02	18%	Replication initiation and membrane attachment protein
CSTER	358	314036	314938	+	primosomal prot. DnaI	99%	7.00E-58	35%	Primosomal protein DnaI
CSTER	359	315044	316354	+	GTP-binding prot. EngA	100%	0.00E+00	67%	GTPase Der
CSTER	368	321054	322394	-	Seryl-tRNA synthetase	100%	0.00E+00	63%	
CSTER	376	329673	330116	+	conserved hyp. prot.	89%	6.00E-33	43%	tRNA threonylcarbamoyladenosine biosynthesis protein TsaE
CSTER	380	332545	333732	+	Transcription elongation factor	90%	2.00E-113	46%	Transcription termination/antitermination protein NusA

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	383	334363	337194	+	Translation initiation factor IF-2	99%	0.00E+00	55%	
CSTER	387	339864	341309	-	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	92%	3.00E-52	29%	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase murE
CSTER	419	364962	365894	+	put. manganese-dependent inorganic pyrophosphatase	99%	3.00E-125	58%	Manganese-dependent inorganic pyrophosphatase ppac
CSTER	430	375355	375579	+	acyl carrier prot.	93%	5.00E-14	49%	Acyl carrier protein
CSTER	432	376667	377593	+	acyl-carrier-protein S-malonyltransferase	95%	3.00E-86	47%	Malonyl CoA-acyl carrier protein transacylase
CSTER	433	377606	378340	+	3-ketoacyl-(acyl-carrier-protein) reductase	99%	6.00E-79	47%	3-oxoacyl-[acyl-carrier-protein] reductase FabG
CSTER	434	378401	379633	+	3-oxoacyl-(acyl carrier protein) synthase II	99%	1.00E-133	48%	3-oxoacyl-[acyl-carrier-protein] synthase 2
CSTER	435	379637	380125	+	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	98%	9.00E-26	37%	Biotin carboxyl carrier protein of acetyl-CoA carboxylase
CSTER	437	380667	382037	+	acetyl-CoA carboxylase biotin carboxylase subunit	98%	0.00E+00	60%	Biotin carboxylase 1
CSTER	438	382043	382909	+	acetyl-CoA carboxylase subunit beta	94%	8.00E-98	51%	Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta
CSTER	439	382906	383676	+	acetyl-CoA carboxylase subunit alpha	76%	2.00E-81	53%	Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha
CSTER	442	385819	387417	+	conserved hyp. prot.	98%	0.00E+00	58%	Ribonuclease Y ymda
CSTER	455	402156	402470	+	50S ribosomal prot. L21	99%	2.00E-44	66%	
CSTER	456	402507	402797	+	50S ribosomal prot. L27	95%	5.00E-47	79%	
CSTER	460	405038	405805	+	Dihydrodipicolinate reductase	98%	7.00E-95	53%	4-hydroxy-tetrahydrodipicolinate reductase dapb
CSTER	461	405802	407010	+	tRNA CCA-pyrophosphorylase	98%	6.00E-86	40%	CCA-adding enzyme
CSTER	475	422256	422813	+	Ribosome recycling factor	100%	2.00E-68	56%	
CSTER	485	430547	432550	+	Methyonyl-tRNA synthetase	98%	0.00E+00	58%	
CSTER	492	438748	439890	+	protease maturation prot. precursor	91%	6.00E-11	26%	Foldase protein PrsA
CSTER	493	440232	442850	+	Alanyl-tRNA synthetase	99%	0.00E+00	52%	
CSTER	513	460453	463104	+	Valyl-tRNA synthetase	99%	0.00E+00	63%	

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	523	469888	471168	+	cell division prot. FtsW	93%	4.00E-59	34%	Putative lipid II flippase FtsW
CSTER	524	471406	472602	+	elongation factor Tu	99%	0.00E+00	76%	
CSTER	525	472851	473609	+	Triosephosphate isomerase	97%	8.00E-109	62%	Triosephosphate isomerase tpia
CSTER	526	473848	474477	+	Thymidylate kinase	96%	4.00E-73	55%	Thymidylate kinase tmk
CSTER	527	474486	475361	+	DNA polymerase III subunit delta'	96%	8.00E-41	33%	DNA polymerase III subunit delta'
CSTER	539	484827	485744	+	Glycyl-tRNA synthetase, alpha subunit	97%	4.00E-169	74%	
CSTER	540	486028	488064	+	Glycyl-tRNA synthetase, beta subunit	97%	0.00E+00	45%	
CSTER	567	512413	512916	+	50S ribosomal prot. L10	100%	7.00E-61	57%	
CSTER	568	512991	513359	+	50S ribosomal prot. L7/L12	100%	4.00E-38	67%	
CSTER	603	548926	550308	+	N-acetylglucosamine-1-phosphate uridyltransferase	97%	3.00E-164	52%	Bifunctional protein GlmU
CSTER	623	566278	566781	+	Dihydrofolate reductase	94%	8.00E-34	38%	Dihydrofolate reductase dfra
CSTER	626	568210	568809	+	GTPase EngB	96%	7.00E-92	63%	Probable GTP-binding protein EngB
CSTER	632	573613	574254	-	put. glycerol-3-phosphate acyltransferase PlsY	90%	3.00E-40	46%	Glycerol-3-phosphate acyltransferase plsy
CSTER	633	574392	576341	+	DNA topoisomerase IV subunit B	97%	0.00E+00	71%	DNA topoisomerase 4 subunit B
CSTER	634	576969	579434	+	DNA topoisomerase IV subunit A	95%	0.00E+00	54%	DNA topoisomerase 4 subunit
CSTER	660	598871	599725	+	methylenetetrahydrofolate dehydrogenase/ methenyltetrahydrofolate cyclohydrolase	99%	2.00E-109	56%	Bifunctional protein FolD
CSTER	668	605570	606469	+	GTPase Era	98%	5.00E-143	64%	GTPase Era
CSTER	670	607313	607927	+	Dephospho-CoA kinase	95%	2.00E-50	45%	Dephospho-CoA kinase coae
	672	609192	609338		50S ribosomal prot. L33	97%	1.00E-16	52%	
CSTER	684	620012	621316	+	Enolase	100%	0.00E+00	70%	Enolase eno
CSTER	731	664470	665690	-	put. cytosine-C5 specific DNA methylase	60%	5.00E-16	31%	Probable BsuMI modification methylase subunit YdiO
CSTER	733	666939	668429	-	Lysyl-tRNA synthetase (class II)	97%	0.00E+00	64%	
CSTER	761	694548	695582	+	DNA polymerase III delta subunit	84%	2.00E-37	33%	Uncharacterized protein YqeN
CSTER	773	704704	706056	+	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase	98%	8.00E-147	49%	UDP-N-acetylmuramoylalanine--D-glutamate ligase murD

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	774	706060	707130	+	N-acetylglucosaminyl transferase	95%	3.00E-30	27%	UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase murG
CSTER	775	707140	708264	+	Cell division protein FtsQ	79%	4.00E-15	26%	Cell division protein DivIB
CSTER	776	708388	709764	+	Cell division protein FtsA	84%	4.00E-103	44%	Cell division protein FtsA
CSTER	777	709793	711115	+	Cell division protein FtsZ	92%	2.00E-123	53%	Cell division protein FtsZ
CSTER	783	714803	717592	+	Isoleucyl-tRNA synthetase	99%	0.00E+00	58%	
CSTER	787	720193	720459	-	50S ribosomal protein L31	100%	1.00E-22	50%	
CSTER	793	725315	726394	+	Peptide chain release factor 1	99%	6.00E-158	59%	
CSTER	796	727911	729161	+	serine hydroxymethyltransferase	97%	2.00E-171	60%	Serine hydroxymethyltransferase glya
CSTER	833	760678	762396	-	put. phosphoglucomutase	98%	0.00E+00	47%	Phosphoglucomutase pgm
CSTER	850	783500	783736	-	30S ribosomal prot. S20	94%	1.00E-13	46%	
CSTER	864	797961	799307	+	Asparaginyl-tRNA synthetases	99%	4.00E-180	56%	
CSTER	903	833853	835661	+	D-fructose-6-phosphate amidotransferase	100%	0.00E+00	59%	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] glmS
CSTER	915	845349	846911	+	Signal recognition particle prot.	95%	2.00E-178	57%	Signal recognition particle protein
CSTER	919	848991	849845	+	ribosomal biogenesis GTPase	98%	8.00E-101	50%	Ribosome biogenesis GTPase rbga
CSTER	923	852605	854749	+	DNA topoisomerase I	98%	0.00E+00	64%	DNA topoisomerase 1
CSTER	994	915311	917623	+	ATP-dependent DNA helicase PcrA	99%	0.00E+00	54%	ATP-dependent DNA helicase PcrA
CSTER	1034	959843	961231	-	branched-chain alpha-keto acid dehydrogenase subunit E2	99%	5.00E-69	34%	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex odhb
CSTER	1036	962412	963383	-	acetoin dehydrogenase complex, E1 component, alpha subunit	81%	3.00E-33	28%	Pyruvate dehydrogenase E1 component subunit alpha pdha
CSTER	1087	1007529	1007888	-	50S ribosomal prot. L20	85%	2.00E-56	80%	
CSTER	1088	1007945	1008145	-	50S ribosomal prot. L35	100%	1.00E-23	62%	
CSTER	1089	1008184	1008714	-	Translation initiation factor 3 (IF-3)	98%	2.00E-63	60%	
CSTER	1102	1017867	1018880	-	Peptide chain release factor 2	91%	2.00E-137	57%	

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1115	1029796	1031136	-	sensor histidine kinase	97%	4.00E-117	71%	Transcriptional regulatory protein YycF
CSTER	1116	1031129	1031836	-	two-component response regulator	68%	2.00E-120	47%	Sensor histidine kinase YycG
CSTER	1123	1036555	1037826	-	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	96%	2.00E-172	59%	UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1 muraa
CSTER	1135	1049456	1050649	-	S-adenosylmethionine synthetase	97%	0.00E+00	67%	S-adenosylmethionine synthase metk
CSTER	1136	1050929	1051867	+	biotin-(acetyl-CoA carboxylase) ligase	92%	4.00E-44	33%	Bifunctional ligase/repressor BirA
CSTER	1138	1052266	1053918	-	DNA polymerase III, gamma/tau subunit	93%	3.00E-137	44%	DNA polymerase III subunit gamma/tau Dnax
CSTER	1144	1056546	1056893	-	50S ribosomal prot. L19	100%	2.00E-61	77%	
CSTER	1164	1073225	1074244	-	6-phosphofructokinase	93%	4.00E-124	61%	ATP-dependent 6-phosphofructokinase pfka
CSTER	1165	1074337	1077447	-	DNA polymerase III, alpha subunit DnaE	96%	0.00E+00	35%	DNA polymerase III subunit alpha dnae
CSTER	1166	1077600	1078385	-	Putative translation factor (SUA5)	42%	3.00E-04	24%	ywlC unknown conserved protein with a putative RNA binding motif TW
CSTER	1175	1085950	1086225	-	histone-like DNA-binding prot.	95%	4.00E-40	72%	DNA-binding protein HU 1
CSTER	1182	1091629	1092501	-	Geranylgeranyl pyrophosphate synthase	89%	5.00E-58	44%	Farnesyl diphosphate synthase ispa yqid
CSTER	1188	1095764	1096462	-	DNA replication prot. DnaD	85%	1.00E-07	22%	DNA replication protein DnaD
CSTER	1193	1101272	1102201	-	ribonuclease Z	99%	2.00E-105	50%	Ribonuclease Z rnz
CSTER	1197	1105450	1106877	-	cell division prot.	88%	2.00E-57	38%	Rod shape-determining protein RodA
CSTER	1230	1131390	1132742	-	Phosphoglucosamine mutase	100%	0.00E+00	63%	Phosphoglucosamine mutase ybbt glnM
CSTER	1248	1154457	1156616	+	ribonucleotide-diphosphate reductase, alpha subunit	98%	0.00E+00	47%	Ribonucleoside-diphosphate reductase subunit alpha nrde
CSTER	1249	1156775	1157737	+	ribonucleotide-diphosphate reductase, beta subunit	97%	8.00E-98	49%	Ribonucleoside-diphosphate reductase subunit beta nrdf
CSTER	1256	1161831	1164284	-	DNA gyrase, A subunit	98%	0.00E+00	62%	DNA gyrase subunit A

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1267	1174041	1176449	-	phenylalanyl-tRNA synthetase, beta subunit	99%	0.00E+00	47%	
CSTER	1270	1177285	1178328	-	Phenylalanyl-tRNA synthetase alpha subunit	98%	3.00E-161	62%	
CSTER	1271	1178950	1182483	-	chromosome segregation ATPase, SMC prot.	99%	0.00E+00	37%	Chromosome partition protein Smc
CSTER	1272	1182486	1183175	-	ribonuclease III	89%	4.00E-68	47%	Ribonuclease 3
CSTER	1273	1183332	1184267	-	Dihydrodipicolinate synthase	93%	5.00E-88	47%	4-hydroxy-tetrahydrodipicolinate synthase dapA
CSTER	1274	1184605	1185681	-	Aspartate-semialdehyde dehydrogenase	99%	1.00E-143	58%	Aspartate-semialdehyde dehydrogenase asd
CSTER	1382	1293065	1294063	-	Thioredoxin reductase	97%	7.00E-113	50%	Ferredoxin--NADP reductase 2 yumc
CSTER	1383	1294065	1294784	-	tRNA (guanine-N(1)-)-methyltransferase	100%	6.00E-87	56%	tRNA (guanine-N(1)-)-methyltransferase
CSTER	1395	1303326	1304261	-	methionyl-tRNA formyltransferase	96%	5.00E-98	49%	
CSTER	1396	1304279	1306675	-	primosome assembly protein PriA	99%	0.00E+00	48%	Primosomal protein N PriA
CSTER	1398	1307153	1307782	-	Guanylate kinase	98%	5.00E-87	59%	Guanylate kinase gmk
CSTER	1399	1308015	1309406	-	cell division prot. FtsY	97%	8.00E-121	55%	Signal recognition particle receptor FtsY
CSTER	1422	1330503	1331339	-	inorganic polyphosphate/ATP-NAD kinase	100%	5.00E-78	44%	NAD kinase 1 ppnk
CSTER	1425	1333019	1333990	+	ribose-phosphate pyrophosphokinase	100%	4.00E-113	54%	Ribose-phosphate pyrophosphokinase prs
CSTER	1426	1333994	1335106	+	put. cysteine desulfurase	98%	2.00E-115	47%	Putative cysteine desulfurase IscS
CSTER	1448	1355769	1356878	-	RNA polymerase sigma factor RpoD	93%	6.00E-167	70%	RNA polymerase sigma factor SigA
CSTER	1449	1356882	1358693	-	DNA primase	68%	9.00E-86	37%	DNA primase DnaG
CSTER	1451	1359076	1359252	-	30S ribosomal prot. S21	98%	3.00E-29	89%	
CSTER	1464	1371306	1372619	-	GTPase ObgE	100%	0.00E+00	67%	GTPase ObgE
CSTER	1480	1387053	1389005	-	DNA gyrase subunit B	100%	0.00E+00	68%	DNA gyrase subunit B
CSTER	1497	1401241	1402143	-	UDP-N-acetylenolpyruvoylglucosamine reductase	98%	1.00E-66	40%	UDP-N-acetylenolpyruvoylglucosamine reductase murB

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1506	1409996	1410268	-	30S ribosomal prot. S16	100%	5.00E-42	67%	
CSTER	1512	1415065	1416078	-	put. lipid kinase	95%	1.00E-96	51%	Diacylglycerol kinase dagK
CSTER	1513	1416088	1418034	-	NAD-dependent DNA ligase	97%	0.00E+00	55%	DNA ligase LigA
CSTER	1516	1419659	1420519	-	methionine aminopeptidase	99%	6.00E-56	37%	
CSTER	1519	1422447	1423709	-	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	96%	6.00E-125	46%	UDP-N-acetylglucosamine 1-carboxyvinyltransferase 1 muraa
CSTER	1522	1426828	1427583	-	1-acyl-sn-glycerol-3-phosphate acyltransferase	92%	1.00E-29	32%	1-acyl-sn-glycerol-3-phosphate acyltransferase plsC
CSTER	1534	1439747	1441120	-	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	98%	2.00E-98	38%	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase murF
CSTER	1544	1447537	1448583	-	D-alanyl-alanine synthetase A	98%	2.00E-95	42%	D-alanine--D-alanine ligase ddl
CSTER	1583	1483475	1484107	-	nicotinic acid mononucleotide adenylyltransferase	98%	2.00E-62	44%	Nicotinate-nucleotide adenylyltransferase nadd
CSTER	1584	1484211	1484528	-	put. RNA-binding prot.	95%	1.00E-24	50%	Probable RNA-binding protein YqeI
CSTER	1585	1484767	1485885	-	GTP-binding prot. YqeH	100%	6.00E-163	59%	Uncharacterized protein YqeH
CSTER	1590	1488021	1489463	-	aspartyl/glutamyl-tRNA amidotransferase subunit B	99%	0.00E+00	63%	
CSTER	1591	1489463	1490929	-	aspartyl/glutamyl-tRNA amidotransferase subunit A	99%	0.00E+00	59%	
CSTER	1592	1490929	1491231	-	aspartyl/glutamyl-tRNA amidotransferase subunit C	97%	1.00E-18	41%	
CSTER	1615	1511378	1512298	-	Thioredoxin reductase	95%	3.00E-126	58%	Thioredoxin reductase trxb
CSTER	1665	1554189	1555211	-	phospho-N-acetylmuramoyl-pentapeptide-transferase	92%	1.00E-66	41%	Phospho-N-acetylmuramoyl-pentapeptide-transferase mray
CSTER	1666	1555213	1557480	-	put. penicillin-binding protein 2X	90%	9.00E-88	32%	Penicillin-binding protein 2B
CSTER	1667	1557484	1557804	-	cell division prot. FtsL				
CSTER	1701	1591284	1592387	-	Alanine racemase	93%	1.00E-98	43%	Alanine racemase 1 alr
CSTER	1702	1592408	1592767	-	put. 4'-phosphopantetheinyl transferase	97%	6.00E-29	41%	Holo-[acyl-carrier-protein] synthase
CSTER	1705	1594955	1597504	-	preprotein translocase subunit SecA	100%	0.00E+00	55%	Protein translocase subunit SecA
CSTER	1726	1616337	1616576	-	30S ribosomal prot. S18	100%	5.00E-32	65%	

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1727	1616618	1617136	-	single-strand DNA-binding prot.	93%	1.00E-41	56%	Single-stranded DNA-binding protein B
CSTER	1728	1617148	1617438	-	30S ribosomal prot. S6	96%	4.00E-40	61%	
CSTER	1745	1633808	1634821	-	put. O-sialoglycoprotein endopeptidase	97%	6.00E-130	55%	tRNA N6-adenosine threonylcarbamoyltransferase tsad
CSTER	1747	1635236	1635922	-	put. glycoprotein endopeptidase	100%	6.00E-44	38%	tRNA threonylcarbamoyladenosine biosynthesis protein TsaB
CSTER	1749	1636421	1638103	+	mRNA degradation ribonucleases J1/J2	99%	0.00E+00	62%	Ribonuclease J1 yqkc
CSTER	1755	1642501	1643700	-	phosphoglycerate kinase	100%	2.00E-124	50%	Phosphoglycerate kinase pgk
CSTER	1762	1648240	1650321	-	translation elongation factor G	100%	0.00E+00	77%	
CSTER	1763	1650542	1651012	-	30S ribosomal prot. S7	100%	1.00E-86	75%	
CSTER	1764	1651031	1651444	-	30S ribosomal prot. S12	98%	3.00E-80	86%	
CSTER	1770	1656078	1656950	-	ribosome-associated GTPase	99%	5.00E-105	51%	Putative ribosome biogenesis GTPase RsgA
CSTER	1776	1660099	1660413	-	put. thioredoxin	92%	1.00E-37	61%	Thioredoxin trxa
CSTER	1787	1670058	1670192	-	50S ribosomal prot. L34	100%	1.00E-15	70%	
CSTER	1790	1673060	1673395	-	ribonuclease P	95%	1.00E-31	47%	Ribonuclease P protein component
CSTER	1793	1677296	1678750	-	Glutamyl-tRNA synthetases	100%	0.00E+00	55%	
CSTER	1797	1681106	1681795	-	50S ribosomal prot. L1	98%	4.00E-99	65%	50S ribosomal protein L1
CSTER	1809	1690036	1691058	+	Glycerol-3-phosphate dehydrogenase	94%	4.00E-122	53%	Glycerol-3-phosphate dehydrogenase [NAD(P)+] gspA
CSTER	1813	1693298	1694431	-	Metal-dependent amidase/aminoacylase/ carboxypeptidase	95%	6.00E-120	47%	N-acetyldiaminopimelate deacetylase ykur
CSTER	1814	1694510	1695208	-	put. 2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase	98%	4.00E-86	60%	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase dapH
CSTER	1821	1699209	1699601	-	single-strand binding prot.	93%	1.00E-25	40%	Single-stranded DNA-binding protein B
CSTER	1844	1713286	1716924	-	DNA-directed RNA polymerase subunit beta'	97%	0.00E+00	68%	DNA-directed RNA polymerase subunit beta
CSTER	1845	1717025	1720606	-	DNA-directed RNA polymerase subunit beta	98%	0.00E+00	71%	DNA-directed RNA polymerase subunit beta'

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1847	1723493	1724749	+	Tyrosyl-tRNA synthetase	99%	0.00E+00	58%	
CSTER	1876	1751169	1752050	-	fructose-bisphosphate aldolase	100%	4.00E-83	45%	Probable fructose-bisphosphate aldolase fbaa
CSTER	1880	1755102	1755488	-	50S ribosomal prot. L17	100%	2.00E-62	73%	
CSTER	1881	1755506	1756444	-	DNA-directed RNA polymerase alpha subunit	99%	7.00E-139	62%	DNA-directed RNA polymerase subunit alpha
CSTER	1882	1756493	1756876	-	30S ribosomal prot. S11	93%	7.00E-69	87%	
CSTER	1883	1756904	1757269	-	30S ribosomal prot. S13	100%	2.00E-58	73%	
CSTER	1884	1757290	1757403	-	50S ribosomal prot. L36	100%	4.00E-16	84%	
CSTER	1885	1757429	1757647	-	Translation initiation factor 1 (IF-1)	100%	1.00E-38	78%	
CSTER	1886	1757765	1758421	-	Adenylate kinase	99%	3.00E-89	59%	Adenylate kinase adk
CSTER	1887	1758553	1759848	-	preprot. translocase subunit SecY	99%	4.00E-140	49%	Protein translocase subunit SecY
CSTER	1888	1759865	1760305	-	50S ribosomal prot. L15	100%	1.00E-70	72%	
CSTER	1889	1760433	1760615	-	50S ribosomal prot. L30	98%	3.00E-20	62%	
CSTER	1890	1760630	1761124	-	30S ribosomal prot. S5	93%	8.00E-70	76%	
CSTER	1891	1761143	1761499	-	50S ribosomal prot. L18	100%	4.00E-55	73%	
CSTER	1892	1761589	1762125	-	50S ribosomal prot. L6	100%	3.00E-69	60%	50S ribosomal protein L6
CSTER	1893	1762252	1762650	-	30S ribosomal prot. S8	100%	1.00E-73	77%	
CSTER	1894	1762773	1762958	-	30S ribosomal prot. S14	100%	1.00E-31	77%	
CSTER	1895	1762976	1763518	-	50S ribosomal prot. L5	99%	3.00E-100	77%	
CSTER	1896	1763545	1763850	-	50S ribosomal prot. L24	99%	7.00E-38	64%	
CSTER	1897	1763931	1764299	-	50S ribosomal prot. L14	100%	5.00E-74	87%	
CSTER	1898	1764324	1764584	-	30S ribosomal prot. S17	97%	4.00E-46	84%	
CSTER	1899	1764612	1764818	-	50S ribosomal prot. L29	78%	1.00E-14	58%	
CSTER	1900	1764828	1765241	-	50S ribosomal prot. L16	93%	1.00E-76	83%	
CSTER	1901	1765245	1765898	-	30S ribosomal prot. S3	100%	1.00E-119	75%	
CSTER	1902	1765911	1766255	-	50S ribosomal prot. L22	96%	7.00E-49	64%	
CSTER	1903	1766271	1766549	-	30S ribosomal prot. S19	100%	2.00E-54	83%	
CSTER	1904	1766643	1767476	-	50S ribosomal prot. L2	100%	5.00E-159	76%	50S ribosomal protein L2
CSTER	1905	1767494	1767790	-	50S ribosomal prot. L23	94%	2.00E-31	60%	
CSTER	1906	1767790	1768413	-	50S ribosomal prot. L4	100%	9.00E-92	60%	50S ribosomal protein L4
CSTER	1907	1768438	1769064	-	50S ribosomal prot. L3	99%	4.00E-115	75%	50S ribosomal protein L3
CSTER	1908	1769181	1769489	-	30S ribosomal prot. S10	100%	1.00E-58	78%	

Genome_part	STER	start	stop	direction		query cover	e-value	aa id	Bacillus subtilis annotation
CSTER	1936	1795296	1795484	+	50S ribosomal prot. L28	100%	7.00E-26	69%	
CSTER	1948	1805428	1807179	-	Aspartyl-tRNA synthetase	98%	0.00E+00	57%	
CSTER	1950	1807719	1808999	-	Histidyl-tRNA synthetase	99%	1.00E-140	49%	
CSTER	1953	1810651	1810833	+	50S ribosomal prot. L32	91%	1.00E-10	48%	
CSTER	1954	1810849	1810998	+	50S ribosomal prot. L33	100%	1.00E-14	55%	
CSTER	1973	1823718	1824329	-	30S ribosomal prot. S4	100%	1.00E-95	69%	
CSTER	1975	1824852	1826213	-	replicative DNA helicase	95%	8.00E-176	56%	Replicative DNA helicase DnaC
CSTER	1976	1826257	1826715	-	50S ribosomal prot. L9	100%	2.00E-39	55%	
CSTER	1979	1830931	1832052	-	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	98%	9.00E-177	70%	tRNA-specific 2-thiouridylase MnmA
CSTER	1986	1838183	1838725	-	Phosphatidylglycerophosphate synthase	93%	2.00E-45	52%	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase pgsA
CSTER	1992	1843910	1845391	-	inositol-5-monophosphate dehydrogenase	99%	0.00E+00	68%	Inosine-5'-monophosphate dehydrogenase guab
CSTER	1993	1845568	1846590	-	Tryptophanyl-tRNA synthetase II	98%	8.00E-42	35%	

Table 5. List of differentially expressed genes identified in deletion strain GEI4.

Operon	Putative pathway/function	Regulator-effector	Gene	log_2 change	*p*-value
STER_0390-STER-0393	Cysteine metabolism	Cmbr/Homr/Mtar	STER_0390	-1.1	2.7E-23
			STER_0391	-1.3	2.2E-22
			STER_0392	-1.4	1.2E-32
			STER_0393	-1.2	2.8E-15
STER_1016-STER_1017	Maltodextrin metabolism		STER_1016	-1.0	0
			STER_1017	-0.8	7.50E-38
STER_1548-STER_1555	Aromatic amino acid biosynthesis	T-box RNA (Trp)	STER_1548	-1.3	2.3E-37
			STER_1549	-1.0	1.5E-26
			STER_1550	-1.3	7.9E-38
			STER_1551	-1.3	1.2E-35
			STER_1552	-1.3	0
			STER_1553	-1.4	0
			STER_1554	-1.3	0
			STER_1555	-1.6	7.8E-73
STER_1960-STER_1963	Membrane proteins		STER_1960	-0.9	0
			STER_1961	-1.0	0
			STER_1962	-1.0	8.3E-317
			STER_1963	-1.1	3.2E-965
STER_0049-STER_0054	Purine biosynthesis	PurR	STER_0049	1.2	5.5E-34
			STER_0050	1.1	1.4E-29
			STER_0051	1.1	0
			STER_0052	1.1	0
			STER_0053	1.1	0
			STER_0054	1.1	6.6E-40
STER_0699-STER_0701	Ethanolamine metabolism		STER_0699	1.6	0
			STER_0700	2.1	9.5E-200
			STER_0701	2.2	7.8E-329
STER_1020-STER_1024	Twin arginine translocase	TatA	STER_1020	1.3	5.3E-14
			STER_1021	1.3	8.8E-12
			STER_1022	1.4	3.3E-12
			STER_1023	1.0	9.6E-6
			STER_1024	0.9	0.0022
STER_1025-STER_1028	Iron homeostasis	PerR	STER_1025	1.1	1.9E-09
			STER_1026	0.9	2.5E-6
			STER_1027	1.2	3E-12
			STER_1028	1.2	1.5E-17
STER_1405-STER_1409	ABC Peptide Transport	Cody	STER_1405	1.6	9.5E-90
			STER_1406	1.6	0
			STER_1407	1.6	3.7E-59
			STER_1408	1.5	0
			STER_1409	1.4	4.2E-54
STER_1821-STER_1823	Stress		STER_1821	2.0	0
			STER_1822	2.2	5.4E-98
			STER_1823	2.0	0

References

1. Darmon E, Leach DF. (2014) Bacterial Genome Instability. *Microbiol. Mol. Biol.* Rev. 78, 1–39.
2. Labrie SJ, Samson JE, and Moineau S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327.
3. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* 54(2):234-244.
4. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–1712.
5. Brouns SJJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960–964.
6. Young JC et al. (2012) Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS ONE* 7(5):e38077.
7. Garneau JE, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67–71.
8. Groenen PM, Bunschoten AE, Van Soolingen D, Van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 10(5):1057–1065.
9. Yin S, et al. (2013) The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl Environ Microbiol* 79(18):5710–5720.
10. Liu F, et al. (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* 77(6):1946–1956.
11. Barrangou R, Horvath P (2012) CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 3, 143–162.
12. Sander JD, and Joung JK. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–355.

13. Bikard D, et al. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 41(15):7429-7437

14. Qi LS, et al. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152(5):1173–1183.

15. Gomaa AA, et al. (2014) Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems. *mBio* 5(1):e00928–13.

16. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169(12):5429–5433.

17. Jansen R, Embden JDA van, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43(6):1565–1575.

18. Bolotin A, et al. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22(12):1554–1558.

19. Horvath P, et al. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1401–1412.

20. Bolotin A, Quinguis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(Pt 8):2551-2561.

21. Deveau H, et al. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1390–1400.

22. Paez-Espino D, et al. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4, 1430.

23. Sun CL, et al. (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* 15(2):463–470.

24. Sapranauskas R, et al. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39(21):9275–9282.

25. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109(39):E2579–2586.

26. Briner AE, et al. (2014) Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Molecular Cell.* 56(2):333-339

27. Bondy-Denomy J, Davidson AR. (2014) To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* 22, 218–225.

28. Horvath P, et al. (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 131(1):62-70.

29. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31(3):233–239.

30. Vercoe RB, et al. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* 9(4):e1003454.

31. Oh JH, van Pijkeren JP. (2014) CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res* 10.1093/nar/gku623

32. Selle K, Barrangou R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol.* In press

33. Kobayashi K, et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci.* U.S.A. 100, 4678–4683.

34. Mahillon J, Chandler M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62(3):725–774.

35. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R (2011) Specialized adaptation of a lactic acid bacterium to the milk environment; the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microbial Cell Factories* 10 (Suppl 1):S22

36. Dandoy D, et al. (2011) The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. *Microb. Cell Fact.* 10 Suppl 1, S21.

37. Deltcheva E, et al. (2011) CRISPR RNA maturation by trans-activating small RNA and host factor RNase III. *Nature* 471, 602–607.

38. Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11(8):1365–1374.

39. Koonin EV, Makarova KS (2007) Evolutionary genomics of lactic acid bacteria. *J Bacteriol* 189(4):1199-1208

40. Makarova KS, et al. (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* 103(42):15611–15616.

41. Goh YJ, et al. (2009) Development and application of a upp-based counterselective gene replacement system for the study of the S-layer protein SlpX of *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 75(10):3093–3105.

42. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 77(1):61–68.

43. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3):733–740.

44. Russell WM, Klaenhammer TR (2001) Efficient System for Directed Integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* Chromosomes via Homologous Recombination. *Appl Environ Microbiol* 67(9):4361–4364.

45. Wei M-Q, et al. (1995) An improved method for the transformation of *Lactobacillus* strains using electroporation. *Journal of Microbiological Methods* 21(1):97–109.

46. Zhang X, Bremer H (1995) Control of the *Escherichia coli* rrnB p1 promoter strength by ppGpp. *Journal of Biological Chemistry* 270(19):11181-11189.

47. Law J, et al. (1995) A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* 177(24):7011–7018.

The foregoing is illustrative of the present invention, and is not to be construed as limiting thereof. The invention is defined by the following claims, with equivalents of the claims to be included therein.

THAT WHICH IS CLAIMED IS:

1. A method of screening a population of bacterial cells for essential genes, non-essential genes, and/or expendable genomic islands, comprising

introducing into said population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby producing a population of transformed bacterial cells; and

determining the presence or absence of a deletion in the population of transformed bacterial cells, wherein the presence of a deletion in the population of transformed bacterial cells indicates that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population means that the target region is comprised within an essential gene.

2. A method of screening a population of bacterial, archaeal, algal or yeast cells for essential genes, non-essential genes, and/or expendable genomic islands, comprising

introducing into the population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells; and

determining the presence or absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells, wherein the presence of a deletion in the population of transformed bacterial, archaeal or yeast cells means that the target region is comprised within a non-essential gene and/or an expendable genomic island, and the absence of a deletion in the population of transformed bacterial, archaeal, algal or yeast cells means that the target region is comprised within an essential gene.

3. A method of killing one or more bacterial cells within a population of bacterial cells, comprising:

introducing into the population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array (crRNA, crDNA) comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population, thereby killing one or more bacterial cells that comprise the target region within the population of bacterial cells.

4. A method of killing one or more cells within a population of bacterial, archaeal, algal or yeast cells, comprising

introducing into the population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing one or more cells within a population of bacterial, archaeal, algal or yeast cells that comprise the target region in their genome.

5. The method of claim 3 or claim 4, the target region is within an essential gene or a non-essential gene.

6. A method of identifying a phenotype associated with a bacterial gene, comprising:

introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of the bacterial cells of said population,

wherein the target region comprises at least a portion of an open reading frame encoding a polypeptide or functional nucleic acid, thereby killing the cells comprising the target region and producing a population of transformed bacterial cells without the target region; and

analyzing the phenotype of the population.

7. The method of claim 6, comprising prior to analyzing, growing individual bacterial colonies from the population of transformed bacterial cells; and

analyzing the phenotype of the individual colonies.

8 A method of identifying a phenotype of a bacterial, archaeal, algal or yeast gene, comprising:

introducing into a population of bacterial, archaeal, algal or yeast cells (a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and (c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, thereby killing the bacterial, archaeal or yeast cells comprising the target region and producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and

analyzing the phenotype of the population of transformed bacterial, archaeal, algal or yeast cells, and/or (i) growing individual bacterial, archaeal, algal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells and (ii) analyzing the phenotype of the individual colonies.

9. The method of any one of claims 1, 3, 6, or 7, wherein the CRISPR array is a Type I, Type II, Type III, Type IV, Type V CRISPR array.

10. The method of claim 9, wherein the repeat-spacer-repeat sequence or the at least one repeat-spacer sequence comprises a repeat that is identical to a repeat from a wild-type Type I CRISPR array, a wild type Type II CRISPR array, a wild type Type III CRISPR array, a wild type Type IV CRISPR array, or a wild type Type V CRISPR array.

11. The method of any one of claims 2, 4, 5 or 8, wherein the repeat-spacer-repeat sequence or the at least one repeat-spacer sequence comprises a repeat that is identical to a repeat from a wild-type Type II CRISPR array.

12. The method of any one of claims 1 to 11 wherein said target region is randomly selected or is specifically selected.

13. The method of claim 12, wherein a randomly selected target region is selected from any at least 10 consecutive nucleotides located adjacent to a PAM sequence in a bacterial, archaeal or yeast genome.

14. The method of claim 12, wherein a specifically selected target region is selected from a gene, open reading frame or a putative open reading frame comprising at least 10 consecutive nucleotides adjacent to a PAM sequence in a bacterial, archaeal, algal or yeast genome.

15. The method of any one of claims 2, 4, 5, 8, 11 to 14, wherein the heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid and the heterologous nucleic acid construct comprising a CRISPR array are comprised in a CRISPR guide (gRNA, gDNA) that optionally further comprises a heterologous nucleic acid construct comprising a polynucleotide encoding Cas9 polypeptide.

16. The method of claim 15, wherein the CRISPR guide is operably linked to a promoter.

17. A method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells, comprising:

introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the cells comprising the target region are killed, thereby selecting one or

more bacterial cells without the target region and having a reduced genome size from the population of bacterial cells.

18. A method of selecting one or more bacterial cells having a reduced genome size from a population of bacterial cells, comprising:

introducing into a population of bacterial cells:

(a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and

(b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby selecting one or more bacterial cells without the target region and having a reduced genome size from the population of transgenic bacterial cells.

19. A method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced the genome size from a population of bacterial, archaeal, algal or yeast cells, comprising:

introducing into a population of bacterial, archaeal, algal or yeast cells

(a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid,

(b) a heterologous nucleic acid construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and the at least one repeat-spacer-repeat sequence

comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and

(c) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of bacterial, archaeal, algal or yeast cells.

20. A method of selecting one or more bacterial, archaeal, algal or yeast cells having a reduced the genome size from a population of bacterial, archaeal or yeast cells, comprising:

introducing into a population of bacterial, archaeal, algal or yeast cells:

(a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and

(b)(i) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array comprising a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of the at least one repeat-spacer sequence and the at least one repeat-spacer-repeat sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of one or more bacterial, archaeal, algal or yeast cells of said population, and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby selecting one or more bacterial, archaeal, algal or yeast cells without the target region and having a reduced genome size from the population of transgenic bacterial, archaeal, algal or yeast cells.

21. A method of identifying in a population of bacteria at least one isolate having a deletion in its genome, comprising:

introducing into a population of bacterial cells a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein cells comprising the target region are killed, thereby producing a population of transformed bacterial cells without the target region; and

growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of transformed bacteria having a deletion in its genome.

22. A method of identifying in a population of bacteria at least one isolate having a deletion in its genome, comprising:

introducing into the population of bacterial cells

(a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial cells or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and

b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome of one or more bacterial cells of said population, wherein the target region is located between the one or more heterologous nucleic acid constructs introduced into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby producing a population of transformed bacterial cells without the target region; and

growing individual bacterial colonies from the population of transformed bacterial cells, thereby identifying at least one isolate from the population of bacteria having a deletion in its genome.

23. A method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome, comprising:

introducing into a population of bacterial, archaeal, algal or yeast cells:

(a) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid,

(b) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of the bacterial, archaeal, algal or yeast cells of said population, and

(c) a Cas9 polypeptide or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein cells comprising the target region are killed, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and

growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population of transformed bacterial, archaeal, algal or yeast cells having a deletion in its genome.

24. A method of identifying in a population of bacterial, archaeal, algal or yeast cells at least one isolate having a deletion in its genome, comprising:

introducing into the population of bacterial, archaeal, algal or yeast cells

(a)(i) one or more heterologous nucleic acid constructs comprising a nucleotide sequence having at least 80 percent identity to at least 300 consecutive nucleotides present in the genome of said bacterial, archaeal, algal or yeast cells, or (ii) two or more heterologous nucleic acid constructs comprising at least one transposon, thereby producing a population of transgenic bacterial, archaeal, algal or yeast cells comprising a non-natural site for homologous recombination between the one or more heterologous nucleic acid constructs

integrated into the genome and the at least 300 consecutive nucleotides present in the genome, or between a first and a second transposon integrated into the genome; and

(b)(i) a heterologous nucleic acid construct comprising a trans-encoded CRISPR (tracr) nucleic acid, (ii) a heterologous nucleic acid construct comprising a CRISPR array comprising (5' to 3') a repeat-spacer-repeat sequence or at least one repeat-spacer sequence, wherein the spacer of said repeat-spacer-repeat sequence or said at least one repeat-spacer sequence comprises a nucleotide sequence that is substantially complementary to a target region in the genome (chromosomal and/or plasmid) of one or more bacterial, archaeal, algal or yeast cells of said population, and (iii) a Cas9 polypeptide and/or a heterologous nucleic acid construct comprising a polynucleotide encoding a Cas9 polypeptide, wherein the target region is located between the one or more heterologous nucleic acid constructs incorporated into the genome and the at least 300 consecutive nucleotides present in the genome and/or between the first transposon and second transposon, and cells comprising the target region are killed, thereby producing a population of transformed bacterial, archaeal, algal or yeast cells without the target region; and

growing individual bacterial, archaeal or yeast colonies from the population of transformed bacterial, archaeal, algal or yeast cells, thereby identifying at least one isolate from the population having a deletion in its genome.

25. The method of any one of claims 17, 18, 21, or 22, wherein the CRISPR array is a Type I, Type II, Type III, Type IV, Type V CRISPR array.

26. The method of claim 25, wherein the repeat-spacer-repeat sequence or the at least one repeat-spacer sequence comprises a repeat that is identical to a repeat from a wild-type Type I CRISPR array, a wild type Type II CRISPR array, a wild type Type III CRISPR array, a wild type Type IV CRISPR array, or a wild type Type V CRISPR array.

27. The method of any one of claims 19, 20, 23, or 24, wherein the repeat-spacer-repeat sequence or the at least one repeat-spacer sequence comprises a repeat that is identical to a repeat from a wild-type Type II CRISPR array.

28. The method of any one of claims 15 to 27 wherein said target region is randomly selected or is specifically selected.

29. The method of claim 28, wherein a randomly selected target region is selected from any at least 10 consecutive nucleotides located adjacent to a PAM sequence in a bacterial, archaeal or yeast genome.

30. The method of claim 28, wherein a specifically selected target region is selected from a gene, open reading frame or a putative open reading frame comprising at least 10 consecutive nucleotides adjacent to a PAM sequence in a bacterial, archaeal, algal or yeast genome.

31. The method of any one of claims 3 to 5 or 9 to 15, wherein the introduced CRISPR array is compatible with a CRISPR-Cas system in the one or more bacterial cells to be killed that is not compatible with the CRISPR Cas system of at least one or more bacterial cells in the population of bacterial cells.

Fig. 1

CRISPR1

	Spacer	Repeat
LacZ	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC	AATCCCCATAGCACGAACAATATAGTTGTT	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC

CRISPR3

Repeat	Spacer	Repeat
LacZ	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	AGTATTTTGAATCTCTTGAAGAATTTTCTGA	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC
ABC	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	GATAACACGAGATAAAACATCCAGCCCACC	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC
prtS	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	GTTGTAGCTTTGAGGTCTGAGAATACACGC	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC
Cu	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC	GATTGCTCAATCAATCGTTTCAGCTGCTAA	GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC

Fig. 2

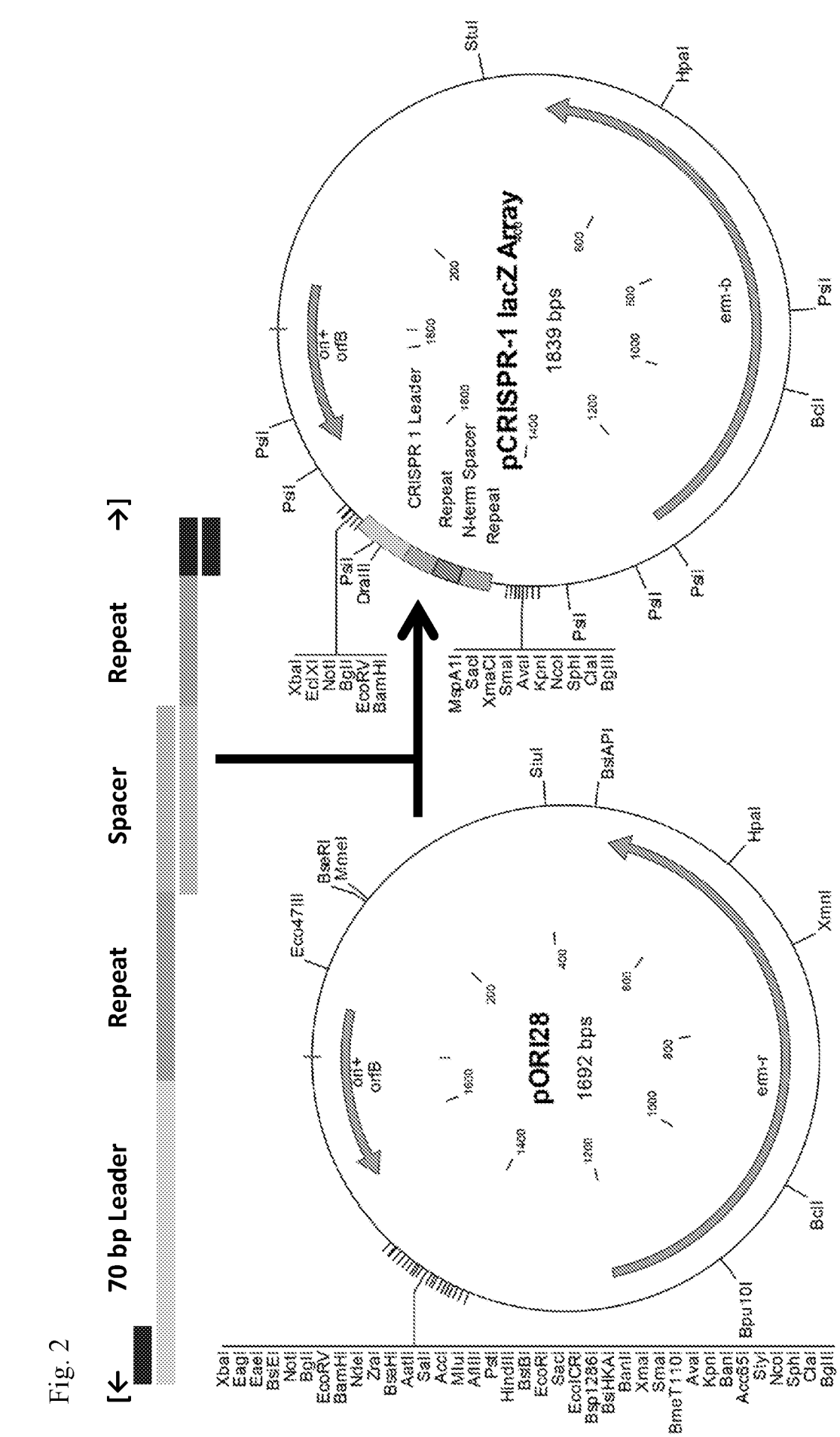

[←
70 bp Leader
Repeat
Spacer
Repeat
→]
pCRISPR-1 lacZ Array
1839 bps
CRISPR 1 Leader
Repeat
N-term Spacer
Repeat
ori+
orfB
em-b
pORI28
1692 bps
em-r
Xbal
EclXI
NotI
BglI
EcoRV
BamHI
PsiI
DraIII
MspA1I
SacI
XmaCI
SmaI
AvaI
KpnI
NcoI
SphI
ClaI
BglII
StuI
HpaI
BclI
Eco47III
BseRI
MmeI
BstAPI
XmnI
Bpu10I

Fig. 3A

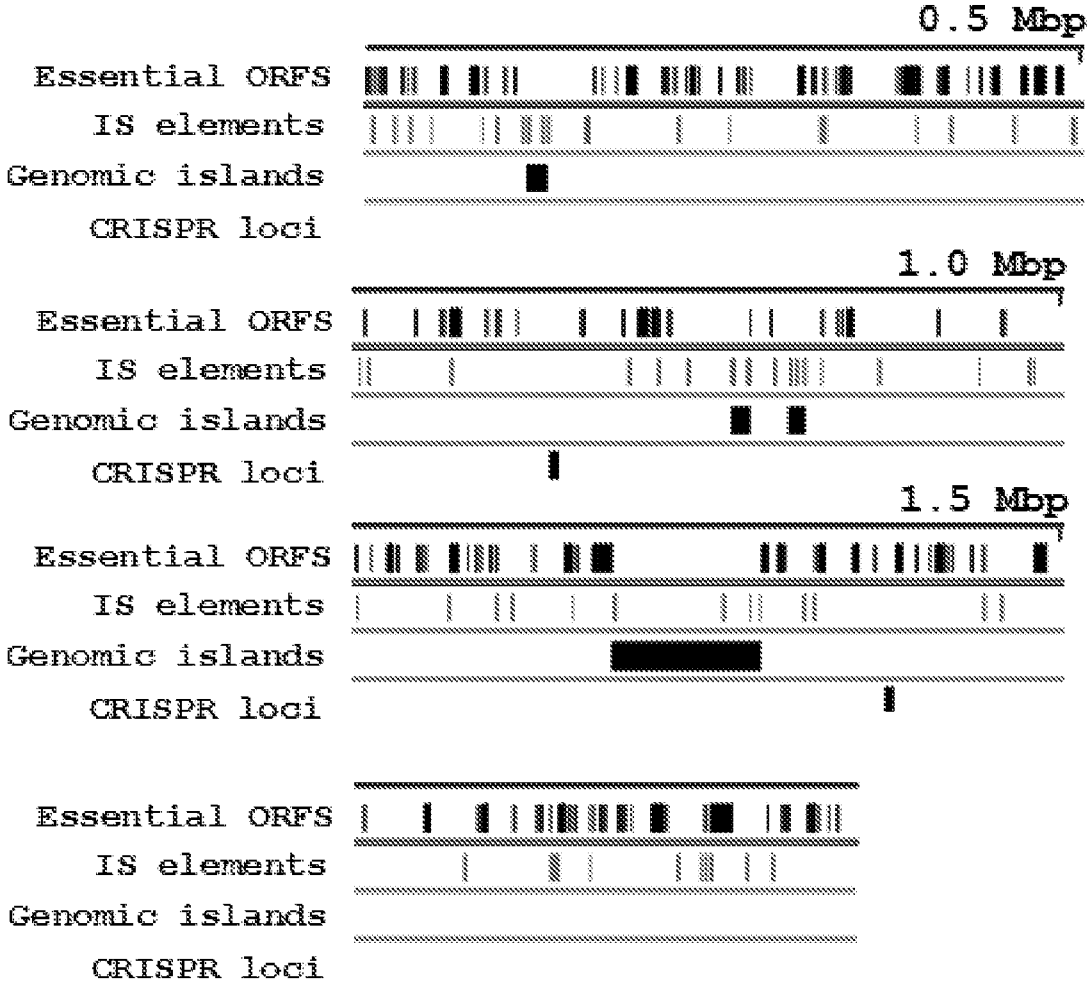

0.5 Mbp
Essential ORFS
IS elements
Genomic islands
CRISPR loci
1.0 Mbp
Essential ORFS
IS elements
Genomic islands
CRISPR loci
1.5 Mbp
Essential ORFS
IS elements
Genomic islands
CRISPR loci
Essential ORFS
IS elements
Genomic islands
CRISPR loci

Fig. 3B

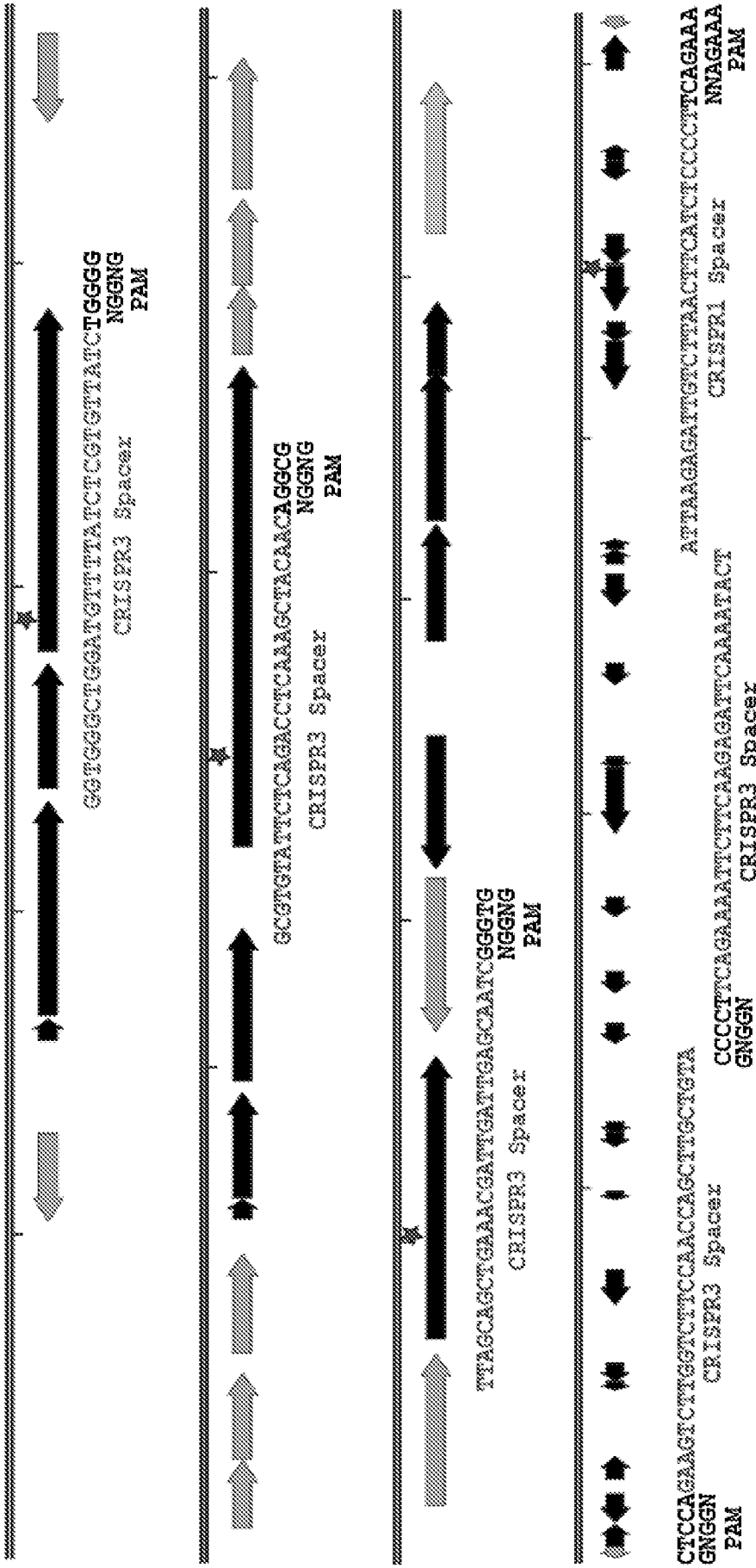

GGTGGGCTGGATGTTTTATCTCGTGTTATCTGGG
NGGNG
PAM
CRISPR3 Spacer
GCGTGTATTCTCAGACCTCAAAGCTACAACAGGCG
NGGNG
PAM
CRISPR3 Spacer
TTAGCAGCTGAAACGATTGATTGAGCAATCGGGTG
NGGNG
PAM
CRISPR3 Spacer
CTCCAGAAGTCTTGGTCTTCCAACCAGCTTGCTGTA
GNGGN
PAM
CRISPR3 Spacer
CCCCTTCAGAAAATTCTTCAAGAGATTCAAAATACT
GNGGN
PAM
CRISPR3 Spacer
ATTAAGAGATTGTCTTTAACTTCATCTCCCTTCAGAAA
NNAGAAA
PAM
CRISPR1 Spacer

Fig. 4

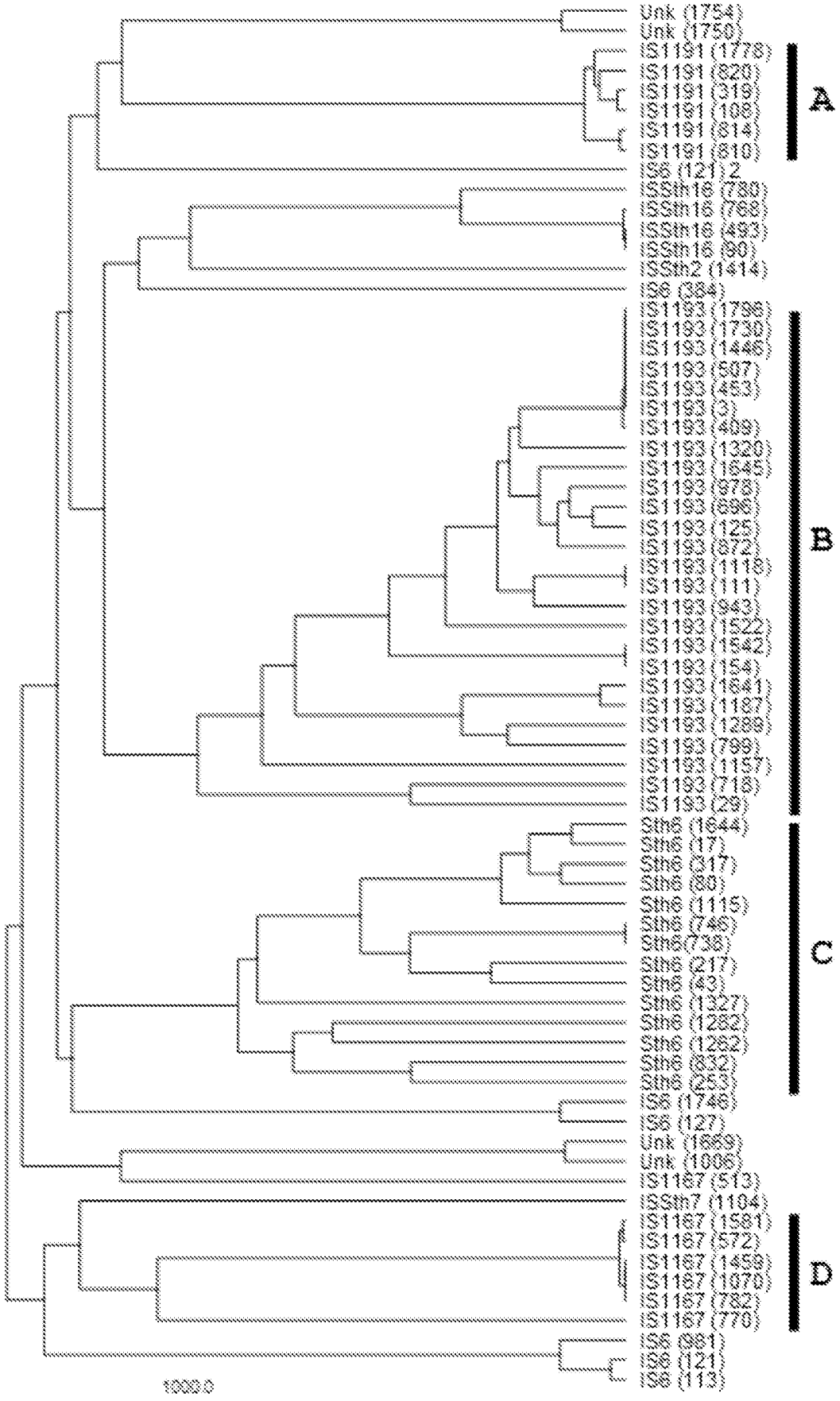

Unk (1754)
Unk (1750)
IS1191 (1778)
IS1191 (820)
IS1191 (319)
IS1191 (108)
IS1191 (814)
IS1191 (810)
IS6 (121).2
ISSth16 (780)
ISSth16 (768)
ISSth16 (493)
ISSth16 (90)
ISSth2 (1414)
IS6 (384)
IS1193 (1796)
IS1193 (1739)
IS1193 (1446)
IS1193 (507)
IS1193 (453)
IS1193 (3)
IS1193 (409)
IS1193 (1320)
IS1193 (1645)
IS1193 (976)
IS1193 (696)
IS1193 (125)
IS1193 (872)
IS1193 (1118)
IS1193 (111)
IS1193 (943)
IS1193 (1522)
IS1193 (1542)
IS1193 (154)
IS1193 (1641)
IS1193 (1187)
IS1193 (1289)
IS1193 (799)
IS1193 (1157)
IS1193 (718)
IS1193 (29)
Sth6 (1644)
Sth6 (17)
Sth6 (317)
Sth6 (80)
Sth6 (1115)
Sth6 (746)
Sth6 (738)
Sth6 (217)
Sth6 (43)
Sth6 (1327)
Sth6 (1282)
Sth6 (1262)
Sth6 (832)
Sth6 (253)
IS6 (1746)
IS6 (127)
Unk (1669)
Unk (1006)
IS1167 (513)
ISSth7 (1104)
IS1167 (1581)
IS1167 (572)
IS1167 (1459)
IS1167 (1070)
IS1167 (782)
IS1167 (770)
IS6 (981)
IS6 (121)
IS6 (113)
1000.0
A
B
C
D

Fig. 5A

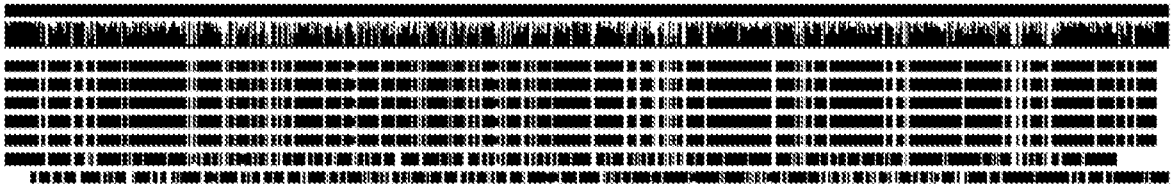


Fig. 5B

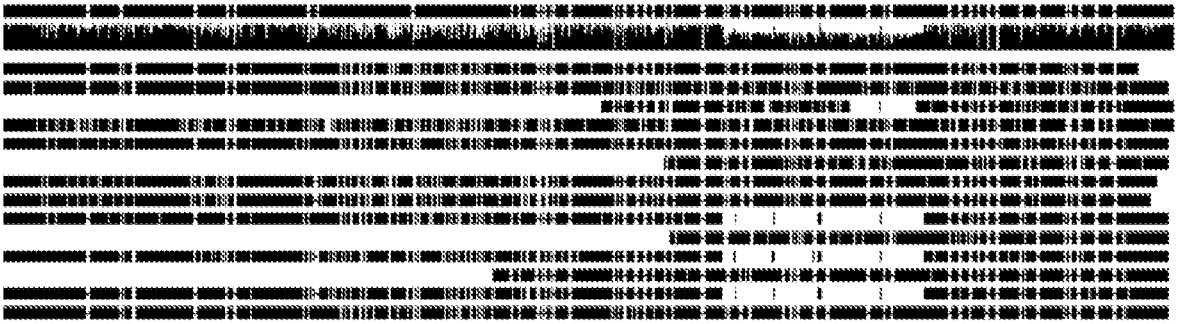


Fig. 5C

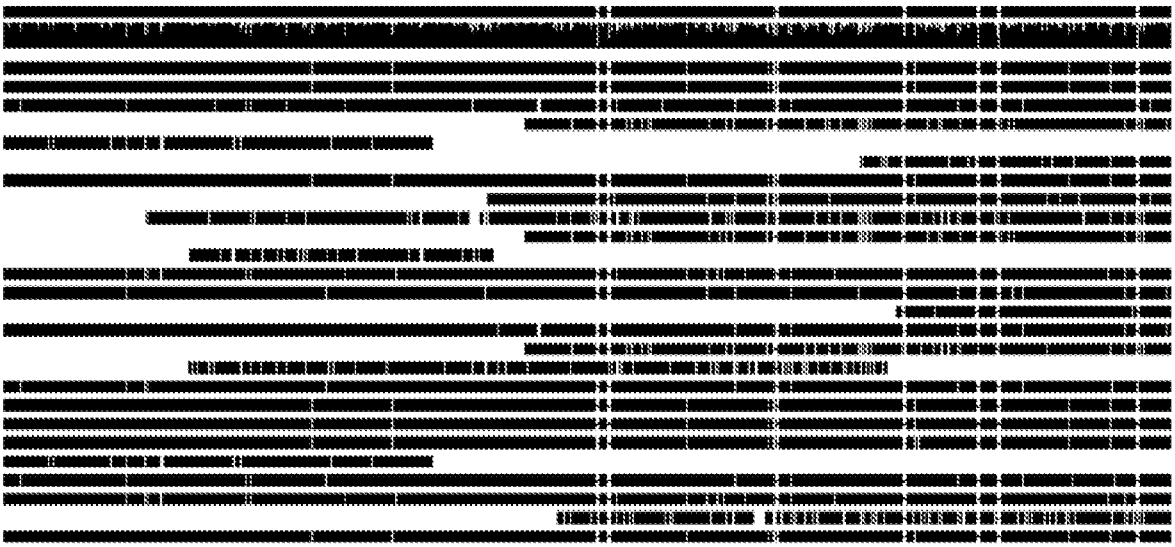


Fig. 5D

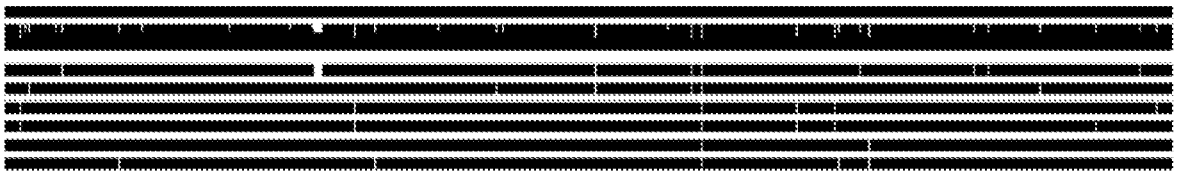


Fig. 6A

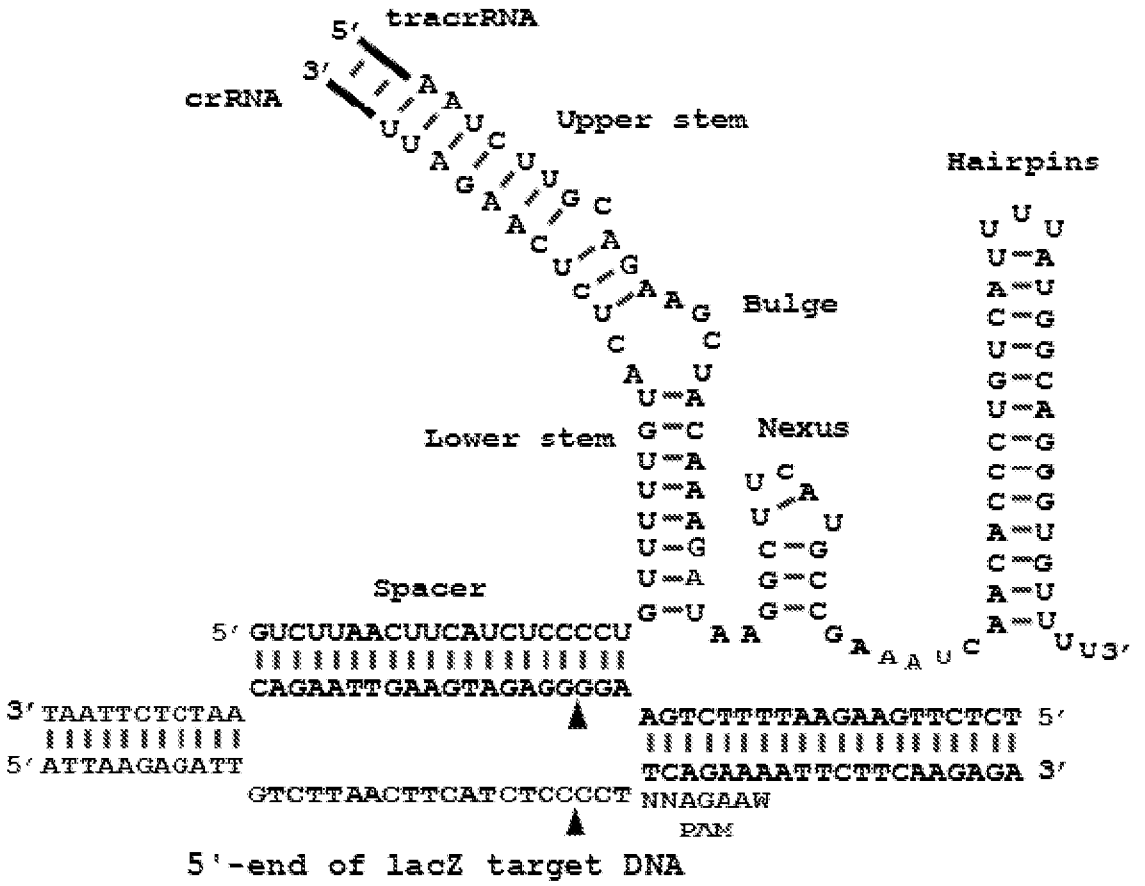

5′ tracrRNA
3′
crRNA
Upper stem
Hairpins
Bulge
Lower stem
Nexus
Spacer
5′ GUCUUAACUUCAUCUCCCCU
CAGAATTGAAGTAGAGGGGA
3′ TAATTCTCTAA
5′ ATTAAGAGATT
AGTCTTTTAAGAAGTTCTCT 5′
TCAGAAAATTCTTCAAGAGA 3′
GTCTTAACTTCATCTCCCCT NNAGAAW
PAM
5'-end of lacZ target DNA

Fig. 6B

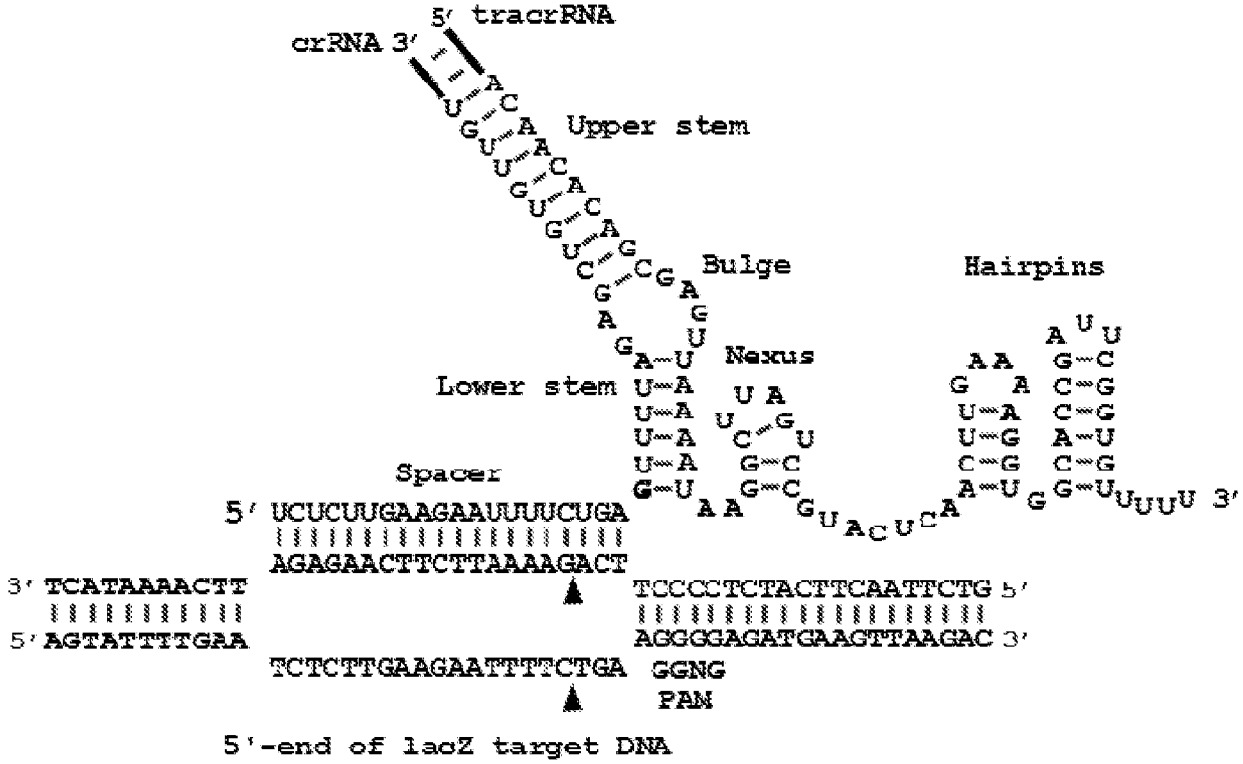
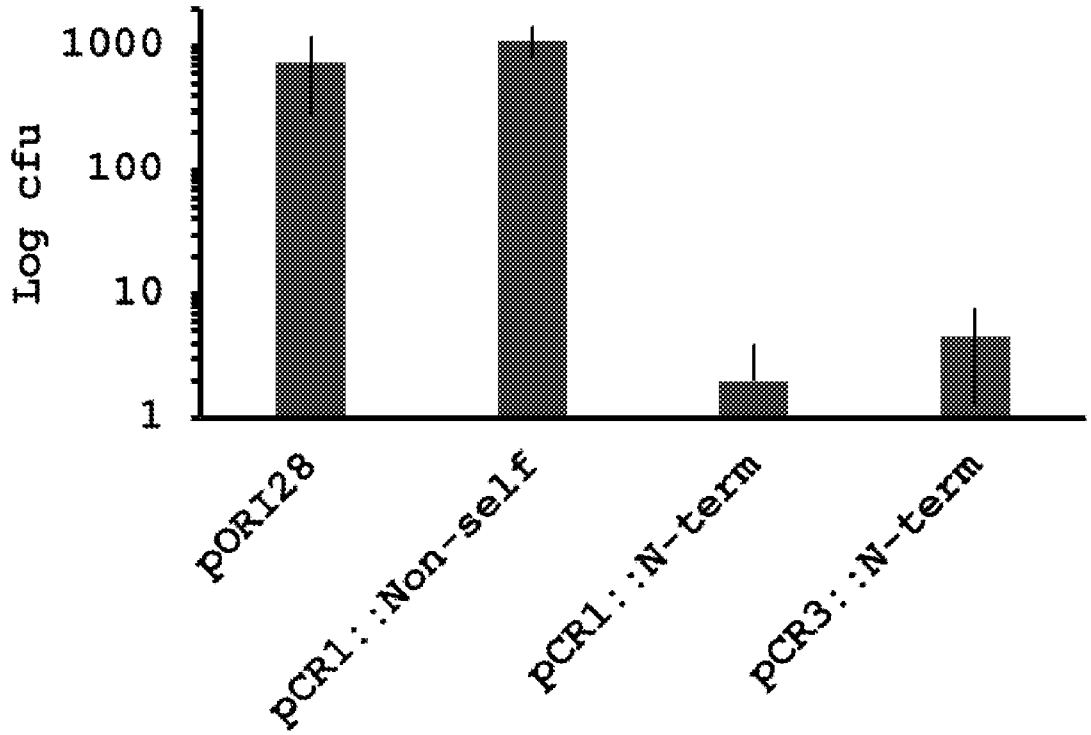

5′ tracrRNA
crRNA 3′
Upper stem
Bulge
Hairpins
Nexus
Lower stem
Spacer
5′ UCUCUUGAAGAAUUUUCUGA
AGAGAACTTCTTAAAAGACT
3′ TCATAAAACTT
5′ AGTATTTTGAA
TCCCCTCTACTTCAATTCTG 5′
AGGGGAGATGAAGTTAAGAC 3′
TCTCTTGAAGAATTTTCTGA GGNG
PAM
5'-end of lacZ target DNA

Fig. 6C


1000
100
10
1
Log cfu
pORI28
pCR1::Non-self
pCR1::N-term
pCR3::N-term

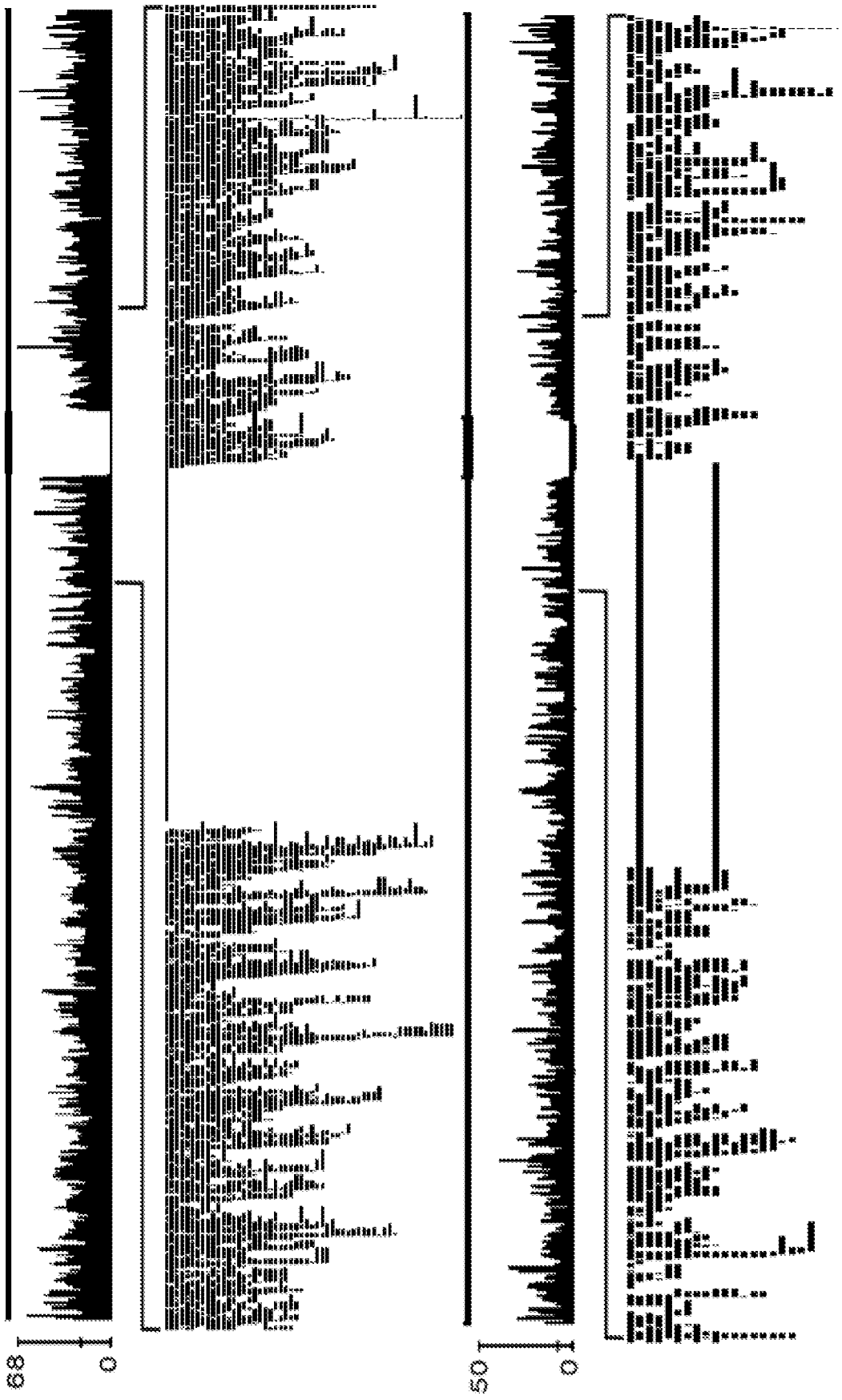
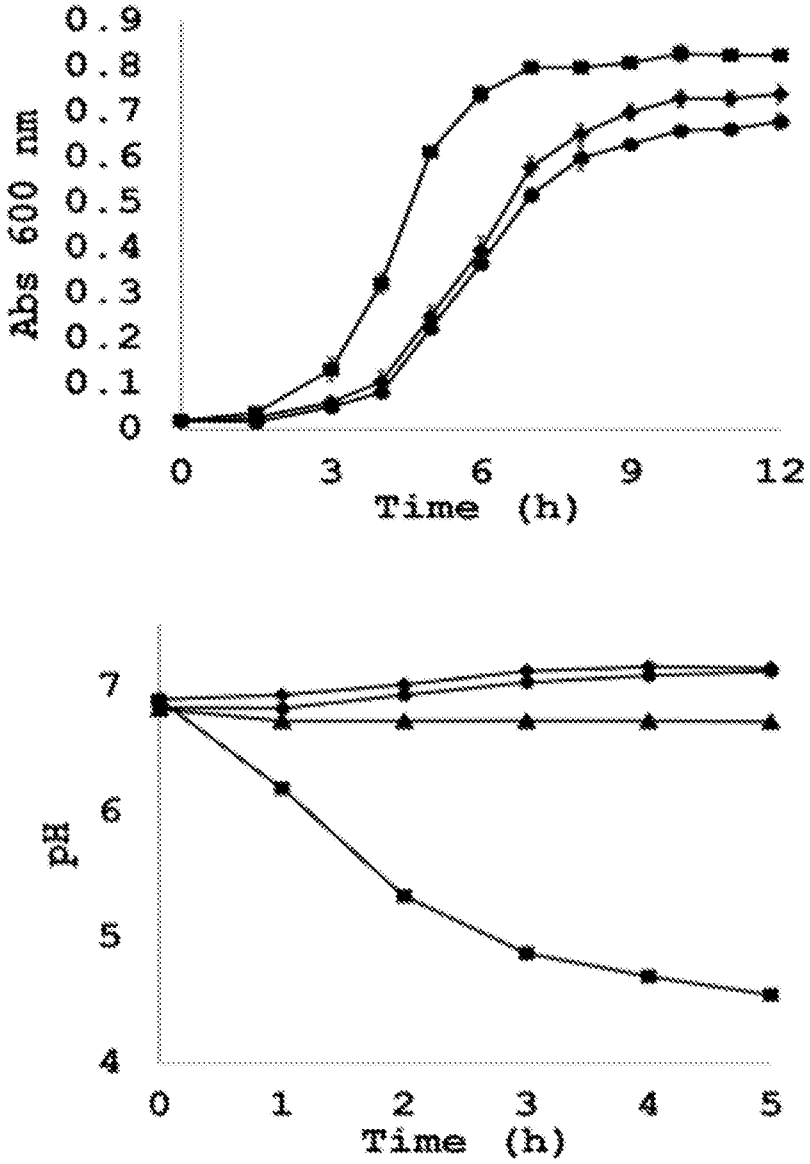


Fig.7A

Fig. 7B


Abs 600 nm
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
3
6
9
12
Time (h)
pH
7
6
5
4
0
1
2
3
4
5
Time (h)

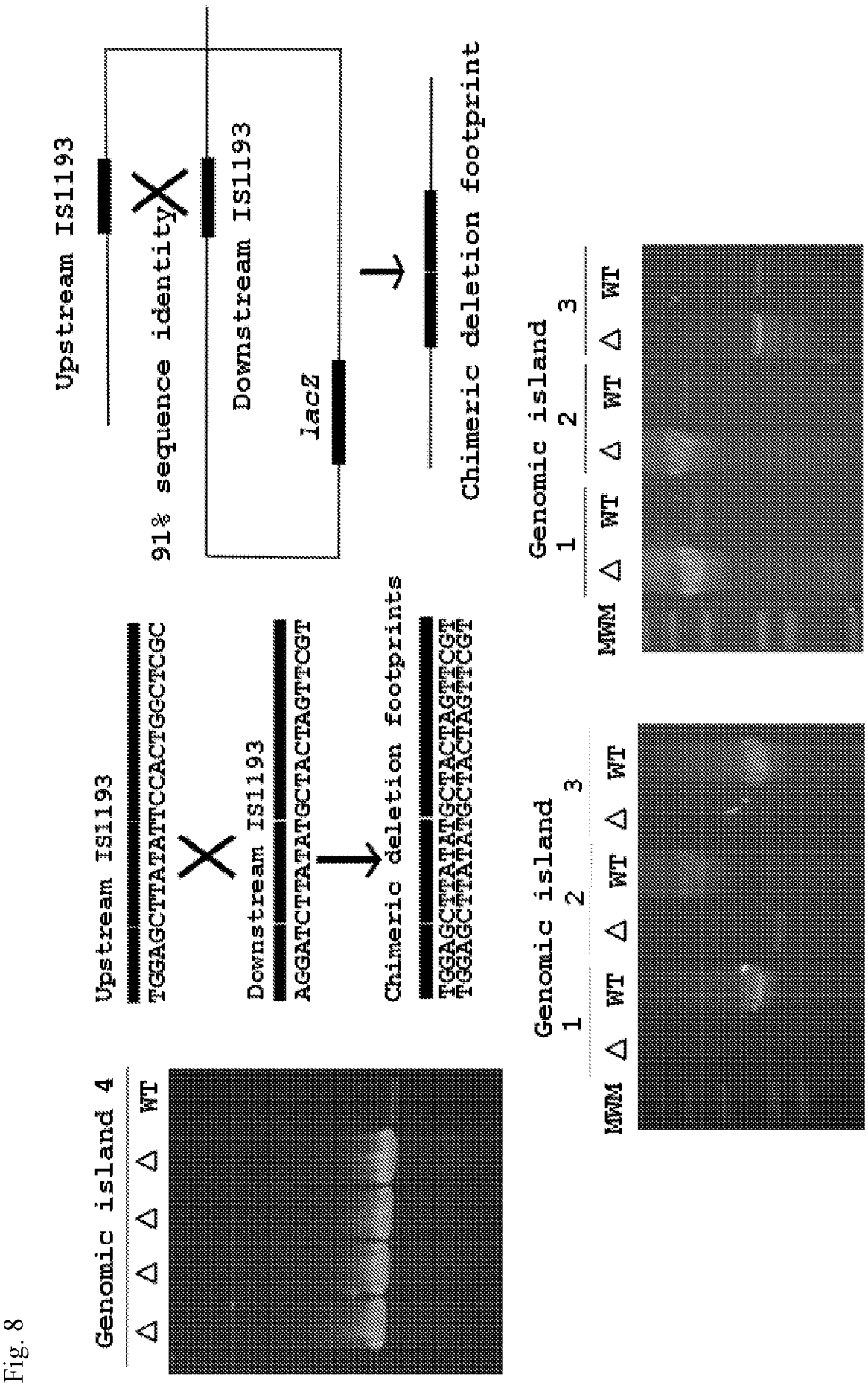

Upstream IS1193
91% sequence identity
Downstream IS1193
lacZ
Chimeric deletion footprint
Upstream IS1193
TGGAGCTTATATTCCACTGGCTCGC
Downstream IS1193
AGGATCTTATATGCTACTAGTTCGT
Chimeric deletion footprints
TGGAGCTTATATGCTACTAGTTCGT
TGGAGCTTATATGCTACTAGTTCGT
Genomic island 4
Δ Δ Δ Δ WT
Genomic island
1 2 3
MWM Δ WT Δ WT Δ WT
Genomic island
1 2 3
MWM Δ WT Δ WT Δ WT

Fig. 8

Fig. 9

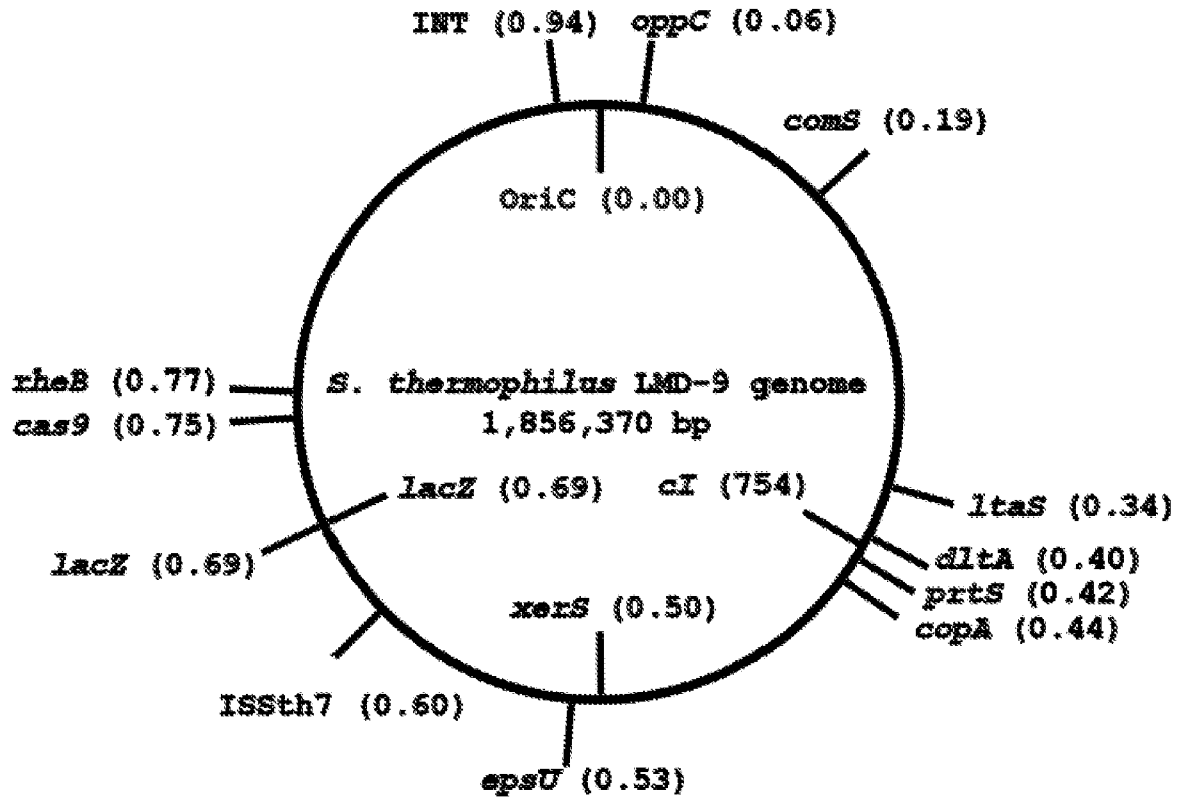
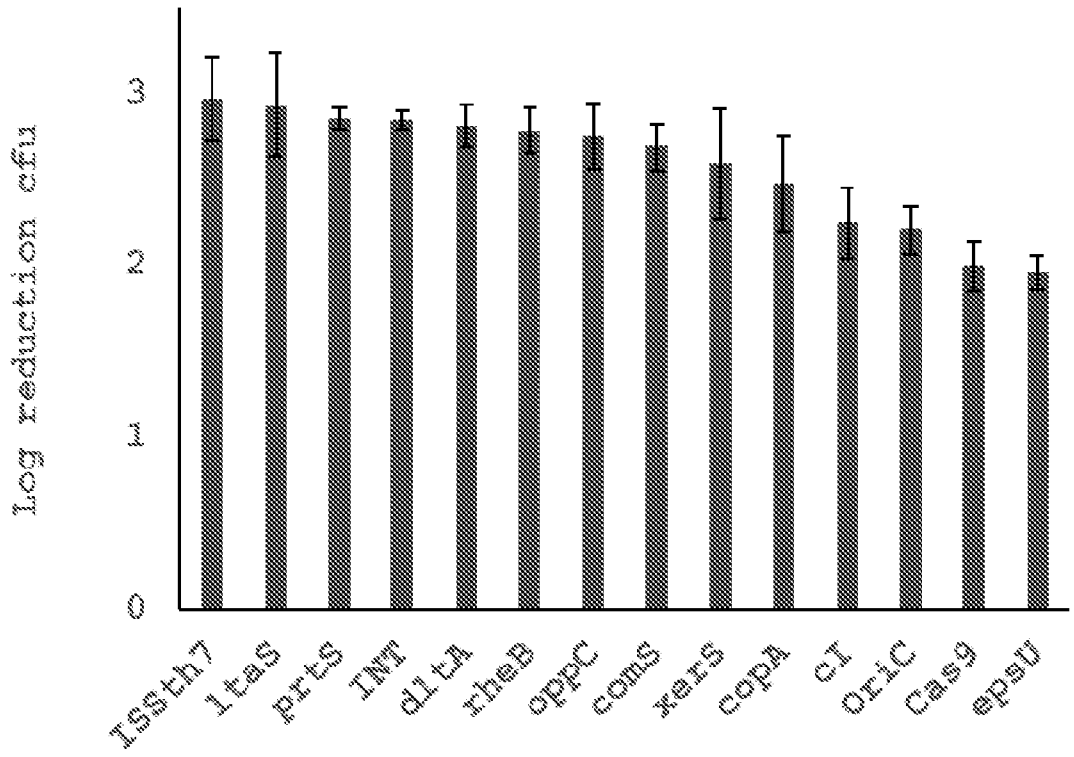

INT (0.94)
oppC (0.06)
comS (0.19)
OriC (0.00)
S. thermophilus LMD-9 genome
1,856,370 bp
rheB (0.77)
cas9 (0.75)
lacZ (0.69)
cI (754)
ltaS (0.34)
dltA (0.40)
prtS (0.42)
copA (0.44)
lacZ (0.69)
xerS (0.50)
ISSth7 (0.60)
epsU (0.53)

Fig. 10


Log reduction cfu
0
1
2
3
lSSth7
ltaS
prtS
tNT
dltA
rheB
oppC
comS
xerS
copA
cI
oriC
cas9
epsU

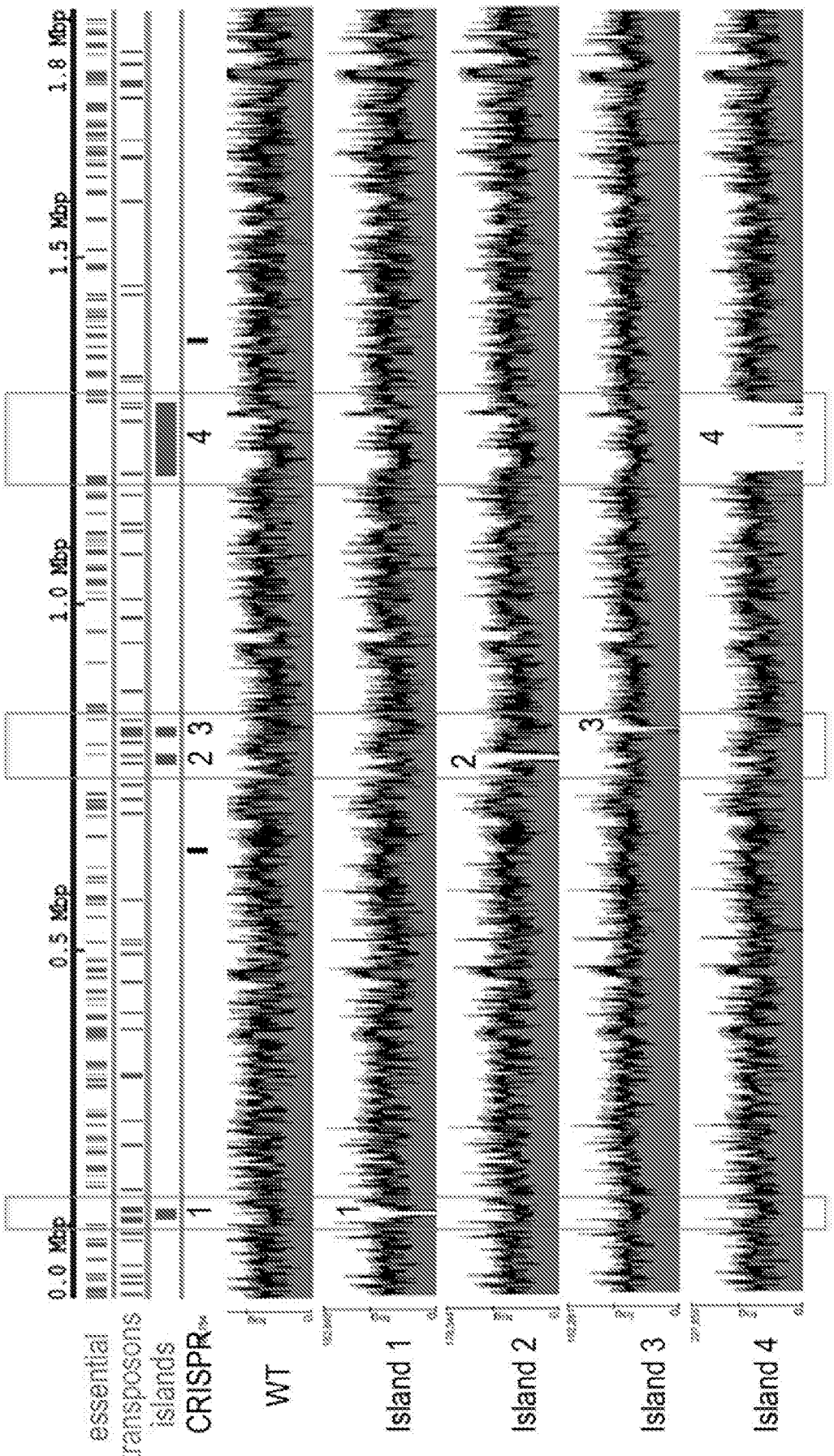

0.0 Mbp
0.5 Mbp
1.0 Mbp
1.5 Mbp
1.8 Mbp
essential transposons islands
CRISPR
WT
Island 1
Island 2
Island 3
Island 4

Fig. 11

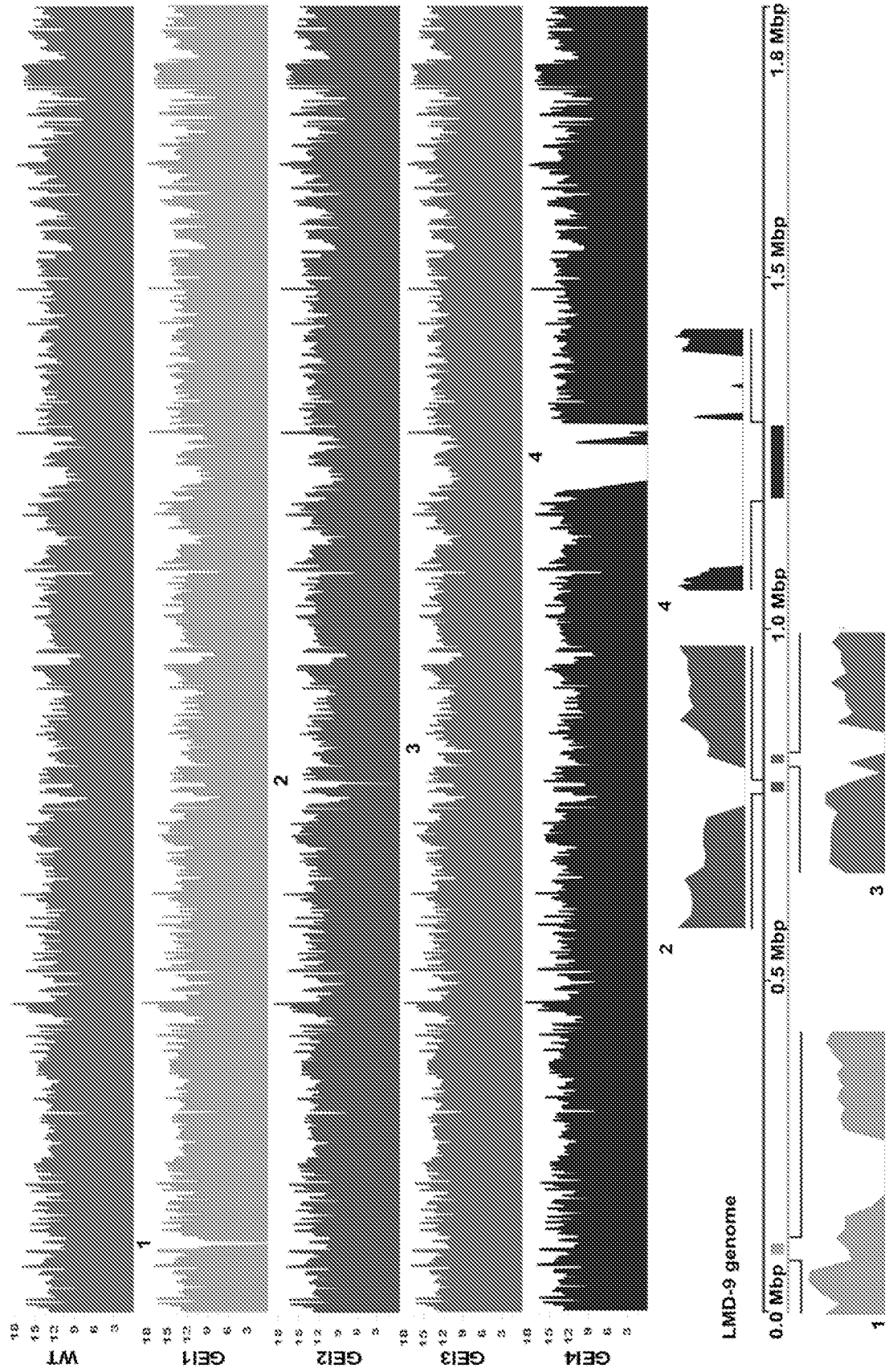
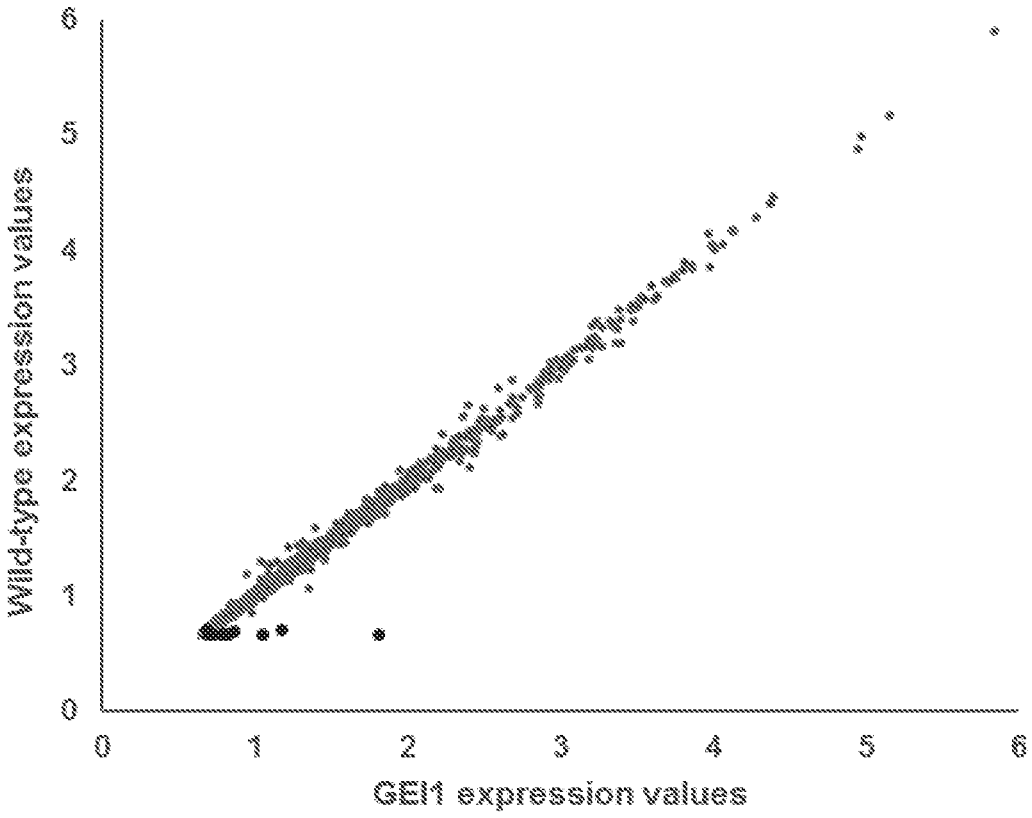

WT
GEI1
GEI2
GEI3
GEI4
LMD-9 genome
0.0 Mbp
0.5 Mbp
1.0 Mbp
1.5 Mbp
1.8 Mbp
1
2
3
4

Fig 12

Fig. 13


Wild-type expression values
0
1
2
3
4
5
6
GEI1 expression values

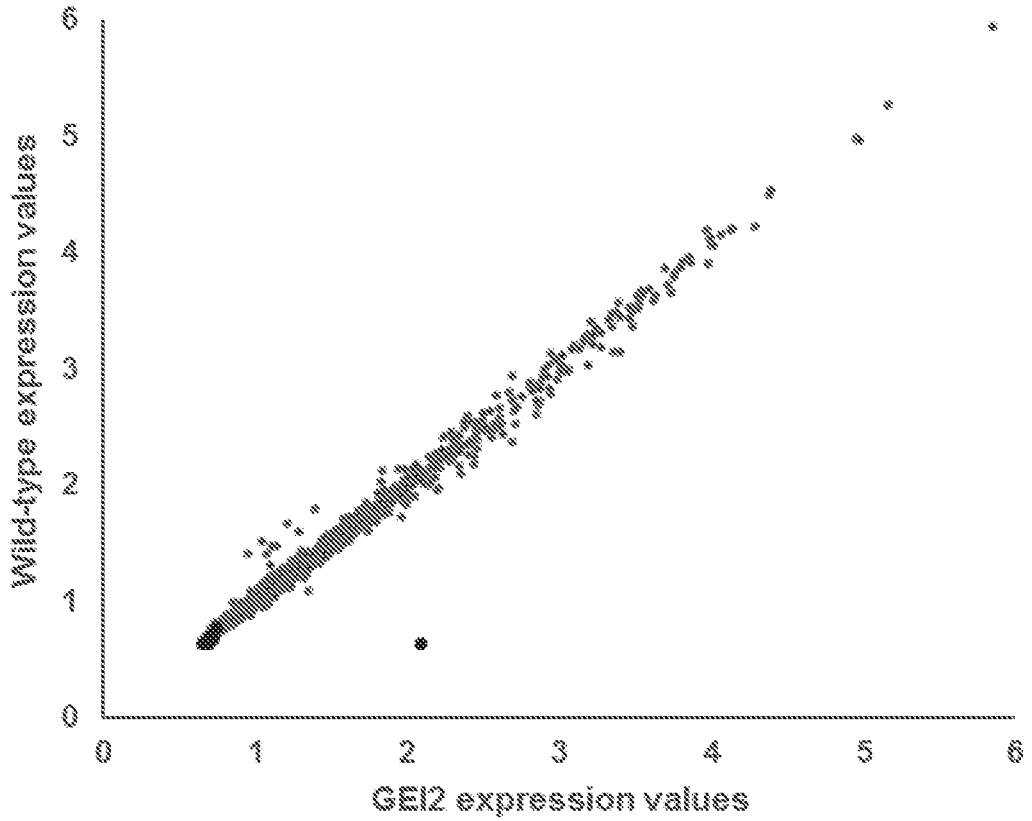
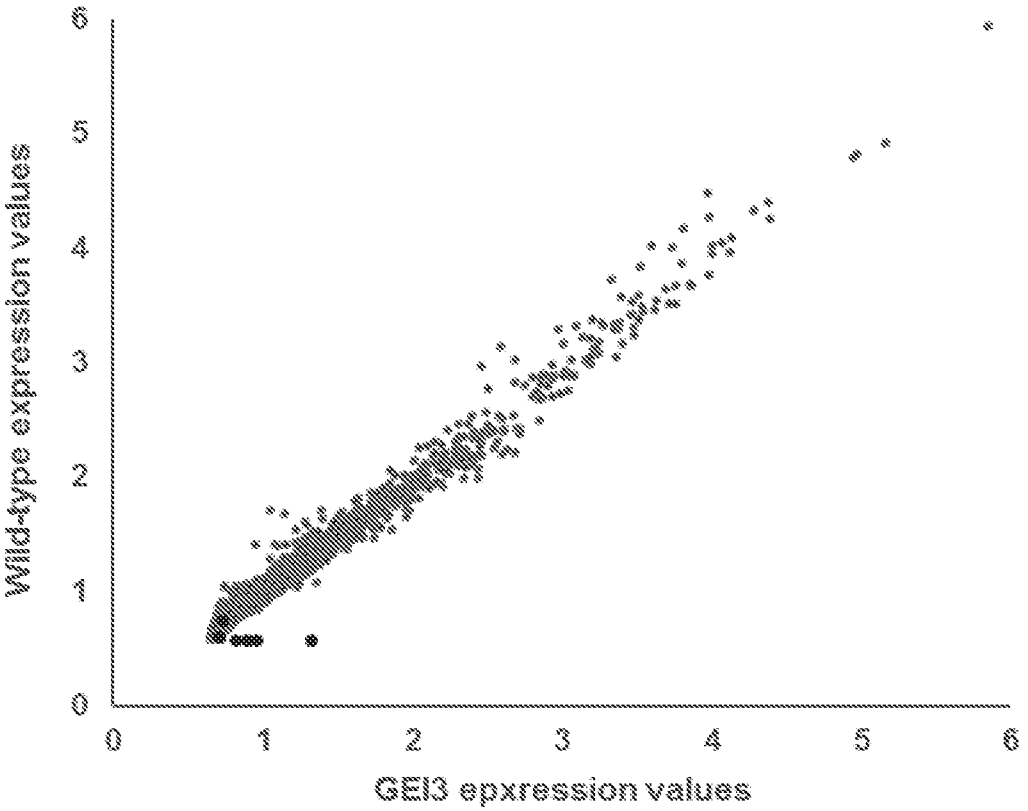
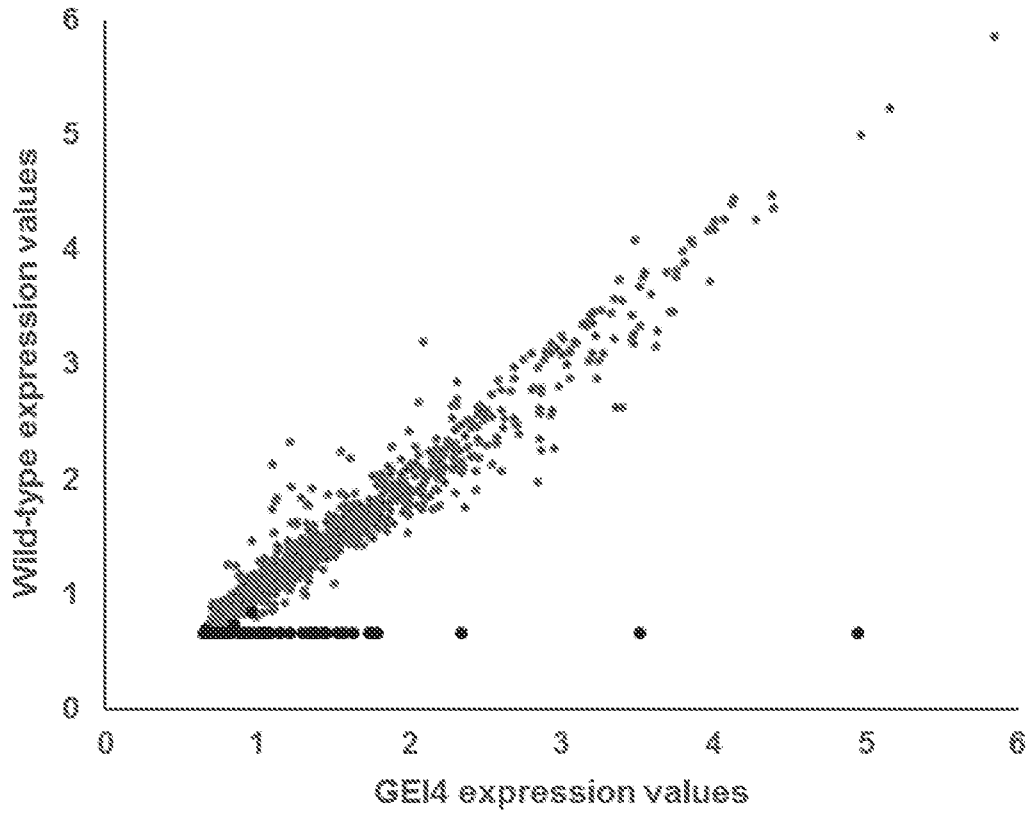

Wild-type expression values
0
1
2
3
4
5
6
GEI2 expression values

Fig. 13, cont'd.


Wild-type expression values
GEI3 epxression values
0
1
2
3
4
5
6


Wild-type expression values
GEI4 expression values
0
1
2
3
4
5
6

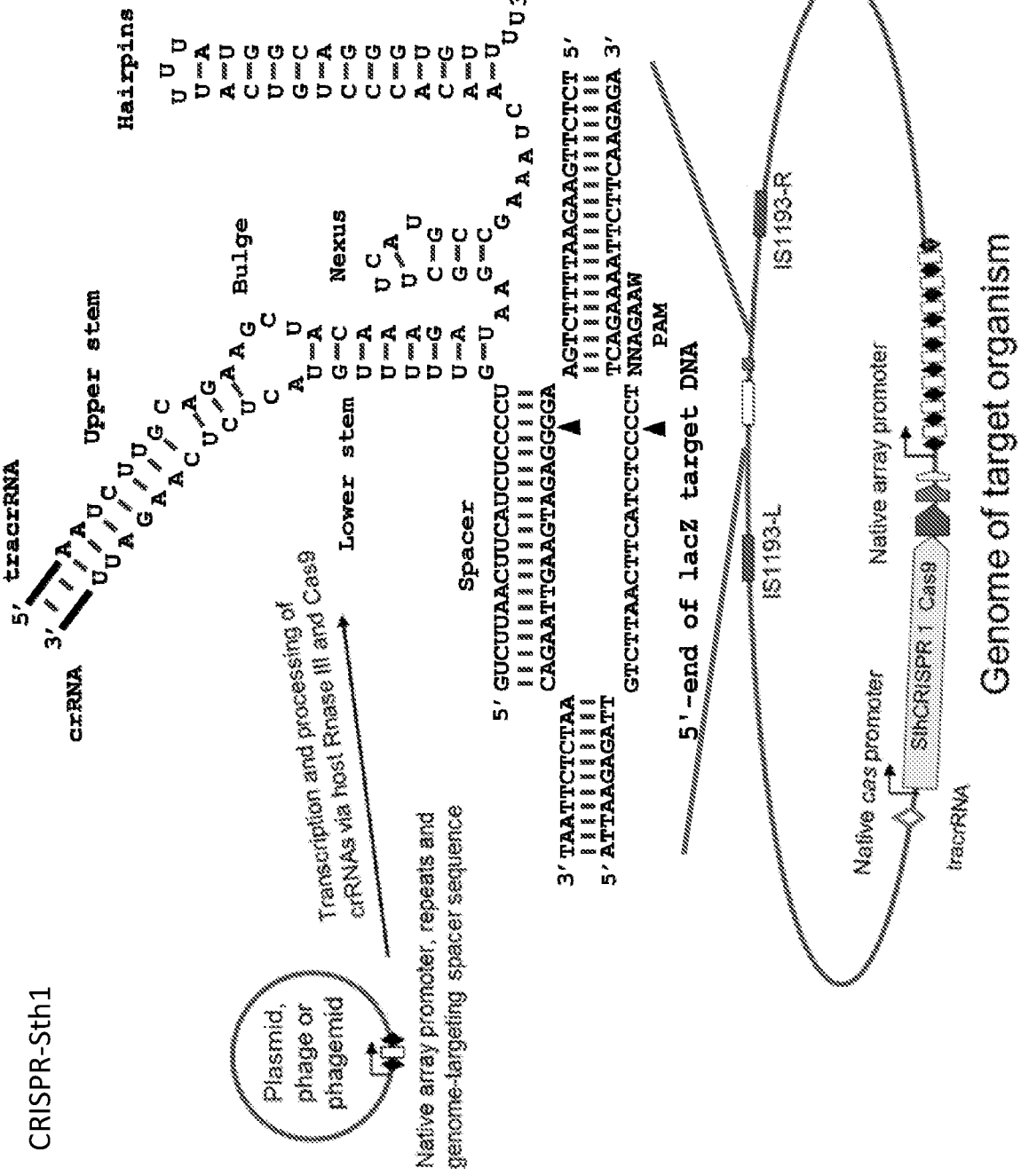

CRISPR-Sth1
5′ tracrRNA
3′
crRNA
Upper stem
Hairpins
Bulge
Lower stem
Nexus
Spacer
5′ GUCUUAACUUCAUCUCCCCU
CAGAATTGAAGTAGAGGGGA
3′ TAATTCTCTAA
5′ ATTAAGAGATT
GTCTTAACTTCATCTCCCCT NNAGAAW
PAM
5′-end of lacZ target DNA
Transcription and processing of crRNAs via host Rnase III and Cas9
Native array promoter, repeats and genome-targeting spacer sequence
Plasmid, phage or phagemid
AGTCTTTTTAAGAAGTTCTCT 5′
TCAGAAAATTCTTCAAGAGA 3′
IS1193-L
IS1193-R
Native cas promoter
Native array promoter
SthCRISPR 1 Cas9
tracrRNA
Genome of target organism

Fig. 14A

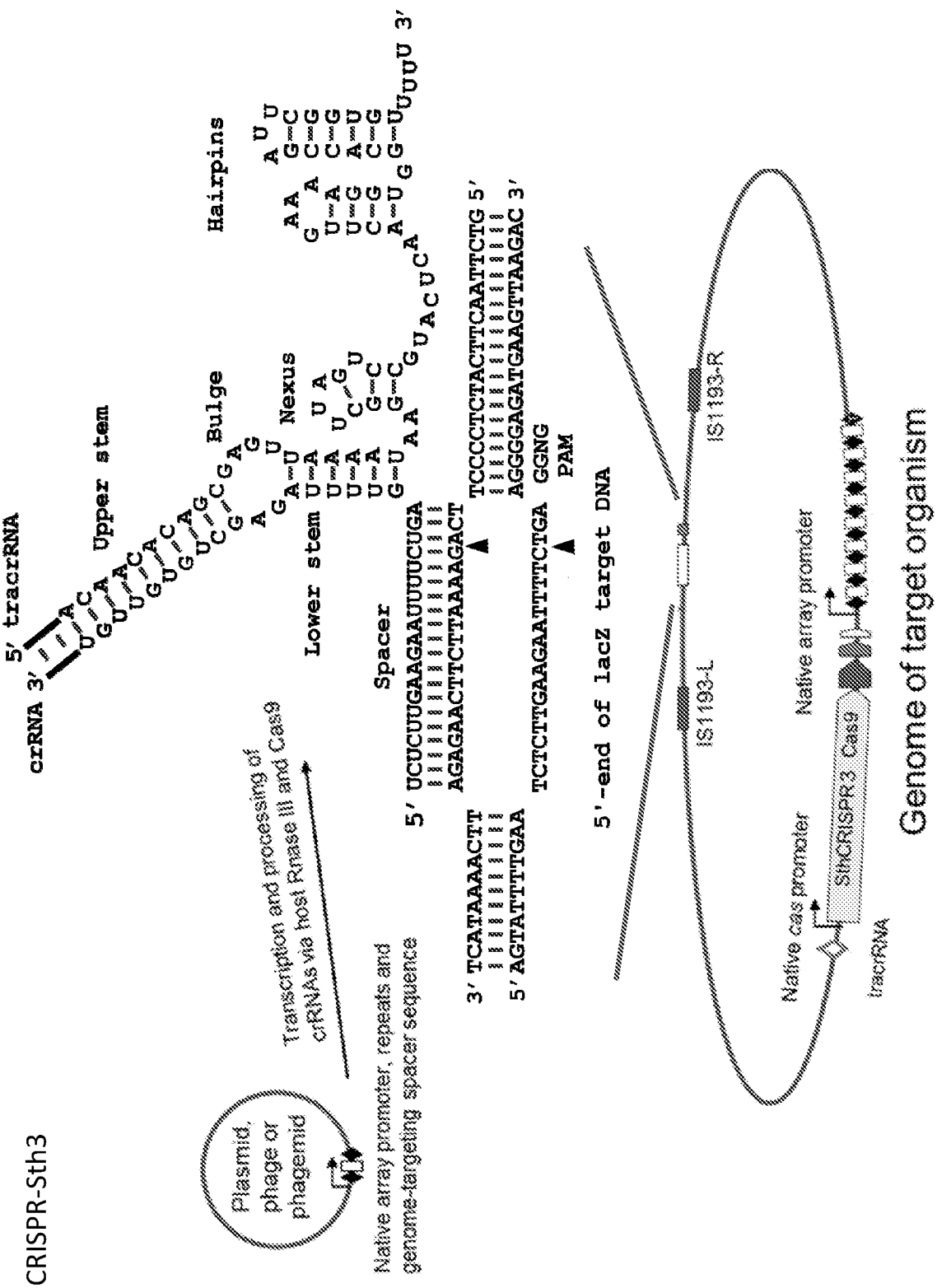

CRISPR-Sth3
5′ tracrRNA
crRNA 3′
Upper stem
Bulge
Hairpins
Nexus
Lower stem
Spacer
5′ UCUCUUGAAGAAUUUUCUGA
AGAGAACTTCTTAAAAGACT
3′ TCATAAAACTT
5′ AGTATTTTGAA
TCCCCTCTACTTCAATTCTG 5′
AGGGGAGATGAAGTTAAGAC 3′
TCTCTTGAAGAATTTTCTGA GGNG
PAM
5′-end of lacZ target DNA
Transcription and processing of crRNAs via host Rnase III and Cas9
Plasmid, phage or phagemid
Native array promoter, repeats and genome-targeting spacer sequence
IS1193-L
IS1193-R
Native array promoter
Native cas promoter
SthCRISPR3 Cas9
tracrRNA
Genome of target organism

Fig. 14B

Fig. 15

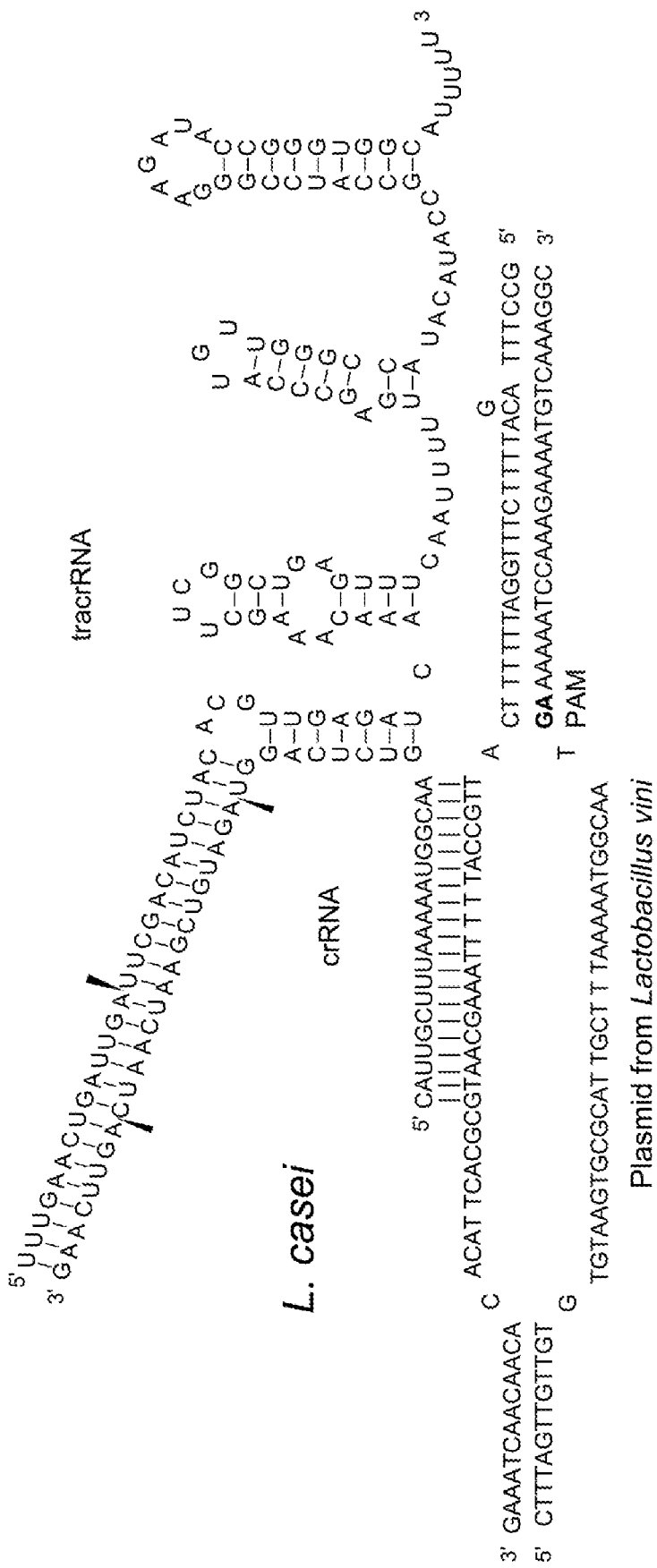

tracrRNA
crRNA
L. casei
5' CAUUGCUUUAAAAAUGGCAA
ACAT TCACGCGTAACGAAATT T T TACCGTT
C
3' GAAATCAACAACA
5' CTTTAGTTGTTGT
G
TGTAAGTGCGCAT TGCT T TAAAAATGGCAA
Plasmid from Lactobacillus vini
CT TT TTTTAGGTTTCT TTTTACA TTTCCG 5'
GA AAAAATCCAAAGAAAATGTCAAAGGC 3'
T PAM

Fig. 15, cont'd.


tracrRNA
L. casei
crRNA
5' CAUUGCUUUAAAAAUGGCAA
ACAT TCACGGCGTAACGAAATT T T TACCGTT
C
3' GAAATCAACAACA
5' CTTTAGTTGTTGT
G
TGTAAGTGCGCAT TGCT T TAAAAATGGCAA
A
CT TT TTTTAGGTTTCTTTTTACAGTTTCCG 5'
GA AAAAATCCAAAGAAAATGTCAAAGGC 3'
T PAM
Plasmid from Lactobacillus vini

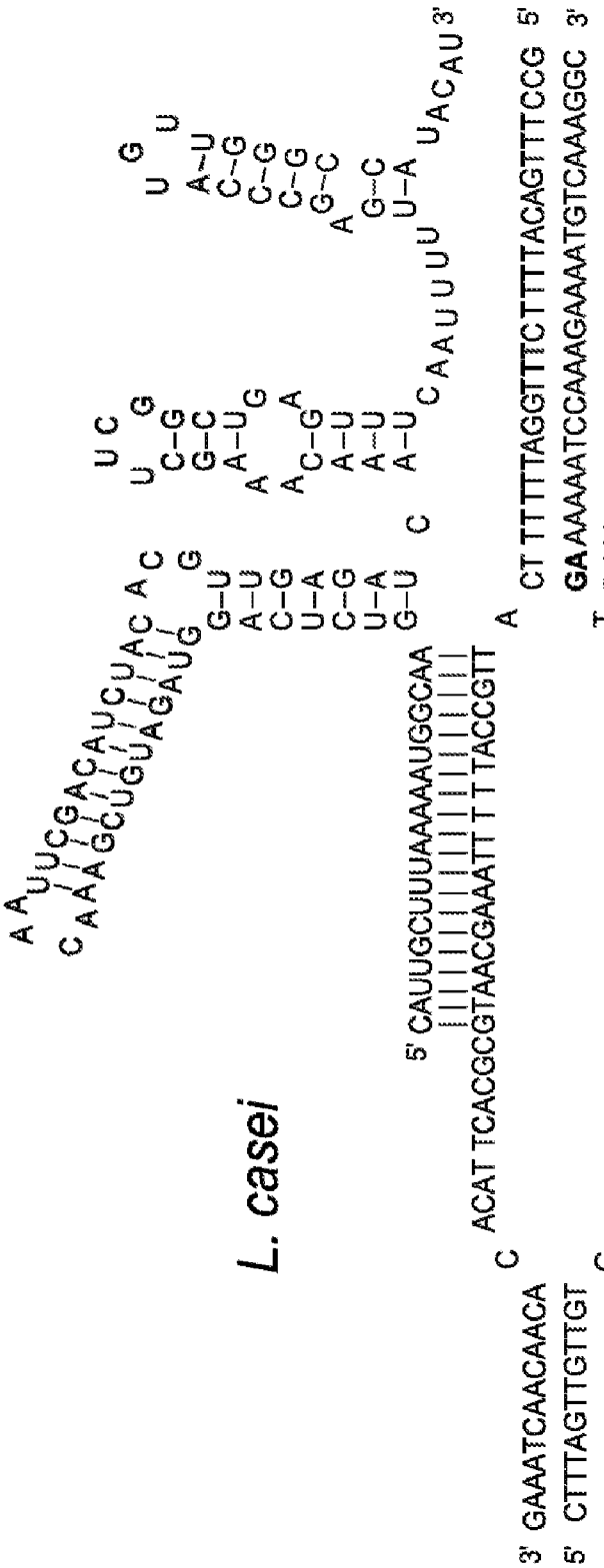

L. casei
5' CAUUGCUUUAAAAAUGGCAA
ACAT TCACGCGTAACGAAATT T T TACCGTT
3' GAAATCAACAACA
5' CTTTAGTTGTTGT
TGTAAGTGCGCAT TGCT T TAAAAATGGCAA
Plasmid from Lactobacillus vini
CT TTTTTAGGTTTCTTTTACAGTTTCCG 5'
GAAAAAATCCAAAGAAAATGTCAAAGGC 3'
PAM

Fig. 15, cont'd.

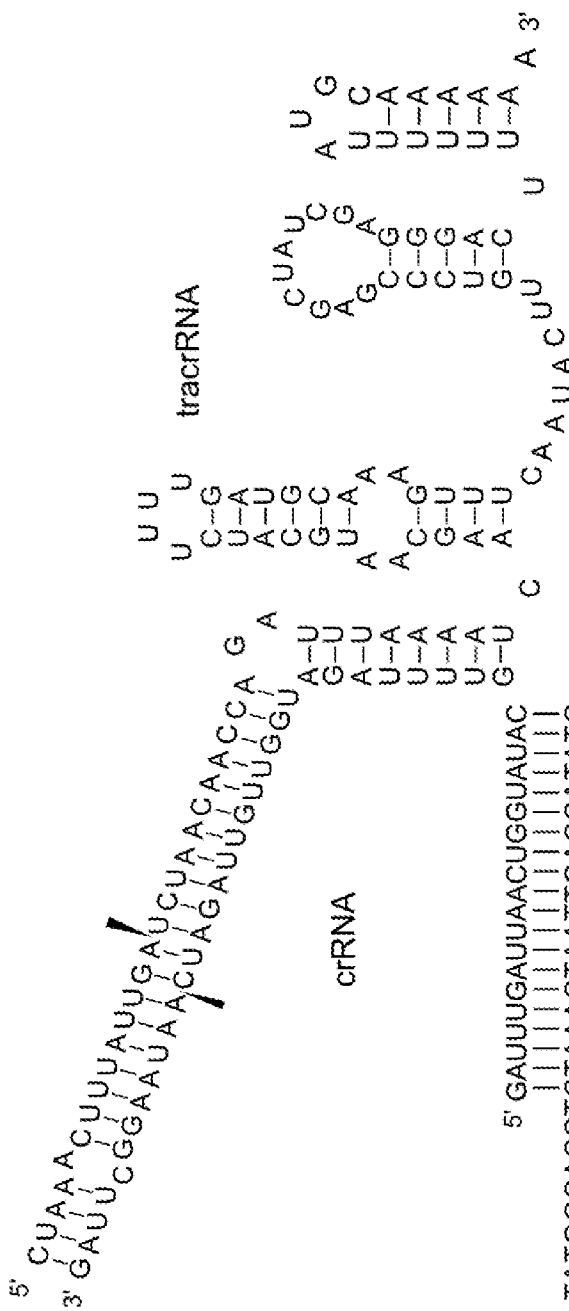
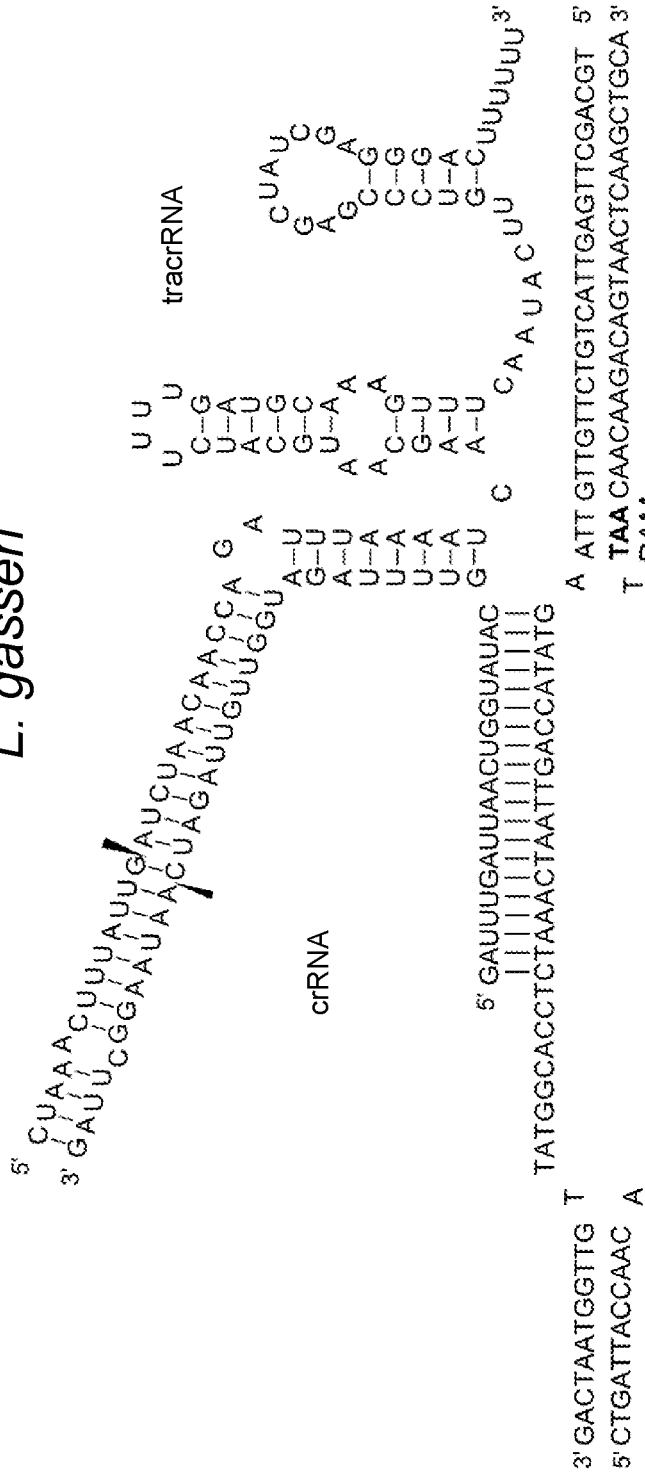

L. gasseri
tracrRNA
crRNA
5'
3'
5' GAUUUGAUUAACUGGUAUAC
TATGGCACCTCTAAACTAATTGACCATATG

Fig. 16

Fig. 16, cont'd.


L. gasseri
tracrRNA
crRNA
5' GAUUUGAUUAACUGGUAUAC
TATGGCACCTCTAAACTAATTGACCATATG
3' GACTAATGGTTG T
5' CTGATTACCAAC A
ATACCGTGGAGATTTGATTAACTGGTATAC
Protospacer
A ATT GTTGTTCTGTCATTGAGTTCGACGT 5'
T TAA CAACAAGACAGTAACTCAAGCTGCA 3'
T PAM

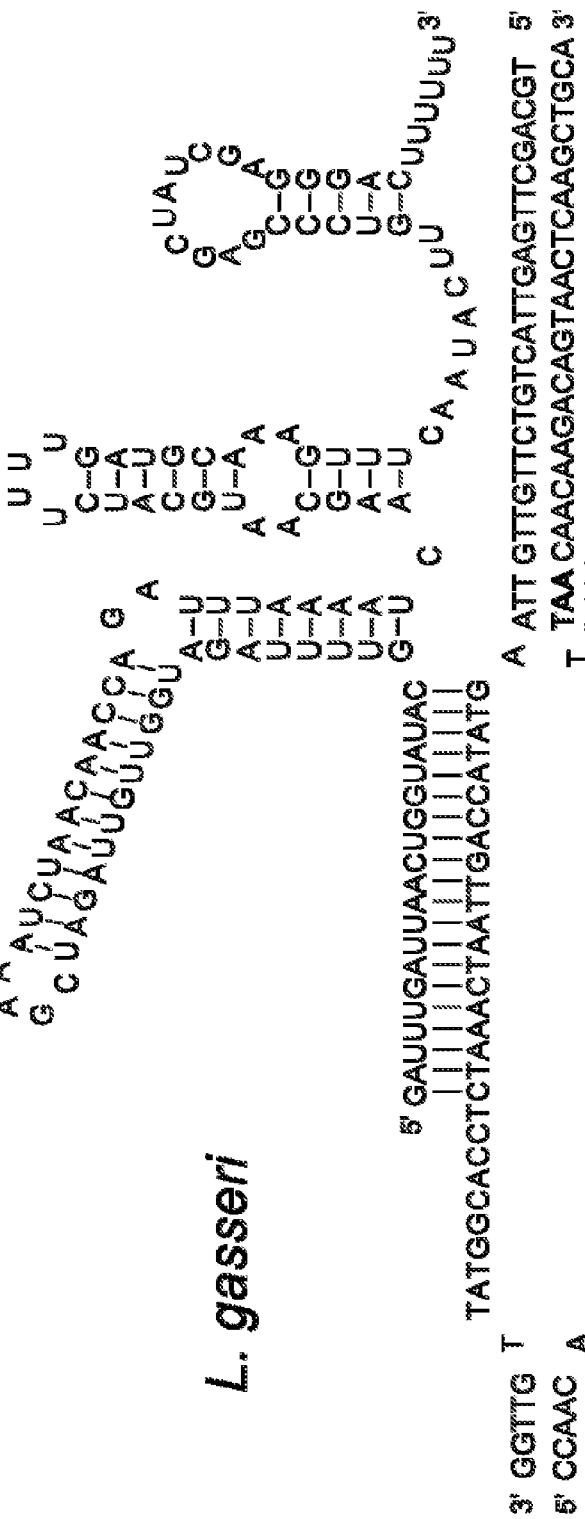

L. gasseri
5' GAUUUGAUUAACUGGUAUAC
TATGGCACCTCTAAACTAATTGACCATATG
3' GGTTG T
5' CCAAC A
ATACCGTGGAGATTTGATTAACTGGTATAC
Protospacer
A ATT GTTGTTCTGTCATTGAGTTCGACGT 5'
T TAA CAACAAGACAGTAACTCAAGCTGCA 3'
PAM

Fig. 16, cont'd.

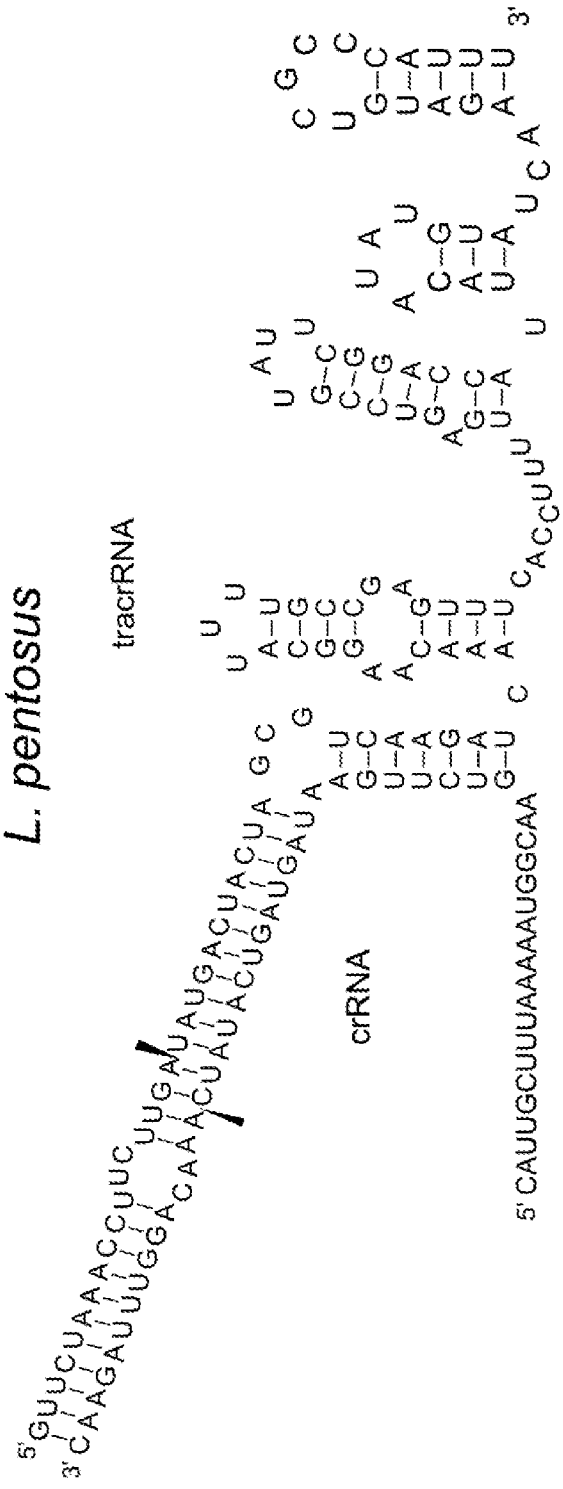


Fig. 17

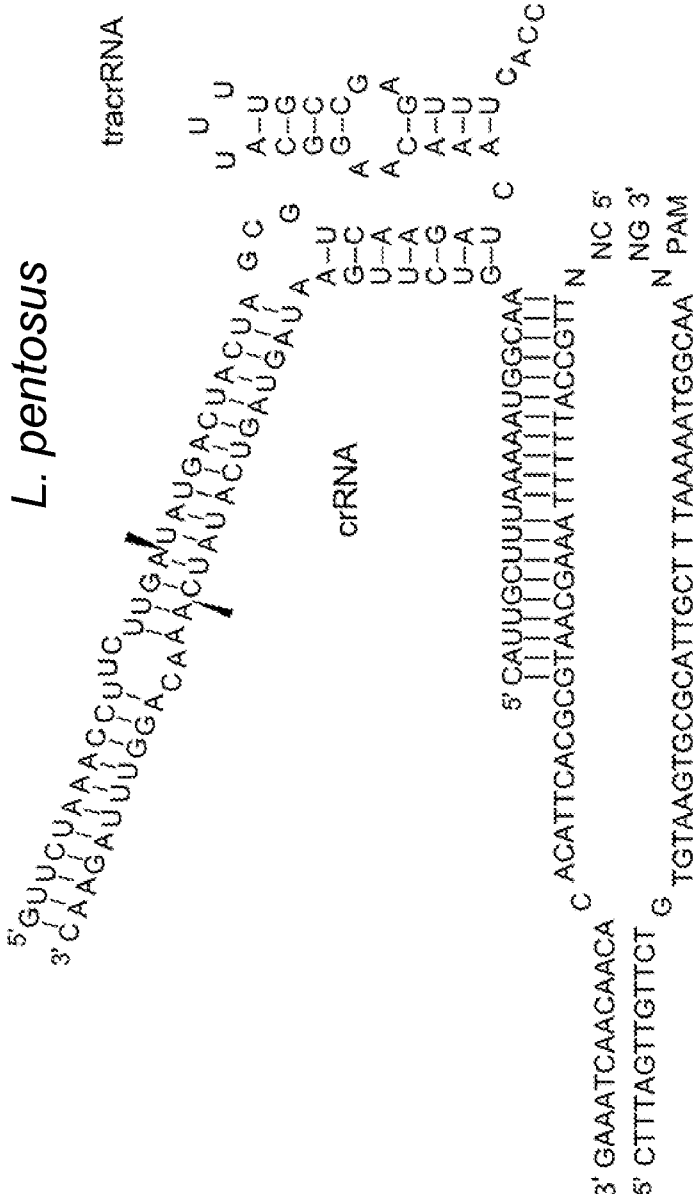

L. pentosus
tracrRNA
crRNA
5' GUUCUAAACUUGAUGACUACUAGC
3' CAAGAUUGGACAACUAUGACUGAUGAUA
5' CAUUGCUUUAAAAUGGCAA
C ACATTCACGCGTAACGAAATTTTTACCGTT N
3' GAAATCAACAACA
5' CTTTAGTTGTTCT G TGTAAGTGCGCATTGCT T TAAAAATGGCAA
NC 5'
NG 3'
N PAM

Fig. 17, cont'd.

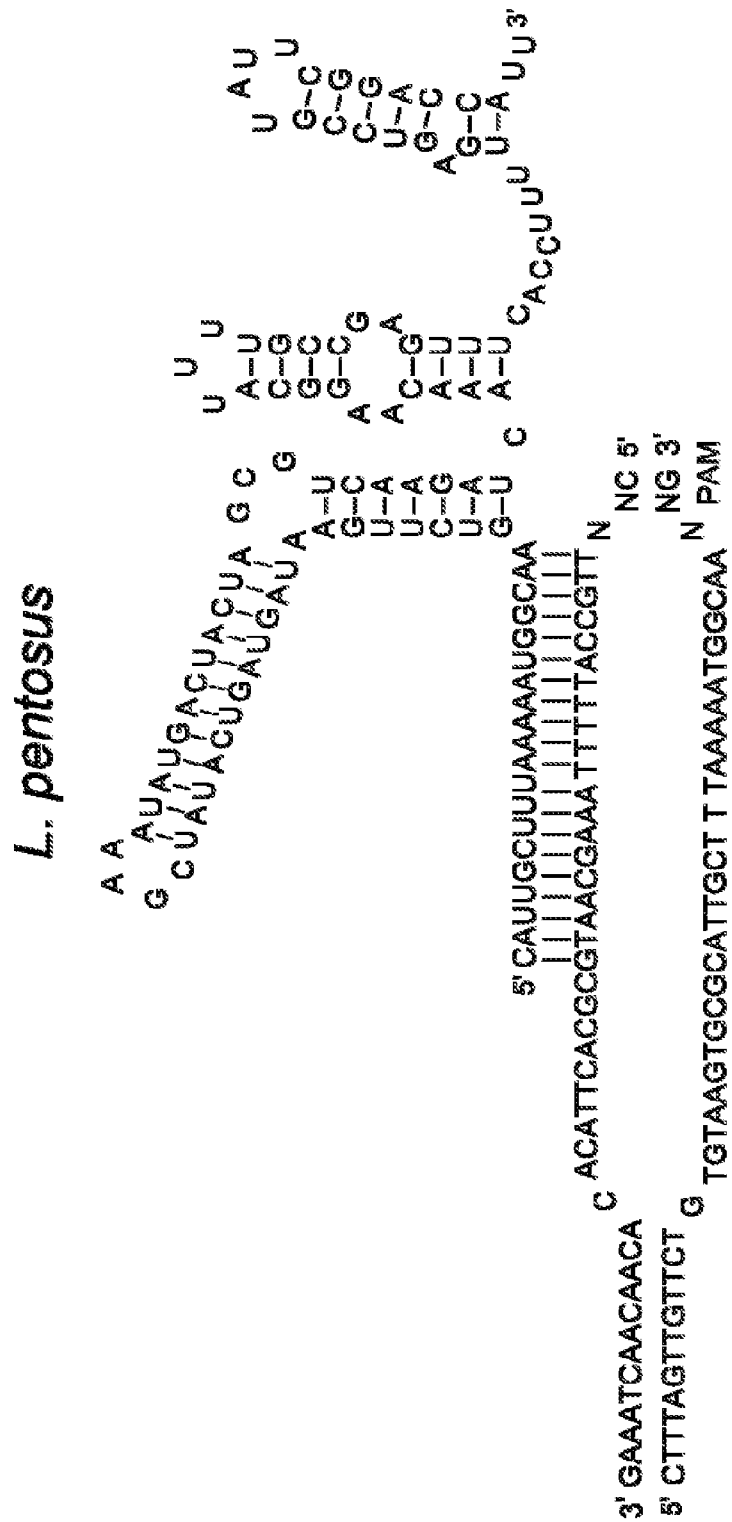

L. pentosus
5' CAUUGCUUUAAAAAUGGCAA
C ACATTCACGCGTAACGAAATTTTTTACCGTT N
3' GAAATCAACAACA
5' CTTTAGTTGTTCT G TGTAAGTGCGCATTGCT T TAAAAATGGCAA
NC 5'
NG 3'
N PAM

Fig. 17, cont'd.

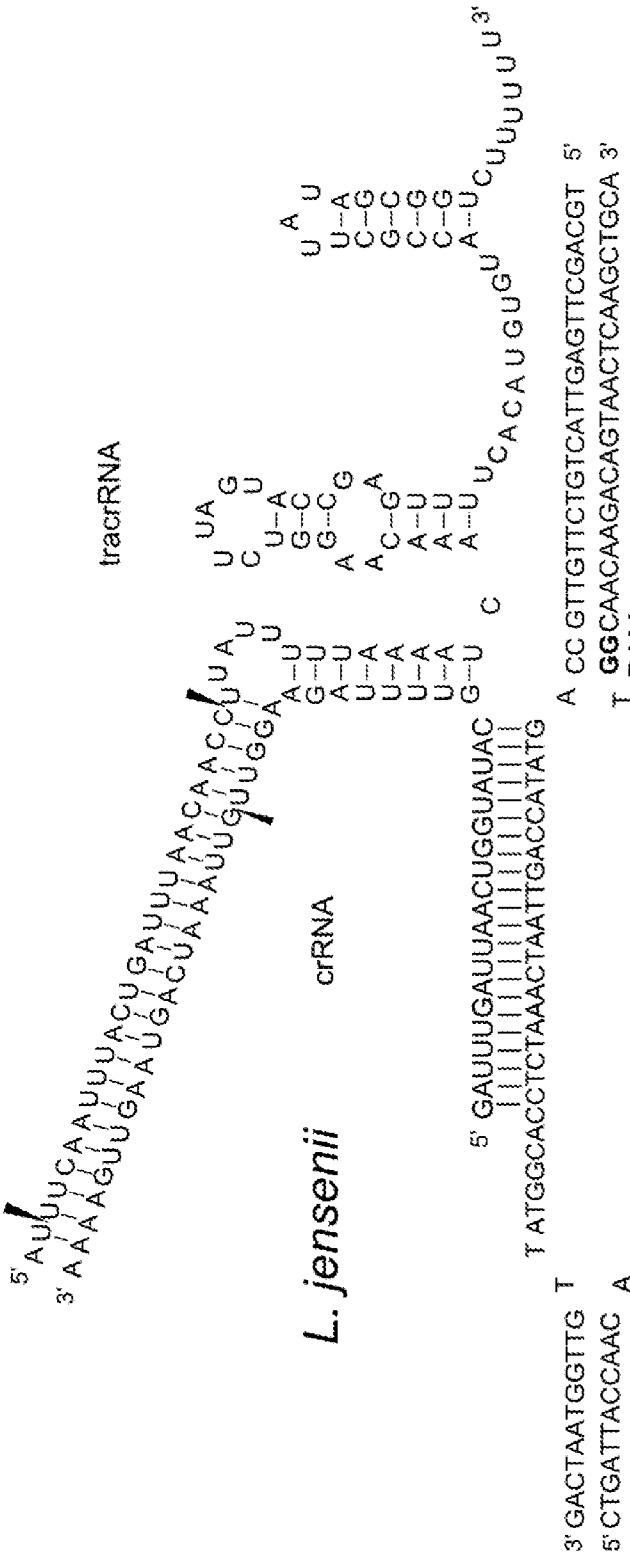
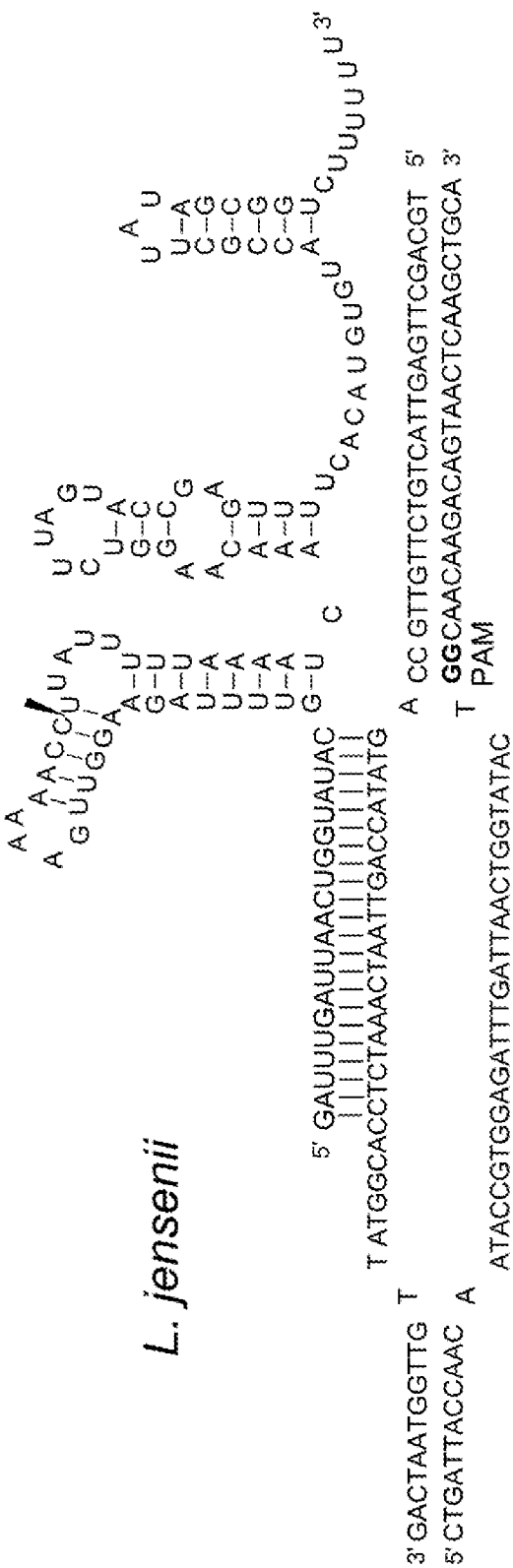

tracrRNA
crRNA
L. jensenii
5' GAUUUGAUUAACUGGUAUAC
T ATGGCACCTCTAAACTAATTGACCATATG
3' GACTAATGGTTG T
5' CTGATTACCAAC A
ATACCGTGGAGATTTGATTAACTGGTATAC
Lactobacillus phage LV-1
A CC GTTGTTCTGTCATTGAGTTCGACGT 5'
T GGCAACAAGACAGTAACTCAAGCTGCA 3'
PAM

Fig. 18

Fig. 18, cont'd.


L. jensenii
5' GAUUUGAUUAACUGGUAUAC
T ATGGCACCTCTAAACTAATTGACCATATG
3' GACTAATGGTTG T
5' CTGATTACCAAC A
A CC GTTGTTCTGTCATTGAGTTCGACGT 5'
GGCAACAAGACAGTAACTCAAGCTGCA 3'
T PAM
ATACCGTGGAGATTTGATTAACTGGTATAC
Lactobacillus phage LV-1

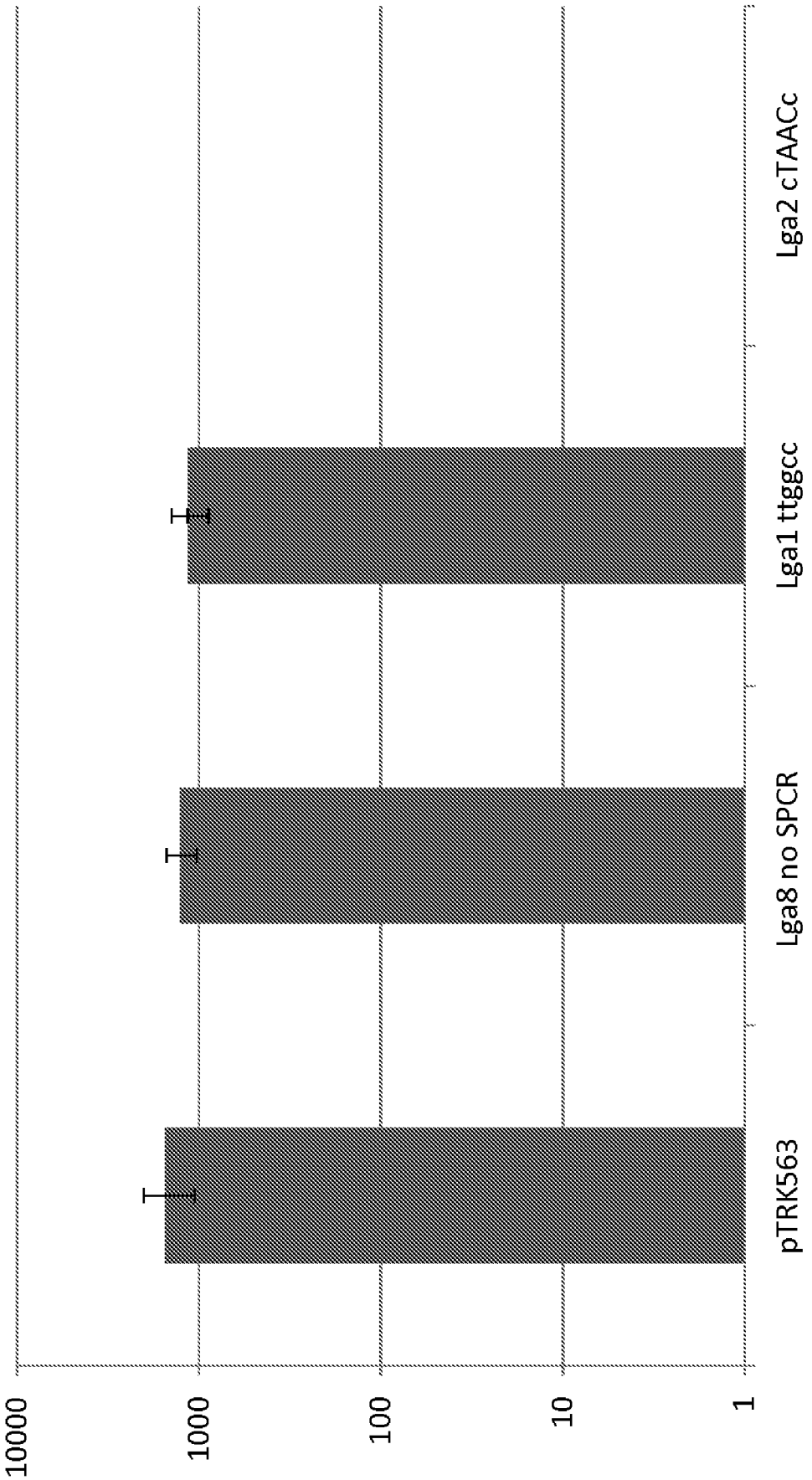

10000
1000
100
10
1
pTRK563
Lga8 no SPCR
Lga1 ttggcc
Lga2 cTAACc

Fig. 19

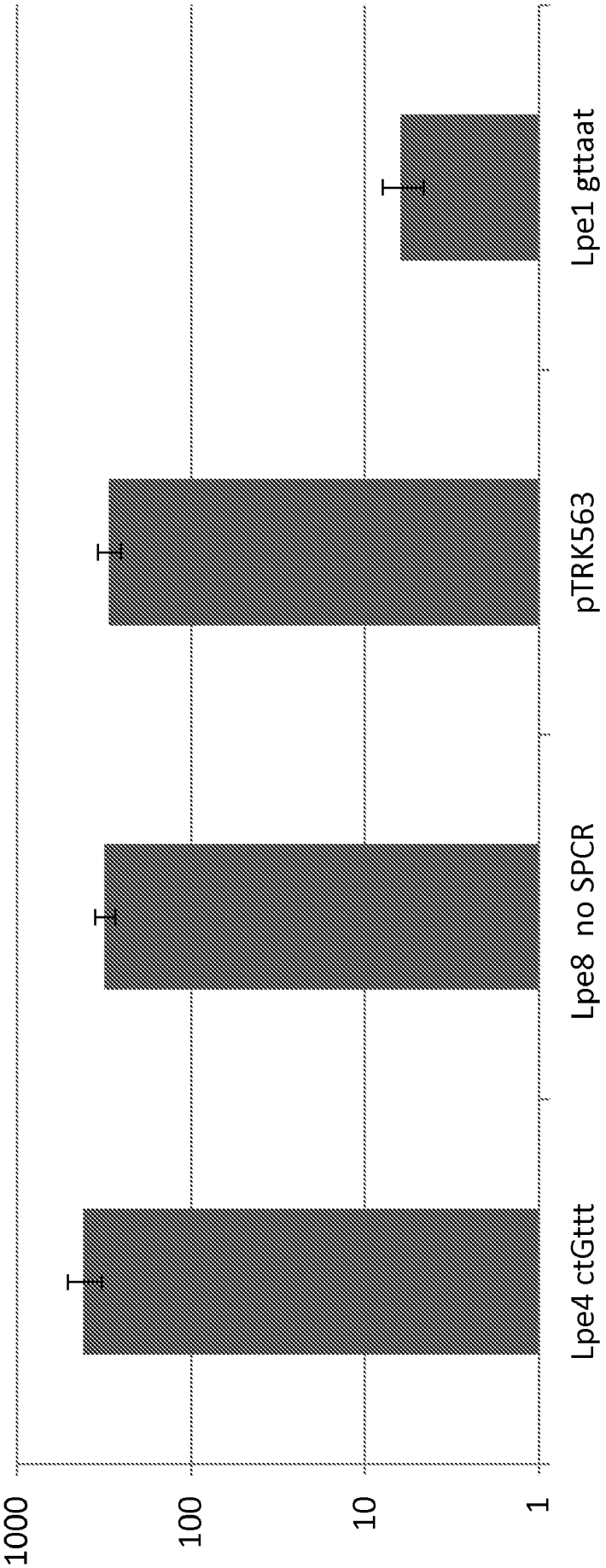


Fig. 20

Fig. 21

Type I LEADER

ATGGGATAGGGATTTTTTAGT

GTTTTCCCCGGCACATGCGGGGGGTGATCC GGCGATGATACGTAGCCGAACTGAGAGGTTGAT GTTTTCCCCGCACATGCGGGGGGTGATCC

REPEAT SPCR 1 REPEAT

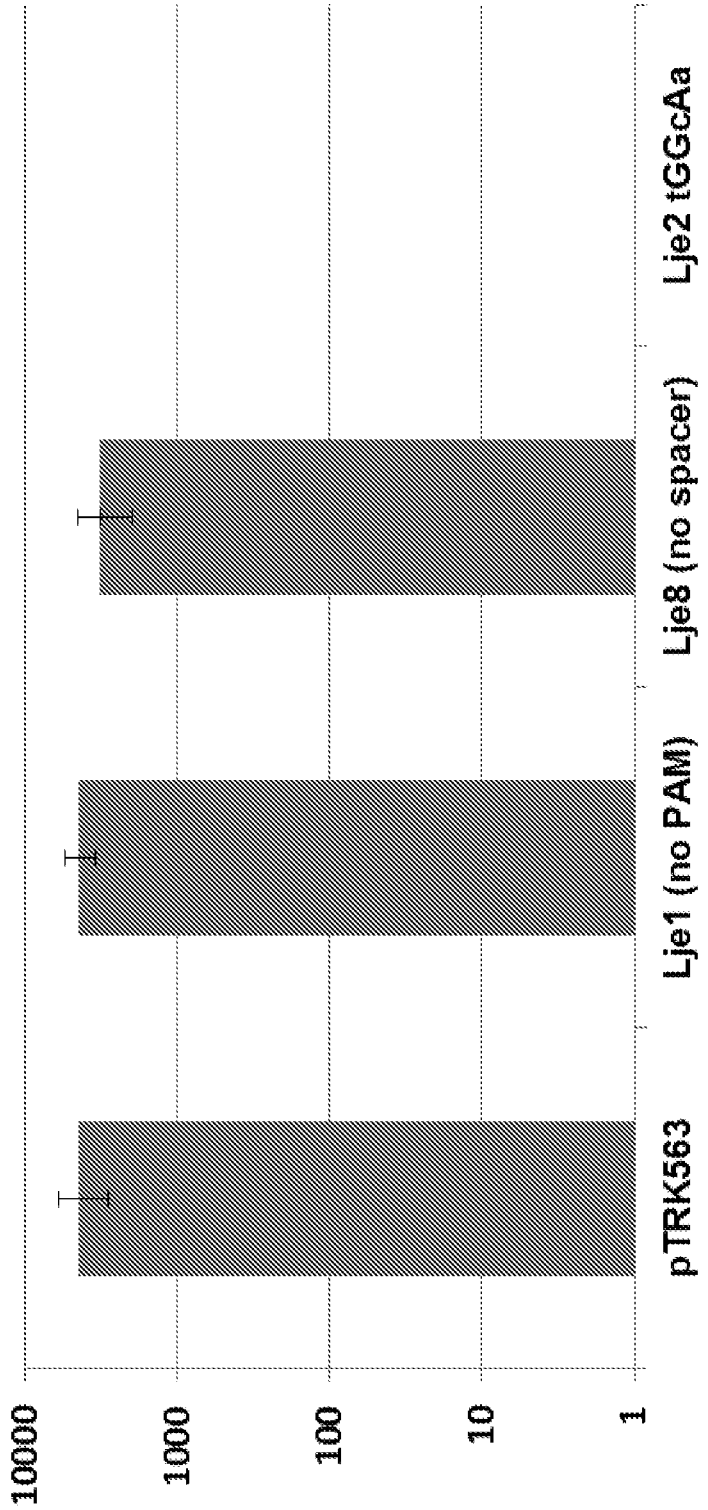

10000
1000
100
10
1
pTRK563
Lje1 (no PAM)
Lje8 (no spacer)
Lje2 tGGcAa

Fig. 22

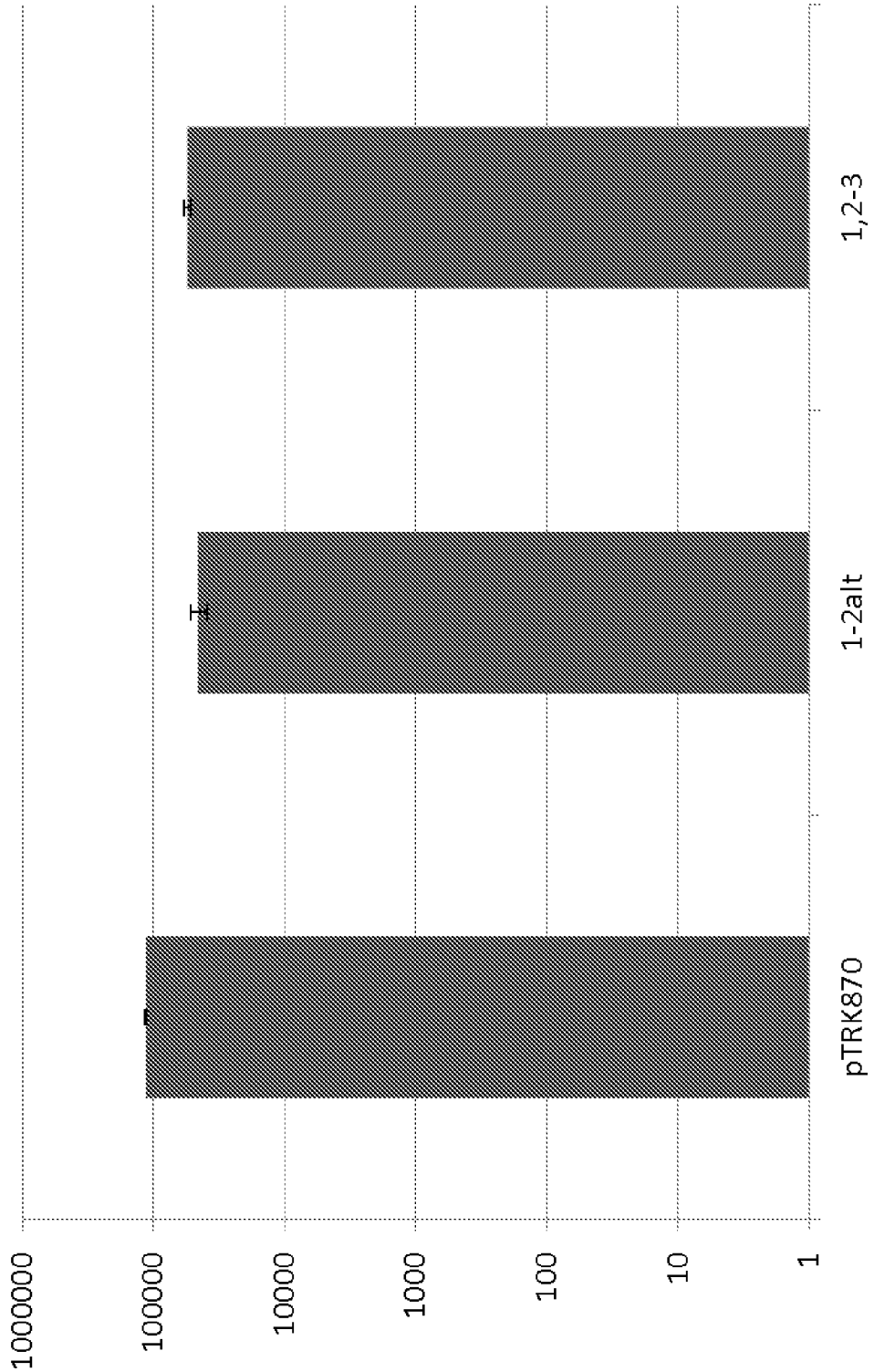


Fig. 23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 2016/034812

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 12-16, 28-31
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 2016/034812

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/02 *(2006.01)*
C12Q 1/68 *(2006.01)*
C12N 1/21 *(2006.01)*
C12N 15/63 *(2006.01)*

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q 1/02, 1/68, C12N 1/21, 15/63, 7/01

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DB WIPO (Patentscope), DB Espacenet, PatFT & AppFT USPTO, RUPAT, EAPATIS, CIPO, PAJ, KIPRIS, K-PION, SIPO, DWPI; PubMed, Scorpus

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	GOMAA AHMED A. et al. "Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems." mBio, 2014, Vol.5, Issue 1: e00928-13. doi: 10.1128/mBio.00928-13, p.1-9, especially abstract, p.1-7	1, 3, 5-7, 17, 21 2, 4, 8-11, 19, 23, 25-27 18, 20, 22, 24
Y	WO 2014/204727 A1 (THE BROAD INSTITUTE INC. et al.) 24.12.2014, abstract, paragraphs [0095], [00130]-[00141], example 1	2, 4, 8-11, 19, 23, 25-27
A	EP 2 860 267 A1 (DUPONT NUTRITION BIOSCIENCES APS) 15.04.2015, abstract	1-11, 17-27

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 August 2016 (30.08.2016)	15 September 2016 (15.09.2016)

Name and mailing address of the ISA/RU:
Federal Institute of Industrial Property,
Berezhkovskaya nab., 30-1, Moscow, G-59,
GSP-3, Russia, 125993
Facsimile No: (8-495) 531-63-18, (8-499) 243-33-37

Authorized officer
T.Babakova

Telephone No. 495 531 65 15

摘要

本申请涉及使用 CRISPR 核酸来筛选细菌、古细菌、藻类和/或酵母中的必需基因和非必需基因以及可消耗基因组岛、来杀灭细菌、古细菌、藻类和/或酵母、来鉴定一种或多种基因的表型、和/或来筛选细菌、古细菌、藻类和/或酵母中的基因组大小减小和/或基因缺失。