



(51) International Patent Classification:

C12N 15/09 (2006.01) C12N 15/63 (2006.01)
C12N 15/10 (2006.01) C12N 15/64 (2006.01)
C12N 15/113 (2010.01) C12N 15/66 (2006.01)

(21) International Application Number:

PCT/US2018/017573

(22) International Filing Date:

09 February 2018 (09.02.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/457,493 10 February 2017 (10.02.2017) US

(71) Applicant: ZYMERGEN INC. [US/US]; 5980 Horton Street, Suite 105, Emeryville, California 94608 (US).

(72) Inventors: CHAIKIND, Brian; 91B Monte Cresta Avenue, Oakland, California 94611 (US). VAN ROSSUM, Hendrik Marinus; 6420 Vallejo Street, Oakland, California 94608 (US).

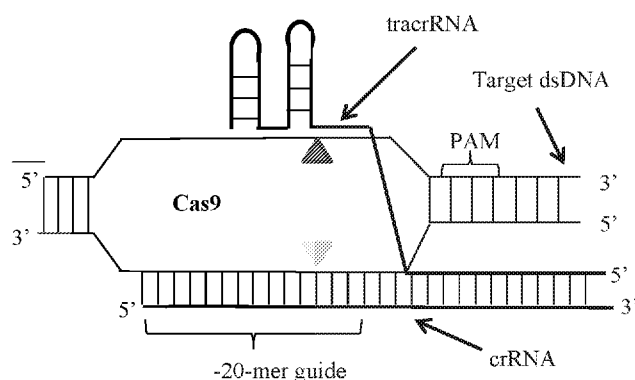
(74) Agent: HOLLY, David Christopher et al.; Cooley LLP, 1299 Pennsylvania Avenue, N.W., Suite 700, Washington, District of Columbia 20004 (US).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

(54) Title: A MODULAR UNIVERSAL PLASMID DESIGN STRATEGY FOR THE ASSEMBLY AND EDITING OF MULTIPLE DNA CONSTRUCTS FOR MULTIPLE HOSTS

FIGURE 1A



CRISPR- Cas9
WT crRNA/tracrRNA

(57) Abstract: The disclosure describes methods, compositions, and kits for high throughput DNA assembly reactions *in vitro*. The disclosure further describes modular CRISPR DNA constructs comprising modular insert DNA parts flanked by cloning tag segments comprising pre-validated CRISPR protospacer/protospacer adjacent motif sequence combinations.



Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

**A MODULAR UNIVERSAL PLASMID DESIGN STRATEGY FOR THE ASSEMBLY
AND EDITING OF MULTIPLE DNA CONSTRUCTS FOR MULTIPLE HOSTS**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Application No. 62/457,493, filed on February 10, 2017, which is hereby incorporated by reference in its entirety for all purposes.

FIELD

[0002] The present disclosure relates to systems, methods, and compositions used for guided genetic sequence editing *in vitro*. The disclosure describes, *inter alia*, methods of using guided sequence editing complexes for improved DNA cloning, assembly of oligonucleotides, and for the improvement of microorganisms.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0003] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: ZYMR_010_01WO_SeqList_ST25.txt, date recorded: January 30, 2018; file size: \approx 800 kilobytes).

GOVERNMENT LICENSE RIGHTS

[0004] This invention was made with Government support under Agreement No. HR0011-15-9-0014, awarded by DARPA. The Government has certain rights in the invention.

BACKGROUND

[0005] A major area of interest in biology is the *in vitro* and *in vivo* targeted modification of genetic sequences. Indeed, one of the most significant bottlenecks to academic and commercial genetic research has been the speed with which novel genetic constructs could be generated or later modified prior to testing.

[0006] The currently available cloning techniques relying on restriction site recognition or DNA hybridization and amplification have proven to be slow, unreliable, and intractable to later modifications. The discovery of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) gene editing systems have provided researchers with additional avenues for genetic modification. Even these new approaches, however, remain impractical for high throughput modular cloning applications.

[0007] CRISPR editing locations for example, are often limited by the location of protospacer adjacent motifs (PAMs). *De novo* CRISPR guide RNA design and gene targeting can be both time consuming and expensive, and is also susceptible to low efficiencies, and potential for off-target mutations.

[0008] Thus, there is a need for improved compositions and methods for targeted alteration of genetic sequences.

SUMMARY OF THE DISCLOSURE

[0009] In some embodiments, the present disclosure teaches methods, compositions, and kits for high-throughput DNA assembly reactions *in vivo* and *in vitro* utilizing modular CRISPR DNA constructs.

[0010] Thus, in some embodiments, the present disclosure teaches CRISPR DNA constructs comprising modular insert DNA parts flanked by cloning tag segments comprising pre-validated CRISPR protospacer/protospacer adjacent motif (PAM) sequence combinations. In some embodiments the present disclosure teaches digesting DNA with CRISPR endonucleases. In some embodiments, the present disclosure teaches digesting DNA with Type II - Class 2 CRISPR endonucleases (e.g. Cas9). In some embodiments, the present disclosure teaches digesting DNA with Type V - Class 2 CRISPR endonucleases. In some embodiments, the present disclosure teaches digesting DNA with Cpf1 endonucleases.

[0011] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct comprising a CRISPR multi-clonal site, said multi-clonal site comprising: a) at least two distinct cloning tags (cTAG), wherein each cTAG comprises: i) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least one of said validated CRISPR landing sites is unique within

the modular CRISPR DNA construct; and b) one or more DNA insert sequences; i) wherein each of said cTAGs are distributed in flanking positions around each of the one or more DNA insert sequences; and ii) wherein at least one of said DNA insert sequences comprises a selection marker.

[0012] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein said modular CRISPR DNA construct is circular.

[0013] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein said modular CRISPR DNA construct is linear.

[0014] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein said modular CRISPR DNA construct is integrated into the genome of an organism.

[0015] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein at least one of said distinct cTAGs comprises at least two validated CRISPR landing sites.

[0016] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein at least one of the CRISPR landing sites is for a Cas9 endonuclease.

[0017] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein at least one of the CRISPR landing sites is for a Cpf1 endonuclease.

[0018] In some embodiments, the present disclosure teaches a recombinant modular CRISPR DNA construct, wherein at least one of said distinct cTAGs comprises a rare (≥ 8 bases long) restriction endonuclease site.

[0019] In some aspects, the disclosure refers to a recombinant modular CRISPR DNA construct as a "MegaModular" construct.

[0020] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, the method comprising: a) forming a mixture comprising: i) a plurality of DNA insert parts, wherein each DNA insert part is flanked by two cloning tags (cTAGs), each cTAG comprising: 1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); ii) one or more CRISPR complexes targeting at least one of said cTAGs present in at least two of the plurality of DNA insert parts, each CRISPR complex comprising: 1) a CRISPR endonuclease, and 2) a guide RNA

or guide RNAs capable of recruiting said CRISPR endonuclease to one of said targeted cTAGs; wherein the mixture is incubated under conditions which allow for digestion of the targeted cTAG(s) in at least two of the plurality of DNA insert parts to generate overhanging ends, and b) incubating the digestion products generated in (a) in conditions which allow for hybridization of compatible overhanging ends and covalent joining of the hybridized ends; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the original insert parts that are ligated in the method.

[0021] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the CRISPR endonuclease is Cpf1.

[0022] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the CRISPR endonuclease is Cas9.

[0023] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the method comprises the step of: i) separating the digested cTAG sequences from the CRISPR complexes prior to ligation, or ii) inactivating the CRISPR complexes prior to ligation.

[0024] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the separation step comprises a DNA purification step.

[0025] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the inactivation step comprises heat or chemical inactivation of said CRISPR complexes.

[0026] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein the two cTAGs for each of the plurality of DNA insert parts form a cTAG pair, and wherein said cTAG pair is unique from all other cTAG pairs of the DNA insert parts that are ligated in the method.

[0027] In some embodiments, the present disclosure teaches a method for preparing a recombinant nucleic acid molecule, wherein at least one of the cTAGs in each cTAG pair is the same as at least one other cTAG in a different cTAG pair.

[0028] In some embodiments, the present disclosure teaches a method for DNA sequence editing, said method comprising: a) introducing into a reaction: i) the modular CRISPR DNA construct of

the present disclosure: ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG; 1) wherein the first insert cTAG comprises the validated CRISPR landing site(s) of one of the distinct cTAGs of the modular CRISPR DNA construct, and the second insert cTAG comprises the validated CRISPR landing site(s) of another distinct cTAG of the modular CRISPR DNA construct; and iii) a first and second CRISPR complex targeting the first and second insert cTAGs, respectively, each CRISPR complex comprising: 1) a CRISPR endonuclease, and 2) a guide RNA capable of recruiting said CRISPR endonuclease to one of said targeted insert cTAGs; wherein the first and second CRISPR complexes cleave the first and second insert cTAGs and their corresponding distinct cTAGs to generate overhanging ends, and b) incubating the replacement DNA insert part and modular CRISPR DNA construct with digested cTAGs generated: (a) under conditions which allow for hybridization of compatible overhanging ends and covalent joining of the hybridized ends; wherein the resulting edited modular CRISPR DNA construct comprises the complete cTAG sequences of the original insert part that is ligated by the method.

[0029] In some embodiments, the present disclosure teaches a method for DNA sequence editing, wherein the reaction of step (b) comprises a functional ligase.

[0030] In some embodiments, the present disclosure teaches a method for DNA sequence editing, said method comprising: a) introducing into a reaction: i) the modular CRISPR DNA construct of the present disclosure; ii) at least two CRISPR complexes targeting two distinct cTAGs in the modular CRISPR DNA construct, each CRISPR complex comprising; 1) a CRISPR endonuclease, and 2) a guide RNA capable of recruiting said CRISPR endonuclease to one of said targeted distinct cTAGs; wherein the first and second CRISPR complexes cleave the two distinct cTAGs in the modular CRISPR DNA construct, wherein the resulting distinct cTAGs comprise overhang ends, and b) introducing into a second reaction: i) the modular CRISPR DNA construct with digested cTAGs generated in (a); and ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG; 1) wherein the first insert cTAG comprises the polynucleotide sequence of one of the undigested distinct cTAGs that is cleaved in (a), and the second insert cTAG comprises the polynucleotide sequence of another undigested distinct cTAG that is cleaved in (a); and 2) wherein the first and second insert cTAGs comprise overhang ends that are compatible with the overhang ends of the distinct cTAGs from (a); under conditions which allow for hybridization of compatible the overhanging ends and covalent joining

of the hybridized ends; wherein the resulting edited modular CRISPR DNA construct comprises the complete sequences of the original undigested distinct cTAGs that were targeted in (a).

[0031] In some embodiments, the present disclosure teaches a method for DNA sequence editing, wherein the reaction of step (b) comprises a functional ligase.

[0032] In some embodiments, the present disclosure teaches a method for DNA sequence editing, wherein the CRISPR endonuclease is Cpf1.

[0033] In some embodiments, the present disclosure teaches a method for DNA sequence editing, wherein step (a) further comprises digesting the two cleaved distinct cTAGs with a single stranded exonuclease, thereby producing the distinct cTAGs with overhang ends. In some aspects, one may add a ligase and polymerase to repair the junctions with a polymerase and ligase after the exonuclease step. In some aspects, this reaction can also be done with Cas9 digested, blunt-end cuts.

[0034] In some embodiments, the present disclosure teaches a method for DNA sequence editing, wherein the CRISPR endonuclease is Cas9.

[0035] In some embodiments, the disclosure provides for a host cell genome comprising a recombinant modular CRISPR DNA construct comprising a CRISPR multi-clonal site, said multi-clonal site comprising: a) at least two distinct cloning tag (cTAG), wherein each cTAG comprises: i) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least one of said validated CRISPR landing sites is unique within the modular CRISPR DNA construct; and b) one or more DNA insert part(s); i) wherein each of said distinct cTAGs are distributed in flanking positions around each of the one or more DNA insert part(s).

[0036] In some embodiments, the disclosure provides for a method for preparing a recombinant nucleic acid molecule, the method comprising: a) incubating a mixture comprising: i) a plurality of DNA insert parts flanked by two cloning tags (cTAGs), each cTAG comprising: 1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); and 2) a rare (≥ 8 base) restriction enzyme recognition site; wherein at least one of the cTAGs of at least two insert parts comprise the same restriction enzyme site; ii) one or more restriction enzymes targeting the rare restriction enzyme sites in at least two

of the plurality of DNA insert parts; under conditions which allow for digestion of the targeted cTAG by the one or more restriction enzymes in at least two of the plurality of DNA insert parts to generate insert parts with digested DNA ends; and b) incubating the DNA insert part(s) with digested DNA ends generated in step (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the original insert parts that are covalently joined in the method.

[0037] In some embodiments, the disclosure provides for a method for DNA sequence editing, said method comprising: a) providing: i) the modular CRISPR DNA construct of claim 1, wherein at least two of the distinct cTAGs comprise a rare (≥ 8 base) restriction enzyme recognition site; ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG; 1) wherein the first insert cTAG comprises the rare restriction enzyme recognition site of one of the distinct cTAGs of the modular CRISPR DNA construct, and the second insert cTAG comprises the rare restriction enzyme recognition site of another distinct cTAG of the modular CRISPR DNA construct; and iii) one or more restriction enzymes targeting the rare restriction enzyme sites in the first and second insert cTAGs; wherein parts (i) and (ii) are each incubated with part (iii) in a single or separate reactions; wherein the one or more restriction enzymes cleave the rare restriction enzyme recognition sites of first and second insert cTAGs and their corresponding distinct cTAGs to generate digested DNA ends, and b) incubating the replacement DNA insert part and modular CRISPR DNA construct with digested DNA ends generated in step (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting edited modular CRISPR DNA construct comprises the complete cTAG sequences of the original insert part that is covalently joined by the method.

[0038] In some embodiments, the disclosure provides for a method for preparing a recombinant nucleic acid molecule, the method comprising: a) incubating a mixture comprising: i) a plurality of DNA insert parts, wherein each DNA insert part is flanked by two cloning tags (cTAGs), each cTAG comprising: 1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least two of the DNA insert parts share the same cTAG; ii) a single stranded DNA (ssDNA) exonuclease; under conditions which allow for digestion of the shared cTAG in the least two DNA insert parts, thereby generating compatible overhang DNA ends in the at least two DNA insert parts, and b) incubating the DNA insert parts with digested cTAGs generated in (a) under conditions which allow for the

hybridization and covalent joining of the compatible overhang DNA ends of the least two DNA insert parts; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the shared cTAG before digestion. This reaction can also be conducted with a polymerase and or ligase that are used to fix junctions. Further, this can be carried out with a predigested vector.

BRIEF DESCRIPTION OF THE FIGURES

[0039] **Figure 1 A – C** illustrates a comparison of the CRISPR/Cas9 and CRISPR/Cpf1 systems of the present disclosure. **Figure 1A** – Cas9 endonucleases are recruited to target dsDNA by tracrRNA and crRNA complexes. **Figure 1B** – Cas9 endonucleases may also be recruited to target dsDNA by artificially fused tracrRNA and crRNA sequences known as single-guide RNAs (sgRNAs). Cas9 endonuclease produces blunt ends. **Figure 1C** – Cpf1 endonucleases only require crRNA guide poly-ribonucleotides. Cpf1 endonuclease cleavage produces double stranded breaks with 5' overhangs.

[0040] **Figure 2 A – C** illustrates an embodiment of the present cloning methods utilizing modular CRISPR constructs of the present disclosure. **Figure 2A** – diagrams a modular CRISPR plasmid that can be easily altered with Cas9 or Cpf1 nucleases, according to the present disclosure. As aforementioned, the modular CRISPR constructs of the disclosure can be termed “MegaModular” constructs. Interchangeable parts represented by numbers are flanked by invariant cTAG sequences represented by letters. Parts may come pre-assembled, or may be assembled *in vitro* based on cTAG sequence identity. Example insert parts are shown on the right of **Figure 2A**. **Figure 2B** – Several strategies such as Cas9, Cpf1, or restriction endonuclease cleavage at cTAGs may be used to replace individual parts without having to reassemble the entire plasmid. cTAG sequences may comprise one or more cloning sites, including, but not limited to Cas9, Cpf1, restriction, and/or recombination sites. **Figure 2C** – Once integrated into the genome of an organism, cTAGs may continue to serve as pre-validated Cas9 or Cpf1 landing sites, enabling replacement, insertion, or removal of genomically integrated DNA with prevalidated and orthogonal gRNA sequences.

[0041] **Figure 3 A – D** illustrates an embodiment of the present cloning methods utilizing modular CRISPR constructs of the present disclosure. **Figure 3A** – diagrams a modular CRISPR plasmid that can be easily altered with Cas9 or Cpf1 nucleases, according to present disclosure.

Interchangeable parts represented by numbers are flanked by invariant cTAG sequences represented by letters. Parts may come pre-assembled, or may be assembled *in vivo* or *in vitro* based on cTAG sequence identity. Example insert parts are shown on the right of **Figure 3A**. **Figure 3B** – Several strategies such as Cas9, Cpf1, or restriction endonuclease cleavage at cTAGs may be used to replace individual parts without having to reassemble the entire plasmid. cTAG sequences may comprise one or more cloning sites, including, but not limited to Cas9, Cpf1, restriction, and/or recombination sites. **Figure 3C** – Illustrates methods of the present disclosure for removing insert parts, or for adding stuffer sequences from existing modular plasmids. **Figure 3D** – Insert parts of the modular plasmids of the present disclosure may serve as sequences for genomic integration of a portion or the whole of the modular CRISPR vectors into the genome of a host cell.

[099] **Figure 4** illustrates the one-pot *in vitro* modular CRISPR cloning of **Example 1**. Specifically, the generation of plasmid 13001009086 by transfer of an insert from one plasmid to another in a one-pot reaction is shown. The details of this reaction are set forth in **Example 1**.

[0042] **Figure 5** illustrates an embodiment of the *in vitro* modular CRISPR cloning methods of **Example 2**. Each panel provides an illustration of the experimental design described in **Example 2**. A chloramphenicol resistance gene was cloned into a kanamycin resistant backbone plasmid to create a dual resistance plasmid. Dual resistance plasmids were then transformed into bacteria, which was subsequently cultured in media augmented with kanamycin and chloramphenicol antibiotics. Resistant colonies indicated successful Cpf1 cloning assemblies.

[0043] **Figure 6** illustrates the results of the *in vitro* modular CRISPR cloning methods of **Example 2**. The y-axis represents the number of recovered colonies growing in media augmented with kanamycin and chloramphenicol. Resistant colonies indicate successful Cpf1 cloning assemblies. The results showed a ligase-dependent assembly of dual resistance plasmids.

[0044] **Figure 7** Depicts the vector map for pJDI427. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG M and cTAG N. Relevant sequence information can be found in SEQ ID NO: 102.

[0045] **Figure 8** Depicts the vector map for pJDI429. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG N and cTAG O. Relevant sequence information can be found in SEQ ID NO: 103.

[0046] **Figure 9** Depicts the vector map for pJDI430. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG P and cTAG N. Relevant sequence information can be found in SEQ ID NO: 104.

[0047] **Figure 10** Depicts the vector map for pJDI431. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG P and cTAG O. Relevant sequence information can be found in SEQ ID NO: 105.

[0048] **Figure 11** Depicts the vector map for pJDI432. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG M and cTAG N. Relevant sequence information can be found in SEQ ID NO: 106.

[0049] **Figure 12** Depicts the vector map for pJDI434. CRISPR landing sites used in the Cpf1C assembly are labeled as cTAG N and cTAG O. Relevant sequence information can be found in SEQ ID NO: 107.

[0050] **Figure 13** Depicts the vector map for pJDI435. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG P and cTAG N. Relevant sequence information can be found in SEQ ID NO: 108.

[0051] **Figure 14** Depicts the vector map for pJDI436. CRISPR landing sites used in the Cpf1 assembly are labeled as cTAG P and cTAG O. Relevant sequence information can be found in SEQ ID NO: 109.

[0052] **Figure 15** illustrates an example gene editing of a modular CRISPR construct, according to the methods of the present disclosure. Specifically, **Figure 15** illustrates a plasmid assembly by restriction enzyme digestion and ligation using the megamodular design of **Example 3**. **Figure 15** shows that a modular CRISPR plasmid backbone p1300283391 and a compatible GFP-containing insert DNA part are each digested with ApaI and PvuI restriction enzymes to create compatible cloning tag ends. The digested backbone and insert are ligated in vitro to create a new modular CRISPR construct

DETAILED DESCRIPTION

Definitions

[0053] While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter.

[0054] The term “a” or “an” refers to one or more of that entity, *i.e.*, can refer to a plural referent. As such, the terms “a” or “an”, “one or more” and “at least one” are used interchangeably herein. In addition, reference to “an element” by the indefinite article “a” or “an” does not exclude the possibility that more than one of the elements is present, unless the context clearly requires that there is one and only one of the elements.

[0055] The term “prokaryotes” is art recognized and refers to cells which contain no nucleus. The prokaryotes are generally classified in one of two domains, the Bacteria and the Archaea. The definitive difference between organisms of the Archaea and Bacteria domains is based on fundamental differences in the nucleotide base sequence in the 16S ribosomal RNA.

[0056] A “eukaryote” is any organism whose cells contain a nucleus and other organelles enclosed within membranes. Eukaryotes belong to the taxon Eukarya or Eukaryota. The defining feature that sets eukaryotic cells apart from prokaryotic cells (the aforementioned Bacteria and Archaea) is that they have membrane-bound organelles, especially the nucleus, which contains the genetic material, and is enclosed by the nuclear envelope.

[0057] The term “Archaea” refers to a categorization of organisms of the division Mendosicutes, typically found in unusual environments and distinguished from the rest of the prokaryotes by several criteria, including the number of ribosomal proteins and the lack of muramic acid in cell walls. On the basis of *ssrRNA* analysis, the Archaea consist of two phylogenetically-distinct groups: Crenarchaeota and Euryarchaeota. On the basis of their physiology, the Archaea can be organized into three types: methanogens (prokaryotes that produce methane); extreme halophiles (prokaryotes that live at very high concentrations of salt (NaCl)); and extreme (hyper) thermophilus (prokaryotes that live at very high temperatures). Besides the unifying archaeal features that distinguish them from Bacteria (*i.e.*, no murein in cell wall, ester-linked membrane lipids, etc.), these prokaryotes exhibit unique structural or biochemical attributes which adapt them to their particular habitats. The Crenarchaeota consists mainly of hyperthermophilic sulfur-dependent prokaryotes and the Euryarchaeota contains the methanogens and extreme halophiles.

[0058] “Bacteria” or “eubacteria” refers to a domain of prokaryotic organisms. Bacteria include at least 11 distinct groups as follows: (1) Gram-positive (gram+) bacteria, of which there are two major subdivisions: (1) high G+C group (Actinomycetes, Mycobacteria, Micrococcus, others) (2) low G+C group (Bacillus, Clostridia, Lactobacillus, Staphylococci, Streptococci, Mycoplasmas); (2) Proteobacteria, *e.g.*, Purple photosynthetic + non-photosynthetic Gram-negative bacteria (includes most “common” Gram-negative bacteria); (3) Cyanobacteria, *e.g.*, oxygenic phototrophs; (4) Spirochetes and related species; (5) Planctomyces; (6) Bacteroides, Flavobacteria; (7) Chlamydia; (8) Green sulfur bacteria; (9) Green non-sulfur bacteria (also anaerobic phototrophs); (10) Radioresistant micrococci and relatives; (11) Thermotoga and Thermosipho thermophiles.

[0059] The terms “genetically modified host cell,” “recombinant host cell,” and “recombinant strain” are used interchangeably herein and refer to host cells that have been genetically modified by the cloning and transformation methods of the present disclosure. Thus, the terms include a host cell (*e.g.*, bacteria, yeast cell, fungal cell, CHO, human cell, etc.) that has been genetically altered, modified, or engineered, such that it exhibits an altered, modified, or different genotype and/or phenotype (*e.g.*, when the genetic modification affects coding nucleic acid sequences of the microorganism), as compared to the naturally-occurring microorganism from which it was derived. It is understood that the terms refer not only to the particular recombinant microorganism in question, but also to the progeny or potential progeny of such a microorganism.

[0060] The term “genetically engineered” may refer to any manipulation of a host cell’s genome (*e.g.* by insertion or deletion of nucleic acids).

[0061] As used herein, “selectable marker” is a nucleic acid segment that allows one to select for a molecule (*e.g.*, a replicon) or a cell that contains it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like. Examples of selectable markers include but are not limited to: (1) nucleic acid segments that encode products which provide resistance against otherwise toxic compounds (*e.g.*, antibiotics); (2) nucleic acid segments that encode products which are otherwise lacking in the recipient cell (*e.g.*, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products which suppress the activity of a gene product; (4) nucleic acid segments that

encode products which can be readily identified (*e.g.*, phenotypic markers such as β -galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that encode products that bind other products which are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that encode nucleic acids that otherwise inhibit the activity of any of the nucleic acid segments resulting in a visible or selectable phenotype (*e.g.*, antisense oligonucleotides); (7) nucleic acid segments that encode products that bind other products that modify a substrate (*e.g.* restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (*e.g.* specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence which can be otherwise non-functional (*e.g.*, for PCR amplification of subpopulations of molecules); and (10) nucleic acid segments, which when absent, directly or indirectly confer resistance or sensitivity to particular compounds.

[0062] As used herein, “counterselectable marker” or a “counterselection marker” is a nucleic acid segment that eliminates or inhibits growth of a host organism upon selection. In some embodiments, the counterselectable markers of the present disclosure render the cells sensitive to one or more chemicals/growth conditions/genetic backgrounds. In some embodiments, the counterselectable markers of the present disclosure are toxic genes. In some embodiments, the counterselectable markers are expressed by inducible promoters.

[0063] As used herein, the term “nucleic acid” refers to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides, or analogs thereof. This term refers to the primary structure of the molecule, and thus includes double- and single-stranded DNA, as well as double- and single-stranded RNA. It also includes modified nucleic acids such as methylated and/or capped nucleic acids, nucleic acids containing modified bases, backbone modifications, and the like. The terms “nucleic acid” and “nucleotide sequence” are used interchangeably.

[0064] As used herein, the term “gene” refers to any segment of DNA associated with a biological function. Thus, genes include, but are not limited to, coding sequences and/or the regulatory sequences required for their expression. Genes can also include non-expressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

[0065] As used herein, the term “homologous” or “homolog” or “ortholog” is known in the art and refers to related sequences that share a common ancestor or family member and are determined based on the degree of sequence identity. The terms “homology,” “homologous,” “substantially similar” and “corresponding substantially” are used interchangeably herein. They refer to nucleic acid fragments wherein changes in one or more nucleotide bases do not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms also refer to modifications of the nucleic acid fragments of the instant disclosure such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the disclosure encompasses more than the specific exemplary sequences. These terms describe the relationship between a gene found in one species, subspecies, variety, cultivar or strain and the corresponding or equivalent gene in another species, subspecies, variety, cultivar or strain. For purposes of this disclosure homologous sequences are compared. “Homologous sequences” or “homologs” or “orthologs” are thought, believed, or known to be functionally related. A functional relationship may be indicated in any one of a number of ways, including, but not limited to: (a) degree of sequence identity and/or (b) the same or similar biological function. Preferably, both (a) and (b) are indicated. Homology can be determined using software programs readily available in the art, such as those discussed in Current Protocols in Molecular Biology (F.M. Ausubel *et al.*, eds., 1987) Supplement 30, section 7.718, Table 7.71. Some alignment programs are MacVector (Oxford Molecular Ltd, Oxford, U.K.), ALIGN Plus (Scientific and Educational Software, Pennsylvania) and AlignX (Vector NTI, Invitrogen, Carlsbad, CA). Another alignment program is Sequencher (Gene Codes, Ann Arbor, Michigan), using default parameters.

[0066] As used herein, the term “nucleotide change” refers to, *e.g.*, nucleotide substitution, deletion, and/or insertion, as is well understood in the art. For example, mutations contain alterations that produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded protein or how the proteins are made.

[0067] As used herein, the term “protein modification” refers to, *e.g.*, amino acid substitution, amino acid modification, deletion, and/or insertion, as is well understood in the art.

[0068] As used herein, the term “at least a portion” or “fragment” of a nucleic acid or polypeptide means a portion having the minimal size characteristics of such sequences, or any larger fragment of the full length molecule, up to and including the full length molecule. A fragment of a polynucleotide of the disclosure may encode a biologically active portion of a genetic regulatory element. A biologically active portion of a genetic regulatory element can be prepared by isolating a portion of one of the polynucleotides of the disclosure that comprises the genetic regulatory element and assessing activity as described herein. Similarly, a portion of a polypeptide may be 4 amino acids, 5 amino acids, 6 amino acids, 7 amino acids, and so on, going up to the full length polypeptide. The length of the portion to be used will depend on the particular application. A portion of a nucleic acid useful as a hybridization probe may be as short as 12 nucleotides; in some embodiments, it is 20 nucleotides. A portion of a polypeptide useful as an epitope may be as short as 4 amino acids. A portion of a polypeptide that performs the function of the full-length polypeptide would generally be longer than 4 amino acids.

[0069] For PCR amplifications of the polynucleotides disclosed herein, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any organism of interest. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook *et al.* (2001) *Molecular Cloning: A Laboratory Manual* (3rd ed., Cold Spring Harbor Laboratory Press, Plainview, New York). See also Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

[0070] The term “primer” as used herein refers to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach, thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product is induced, *i.e.*, in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension

products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and composition (A/T vs. G/C content) of primer. A pair of bi-directional primers consists of one forward and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

[0071] The terms “stringency” or “stringent hybridization conditions” refer to hybridization conditions that affect the stability of hybrids, *e.g.*, temperature, salt concentration, pH, formamide concentration and the like. These conditions are empirically optimized to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid sequence. The terms as used include reference to conditions under which a probe or primer will hybridize to its target sequence, to a detectably greater degree than other sequences (*e.g.* at least 2-fold over background). Stringent conditions are sequence dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent 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 a complementary target sequence hybridizes to a perfectly matched probe or primer. Typically, stringent conditions will be those in which the salt concentration is 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 at least about 30° C for short probes or primers (*e.g.* 10 to 50 nucleotides) and at least about 60° C for long probes or primers (*e.g.* greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or “conditions of reduced stringency” include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37° C and a wash in 2×SSC at 40° C. Exemplary high stringency conditions include hybridization in 50% formamide, 1M NaCl, 1% SDS at 37° C, and a wash in 0.1×SSC at 60° C. Hybridization procedures are well known in the art and are described by *e.g.* Ausubel *et al.*, 1998 and Sambrook *et al.*, 2001. In some embodiments, stringent conditions are hybridization in 0.25 M Na_2HPO_4 buffer (pH 7.2) containing 1 mM Na_2EDTA , 0.5-20% sodium dodecyl sulfate at 45° C, such as 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19% or 20%, followed by a wash in 5×SSC, containing 0.1% (w/v) sodium dodecyl sulfate, at 55° C to 65° C.

[0072] As used herein, the term “substantially identical” refers to two polynucleotide sequences which vary in no more than 1, 2, 3, 4, 5, 6, or 7 nucleotides. When used in the context of cTAGs, the term substantially identical denotes two cTAGs that would be identical, except for a mutation in the PAM or protospacer region of on one of the cTAGs designed to abrogate CRISPR cleavage in at least one CRISPR landing site. When the term substantially identical is used in conjunction with the term “partial” sequence or cTAG, the combination refers to the comparison between two substantially identical cTAGs as described above, wherein one of the cTAGs has been digested by a CRISPR endonuclease. Thus the term would be used to indicate that the cTAG being described was identical to a second cTAG (in its undigested form), except for the mutation in the PAM or protospacer region.

[0073] As used herein, the term “promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence may consist of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter.

[0074] As used herein, the term “heterologous” refers to a nucleic acid sequence which is not naturally found in the particular organism.

[0075] As used herein, the term “endogenous,” “endogenous gene,” refers to the naturally occurring copy of a gene.

[0076] As used herein, the term “naturally occurring” refers to a gene or sequence derived from a naturally occurring source. In some embodiments a naturally occurring gene refers to a gene of a wild type (non-transgene) gene, whether located in its endogenous setting within the source organism, or if placed in a “heterologous” setting, when introduced in a different organism. Thus, for the purposes of this disclosure, a “non-naturally occurring” sequence is a sequence that has been synthesized, mutated, or otherwise modified to have a different sequence from known natural sequences. In some embodiments, the modification may be at the protein level (*e.g.*, amino acid substitutions). In other embodiments, the modification may be at the DNA level, without any effect on protein sequence (*e.g.*, codon optimization). In some embodiments, the non-naturally occurring sequence may be a construct.

[0077] As used herein, the term “exogenous” is used interchangeably with the term “heterologous,” and refers to a substance coming from some source other than its native source. For example, the terms “exogenous protein,” or “exogenous gene” refer to a protein or gene from a non-native source or location, and that have been artificially supplied to a biological system. Artificially mutated variants of endogenous genes are considered “exogenous” for the purposes of this disclosure.

[0078] As used herein, the phrases “recombinant construct”, “expression construct”, “chimeric construct”, “construct”, and “recombinant DNA construct” are used interchangeably herein. A recombinant construct comprises an artificial combination of nucleic acid fragments, *e.g.*, regulatory and coding sequences that are not found together in nature. For example, a chimeric construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. Such construct may be used by itself or may be used in conjunction with a vector. If a vector is used then the choice of vector is dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments of the disclosure. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones *et al.*, (1985) EMBO J. 4:2411-2418; De Almeida *et al.*, (1989) Mol. Gen. Genetics 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others. Vectors can be plasmids, viruses, bacteriophages, pro-viruses, phagemids, transposons, artificial chromosomes, and the like, that replicate autonomously or can integrate into a chromosome of a host cell. A vector can also be a naked RNA polynucleotide, a naked DNA polynucleotide, a polynucleotide composed of both DNA and RNA within the same strand, a poly-lysine-conjugated DNA or RNA, a peptide-conjugated DNA or RNA, a liposome-conjugated DNA, or the like, that is not autonomously replicating. As used herein, the term “expression” refers to the production of a functional end-product *e.g.*, an mRNA or a protein (precursor or mature).

[0079] The term “operably linked” means in the context the sequential arrangement of the promoter polynucleotide according to the disclosure with a further oligo- or polynucleotide, resulting in transcription of said further polynucleotide. In some embodiments, the promoter sequences of the present disclosure are inserted just prior to a gene’s 5’UTR, or open reading frame. In other embodiments, the operably linked promoter sequences and gene sequences of the present disclosure are separated by one or more linker nucleotides. The term “operably linked” in the context of CRISPR protospacers and protospacer adjacent motifs (PAMs) refers to a proximately placed protospacer/PAM combination sequence that is capable of being cleaved at high efficiency by a CRISPR endonuclease complex.

[0080] The term “CRISPR RNA” or “crRNA” refers to the RNA strand responsible for hybridizing with target DNA sequences, and recruiting CRISPR endonucleases. crRNAs may be naturally occurring, or may be synthesized according to any known method of producing RNA.

[0081] The term “guide sequence” or “spacer” refers to the portion of a crRNA or guide RNA (gRNA) that is responsible for hybridizing with the target DNA.

[0082] The term “protospacer” refers to the DNA sequence targeted by a crRNA or guide strand. In some embodiments, the protospacer sequence hybridizes with the crRNA guide sequence of a CRISPR complex.

[0083] The term “seed region” refers to the ribonucleic sequence responsible for initial complexation between a DNA sequence CRISPR ribonucleoprotein complex. Mismatches between the seed region and a target DNA sequence have a stronger effect on target site recognition and cleavage than the remainder of the crRNA/sgRNA sequence. In some embodiments a single mismatch in the seed region of a crRNA/gRNA can render a CRISPR complex inactive at that binding site. In some embodiments, the seed regions for Cas9 endonucleases are located along the last ~12 nt of the 3’ portion of the guide sequence, which correspond (hybridize) to the portion of the protospacer target sequence that is adjacent to the PAM. In some embodiments, the seed regions for Cpf1 endonucleases are located along the first ~5 nt of the 5’ portion of the guide sequence, which correspond (hybridize) to the portion of the protospacer target sequence adjacent to the PAM.

[0084] The term “tracrRNA” refers to a small trans-encoded RNA. TracrRNA is complementary to and basepairs with crRNA to form a crRNA/tracrRNA hybrid, capable of recruiting CRISPR endonucleases to target sequences.

[0085] The term “Guide RNA” or “gRNA” as used herein refers to an RNA sequence or combination of sequences capable of recruiting a CRISPR endonuclease to a target sequence. Thus as used herein, a guide RNA can be a natural or synthetic crRNA (*e.g.*, for Cpf1), a natural or synthetic crRNA/tracrRNA hybrid (*e.g.*, for Cas9), or a single-guide RNA (sgRNA).

[0086] The term “CRISPR landing site” as used herein, refers to a DNA sequence capable of being targeted by a CRISPR complex. Thus, in some embodiments, a CRISPR landing site comprises a proximately placed protospacer/Protospacer Adjacent Motif combination sequence that is capable of being cleaved a CRISPR endonuclease complex. The term “validated CRISPR landing site” refers to a CRISPR landing site for which there exists a guide RNA capable of inducing high efficiency cleaving of said sequence. Thus, the term validated should be interpreted as meaning that the sequence has been previously shown to be cleavable by a CRISPR complex. Each “validated CRISPR landing site” will by definition confirm the existence of a tested guide RNA associated with the validation.

[0087] The term “sticky end(s)” refers to double stranded polynucleotide molecule end that comprises a sequence overhang. In some embodiments, the sticky end can be a dsDNA molecule end with a 5' or 3' sequence overhang. In some embodiments, the sticky ends of the present disclosure are capable of hybridizing with compatible sticky ends of the same or other molecules. Thus, in one embodiment, a sticky end on the 3' of a first DNA fragment may hybridize with a compatible sticky end on a second DNA fragment. In some embodiments, these hybridized sticky ends can be sewn together by a ligase. In other embodiments, the sticky ends might require extension of the overhangs to complete the dsDNA molecule prior to ligation. The term “genetic scar(s)” refers to any undesirable sequence introduced into a nucleic acid sequence by DNA manipulation methods. For example, in some embodiments, the present disclosure teaches genetic scars such as restriction enzyme binding sites, sequence adapters or spacers to accommodate cloning, TA-sites, scars left over from NHEJ, etc. In some embodiments, the present disclosure teaches methods of scarless cloning and gene editing.

[0088] As used herein the term “targeted” refers to the expectation that one item or molecule will interact with another item or molecule with a degree of specificity, so as to exclude non-targeted items or molecules. For example, a first polynucleotide that is targeted to a second polynucleotide, according to the present disclosure has been designed to hybridize with the second polynucleotide in a sequence specific manner (*e.g.*, *via* Watson-Crick base pairing). In some embodiments, the selected region of hybridization is designed so as to render the hybridization unique to the one, or more targeted regions. A second polynucleotide can cease to be a target of a first targeting polynucleotide, if its targeting sequence (region of hybridization) is mutated, or is otherwise removed/separated from the second polynucleotide.

[0089] The disclosure refers to the taught and described universal modular CRISPR DNA constructs or designs as a “MegaModular” construct or design.

DNA Nucleases

[0090] In some embodiments, the present disclosure teaches methods and compositions for gene editing/cloning utilizing DNA nucleases. CRISPR complexes, transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs), and FokI restriction enzymes are some of the sequence-specific nucleases that have been used as gene editing tools. These enzymes are able to target their nuclease activities to desired target loci through interactions with guide regions engineered to recognize sequences of interest. In some embodiments, the present disclosure teaches CRISPR-based gene editing methods

[0091] The principles of *in vivo* CRISPR-based editing largely rely on natural cellular DNA repair systems. Double-stranded dsDNA breaks introduced by nucleases are repaired by either non-homologous end-joining (NHEJ) or homology-directed repair (HDR), or single strand annealing, (SSA), or microhomology end joining (MMEJ).

[0092] HDR relies on a template DNA containing sequences homologous to the region surrounding the targeted site of DNA cleavage. Cellular repair proteins use the homology between the exogenously supplied or endogenous DNA sequences and the site surrounding a DNA break to repair the dsDNA break, replacing the break with the sequence on the template DNA. Failure to integrate the template DNA however, can result in NHEJ, MMEJ, or SSA. NHEJ, MMEJ and SSA are error-prone processes that are often accompanied by insertion or deletion of nucleotides (indels) at the target site, resulting in genetic knockout (silencing) of the targeted region of the

genome due to frameshift mutations or insertions of a premature stop codon. Cpf1-mediated editing can also function *via* traditional hybridization of overhangs created by the endonuclease, followed by ligation.

[0093] CRISPR endonucleases are also useful for *in vitro* DNA manipulations, as discussed in later sections of this disclosure.

CRISPR Systems

[0094] CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (cas) endonucleases were originally discovered as adaptive immunity systems evolved by bacteria and archaea to protect against viral and plasmid invasion. Naturally occurring CRISPR/Cas systems in bacteria are composed of one or more Cas genes and one or more CRISPR arrays consisting of short palindromic repeats of base sequences separated by genome-targeting sequences acquired from previously encountered viruses and plasmids (called spacers). (Wiedenheft, B., *et al.* Nature. 2012; 482:331; Bhaya, D., *et al.*, Annu. Rev. Genet. 2011; 45:231; and Terms, M.P. *et al.*, Curr. Opin. Microbiol. 2011; 14:321). Bacteria and archaea possessing one or more CRISPR loci respond to viral or plasmid challenge by integrating short fragments of foreign sequence (protospacers) into the host chromosome at the proximal end of the CRISPR array. Transcription of CRISPR loci generates a library of CRISPR-derived RNAs (crRNAs) containing sequences complementary to previously encountered invading nucleic acids (Haurwitz, R.E., *et al.*, Science. 2012:329;1355; Gesner, E.M., *et al.*, Nat. Struct. Mol. Biol. 2001:18;688; Jinek, M., *et al.*, Science. 2012:337; 816-21). Target recognition by crRNAs occurs through complementary base pairing with target DNA, which directs cleavage of foreign sequences by means of Cas proteins. (Jinek *et al.* 2012 “A Programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” Science. 2012:337; 816-821).

[0095] There are at least five main CRISPR system types (Type I, II, III, IV and V) and at least 16 distinct subtypes (Makarova, K.S., *et al.*, Nat Rev Microbiol. 2015. Nat. Rev. Microbiol. 13, 722–736). CRISPR systems are also classified based on their effector proteins. Class 1 systems possess multi-subunit crRNA-effector complexes, whereas in class 2 systems all functions of the effector complex are carried out by a single protein (*e.g.*, Cas9 or Cpf1). In some embodiments, the present disclosure teaches using type II and/or type V single-subunit effector systems. Thus, in some embodiments, the present disclosure teaches using class 2 CRISPR systems.

CRISPR/Cas9

[0096] In some embodiments, the present disclosure teaches methods of gene editing using a Type II CRISPR system. In some embodiments, the Type II CRISPR system uses the Cas9 enzyme. Type II systems rely on a **i**) single endonuclease protein, **ii**) a transactivating crRNA (tracrRNA), and **iii**) a crRNA where a ~20-nucleotide (nt) portion of the 5' end of crRNA is complementary to a target nucleic acid. The region of a CRISPR crRNA strand that is complementary to its target DNA protospacer is hereby referred to as "guide sequence."

[0097] In some embodiments, the tracrRNA and crRNA components of a Type II system can be replaced by a single-guide RNA (sgRNA). The sgRNA can include, for example, a nucleotide sequence that comprises an at least 12-20 nucleotide sequence complementary to the target DNA sequence (guide sequence) and can include a common scaffold RNA sequence at its 3' end. As used herein, "a common scaffold RNA" refers to any RNA sequence that mimics the tracrRNA sequence or any RNA sequences that function as a tracrRNA.

[0098] Cas9 endonucleases produce blunt end DNA breaks and are recruited to target DNA by a combination of a crRNA and a tracrRNA oligos, which tether the endonuclease *via* complementary hybridization of the RNA CRISPR complex. (*see* solid triangle arrows in **Figure 1A**).

[0099] In some embodiments, DNA recognition by the crRNA/endonuclease complex requires additional complementary base-pairing with a protospacer adjacent motif (PAM) (*e.g.*, 5'-NGG-3') located in a 3' portion of the target DNA, downstream from the target protospacer. (Jinek, M., *et al.*, *Science*. 2012:337;816-821). In some embodiments, the PAM motif recognized by a Cas9 varies for different Cas9 proteins.

[0100] In some embodiments, one skilled in the art can appreciate that the Cas9 disclosed herein can be any variant derived or isolated from any source. For example, in some embodiments, the Cas9 peptide of the present disclosure can include one or more of SEQ ID Nos selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, and SEQ ID NO: 6. In other embodiments, the Cas9 peptide of the present disclosure can include one or more of the mutations described in the literature, including but not limited to the functional mutations described in: Fonfara *et al.* *Nucleic Acids Res.* 2014 Feb;42(4):2577-90; Nishimasu H. *et al.* *Cell.* 2014 Feb 27;156(5):935-49; Jinek M. *et al.* *Science.* 2012 337:816-21; and Jinek M. *et al.* *Science.* 2014 Mar 14;343(6176); *see also* U.S. Pat. App. No. 13/842,859, filed March 15, 2013,

which is hereby incorporated by reference; further, *see* U.S. Pat. Nos. 8,697,359; 8,771,945; 8,795,965; 8,865,406; 8,871,445; 8,889,356; 8,895,308; 8,906,616; 8,932,814; 8,945,839; 8,993,233; and 8,999,641, which are all hereby incorporated by reference. Thus, in some embodiments, the systems and methods disclosed herein can be used with the wild type Cas9 protein having double-stranded nuclease activity, Cas9 mutants that act as single stranded nickases, or other mutants with modified nuclease activity.

CRISPR/Cpf1

[0101] In other embodiments, the present disclosure teaches methods of gene editing using a Type V CRISPR system. In some embodiments, the present disclosure teaches methods of using CRISPR from *Prevotella* and *Francisella 1* (Cpf1).

[0102] The Cpf1 CRISPR systems of the present disclosure comprise i) a single endonuclease protein, and ii) a crRNA, wherein a portion of the 3' end of crRNA contains the guide sequence complementary to a target nucleic acid. In this system, the Cpf1 nuclease is directly recruited to the target DNA by the crRNA (*see* solid triangle arrows in **Figure 1B**). In some embodiments, guide sequences for Cpf1 must be at least 12nt, 13nt, 14nt, 15nt, or 16nt in order to achieve detectable DNA cleavage, and a minimum of 14nt, 15nt, 16nt, 17nt, or 18nt to achieve efficient DNA cleavage.

[0103] The Cpf1 systems of the present disclosure differ from Cas9 in a variety of ways. *First*, unlike Cas9, Cpf1 does not require a separate tracrRNA for cleavage. In some embodiments, Cpf1 crRNAs can be as short as about 42-44 bases long—of which 23-25 nt is guide sequence and 19 nt is the constitutive direct repeat sequence. In contrast, the combined Cas9 tracrRNA and crRNA synthetic sequences can be about 100 bases long. In some embodiments, the present disclosure will refer to a crRNA for Cpf1 as a “guide RNA.”

[0104] *Second*, Cpf1 prefers a “TTN” PAM motif that is located 5' upstream of its target. This is in contrast to the “NGG” PAM motifs located on the 3' of the target DNA for Cas9 systems. In some embodiments, the uracil base immediately preceding the guide sequence cannot be substituted (Zetsche, B. *et al.* 2015. “Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System” *Cell* 163, 759-771, which is hereby incorporated by reference in its entirety for all purposes).

[0105] *Third*, the cut sites for Cpf1 are staggered by about 3-5 bases, which create “sticky ends” (Kim *et al.*, 2016. “Genome-wide analysis reveals specificities of Cpf1 endonucleases in human

cells” published online June 06, 2016). These sticky ends with ~3-5 nt overhangs are thought to facilitate NHEJ-mediated-ligation, and improve gene editing of DNA fragments with matching ends. The cut sites are in the 3' end of the target DNA, distal to the 5' end where the PAM is. The cut positions usually follow the 18th base on the non-hybridized strand and the corresponding 23rd base on the complementary strand hybridized to the crRNA (**Figure 1B**).

[0106] *Fourth*, in Cpf1 complexes, the “seed” region is located within the first 5 nt of the guide sequence. Cpf1 crRNA seed regions are highly sensitive to mutations, and even single base substitutions in this region can drastically reduce cleavage activity (*see Zetsche B. et al. 2015 “Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System” Cell 163, 759-771*). Critically, unlike the Cas9 CRISPR target, the cleavage sites and the seed region of Cpf1 systems do not overlap. Additional guidance on designing Cpf1 crRNA targeting oligos is available on (*Zetsche B. et al. 2015. “Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System” Cell 163, 759-771*).

[0107] Persons skilled in the art will appreciate that the Cpf1 disclosed herein can be any variant derived or isolated from any source. For example, in some embodiments, the Cpf1 peptide of the present disclosure can include one or more of SEQ ID Nos selected from SEQ ID NO: 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, or any variants thereof.

Ligases

[0108] In some embodiments, the present disclosure teaches methods of cleaving target DNA *via* targeted Cpf1 complexes, and then ligating the resulting sticky ends with DNA inserts. In some embodiments, the present disclosure teaches methods of providing a Cpf1 complex to cleave the target DNA, and a ligase to “sew” the DNA back together. In other embodiments, the present disclosure teaches modified Cpf1 complexes that include a tethered ligase enzyme.

[0109] As used herein, the term “ligase” can comprise any number of enzymatic or non-enzymatic reagents. For example, ligase is an enzymatic ligation reagent or catalyst that, under appropriate conditions, forms phosphodiester bonds between the 3'-OH and the 5'-phosphate of adjacent nucleotides in DNA molecules, RNA molecules, or hybrids.

[0110] In some embodiments, the present disclosure teaches the use of enzymatic ligases. Compatible temperature sensitive enzymatic ligases, include, but are not limited to, bacteriophage T4 ligase and *E. coli* ligase. Thermostable ligases include, but are not limited to, Afu ligase, Taq ligase, Tfl ligase, Tth ligase, Tth HB8 ligase, *Thermus* species AK16D ligase and Pfu ligase (see for example Published P.C.T. Application WO/2000/026381, Wu *et al.*, *Gene*, 76(2):245-254, (1989), and Luo *et al.*, *Nucleic Acids Research*, 24(15): 3071-3078 (1996)). The skilled artisan will appreciate that any number of thermostable ligases can be obtained from thermophilic or hyperthermophilic organisms, for example, certain species of eubacteria and archaea; and that such ligases can be employed in the disclosed methods and kits. In some embodiments, reversibly inactivated enzymes (see for example U.S. Pat. No. 5,773,258) can be employed in some embodiments of the present teachings.

[0111] In other embodiments, the present disclosure teaches the use of chemical ligation agents. Chemical ligation agents include, without limitation, activating, condensing, and reducing agents, such as carbodiimide, cyanogen bromide (BrCN), N-cyanoimidazole, imidazole, 1-methylimidazole/carbodiimide/cystamine, dithiothreitol (DTT) and ultraviolet light. Autoligation, *i.e.*, spontaneous ligation in the absence of a ligating agent, is also within the scope of the teachings herein. Detailed protocols for chemical ligation methods and descriptions of appropriate reactive groups can be found in, among other places, Xu *et al.*, *Nucleic Acid Res.*, 27:875-81 (1999); Gryaznov and Letsinger, *Nucleic Acid Res.* 21:1403-08 (1993); Gryaznov *et al.*, *Nucleic Acid Res.* 22:2366-69 (1994); Kanaya and Yanagawa, *Biochemistry* 25:7423-30 (1986); Luebke and Dervan, *Nucleic Acids Res.* 20:3005-09 (1992); Sievers and von Kiedrowski, *Nature* 369:221-24 (1994); Liu and Taylor, *Nucleic Acids Res.* 26:3300-04 (1999); Wang and Kool, *Nucleic Acids Res.* 22:2326-33 (1994); Purmal *et al.*, *Nucleic Acids Res.* 20:3713-19 (1992); Ashley and Kushlan, *Biochemistry* 30:2927-33 (1991); Chu and Orgel, *Nucleic Acids Res.* 16:3671-91 (1988); Sokolova *et al.*, *FEBS Letters* 232:153-55 (1988); Naylor and Gilham, *Biochemistry* 5:2722-28 (1966); and U.S. Pat. No. 5,476,930.

[0112] In some embodiments, the methods, kits and compositions of the present disclosure are also compatible with photoligation reactions. Photoligation using light of an appropriate wavelength as a ligation agent is also within the scope of the teachings. In some embodiments, photoligation comprises probes comprising nucleotide analogs, including but not limited to, 4-thiothymidine, 5-vinyluracil and its derivatives, or combinations thereof. In some embodiments,

the ligation agent comprises: (a) light in the UV-A range (about 320 nm to about 400 nm), the UV-B range (about 290 nm to about 320 nm), or combinations thereof, (b) light with a wavelength between about 300 nm and about 375 nm, (c) light with a wavelength of about 360 nm to about 370 nm; (d) light with a wavelength of about 364 nm to about 368 nm, or (e) light with a wavelength of about 366 nm. In some embodiments, photoligation is reversible. Descriptions of photoligation can be found in, among other places, Fujimoto *et al.*, Nucl. Acid Symp. Ser. 42:39-40 (1999); Fujimoto *et al.*, Nucl. Acid Res. Suppl. 1:185-86 (2001); Fujimoto *et al.*, Nucl. Acid Suppl., 2:155-56 (2002); Liu and Taylor, Nucl. Acid Res. 26:3300-04 (1998) and on the world wide web at: sbchem.kyoto-u.ac.jp/saito-lab.

Universal Modular CRISPR DNA Constructs and Uses Thereof

[0113] In some embodiments, the present invention describes a strategy for the modular assembly of DNA constructs. In some embodiments the DNA assembly methods of the present disclosure are applicable to any construct, including plasmids, small linear DNA, and transformed chromosomal loci.

[0114] In aspects, the inventors refer to such a universal modular CRISPR DNA Construct as a “MegaModular” design.

Shortcomings in Traditional DNA Editing and Assembly Techniques

[0115] Traditional multicomponent DNA cloning strategies are limited in their ability to effectively assemble and modify multi-component DNA constructs with complex sequences. For example, restriction enzyme cloning is limited by the availability of unique restriction enzyme recognition sites that are appropriately located at the cloning junctures at each of the DNA inserts, and their destination sites within a final vector. Gateway cloning technologies are similarly limited by the relatively small number of unique recombination sites available for multi-component assemblies.

[0116] Another downside to traditional DNA assembly techniques is that their ability to edit sequences is often restricted to the time of construction. For example, the products of efficient assembly strategies such as Ligase Cycling Reactions (LCR) are not easily modified once the initial assembly is completed (Kok, S, *et al.*, 2014 “Rapid and Reliable DNA assembly *via* Ligase Cycling Reaction” ACS Synth. Biol., 3 (2): 97-106). Similar concerns arise with traditional

restriction enzyme cloning, whose common restriction recognition sites cease to function as unique cloning points once a polynucleotide containing the restriction sites is inserted into the construct being assembled, or when said construct is integrated into a chromosome full of said sites. Vectors produced through sequential restriction cloning thus provide very few options for fixing or updating sequences once the cloning process is well under way.

[0117] Even newer technologies, such as the traditional CRISPR DNA assembly techniques continue to suffer from similar complexity, the ease of iterating on a previous assembled construct/vector design, and speed limitations (Wang, JW. *et al.*, 2015 “CRISPR/Cas9 nuclease combined with Gibson assembly for seamless cloning” *BioTechniques*, Vol 58, No. 4:161-170). CRISPR cloning requires the design of a functional guide RNA targeted next to a compatible protospacer adjacent motif (PAM). Availability of suitable PAM sequences within target sites results in a significant design limitation to the number of possible DNA insertion locations within a genome or construct.

[0118] Moreover, the design and testing of guide RNA sequences imposes significant technical challenges for multi-component assemblies. Persons having skill in the art will recognize for example, that not all gRNA sequences are functional, and that effective implementation of a CRISPR DNA assembly may sometimes require the design and validation of multiple gRNA sequence variants. These limitations are particularly cumbersome in multi-component assemblies, where failure of a single gRNA sequence to successfully produce a desired modification can trigger the need to redesign subsequent assembly components that no longer fall within the original cloning plan. Applying techniques that require multiple custom guide RNAs for every junction of a multicomponent assembly can thus also be very expensive, cumbersome, and impractical.

Modular CRISPR Tag Assembly Vectors and Methods of Using Such

[0119] In some embodiments, the present disclosure teaches methods for DNA assembly that overcome many of the limitations associated with the aforementioned traditional techniques described above. In some embodiments, the present disclosure also teaches modular CRISPR assembly constructs, compositions, and kits for use with the methods of the present invention.

[0120] In some embodiments, the present disclosure teaches DNA constructs comprising one or more CRISPR multi-clonal sites (cMCS). In some embodiments, the cMCS of the present disclosure represent only a portion of the DNA constructs described (*i.e.*, only a portion of the

construct is readily editable according to the methods of the present disclosure). In other embodiments, the cMCS of the present disclosure are located on key positions within the entire construct, such that the entire DNA construct is readily editable. Thus, in some embodiments all the functional parts of the modular cTAG vectors (e.g., all origins, markers, cargo, elements required for assembly) are comprised within insert DNA parts and can be readily exchanged *via* the gene editing methods of the present disclosure.

[0121] In some embodiments, the cMCS of the present disclosure comprise one or more cloning tags (cTAG), each comprising at least one validated CRISPR targeting site. In some embodiments, the cMCS of the present disclosure further comprises DNA insert parts, each flanked by a pair of cTAGs, such that digestion of the cMCS with one or more CRISPR endonuclease targeting one or more cTAGs, will release said flanked insert part, allowing for insertion of a compatible donor DNA part.

[0122] **Figures 2 and 3** of this specification illustrate an embodiment of a modular CRISPR assembly plasmid construct, according to the methods of the present disclosure. The disclosed example plasmid contains a series of DNA insertions (Parts 1-8 in **Figure 2A**), each flanked by a pair of cTAGs (Tags A-H) in **Figure 2A**. Digestion of cTAGs A and B of this example with the appropriate CRISPR/guide sequence complexes will release Part 2 of the plasmid, allowing for insertion of a replacement part 2 insert with the desired characteristics.

[0123] Persons having skill in the art will immediately recognize the advantages of the presently described vector system, which allows for the sequence-specific modular cloning/editing of vectors *in vivo* and *in vitro*. The sections below will outline the various aspects of the disclosed modular cloning vectors, as well as their various applications to molecular biology, gene therapy, and gene editing.

Modular CRISPR Vector Insert Parts

[0124] In some embodiments, the insert parts of the present disclosure are donor DNA sequences for homologous recombination insertion following a CRISPR digestion. Thus, in some embodiments, insert part sequences of the present disclosure comprise an insert sequence of interest, flanked by sequences with sufficient homology to the ends of the digested modular CRISPR construct, so as to trigger homologous recombination, hybridization and insertion of the sequence.

[0125] In other embodiments, the insert parts of the present disclosure are donor DNA sequences capable of hybridizing and ligation *via* sticky ends (*e.g.*, following a Cpf1 digestion, restriction enzyme digestion, Gibson assembly, or other hybridization-based assembly, including LCR). Thus, in some embodiments, insert part sequences of the present disclosure comprise an insert sequence of interest, flanked by sequences with sufficient homology to the ends of the digested modular CRISPR construct, so as to allow for hybridization of sticky ends.

[0126] In yet other embodiments, the insert parts of the present disclosure are donor DNA sequences for blunt end ligation.

[0127] In some embodiments, the modular CRISPR DNA constructs of the present disclosure are compatible with any insert part sequence. Thus, the parts of the present vectors can comprise, without limitation, selectable markers, origins of replication, promoters, terminator sequences; other regulatory sequences, barcodes, recombination sites, or other sequences of interest to the user. In some embodiments, the insert parts of the present disclosure can comprise homology sequences for triggering homologous recombination and insertion into one or more genetic loci. In some embodiments, said homologous recombination insert parts will precede and follow other insert parts that will be also be inserted into the genome *via* the recombination event.

[0128] In some embodiments, the present disclosure teaches that each insert part comprises a single sequence (*e.g.*, only a promoter or only a gene of interest, see **Figure 2A**, part 8). In other embodiments, the present disclosure teaches that one or more insert parts may contain multiple elements, such as promoter-gene of interest (GOI) combinations, multi-subunit chimeric protein fusions, or even entire constructs (see **Figure 2A**, part 5, comprising a promoter-GOI-terminator combination).

[0129] In some embodiments, the present disclosure teaches uncombined individual insert parts. That is, in some embodiments, the present disclosure teaches one or a plurality of unconnected insert parts (see **Figure 2A**, right side showing a list of uncombined insert parts). In some embodiments, the present disclosure teaches methods of assembling said plurality of parts into one or more modular CRISPR constructs. In some aspects, the disclosure teaches kits for assembling a MegaModular construct.

[0130] In other embodiments, the present disclosure teaches partial- or fully- assembled modular CRISPR DNA constructs. For example, in some embodiments the present disclosure teaches

modular CRISPR DNA constructs comprising 1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, or more assembled insert parts, and any ranges therebetween. The disclosure also teaches kits comprising said insert parts.

[0131] In some embodiments, said assembled or partially assembled modular CRISPR DNA constructs are linear. In some embodiments, said assembled or partially assembled modular CRISPR DNA constructs are circular (*e.g.*, a plasmid). In some embodiments, said assembled or partially assembled modular CRISPR DNA constructs are integrated into genomic DNA.

[0132] In some embodiments, the constructs of the present disclosure will initially contain only short spacer sequences as placeholders for further cloning (*see* “stuffer” sequence in **Figure 3C**). In some embodiments the insert part placeholders are small, randomized sequences. In other embodiments, the vectors of the present disclosure will initially comprise one or more pre-selected insert DNA parts. For example, in some embodiments, the modular CRISPR constructs will initially comprise at least one selection marker, and/or at least one origin of replication.

[0133] Suitable selectable markers include, but are not limited to, genes that confer antibiotic resistance, genes that encode fluorescent proteins, tRNA genes, auxotrophic markers, toxic genes, phenotypic markers, antisense oligonucleotides, restriction endonucleases, restriction endonuclease cleavage sites, enzyme cleavage sites, protein binding sites, and sequences complementary to PCR primer sequences.

[0134] Suitable antibiotic resistance genes include, but are not limited to, a chloramphenicol resistance gene, an ampicillin resistance gene, a tetracycline resistance gene, a Zeocin resistance gene, a spectinomycin resistance gene and a kanamycin resistance gene.

[0135] In certain embodiments of the present invention, the counterselectable marker is a toxic gene. Suitable toxic genes include, but are not limited to, a *ccdB* gene, a gene encoding a *tus* protein which binds one or more *ter* sites, a *kicB* gene, a *sacB* gene, an *ASK1* gene, a Φ X174 E gene and a *DpnI* gene. In some embodiments, the presence of a toxic selectable marker serves as an indicator that an insertion was not conducted, or was unsuccessful. Toxic selectable markers

may also serve to decrease background of unmodified parent vectors of positive cells, by causing death to cells harboring unmodified vectors with the toxic gene still in place.

[0136] In additional embodiments of the methods of the present invention, the modular CRISPR constructs may comprise both one or more toxic genes and one or more antibiotic resistance genes.

[0137] In some embodiments, the modular CRISPR constructs will initially comprise at least one regulatory sequence. In some embodiments, the present disclosure teaches vectors comprising, without limitation, Matrix Attachment Regions, expression insulator sequences, expression enhancer sequences, promoters, 5' UTRs, 3' UTRs, terminator sequences, stop codons, start codons, etc. In some embodiments, the modular CRISPR constructs will initially comprise sequences for facilitating chromosomal insertion of said construct (*e.g.*, t-DNA borders, Cre/Lox, or homology ends to chromosomal sequences). In some embodiments, the sequences for chromosomal insertion are positioned so as to insert the entire modular CRISPR construct into the genome of an organism. In other embodiments, the sequences for chromosomal insertion are positioned so as to insert only a portion of the modular CRISPR construct (see **Figure 3D**).

[0138] In some embodiments, the insert parts of the present disclosure can even comprise additional cTAGs. The addition of cTAGs, through insert parts, can increase the complexity of available cloning schemes, and can also expand the size of the construct by expanding the number of available insert parts that can be replaced.

[0139] In some embodiments, the insert parts of the present disclosure can comprise a traditional cloning site. For example, in some embodiments, the present disclosure teaches insert parts comprising gateway recombination sites, restriction sites, Cre/Lox sites, or other traditional cloning sites).

[0140] In some embodiments, the present disclosure teaches methods of producing insert parts from traditional DNA constructs. That is, in some embodiments, the present disclosure teaches methods of adding cTAGs to traditional DNA constructs (*e.g.*, to oligos, PCR fragments, plasmids, or other available DNA segment). In some embodiments, the present disclosure teaches methods of adding cTAGs to a single component, such as a gene of interest (GOI), promoter. In other embodiments, the present disclosure teaches methods of adding cTAGS to multi-element constructs.

[0141] Persons having skill in the art will recognize methods for constructing insert parts. For example, in some embodiments, the cTAGs may be incorporated into a DNA molecule *via* PCR amplification with primers comprising said cTAGs. In other embodiments the cTAGs may be incorporated *via* traditional cloning techniques (*e.g.*, restriction enzymes, Gibson, or other assembly method). In yet other embodiments, the cTAGs can be incorporated *via* blunt-end ligation.

[0142] In some embodiments, the insert parts of the present disclosure can have a wide species compatibility spectrum (*e.g.*, a marker may contain both prokaryotic and eukaryotic expression sequences to make it effective in multiple organisms). In other embodiments, the insert parts of the present disclosure are designed to have limited applicability to organisms within a single species/genus/family/order/class/phylum/kingdom or domain. In some embodiments for example, an origin of replication part may be capable of maintaining a plasmid in only a single species, or a group of species. In other embodiments, a fluorescent marker may be codon optimized to function across both prokaryotic and eukaryotic domains.

[0143] In some embodiments Cas9 endonucleases cleave 3-4 nucleotides upstream from the PAM of a target sequence. cTAG digestion by a Cas9 complex can thus result in loss of cTAG functionality through the loss of the PAM sequence, or protospacer sequence of the target. In some embodiments, the present disclosure teaches methods of maintaining the functionality of said cTAG sequences by designing donor insert sequences such that they reconstitute the cTAG sequence upon insertion (*e.g.*, through insertion of the previously lost PAM or protospacer sequence). Similar provisions are envisioned for sequences cleaved through Cpf1 endonucleases.

[0144] **Figure 2B** illustrates the presently disclosed concept of cTAG repair. Cleavage of insert part 2 with a Cas9 endonuclease also results in loss of a portion of cTAGs A and B. Subsequent insertions of any one of insert parts 2a-2d *via* homologous recombination results in a restoration of the full cTAG sequence.

[0145] Persons having skill in the art will recognize the nearly infinite options for insert parts. The foregoing list of inserts was intended as illustrative, and should in no way be construed as limiting the applicability of the presently disclosed methods, kits, and constructs.

Modular CRISPR Cloning Tags

[0146] In some embodiments the modular CRISPR constructs of the present disclosure comprise one or more cloning tags (cTAGs). In some embodiments, the modular CRISPR constructs of the present disclosure comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 or more cTAGs.

[0147] In some embodiments, the present disclosure teaches that each cTAG comprises at least one validated CRISPR protospacer/PAM combination sequence (“CRISPR landing site”). That is, in some embodiments, cTAGs comprise at least one experimentally validated, high efficiency CRISPR landing site. In some embodiments, the cTAGs of the present disclosure may be validated by wet bench experimentation (*e.g.*, *in vitro* cleavage of the cTAG sequence with a CRISPR complex targeting said CRISPR landing site). In other embodiments, the cTAG validation may be assumed from reports of cleavage in peer-reviewed journals.

[0148] In some embodiments, the cTAGs of the present disclosure comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more CRISPR landing sites. In some embodiments, the CRISPR landing sites overlap with each other. In other embodiments, the CRISPR landing sites occupy distinct non-overlapping regions within the cTAG. In some embodiments, the CRISPR landing sites can be specific for either Cas9 or Cpf1 endonuclease cleavage. In some embodiments, the CRISPR landing sites can be specific to any other current or yet to be discovered CRISPR endonuclease.

[0149] In other embodiments, the present disclosure teaches that multiple cloning sites in a single cTAG can be designed to function across different organisms. Thus in some embodiments, cTAG Cpf1 landing sites may be preferred in organisms lacking or downregulating HR machinery. In other embodiments, restriction sites of a cTAG may be preferred for initial *in vitro* cloning, while Cas9 or Cpf1 landing sites may be preferred for more complex editing occurring *in vivo* in selected eukaryotic organisms.

[0150] In some embodiments, the present disclosure teaches that cTAGs may comprise one or more non-CRISPR cloning sequences. For example, in some embodiments, the cTAGs of the present disclosure may comprise one or more elements selected from the group consisting of a

restriction enzyme site, a recombination site, a topoisomerase site, a splicing site, and a Cre-Lox site.

[0151] In some embodiments, suitable restriction enzyme sites include, without limitation, sites recognized by restriction enzymes selected from the group consisting of AaII, AarI, AasI, AatII, Acc65I, AccB7I, AccI, AccIII, Acil, AclI, Acul, Adel, Afel, AfIII, AfIII, AgeI, AhdI, AleI, AloI, AluI, Alw21I, Alw26I, Alw44I, AlwI, AlwNI, ApaI, ApaLI, ApeKI, ApoI, AscI, AseI, AsiSI, AvaI, Avall, AvrII, BaeI, Ball, BamHI, BanI, BanII, BbsI, BbuI, BbvCI, BbvI, BccI, BceAI, BcgI, BciVI, BclI, Bcni, BcuI, Bfal, Bfil, Bfml BfrBI, BfuAI, BfuCI, BfuI, BglI, BglIII, BlpI, BmeI 390I, BmeI 580I, BmgBI, BmrI, BmtI, BoxI, BpiI, BpII, Bpml, Bpu10I, Bpu1102I, BpuEI, BsaAI, BsaBI, BsaHI, BsaI, BsaII, BsaMI, BsaWI, BsaXI, BseDI, BseGI, BseJI, BseLI, BseMI, BseMII, BseNI, BseRI, BseSI, BseXI, BseYI, BsgI, Bsh1236I, Bsh1285I, BshNI, BshTI, BsiEI, BsiHKAI, BsiWI, BslI, BsmAI, BsmBI, BsmFI, BsmI, BsoBI, Bsp19I, Bsp120I, Bsp1286I, Bsp1407I, Bsp143I, Bsp143II, Bsp68I, BspCNI, BspDI, BspEI, BspHI, BspLI, BspMI, BspPI, BspQI, BspTI, BsrBI, BsrDI, BsrFI, BsrGI, BsrI, BsrSI, BssHII, BssKI, BssSI, Bst1107I, Bst98I, BstAPI, BstBI, BstEII, BstF5I, BstNI, BstOI, BstUI, BstXI, BstYI, BstZI, BstZ17I, BsuI 5I, Bsu36I, BsuRI, BtgI, BtgZI, BtsCI, BtsI, BveI, Cac8I, Cail, CfoI, Cfr10I, Cfr13I, Cfr42I, Cfr9I, CfrI, ClaI, CpoI, Csp45I, Csp6I, CspI, CspCI, CviaII, CviKI-1, CviQI, DdeI, DpnI, DpnII, DraI, DraIII, DrdI, EaeI, EagI, Eam1104I, Eam1105I, EarI, EciI, Ecl136II, EclHKI, Eco105I, Eco130I, Eco147I, Eco24I, Eco31I, Eco32I, Eco47I, Eco47III, Eco52I, Eco57I, Eco57MI, Eco72I, Eco81I, Eco88I, Eco91I, EcolCRI, EcoNI, EcoO109I, EcoP15I, EcoRI, EcoRV, EheI, Esp3I, FatI, FauI, Fnu4HI, FokI, FseI, FspI, FspAI, GsuI, HaeII, HaeIII, HgaI, HhaI, Hin1I, Hin4I, Hin6I, HincII, HindIII, HinfI, HinPII, HpaI, HpaII, HphI, Hpy166II, Hpy188I, Hpy188III, Hpy8I, Hpy99I, HpyAV, HpyCH4III, HpyCH4IV, HpyCH4V, HpyF10VI, Hsp92I, Hsp92II, I-PpoI, KasI, Kpn2I, KpnI, KspAI, LweI, MbiI, MboI, MboII, MfeI, MisI, MluI, MlyI, MmeI, MnlI, Mph1103I, MscI, MseI, MslI, MspAII, MspI, MssI, MunI, Mva1269I, MvaI, MwoI, NaeI, NarI, NciI, NcoI, NdeI, NdeII, NgoMIV, NheI, NheI-HF, NlaIII, NlaIV, NmeAIII, NmuCI, NotI, NruI, NsbI, NsiI, NspI, OliI, PacI, PaeI, Paer7I, PagI, Paul, PciI, PdiI, PdmI, Pfl23II, PflFI, PflMI, PfoI, PhoI, PleI, PmeI, PmlI, PpiI, PpuMI, PshAI, Psil, Psp1406I, Psp5II, PspGI, PspOMI, PspXI, PstI, PsuI, PsyI, PvuI, PvuII, PvuII-HF, RsaI, RsrII, SacI, SacII, Sall, Sall-HF, SapI, SatI, Sau3AI, Sau96I, SbfI, ScaI, ScaI-HF, SchI, ScrFI, SdaI, SduI, SexAI, SfaNI, SfcI, SfiI, SfoI, Sgfi, SgrAI, SinI, SmaI, SmiI, SmlI, SmuI, SnaBI, SpeI, SphI, SphI-HF, SspI, StuI, StyD4I, StyI, SwaI, Taal, TaiI, TaqoI, TaqI, TasI, TatI,

TauI, TfiI, TliI, TruII, Tru9I, TseI, Tsp45I, Tsp509I, TspMI, TspRI, Tth111I, TurboNaeI, TurboNarI, Van9II, VspI, XagI, XapI, XbaI, XceI, XcmI, XhoI, XhoII, XmaI, XmaII, XmiI, XmnI, and ZraI. Aspects also include homing endonucleases such as: I-SceI, I-CeuI, and PI-PspI. The corresponding cleavage sites for these enzymes are known in the art.

[0152] In some embodiments the present disclosure teaches the use of rare restriction enzymes, recognizing sites greater than or equal to eight nucleotides in length (≥ 8 restriction enzymes). In some embodiments, the present disclosure teaches use of a single rare restriction site in each cTAG. In other embodiments, the cTAGs of the present disclosure may comprise two or more restriction sites. **Table 1** below provides a list of cTAGs according to the present invention, each with their rare restriction enzyme sites bolded.

Table 1: Example cTAG sequences, CRISPR landing sites, and rare restriction enzyme sites (bold sequence portions are restriction sites)

TagA	ACTGGGTGGAATCCCTTCTGCAGCACCTGGATT ACCCTGTTATCCCT AGT	I-SceI	SEQ ID NO: 65
TagB	TAATGAGTAGTCCTCATCTCCCTCAAGCAGGCG CCGGCGG CTACTGCCATC	MreI	SEQ ID NO: 66
TagC	CATATAATCTCCCTCAAGCAGGCC CCCGCTGGCGCGCGCG GAATGTTAGGAA	MauBI	SEQ ID NO: 67
TagD	GCCTATAATGTGAAGAGCTTCACTGAGTAGGG CCCGGG CTGTAAACGGTT	SrfI	SEQ ID NO: 68
TagE	ATTCGCTAGCAGATGTAGTGT TCCACAGGGCGATCG CTGATATGGGTC	AsiSI	SEQ ID NO: 69
TagF	ACTACCTAGCTGCATTTTCAGGAGGAAGCGATGGG CGGCCG CACACCTTC	NotI	SEQ ID NO: 70
TagG	TGATAATGGGTGAGTGAGTGTGTGCGTGTGG GGCGCGC CCAGATGGGAACA	AscI	SEQ ID NO: 71

TagH	ACTCCAGTCTTTCTAGAAGATGGCAAACAGCTATTATGGGTATTATGGGT	PI-PspI	SEQ ID NO: 72
TagI	TAGTGGACGGGGCCACTAGGGACAGGATTGGCCTGCAGGATTCCCGTCAA	SbfI	SEQ ID NO: 73
TagJ	TGAACTAAGGCGGCTGCACAACCAGTGGAG GCCTAAATGATC	none	SEQ ID NO: 74

[0153] In some embodiments, suitable recombination sites for use in the present invention include, but are not limited to: attB sites, attP sites, attL sites, attR sites, lox sites, psi sites, tnpI sites, dif sites, cer sites, frt sites, and mutants, variants and derivatives thereof. In certain embodiments of the present invention, the topoisomerase recognition site, if present, is recognized and bound by a type I topoisomerase, which may be a type IB topoisomerase. Suitable types of type IB topoisomerase include, but are not limited to, eukaryotic nuclear type I topoisomerase and poxvirus topoisomerase. In some embodiments, suitable types of poxvirus topoisomerase include, but are not limited to, poxvirus topoisomerase produced by or isolated from a virus such as vaccinia virus, Shope fibroma virus, ORF virus, fowlpox virus, molluscum contagiosum virus and Amsacta morreientomopoxvirus.

[0154] In some embodiments, cTAG arrangement of CRISPR and non-CRISPR cloning sites can be ordered according to user preference. In some embodiments, the present disclosure teaches that CRISPR binding sites should be ordered so as to be the furthest away from insert parts. In one illustrative embodiment, a cTAG could be arranged as follows from 5'-3': (Part I)-[R1-A1-C-A2-R2]-(Part II), where R= restriction site, A= recombinase site, and C=CRISPR landing site. In some embodiments, C may include multiple overlapping, or sequential CRISPR and/or restriction landing sites. In some embodiments, the arrangement of cloning sites on a cTAG of the present disclosure will be symmetrical (*i.e.*, provide for a symmetrical order of types of cloning sites).

[0155] In other embodiments, arrangement of cloning sites on a cTAG of the present disclosure may be non-symmetrical. For example, in another illustrative embodiment, a cTAG could be arranged as follows from 5'-3': (Part I)-[R1-A1-C1-C2]-(Part II), where R= restriction site, A= recombinase site, and C1-2=CRISPR landing site(s). In yet other embodiments, a cTAG could be

arranged as follows from 5'-3': i) (Part I)-[R1- C1-C2]-(Part II), ii) (Part I)-[R1-C1]-(Part II), iii) (Part I)-[C1-C2]-(Part II), or their reverse order, wherein R= restriction site, A= recombinase site, and C1-2=CRISPR landing site(s).

[0156] Persons having skill in the art will recognize the advantages and applications of various cTAG arrangements. For example, in single-tag embodiments, the modular construct would allow for insertion with the digestion of a single CRISPR endonuclease, but would not (without more, for example further digestion of additional cTAGs) allow for removal or replacement of said insertion, due to the lack of a second flanking cTAG site. In some embodiments, the present disclosure teaches that inserted parts may themselves contain additional cTAGs, to expand the number of possible insert part locations within the cMCS.

[0157] In other embodiments, the present disclosure teaches methods of removing one or more insert parts from the modular CRISPR constructs. In some embodiments, two or more of the cTAGs of a modular CRISPR construct comprise restriction enzyme binding sites capable of creating compatible ends. In some embodiments, the restriction enzyme sites are identical. In other embodiments the restriction enzyme sites are distinct, but the resulting digestion of said sites produces compatible ends for hybridization and ligation. In some embodiments, the restriction sites for deletion of portions of a modular CRISPR construct are placed on other ends of two or more cTAGS, such that the resulting ligated construct will still maintain the same ratio of insert parts to cTAGS.

[0158] In some embodiments, the present disclosure teaches that the restriction enzyme sites used for deletions within the modular CRISPR constructs of the present disclosure can be any restriction enzyme that results in compatible ends. In other embodiments, the present disclosure teaches that the restriction enzyme sites used for deletions within the modular CRISPR constructs of the present disclosure can be any rare 8 \geq base restriction enzyme that result in compatible ends. In selected embodiments, the present disclosure teaches that the restriction enzyme sites used for deletions within the modular CRISPR constructs of the present disclosure can be I-SceI and PI-PspI.

[0159] In some embodiments, the present disclosure teaches modular CRISPR constructs with two cTAGs flanking each insert part, so as to create a cTAG pair. In some embodiments, the aforementioned cTAG pairs allow for the selective cutting/replacement of insert parts. For

example, as illustrated in **Figure 2B**, digestion of the modular CRISPR plasmid with endonucleases targeting cTAGs A and B would result in the specific removal of insert part 2.

[0160] As discussed above, selected embodiments of the present disclosure provide for replacement insert parts which restore cTAG function following endonuclease cleavage. Thus, as illustrated in **Figure 2B**, replacement insert parts 2a-2d comprise sequences that will restore cTAG A and B function upon insertion into the modular CRISPR plasmid.

[0161] In some embodiments, the present disclosure teaches that cTAGs can also control insert part directionality. Sequence homology between cTAG ends in insert parts and cleaved cTAGs in the modular CRISPR construct will determine insertion directionality for Cas9 cleaved sequences, either through homologous recombination or hybridization (*e.g.*, in Gibson approaches). Insertion directionality in Cpf1 sequences may also be controlled *via* Watson crick hybridization of Cpf1 sticky ends on either cTAG.

[0162] In some embodiments, the present disclosure also provides for alternative cTAG arrangements. For example, in some embodiments, the modular CRISPR constructs of the present disclosure may be designed such as to provide functionality for the use of nested cTAGs.

[0163] In some embodiments, the present disclosure teaches component-based CRISPR assemblies based on shared overlapping “tag” regions that enable multicomponent assembly *in vitro* and *in vivo*. In some embodiments, the tags of the present disclosure comprise CRISPR landing sites to facilitate future cloning or *in vitro* DNA assembly from DNA constructs. If DNA constructs are integrated into the genome of a host organism, preselected Cas9 or Cpf1 landing sites may facilitate facile genetic alterations. In a single suite of experiments, the assembly strategy enables construction of DNA plasmids that can be used in multiple organisms, containing multiple numbers and types of DNA components.

[0164] In some embodiments, this assembly strategy can be used to assemble and quickly reassemble plasmids encoding any desired set of DNA components, including metabolic pathways. In other embodiments, designing cTAGs into integrating plasmids can also be used to swap DNA components directly in and out of the genome of host organisms, circumventing the need to clone future plasmids.

cTAG Sequence Design Algorithm

[0165] In some embodiments, the present disclosure teaches algorithms designed to facilitate CRISPR landing sites within cTAGs. In some embodiments, the CRISPR landing sites are sequences identified from existing sequences. Thus, in some embodiments, the present disclosure teaches use of software programs is designed to identify candidate CRISPR target sequences on both strands of an input DNA sequence based on desired guide sequence length and a CRISPR motif sequence (PAM, protospacer adjacent motif) for a specified CRISPR enzyme. For example, target sites for Cpf1 from *Francisella novicida* U112, with PAM sequences TTN, may be identified by searching for 5'-TTN-3' both on the input sequence and on the reverse-complement of the input. The target sites for Cpf1 from *Lachnospiraceae bacterium* and *Acidaminococcus sp.*, with PAM sequences TTTN, may be identified by searching for 5'-TTTN-3' both on the input sequence and on the reverse complement of the input. Likewise, target sites for Cas9 of *S. thermophilus* CRISPR1, with PAM sequence NNAGAAW, may be identified by searching for 5'-Nx-NNAGAAW-3' both on the input sequence and on the reverse-complement of the input. The PAM sequence for Cas9 of *S. pyogenes* is 5'-NGG-3'.

[0166] Likewise, target sites for Cas9 of *S. thermophilus* CRISPR, with PAM sequence NGGNG, may be identified by searching for 5'-N, —NGGNG-3' both on the input sequence and on the reverse-complement of the input.

[0167] In other embodiments, the present disclosure teaches methods of designing CRISPR landing sites from scratch. Persons having skill in the art will readily be able to design CRISPR landing sites in conjunction with the guide RNAs of the present disclosure, wherein the resulting protospacer sequence is combined with the PAM motif appropriate to the desired CRISPR endonuclease, as described above.

[0168] In some embodiments, the present disclosure teaches cTAGs comprising a sequence selected from the group consisting of: SEQ ID NO. 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 78, 79, 80, 81, and combinations thereof.

[0169] Since multiple occurrences in the genome of the DNA target site may lead to nonspecific genome editing, after identifying all potential sites, the present disclosure teaches, in some embodiments, filtering out sequences based on the number of times they appear in the relevant reference genome or modular CRISPR construct. For those CRISPR enzymes for which sequence specificity is determined by a 'seed' sequence (such as the first 5 nt of the guide sequence for

Cpf1-mediated cleavage) the filtering step may also filter out different sequences with the same seed.

[0170] In some embodiments algorithmic tools can also identify potential off target sites for a particular guide sequence. For example, in some embodiments Cas-Offinder can be used to identify potential off target sites for Cpf1 (*see* Kim *et al.*, 2016. “Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells” published online June 06, 2016). Any other publicly available CRISPR design/identification tool may also be used, including for example the Zhang lab’s crispr.mit.edu tool (*see* Hsu, *et al.* 2013 “DNA targeting specificity of RNA-guided Cas9 nucleases” *Nature Biotech* 31, 827-832).

[0171] In some embodiments, the user may be allowed to choose the length of the seed sequence. The user may also be allowed to specify the number of occurrences of the seed:PAM sequence in a genome for purposes of passing the filter. The default is to screen for unique sequences. Filtration level is altered by changing both the length of the seed sequence and the number of occurrences of the sequence in the genome. The program may, in addition, or alternatively, provide the sequence of a guide sequence complementary to the reported target sequence(s) by providing the reverse complement of the identified target sequence(s).

Modular CRISPR DNA Construct Cloning

[0172] In some embodiments, the present disclosure teaches methods for preparing new recombinant nucleic acid molecules using the modular CRISPR DNA constructs of the present disclosure. In some embodiments, the present disclosure teaches methods of DNA part assembly. Descriptions of each method are provided below.

DNA Assembly Methods

[0173] In some embodiments, the present disclosure teaches methods for the modular assembly of DNA parts. In some embodiments, the DNA assembly methods of the present disclosure are conducted *in vitro*. Thus, in some embodiments, the present disclosure teaches the steps of **i)** forming a mixture comprising at least two insert part DNAs together with at least one CRISPR complex, and **ii)** allowing said mixture to incubate in conditions for CRISPR digestion of the insert DNAs, **iii)** followed by hybridizing the compatible sticky ends from the digestion of each of the two insert part DNAs, and **iv)** ligating said hybridized ends to one another to create the new

recombinant nucleic acid. Thus, in some embodiments, the insert part DNAs of the present disclosure are digested together. In other embodiments, the present disclosure teaches methods of digesting each insert part DNA individually, with the same or different CRISPR complexes. In some embodiments, at least one insert part is not digested by a CRISPR complex. In some embodiments, the present disclosure teaches that an exonuclease treatment is conducted prior to the hybridization of step **iii**) (for dual CRISPR digestions as described in later sections).

[0174] In yet other embodiments, the present disclosure teaches Gibson-like joining of insert parts, by exposing the insert part ends to an ssDNA exonuclease, and hybridizing the resulting sticky ends followed by an optional fill with polymerase, and ligation. In some embodiments one or more insert parts are exposed to a dsDNA exonuclease prior to the ssDNA exonuclease treatment. In some embodiments, the present disclosure teaches Gibson-like joining of insert parts or modular CRISPR vectors that have been digested by one or more CRISPR endonuclease (e.g., dual CRISPR digestions, as described in later sections).

[0175] The sections below provide a series of illustrative examples demonstrating the various ways in which the insert parts and modular CRISPR constructs of the present disclosure can be assembled and edited. The list of techniques described below provides an illustrative series of examples highlighting the utility of the sequences of the present disclosure, but is not intended to be limiting. Persons having skill in the art will recognize other techniques that allow for the assembly and editing of insert parts according to the present disclosure.

[0176] In some embodiments, the present disclosure describes methods involving Cpf1 and/or Cas9 CRISPR endonucleases. Reference to these specific CRISPR endonucleases is illustrative, and is not intended to be limiting, unless specified in a claim. Persons having skill in the art will immediately recognize the applicability of other existing—or heretofore undiscovered CRISPR endonucleases to the constructs and methods of the present disclosure. References to Cpf1 may be interpreted as encompassing use of any presently known or undiscovered CRISPR endonuclease capable of catalyzing staggered DNA cleavage to produce sticky DNA ends. References to Cas9 may similarly be interpreted as encompassing use of any presently known or undiscovered CRISPR endonuclease capable of catalyzing blunt end cleavage of dsDNA.

In vitro Cpf1

[0177] In some embodiments, the *in vitro* DNA assemblies of the present disclosure are conducted with Cpf1 CRISPR complexes as described below. *First*, two or more insert parts are incubated with a Cpf1 CRISPR complex targeting the cTAG that is common between the at least two insert parts. In some embodiments, the insert parts are incubated together in a single mixture. In other embodiments, the insert parts are incubated in different mixtures.

[0178] *Second*, in some embodiments, the digested products are purified to remove active CRISPR nuclease. In some embodiments, the purification involves separation of the active Cpf1 complex from the digested insert parts. In some embodiments, this can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by Cpf1 inactivation, such as through heat or chemical inactivation.

[0179] *Third*, the digested insert parts are incubated in conditions appropriate for hybridization of the compatible sticky ends created by the Cpf1 complex. Hybridized ends are then ligated according to any known ligation methods, including those described in earlier portions of this disclosure.

In vitro Cas9

[0180] In other embodiments, the *in vitro* DNA assemblies of the present disclosure are conducted with Cas9 CRISPR complexes as described below. *First*, two or more insert parts are incubated with a Cas9 CRISPR complex targeting the cTAG that is common between the at least two insert parts. In some embodiments, the insert parts are incubated together in a single mixture. In other embodiments, the insert parts are incubated in different mixtures.

[0181] *Second*, in some embodiments, the digested products are purified to remove active CRISPR nuclease. In some embodiments, the purification involves separation of the active Cas9 complex from the digested insert parts. In some embodiments, this can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by Cas9 inactivation, such as through heat or chemical inactivation.

[0182] In some embodiments, the third step for Cas9 digested products is to incubate the insert parts in conditions appropriate for blunt end-ligation.

Dual CRISPR Assemblies

[0183] In other embodiments, the present disclosure also teaches Gibson-assembly type methods for assembling the pieces of CRISPR-digested insert parts with at least one shared cTAG sequence (e.g., assembly of compatible cTAGs digested at different CRISPR landing sites). Thus, in some embodiments, the present disclosure teaches dual CRISPR digestion assemblies as described below.

[0184] *First*, two or more insert parts are incubated with two CRISPR complexes targeting two different CRISPR landing sites flanking each part within the aforementioned cTAGs that are common between the at least two insert parts.

[0185] In some embodiments, the two different CRISPR landing sites are digested together. In other embodiments, one insert part DNA is digested with one CRISPR complex targeting one CRISPR landing site, and the other insert part DNA is digested with a different CRISPR complex targeting the second CRISPR landing target site in separate vessels. In each case, the result of these digestions will be that the shared cTAG in each of the two insert DNA cTAGs will comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 bp of sequence overlap with each other.

[0186] For example, in an illustrative embodiment, the shared cTAG between two insert DNA parts would be arranged as follows from 5'-3': (Part I)-[R1-C1-C2]-(Part II), where R= restriction site, C1= a first CRISPR landing site and C2= a second CRISPR landing site. In this illustrative embodiment, the first insert DNA part with a 3' shared cTAG would be digested with a CRISPR complex targeting C2 and the second insert DNA part with a 5' shared cTAG would be digested with a CRISPR complex targeting C1. This would result in two DNA insert parts with overlapping sequence spanning C1-C2.

[0187] *Second*, in some embodiments, the digested products are purified to remove active CRISPR nuclease. In some embodiments, the purification involves separation of the active CRISPR complex from the digested insert parts. In some embodiments, this can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by CRISPR inactivation, such as through heat or chemical inactivation.

[0188] *Third*, in some embodiments, the CRISPR-digested insert parts are incubated with a ssDNA exonuclease to create overlapping sticky ends between the two insert DNA parts.

[0189] *Fourth*, the digested insert parts are incubated in conditions appropriate for hybridization of the compatible sticky ends created by the CRISPR complex/exonuclease digestions. Hybridized ends are then ligated according to any known ligation methods, including those described in earlier portions of this disclosure. In some embodiments, the hybridized parts are incubated with a polymerase to fill in any missing sequence gaps prior to ligation.

Bridging Assemblies

[0190] In other embodiments, the present disclosure teaches Gibson-assembly of Cas9 digested parts, through the addition of a third DNA sequence comprising a bridging sequence that overlaps with the digested cTAG sequences of the insert parts.

[0191] In this illustrative example, both insert parts are digested with the same Cas9 CRISPR complex targeting the same CRISPR landing site. In this embodiment, the resulting digested cTAGs would have no sequence overlap. Thus, in some embodiments, the *third* step is for Cas9 digested insert parts to be further digested with an ssDNA exonuclease to create either 3' or 5' overhang. The exonuclease digested insert parts are then incubated in conditions appropriate for hybridization of the compatible sticky ends created by the combination of the CRISPR complex and exonuclease digestions with the bridging sequence. Hybridized ends are then ligated according to any known ligation methods, including those described in earlier portions of this disclosure. In some embodiments the exonuclease digestion of the present disclosure is conducted before the second step.

In vitro HDR

[0192] In other embodiments, the present disclosure teaches *in vitro* methods of assembling the ends of insert part DNAs digested by a Cas9 or Cpf1 endonuclease with an HDR complex, thereby triggering recombination of said digested insert parts.

In vivo Homologous Recombination

[0193] In some embodiments, the *in vivo* DNA assemblies of the present disclosure are conducted with Cpf1 or Cas9 CRISPR complexes as described below. In one embodiment, two or more insert parts with at least one shared cTAG are introduced into a host cell. In some embodiments, the presence of DNA insert parts with homologous shared cTAG sequences will be sufficient to trigger homologous recombination assembly (*e.g.*, yeast homologous recombination).

[0194] For example, in some embodiments, at least one shared cTAG sequence between the two insert DNA parts could be assembled to produce a linear construct. In this illustrative embodiment the two remaining outer cTAGs could also be designed to recombine with cTAGs of another vector within the cell (e.g., insertion into an existing plasmid, or a chromosome). In other embodiments, the two parts could be further assembled into a circular construct through the recombination of a second shared cTAG between the two insert DNA parts. The assembled construct can be either used in the organism that was used for the assembly, or can, in some embodiments, be purified and transformed into a second organism (e.g., assembly in yeast, and subsequent transformation into bacteria).

[0195] In other embodiments, one or more insert parts with a shared cTAG can be digested prior to introduction into the host cell. Thus, in some embodiments, the present disclosure teaches CRISPR digestions to release insert parts from larger vectors prior to *in vivo* assembly of the released parts. In some embodiments, the digestion is carried out with Cas9. In other embodiments the digestion is carried out with Cpf1. In other embodiments the digestion is carried out with restriction endonucleases. In some embodiments the CRISPR digestions of insert parts are conducted *in vitro*. In some embodiments, the digested products are purified to remove active CRISPR endonuclease prior to transformation of the insert parts into the assembly host cell.

[0196] In some embodiments, the purification step can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by CRISPR inactivation, such as through heat or chemical inactivation.

In vivo ligation

[0197] In some embodiments, the present disclosure teaches methods of protecting insert parts from re-cleavage by the CRISPR endonuclease. In some embodiments, the insert parts of the present disclosure may be protected from endonuclease cleavage *via* chemical modification of the DNA sequence. For example, in some embodiments, the present disclosure teaches phosphorothioate oligonucleotides.

[0198] In some embodiments, the methods of the present disclosure are especially useful for multi-part DNA assemblies.

[0199] Figure 2A of the specification provides an illustrative example of a multi-part DNA assembly, according to the methods of the present disclosure. In this example, a series of eight DNA parts (parts 1-8), each with two cTAGs (tags A-H) are combined *in vitro* and are then able to self-assemble (either *via* homologous recombination *in vivo*, or *via* ligation, as described above).

DNA Editing Methods

[0200] In some embodiments, the present disclosure teaches methods for the editing of modular CRISPR DNA constructs. In some embodiments, the DNA editing methods of the present disclosure apply the same principles of the DNA assembly methods described above, but do so for the purposes of editing one or more pre-existing modular CRISPR DNA constructs.

[0201] In some embodiments, the DNA editing methods of the present disclosure are conducted *in vitro*. Thus, in some embodiments, the present disclosure teaches the steps of **i)** forming a mixture comprising a modular CRISPR DNA construct, and at least one insert DNA part, together with at least one CRISPR complex, and **ii)** allowing said mixture to incubate in conditions for CRISPR digestion of the cTAGS of the insert DNA, and its corresponding modular CRISPR DNA construct cTAGs, followed by **iii)** hybridizing the compatible sticky ends (if Cpf1) produced by the digestion of each of the aforementioned cTAGs, and **iv)** ligating said hybridized ends (or blunt ends, if Cas9 is used) to one another to create the new recombinant nucleic acid. In some embodiments, an exonuclease treatment is conducted prior to the hybridization of step **iii)** (for dual CRISPR digestions as described in later sections). In some embodiments, the digestions of the present disclosure are conducted separately for the insert part DNA and modular CRISPR DNA construct. In some embodiments, only the modular CRISPR DNA construct is digested with a CRISPR complex.

In vitro Cpf1

[0202] In some embodiments, the *in vitro* DNA editing methods of the present disclosure are conducted with Cpf1 CRISPR complexes as described below. *First*, a modular CRISPR DNA construct, and at least one insert DNA part are incubated with a Cpf1 CRISPR complex targeting the cTAGs of the insert parts, and their corresponding tags within the modular CRISPR DNA construct. In some embodiments, the digestion of the modular CRISPR DNA and the insert part DNA is conducted in separate reactions.

[0203] *Second*, in some embodiments, the digested products are purified to remove active CRISPR nuclease. In some embodiments, the purification involves separation of the active Cpf1 complex from the digested nucleotides. In some embodiments, this can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by Cpf1 inactivation, such as through heat or chemical inactivation.

[0204] *Third*, the digested modular CRISPR DNA construct and insert parts are incubated in conditions appropriate for hybridization of the compatible sticky ends created by the Cpf1 complex. Hybridized ends are then ligated according to any known ligation methods, including those described in earlier portions of this disclosure.

In vitro Cas9

[0205] In other embodiments, the *in vitro* DNA editing methods of the present disclosure are conducted with Cas9 CRISPR complexes as described below. *First*, a modular CRISPR DNA construct, and at least one insert DNA part are incubated with a Cas9 CRISPR complex targeting the cTAGs of the insert parts, and their corresponding tags within the modular CRISPR DNA construct. In some embodiments, the digestion of the modular CRISPR DNA and the insert part DNA are conducted in separate reactions.

[0206] *Second*, in some embodiments, the digested products are purified to remove active CRISPR nuclease. In some embodiments, the purification involves separation of the active Cas9 complex from the digested nucleotides. In some embodiments, this can be accomplished through a DNA purification, such as a gel or column purification. In other embodiments, the purification can be accomplished by Cas9 inactivation, such as through heat or chemical inactivation.

[0207] In some embodiments, the *third* step for Cas9 digested products is to incubate the insert parts in conditions appropriate for blunt end-ligation.

Gibson editing

[0208] In other embodiments, the present disclosure also teaches Gibson-assembly type methods for editing the sequences of CRISPR-digested constructs and/or undigested insert parts containing intact overlapping cTAG sequence. Thus, in some embodiments, the *third* step is for Cas9 digested modular CRISPR DNA construct and insert part(s) to be further digested with an ssDNA exonuclease to create either a 3' or 5' overhang. In some embodiments, the present disclosure

teaches dsDNA exonuclease digestion to shorten the non-CRISPR digested insert parts prior to the ssDNA digestion.

[0209] The exonuclease digested DNA sections are then incubated under conditions appropriate for hybridization of the compatible sticky ends created by the combination of the CRISPR complex and exonuclease digestions. Hybridized ends are then ligated according to any known ligation methods, including those described in earlier portions of this disclosure. In some embodiments the hybridized DNA is incubated with a polymerase to fill in missing DNA sections prior to ligation. In some embodiments the exonuclease digestion of the present disclosure is conducted before the CRISPR inactivation step.

[0210] In some embodiments, the ligation of digested sequences can occur *in vitro*.

[0211] In other embodiments, the present disclosure teaches *in vitro* methods of assembling the ends of a modular CRISPR DNA construct digested by a Cas9 or Cpf1 endonuclease, and at least one undigested insert with an HDR complex, thereby triggering recombination of said digested modular CRISPR DNA construct, and at least one insert DNA part.

[0212] In some embodiments of the DNA editing methods of the present disclosure, the DNA insert parts are comprised within a second modular CRISPR DNA construct. Thus, in some embodiments, the DNA editing methods of the present disclosure comprise the transfer of a DNA insert part from one modular CRISPR DNA construct to another.

Expression, Purification, and Delivery

[0213] In some embodiments, the present disclosure teaches methods and compositions of vectors, constructs, and nucleic acid sequences encoding CRISPR complexes. In some embodiments, the present disclosure teaches plasmids for transgenic or transient expression of the Cas9 or Cpf1 proteins. In some embodiments the present disclosure teaches a plasmid encoding chimeric Cas9 or Cpf1 proteins comprising in-frame sequences for protein fusions of one or more of the other polypeptides described herein, including, but not limited to a ligase, a linker, and an NLS.

[0214] In some embodiments the plasmids and vectors of the present disclosure will encode for the Cas9/Cpf1 protein(s) and also encode the crRNA/tracrRNA/sgRNA, and/or donor insert sequences of the present disclosure. In other embodiments, the different components of the engineered complex can be encoded in one or more distinct plasmids.

[0215] In some embodiments, the plasmids of the present disclosure can be used across multiple species. In other embodiments, the plasmids of the present disclosure are tailored to the organism being transformed. In some embodiments, the sequences of the present disclosure will be codon-optimized to express in the organism whose genes are being edited. Persons having skill in the art will recognize the importance of using promoters providing adequate expression for gene editing. In some embodiments, the plasmids for different species will require different promoters.

[0216] In some embodiments, the plasmids and vectors of the present disclosure are selectively expressed in the cells of interest. Thus in some embodiments, the present application teaches the use of ectopic promoters, tissue-specific promoters, developmentally-regulated promoters, or inducible promoters. In some embodiments, the present disclosure also teaches the use of terminator sequences.

[0217] In some embodiments, the present disclosure also teaches methods of expressing and purifying Cpf1 and/or Cas9 endonuclease protein. In some embodiments, the present disclosure teaches that the proteins of the present disclosure may be produced by any of the commercially available protein production and purification kits or services. For example, in some embodiments, the present disclosure teaches methods of cloning Cas9 and/or Cpf1 into a vector with a polyhistidine (His), glutathione s-transferase (GST), or other purification tag chimeric fusion. In some embodiments the present disclosure teaches a variety of prokaryotic and eukaryotic organisms, and cell-free protein production systems. For example, in some embodiments, the present disclosure teaches expression of protein expression plasmids in *E. coli* BL21. In some embodiments, the protein production system will be inducible, to reduce the effects of protein toxicity. For example, in some embodiments, the present disclosure teaches methods of using the IPTG or an arabinose induction system.

[0218] In some embodiments, the present disclosure also teaches various protein purification schemes, including affinity tags (His-Nickel, GST-Glutathione, etc.). In some embodiments, the present disclosure teaches both native and denaturing conditions for protein purification.

[0219] In other embodiments, the present disclosure teaches production of Cas9 and/or Cpf1 *via* one or more protein production services, including, but not limited to GenScript®, ThermoFisher®, and NovoProtein®.

Transformation

[0220] In some embodiments, the present disclosure teaches the use of transformation of the plasmids and vectors disclosed herein. Persons having skill in the art will recognize that the plasmids of the present disclosure can be transformed into cells through any known system as described in other portions of this specification. For example, in some embodiments, the present disclosure teaches transformation by particle bombardment, chemical transformation, agrobacterium transformation, nano-spike transformation, electroporation and virus transformation.

[0221] In some embodiments, the vectors of the present disclosure may be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, I., 1986 “Basic Methods in Molecular Biology”). Other methods of transformation include for example, lithium acetate transformation and electroporation *See, e.g.,* Gietz *et al.*, *Nucleic Acids Res.* 27:69-74 (1992); Ito *et al.*, *J. Bacterol.* 153:163-168 (1983); and Becker and Guarente, *Methods in Enzymology* 194:182-187 (1991). In some embodiments, transformed host cells are referred to as recombinant host strains.

[0222] In some embodiments, the present disclosure teaches high throughput transformation of cells using the 96-well plate robotics platform and liquid handling machines of the present disclosure.

[0223] In some embodiments, the present disclosure teaches methods for getting exogenous protein (Cpf1/Cas9 and DNA ligase), RNA (crRNA/tracrRNA/GuideRNA), and DNA (insert DNA part or modular CRISPR construct) into the cell. Various methods for achieving this have been described previously including direct transfection of protein/RNA/DNA or DNA transformation followed by intracellular expression of RNA and protein (Dicarlo, J. E. *et al.* “Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems.” *Nucleic Acids Res* (2013). doi:10.1093/nar/gkt135; Ren, Z. J., Baumann, R. G. & Black, L. W. “Cloning of linear DNAs in vivo by overexpressed T4 DNA ligase: construction of a T4 phage hoc gene display vector.” *Gene* **195**, 303–311 (1997); Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. “Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.” *Elife* **3**, e04766 (2014)).

[0224] In some embodiments, the present disclosure teaches screening transformed cells with one or more selection markers as described above. In one such embodiment, cells transformed with a vector comprising a kanamycin resistance marker (KanR) are plated on media containing effective amounts of the kanamycin antibiotic. Colony forming units visible on kanamycin-laced media are presumed to have incorporated the vector cassette into their genome. Insertion of the desired sequences can be confirmed *via* PCR, restriction enzyme analysis, and/or sequencing of the relevant insertion site.

[0225] In other embodiments, a portion, or the entire complexes of the present disclosure can be delivered directly to cells. Thus, in some embodiments, the present disclosure teaches the expression and purification of the polypeptides and nucleic acids of the present disclosure. Persons having skill in the art will recognize the many ways to purify protein and nucleic acids. In some embodiments, the polypeptides can be expressed *via* inducible or constitutive protein production systems such as the bacterial system, yeast system, plant cell system, or animal cell systems. In some embodiments, the present disclosure also teaches the purification of proteins and or polypeptides *via* affinity tags, or custom antibody purifications. In other embodiments, the present disclosure also teaches methods of chemical synthesis for polynucleotides.

[0226] In some embodiments, persons having skill in the art will recognize that viral vectors or plasmids for gene expression can be used to deliver the complexes disclosed herein. Virus-like particles (VLP) can be used to encapsulate ribonucleoprotein complexes, and purified ribonucleoprotein complexes disclosed herein can be purified and delivered to cells *via* electroporation or injection.

Kits

[0227] In some embodiments, the disclosure provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a modular CRISPR DNA construct and instructions for using the kit and any necessary reagents or reactants. In some embodiments, the vector system comprises (a) a modular CRISPR DNA construct (b) a CRISPR complex, including a CRISPR endonuclease protein, and necessary target guide RNA(s) (or sequences encoding said items), and optionally (c) insert DNA parts, as describe supra in this application.

[0228] Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a *vial*, a bottle, or a tube, or host cell, or plasmid. In some embodiments, the kit includes instructions in one or more languages, for example in more than one language.

[0229] In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein (*e.g.*, purified Cpf1 endonuclease). Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (*e.g.* in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a crRNA sequence for insertion into a vector so as to operably link the crRNA sequence and a regulatory element.

EXAMPLES

[0230] The following examples are given for the purpose of illustrating various embodiments of the disclosure and are not meant to limit the present disclosure in any fashion. Changes therein and other uses which are encompassed within the spirit of the disclosure, as defined by the scope of the claims, will occur to those skilled in the art.

Example 1: One-Pot *in vitro* Modular CRISPR Cloning

[0231] This example describes the generation of plasmid 13001009086 (SEQ ID NO: 82) by transfer of an insert from one plasmid to another in a one-pot reaction. *See, Figure 4.*

[0232] Both plasmids carry cloning tags flanking the region of interest (cTAG K [SEQ ID NO: 78] /cTAG L [SEQ ID NO: 79] and cTAG K' [SEQ ID NO: 80] /cTAG L' [SEQ ID NO: 81]). In order to drive the cloning reaction towards the edited plasmid, the Cpf1 spacers are in opposite orientations on the recipient and donor plasmids (K/K' and L/L' respectively). This inside-out/outside-in digest removes the Cpf1 spacer in the final product, eliminating re-cutting of the desired product (see, curved arrows in **Figure 4**, depicting inside-out digestion in the '485 plasmid and outside-in digestion in the '784 plasmid). The Cas9 spacers remain, enabling iterative editing at this site. Thus, the MegaModular construct allows for a rapid single-pot reaction scheme that enables iterative editing.

[0233] Cpf1 protein was synthesized by Genscript and the crRNAs by Synthego. For the one-pot cleavage/ligation reaction, the Cpf1 protein complexed with the crRNAs (crRNA 1 and crRNA 3), was added to the plasmids (13000789485 – SEQ ID NO: 83 and 13000823784 – SEQ ID NO: 84) and DNA ligase in buffer containing ATP. These components were cycled at temperatures optimized for cleavage and ligation.

[0234] The reaction was transformed into *E. coli* and positive clones were sequenced to confirm insertion of the new insert and loss of the Cpf1 spacers.

[0235] For deletions, the Cpf1 sites within cloning tags used must generate compatible overhangs to allow for plasmid closure. cTAG L' was designed to contain two Cpf1 spacers, one for insertion where the overhang is incompatible with cTAG K' and a second one for deletion where the overhang is compatible with cTAG K'.

Example 2: *In vitro* Modular CRISPR cloning

[0236] This example was designed to demonstrate the flexibility of CRISPR cloning. As an initial step, several resistance plasmids encoding for Kanamycin or Chloramphenicol resistance genes were created from source vectors pzHR039 (SEQ ID No: 100) and 13000223370 (SEQ ID No: 101), respectively. The Kanamycin resistance plasmids were each designed so as to include various CpfI landing sites flanking the GFP gene (when digested, these plasmids produce “the kanamycin resistant plasmid backbone”). The Chloramphenicol resistance plasmids were each designed so as to include various CpfI landing sites flanking the Chloramphenicol resistance gene (when digested, these plasmids produce “the chloramphenicol resistant insert”). Sequences, and vector maps for each plasmid used in this Example are disclosed in **Table 2**.

[0237] Each Kanamycin and Chloramphenicol resistant plasmid was initially linearized with type-II restriction enzymes KpnI-HF and PvuI-HF, respectively (both commercially available from NEB). The location of the KpnI and PvuI restriction sites on each plasmid are noted in the vector maps provided in **Figures 7-14**. After linearization, the resistance plasmids were no longer capable of self-replication in a bacterial host system.

[0238] Linearized resistance plasmids were then mixed with a pre-incubated mixture of 15 ug (1.58 uM final concentration) of CpfI enzyme and 2 uL of 5 uM of each guide RNA described below (0.167 uM final concentration) in a 60 uL reaction to form active CRISPR complexes.

[0239] The CpfI enzyme used in this Example was commercially obtained from IDT. The CpfI was sourced from *Acidaminococcus sp.* CpfI (AsCpfI). The enzyme was further modified to comprise 1 N-terminal nuclear localization sequence (NLS) and 1 C-terminal NLSs, as well as 3 N-terminal FLAG tags and a C-terminal 6-His tag.

[0240] The guide RNAs used in this example were custom ordered from IDT. Each guide RNA was designed to target a different CRISPR landing site located within the linearized resistance plasmid. In this Example, the CpfI landing sites of the backbone plasmid were excised, but restored upon ligation of the insert. **Table 2** provides the guide sequence portion of each guide RNA used. The CRISPR complexes in the mixture were thus designed to cleave out the GFP gene from each kanamycin resistant plasmid to generate kanamycin resistant plasmid backbones (see **Figure 5**, second panel). The CRISPR complexes in the mixture were also designed to cleave out the chloramphenicol resistance gene from the chloramphenicol resistance plasmid to generate chloramphenicol resistant inserts (see **Figure 5**, second panel). The kanamycin resistant plasmid

backbone and the chloramphenicol resistant insert of each reaction were similarly designed to generate compatible overhangs that would result in hybridization of the ends to produce a “dual resistant” kanamycin and chloramphenicol plasmid.

[0241] The linearized resistance plasmid mixtures comprising the CpfI and guide RNAs were allowed to incubate for 3 hours at 37 degrees Celsius in the manufacturer’s recommended CpfI buffer. Selected reactions were run on agarose gels and the resulting fragments were purified using standard DNA extraction kits (Zymo Research kit, used according to manufacturer’s instructions). Purified (control) and unpurified (test).

[0242] DNA fragments comprising the kanamycin resistant plasmid backbone and the chloramphenicol resistant insert, each comprising two compatible CpfI sticky ends were combined in a new reactions with or without a T4 DNA ligase (commercially available from NEB) and transformed into NEB10-B cells (commercially available from NEB). Transformed cells were plated on media augmented with both Kanamycin and Chloramphenicol designed to prevent the growth of any cells that did not contain functional resistance plasmids.

[0243] Individual colonies were sent for sequencing to confirm junctions of CpfI cloning. Recovered colonies were also validated *via* PCR using primers described in **Table 2**. **Figure 5** illustrates the general experimental design described above, except that the plasmids were linearized prior to CpfI digestion, as described above.

Table 2: List of sequences used in this Example 2

Component	Description	SEQ ID NO
GFP CpfI cTAG M fwd	5' CAGCACCTGGATTACCCTGTTATCCCTAGT TTTGGGTTAAAGATGGTTAAATGATTCCG AAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 86
GFP CpfI cTAG N fwd	5' CAGCACCTGGATTACCCTGTTATCCCTAGT TTTGGGATGTAAAGAGTCCCTATCTTCG AAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 87

Component	Description	SEQ ID NO
GFP CpfI cTAG P fwd	5' CAGCACCTGGATTACCCTGTTATCCCTAGT TTTGAGGAGTGTCAGTCTCCGTGAACT CGAAAATAATAAAGGGAAAATCA 3'	SEQ ID No: 88
GFP CpfI cTAG O rvs	5' CGCTTCCTCCTGAAAATGCAGCTAGGTAGT TTTGACCGCCCCCCCCCATACCCCAATCG ACATGCCGAACTCAGAAGTGA 3'	SEQ ID No: 89
GFP CpfI cTAG N rvs	5' CGCTTCCTCCTGAAAATGCAGCTAGGTAGT TTTGGGATGTTAAGAGTCCCTATCTTCG ACATGCCGAACTCAGAAGTGA 3'	SEQ ID No: 90
CAT01 CpfI cTAG M fwd	5' TTTGGGTTAAAGATGGTTAAATGATTTCG ACATACACATAAAGTAGCTTGCG 3'	SEQ ID No: 91
CAT01 CpfI cTAG N fwd	5' TTTGGGATGTTAAGAGTCCCTATCTTCGA CATACACATAAAGTAGCTTGCG 3'	SEQ ID No: 92
CAT01 CpfI cTAG P fwd	5' TTTGAGGAGTGTCAGTCTCCGTGAACT CGACATACACATAAAGTAGCTTGCG 3'	SEQ ID No: 93
CAT01 CpfI cTAG N rvs	5' TTTGGGATGTTAAGAGTCCCTATCTTCG ACTGGAAGGACAAGGGGGACC 3'	SEQ ID No: 94

Component	Description	SEQ ID NO
CAT01 CpfI cTAG O rvs	5' TTTGACCGCCCCCCCCCATAACCCCAATCG ACTGGAAGGACAAGGGGGACC 3'	SEQ ID No: 95
CpfI cTAG M	5' TTTGGGTAAAGATGGTTAAATGAT 3'	SEQ ID No: 96
RNA targeting cTAG M	5' UAAUUUCUACUCUUGUAGAUGGUUAAAGAU GGUAAAUGAU 3'	SEQ ID NO: 110
CpfI cTAG N	5' TTTGGGATGTAAAGAGTCCCTATCT 3'	SEQ ID No: 97
RNA targeting cTAG N	5' UAAUUUCUACUCUUGUAGAUGGAUGUUAAG AGUCCCUAUCU 3'	SEQ ID NO: 111
CpfI cTAG O	5' TTTGACCGCCCCCCCCCATAACCCCAA 3'	SEQ ID No: 98
RNA targeting cTAG O	5' UAAUUUCUACUCUUGUAGAUAACCGCCCCC CCAUACCCCAA 3'	SEQ ID NO: 112
CpfI cTAG P	5' TTTGAGGAGTGTTTCAGTCTCCGTGAAC 3'	SEQ ID No: 99
RNA targeting cTAG P	5' UAAUUUCUACUCUUGUAGAUAAGGAGUGUUC AGUCUCCGUGAAC 3'	SEQ ID NO: 113
pzHR039	Source for Kanamycin resistance and GFP. See listing for full sequence	SEQ ID NO: 100

Component	Description	SEQ ID NO
13000223370	Source for Chloramphenicol resistance and GFP. See listing for full sequence	SEQ ID NO: 101
pJDI427	GFP Cpf1 cTAGs M and N KanR CENARS TRP1 see listing for full sequence	SEQ ID NO: 102 Figure 7
pJDI429	GFP Cpf1 cTAGs N and O KanR CENARS TRP1 see listing for full sequence	SEQ ID NO: 103 Figure 8
pJDI430	GFP Cpf1 cTAGs N and P KanR CENARS TRP1 see listing for full sequence	SEQ ID NO: 104 Figure 9
pJDI431	GFP Cpf1 cTAGs O and P KanR CENARS TRP1 see listing for full sequence	SEQ ID NO: 105 Figure 10
pJDI432	pJET AmpR CmR Cpf1 cTAGs M and N see listing for full sequence	SEQ ID NO: 106 Figure 11
pJDI434	pJET AmpR CmR blunt Cpf1 cTAGs N and O see listing for full sequence	SEQ ID NO: 107 Figure 12
pJDI435	pJET AmpR CmR blunt Cpf1 cTAGs N and P see listing for full sequence	SEQ ID NO: 108 Figure 13
pJDI436	pJET AmpR CmR blunt Cpf1 cTAGs O and P see listing for full sequence	SEQ ID NO: 109 Figure 14

***non-underlined portion of guide RNA for SEQ ID NOs: 110-113 is the chemically modified Alt-R RNA from IDT. The homologous region of sequence to the respective cTAGs (*i.e.* M-P) is underlined.

[0244] The results of this experiment are shown in **Table 3** and **Figure 6**. Reaction numbers for each transformation are shown along the top row, with guide RNAs used listed along the left-hand

column of **Table 3**. The comparison of identical Cpf1 reactions with and without ligase showed a 9.9-fold increase in transformants in the presence of ligase enzyme, indicating that colony growth was due to formation of the double kanamycin and chloramphenicol resistant plasmid after Cpf1 digestion. The no-ligase reactions are matched controls designed to establish that the reactions are specific, and were not simply due to the presence of contaminating levels of undigested resistance plasmids.

[0245] Sixteen individual colonies were Sanger sequenced to verify both the upstream and downstream cloning junctions. In seven of seven upstream sequenced junctions, and eight of nine downstream junctions, the Cpf1 mediated clones from the reactions with T4 DNA ligase indicated faithful digestion and ligation.

[0246] Reactions 71 and 72 were transformed with Cpf1 digested plasmids that were not subjected to DNA gel purification steps. Cpf1 enzyme however was heat inactivated according to supplier's instructions before addition of T4 DNA ligase (reaction 72). Reactions 71 and 72 exhibited the same ligase-dependency.

Table 3: Resistant Transformant Colonies Comprising Cpf1-edited vectors

	55	56	59	60	67	68	71*	72*
Guides M + N	yes	yes						
Guides N + O			yes	yes				
Guides P + O					yes	yes	yes	yes
T4 Ligase	No	Yes	No	Yes	No	Yes	No	Yes
# of transformants (Kan Resistant Colonies)	1	20	0	141	0	12	2	95

*Plates 71 and 72 were transformed with digested DNA that had not undergone DNA gel purification after Cpf1 digestion.

[0247] The disclosure of PCT/US2017/042245 (WO 2018/013990 A1, claiming priority to U.S. Provisional App. No. 62/362,909) is incorporated herein in its entirety.

Example 3: Plasmid Assembly by Restriction Enzyme Digestion and Ligation Using the MegaModular Design

[0248] This example describes the genetic editing of a modular CRISPR vector, according to the methods of the present disclosure. **Figure 15** illustrates the genetic editing of modular CRISPR plasmid 13000444591 described in this example. The plasmid backbone was first prepared by removing a “stuffer” insert DNA part from a previously constructed plasmid. The stuffer insert DNA part was removed by digesting the stuffer part’s flanking cloning tags (cTAGs) D (SEQ ID NO: 68) and E (SEQ ID NO: 69) with restriction enzymes *ApaI* and *PvuI*. The resulting fragments were separated *via* gel electrophoresis, and the desired 8.3 kb fragment corresponding to the plasmid backbone was excised from the gel and extracted using standard silica membrane columns.

[0249] To generate the new insert for the modular CRISPR vector, a desired insert DNA part flanked by cTAG D and cTAG E, was PCR amplified using universal cTAG oligos tagD_FWD (SEQ ID NO: 75) and tagE_REV (SEQ ID NO: 76). The resulting insert contained a GFP marker gene flanked by cTAG D and cTAG E. The resulting PCR fragment was digested with the *ApaI* enzyme that cuts within cTAG D and the *PvuI* enzyme that cuts within cTAG E sequence. The digested insert DNA part was purified using standard silica membrane columns.

[0250] The purified modular CRISPR vector backbone and insert DNA part were combined into a single reaction with a ligase to generate a circular plasmid. The sequence for the resulting edited GFP-containing plasmid 13000444591 is provided in (SEQ ID NO: 77).

Example 4: Plasmid Assembly by Yeast Homologous Recombination Using the MegaModular Design

[0251] Plasmid 13000283399 (SEQ ID NO: 85) was assembled by yeast homologous recombination of PCR fragments flanked by MegaModular tags. The desired constructs for assembly were amplified by PCR in such a way that they were flanked by specific MegaModular tags. These tags allowed for directional assembly of fragments in *Saccharomyces cerevisiae* as the tags themselves served as the overlapping homologous region for homologous recombination.

Specifically, 5 fragments were amplified *via* PCR flanked by MegaModular tags as follows: tag A – Fragment 1 – tag B; tag B – Fragment 2 – tag C; tag C – Fragment 3 – tag D; tag D – Fragment 4 – tag E; and tag E – Fragment 5 – tag F. These fragments, along with a linearized assembly vector containing a yeast origin of replication and a TRP auxotrophic selection marker as well as tag A at one end and tag F at the other, were transformed into *S. cerevisiae*. Circularized, assembled plasmids were selected by *S. cerevisiae* growth in media lacking tryptophan. These plasmids were recovered and amplified in *Escherichia coli*, and correct conformation was confirmed by sequencing.

INCORPORATION BY REFERENCE

[0252] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes. However, mention of any reference, article, publication, patent, patent publication, and patent application cited herein is not, and should not, be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

CLAIMS**What is claimed is:**

1. A recombinant modular CRISPR DNA construct comprising a CRISPR multi-clonal site, said multi-clonal site comprising:
 - a) at least two distinct cloning tags (cTAGs), wherein each cTAG comprises:
 - i) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least one of said validated CRISPR landing sites is unique within the modular CRISPR DNA construct; and
 - b) one or more DNA insert part(s);
 - i) wherein each of said distinct cTAGs are distributed in flanking positions around each of the one or more DNA insert part(s).
2. The modular CRISPR DNA construct of claim 1, wherein said modular CRISPR DNA construct is circular.
3. The modular CRISPR DNA construct of claim 1, wherein said modular CRISPR DNA construct is linear.
4. The modular CRISPR DNA construct of claim 1, wherein said modular CRISPR DNA construct is integrated into the genome of an organism.
5. The modular CRISPR DNA construct of any one of claims 1-4, wherein at least one of said distinct cTAGs comprises at least two validated CRISPR landing sites.
6. The modular CRISPR DNA construct of claim 5, wherein at least one of the CRISPR landing sites is for a Cas9 endonuclease.
7. The modular CRISPR DNA construct of claim 6, wherein at least one of the CRISPR landing sites is for a Cpf1 endonuclease.

8. The modular CRISPR DNA construct of claim 1, wherein at least one of said distinct cTAGs comprises a rare (≥ 8 bases long) restriction endonuclease site.
9. A method for preparing a recombinant nucleic acid molecule, the method comprising:
- a) incubating a mixture comprising:
 - i) a plurality of DNA insert parts, wherein each DNA insert part is flanked by two cloning tags (cTAGs), each cTAG comprising:
 - 1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); and
 - ii) one or more CRISPR complexes targeting at least one of said cTAGs present in at least two of the plurality of DNA insert parts, each CRISPR complex comprising:
 - 1) a CRISPR endonuclease, and
 - 2) a guide RNA capable of recruiting said CRISPR endonuclease to one of said targeted cTAGs;
 - under conditions which allow for digestion of the targeted cTAG(s) in at least two of the plurality of DNA insert parts to generate digested DNA ends, and
 - b) incubating the DNA insert parts with digested cTAGs generated in (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the original insert parts that are ligated in the method.
10. The method of claim 9, wherein the digested DNA ends are sticky ends with overhang sequences capable of hybridizing to each other prior to the covalent joining of the digested ends.
11. The method of claim 10, wherein the CRISPR endonuclease is Cpf1.
12. The method of claim 9, wherein the digested DNA ends are blunt ends.
13. The method of claim 12, wherein the blunt ends are further digested with a ssDNA exonuclease to create sticky ends with overhang sequences, and wherein step b) further comprises

adding a bridging DNA sequence capable of hybridizing to the overhang sequences prior to the covalent joining of the digested ends.

14. The methods of claim 9, 12, or 13, wherein the CRISPR endonuclease is Cas9.

15. The method of any one of claims 9 to 13, wherein the method further comprises the step of:

- i) separating the digested cTAG sequences from the CRISPR complexes prior to step (b), or
- ii) inactivating the CRISPR complexes prior to step (b).

16. The method of claim 15, wherein the separation step comprises a DNA purification step.

17. The method of claim 15, wherein the inactivation step comprises heat or chemical inactivation of said CRISPR complexes.

18. The method of any one of claims 9 to 13, wherein the two cTAGs of each of the plurality of DNA insert parts form a cTAG pair, and wherein said cTAG pair is unique from all other cTAG pairs of the DNA insert parts that are ligated in the method.

19. The method of claim 18, wherein at least one of the cTAGs in each cTAG pair is the same as at least one other cTAG in a different cTAG pair.

20. A method for DNA sequence editing, said method comprising:

a) providing:

i) the modular CRISPR DNA construct of claim 1:

ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG;

1) wherein the first insert cTAG comprises the validated CRISPR landing site(s) of one of the distinct cTAGs of the modular CRISPR DNA construct, and the second insert cTAG comprises the validated CRISPR landing site(s) of another distinct cTAG of the modular CRISPR DNA construct; and

iii) a first and second CRISPR complex targeting the first and second insert cTAGs, respectively, each CRISPR complex comprising:

- 1) a CRISPR endonuclease, and
- 2) a guide RNA capable of recruiting said CRISPR endonuclease to one of said targeted insert cTAGs;

wherein parts (i) and (ii) are each incubated with part (iii) in a single or separate reactions; wherein the first and second CRISPR complexes cleave the first and second insert cTAGs and their corresponding distinct cTAGs to generate digested DNA ends, and

b) incubating the replacement DNA insert part and modular CRISPR DNA construct with digested DNA ends generated in step (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting edited modular CRISPR DNA construct comprises the complete cTAG sequences of the original insert part that is covalently joined by the method.

21. The method for DNA sequence editing of claim 20, wherein the reaction of step (b) comprises a functional ligase.

22. The method of claim 20, wherein the method further comprises the step of:

- i) separating the cleaved first and second insert cTAGs and their corresponding distinct cTAGs from the CRISPR complexes prior to step (b), or
- ii) inactivating the CRISPR complexes prior to step (b).

23. The method of claim 22, wherein the separation step comprises a DNA purification step.

24. The method of claim 22, wherein the inactivation step comprises heat or chemical inactivation of said CRISPR complexes.

25. The method of claim 20, wherein the digested DNA ends are sticky ends with overhang sequences capable of hybridizing to each other prior to the covalent joining of the digested ends.

26. The method of claim 20, wherein the CRISPR endonuclease is Cpf1.
27. The method of claim 20, wherein the digested DNA ends are blunt ends.
28. The method of claim 27, wherein the blunt ends are further digested with a ssDNA exonuclease to create sticky ends with overhang sequences, and wherein step b) further comprises adding a bridging DNA sequence capable of hybridizing to the overhang sequences prior to the covalent joining of the digested ends.
29. The method of claim 20, 27, or 28, wherein the CRISPR endonuclease is Cas9.
30. A method for DNA sequence editing, said method comprising:
- a) providing for a first reaction:
 - i) the modular CRISPR DNA construct of claim 1;
 - ii) at least two CRISPR complexes each targeting at least one distinct cTAG in the modular CRISPR DNA construct, each CRISPR complex comprising:
 - 1) a CRISPR endonuclease, and
 - 2) a guide RNA capable of recruiting said CRISPR endonuclease to one of said targeted distinct cTAGs;wherein the first and second CRISPR complexes cleave the targeted distinct cTAGs in the modular CRISPR DNA construct thereby producing digested cTAGs, and
 - b) providing for a second reaction:
 - i) the modular CRISPR DNA construct with digested cTAGs generated in step (a); and
 - ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG;
 - 1) wherein the first insert cTAG comprises a sequence that is identical or substantially identical to one of the undigested distinct cTAGs that is cleaved in step (a), and the second insert cTAG comprises a sequence that is identical or substantially identical to another of the undigested distinct cTAG that is cleaved in step (a);

under conditions which allow for the joining of the modular CRISPR DNA construct of part (i) and the replacement DNA insert part of part (ii) to form an edited modular CRISPR DNA construct; wherein the resulting edited modular CRISPR DNA construct comprises a reconstituted cTAG that is identical or substantially identical to the undigested distinct cTAGs that were targeted in step (a).

31. The method for DNA sequence editing of claim 30, wherein the joining of step (b) is a ligation, and wherein the replacement DNA insert part is also digested by the CRISPR complexes prior to ligation.

32. The method for DNA sequence editing of claim 30, wherein the CRISPR endonuclease is Cpf1.

33. The method for DNA sequence editing of claim 30, wherein the joining of step (b) is a homologous recombination.

34. The method for DNA sequence editing of claim 30 or 33, wherein the CRISPR endonuclease is Cas9.

35. The method for DNA sequence editing of claim 30, wherein the first and second reactions are not separate.

36. The method for DNA sequence editing of claim 33 or 35, wherein the DNA insert part is not digested by the CRISPR endonuclease.

37. The method for DNA sequence editing of claim 33 or 35, wherein the modular CRISPR DNA construct with digested cTAGs and the undigested replacement DNA insert part are both digested with a single stranded DNA (ssDNA) exonuclease, thereby making compatible overhang DNA ends capable of hybridizing to each other.

38. The method for DNA sequence editing of claim 37, wherein the second reaction further comprises a polymerase or ligase capable of filling in any sequence gaps prior to the covalent joining of the compatible overhang DNA ends.

39. The method for DNA sequence editing of claim 36, wherein the insert part has been chemically modified to prevent CRISPR digestion.

40. The method for DNA sequence editing of claim 36, wherein the first and second insert cTAGs comprise mutated PAMs or mutated protospacers in their validated CRISPR landing sites.

41. A host cell genome comprising a recombinant modular CRISPR DNA construct comprising a CRISPR multi-clonal site, said multi-clonal site comprising:

a) at least two distinct cloning tag (cTAG), wherein each cTAG comprises:

i) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least one of said validated CRISPR landing sites is unique within the modular CRISPR DNA construct; and

b) one or more DNA insert part(s);

i) wherein each of said distinct cTAGs are distributed in flanking positions around each of the one or more DNA insert part(s).

42. A method for preparing a recombinant nucleic acid molecule, the method comprising:

a) incubating a mixture comprising:

i) a plurality of DNA insert parts flanked by two cloning tags (cTAGs), each cTAG comprising:

1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); and

2) a rare (≥ 8 base) restriction enzyme recognition site; wherein at least one of the cTAGs of at least two insert parts comprise the same restriction enzyme site;

- ii) one or more restriction enzymes targeting the rare restriction enzyme sites in at least two of the plurality of DNA insert parts;
under conditions which allow for digestion of the targeted cTAG by the one or more restriction enzymes in at least two of the plurality of DNA insert parts to generate insert parts with digested DNA ends; and
- b) incubating the DNA insert parts with digested DNA ends generated in step (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the original insert parts that are covalently joined in the method.

43. A method for DNA sequence editing, said method comprising:

- a) providing:
 - i) the modular CRISPR DNA construct of claim 1, wherein at least two of the distinct cTAGs comprise a rare (≥ 8 base) restriction enzyme recognition site;
 - ii) a replacement DNA insert part, wherein said replacement DNA insert part is flanked by a first and second insert cTAG;
 - 1) wherein the first insert cTAG comprises the rare restriction enzyme recognition site of one of the distinct cTAGs of the modular CRISPR DNA construct, and the second insert cTAG comprises the rare restriction enzyme recognition site of another distinct cTAG of the modular CRISPR DNA construct; and
 - iii) one or more restriction enzymes targeting the rare restriction enzyme sites in the first and second insert cTAGs;
wherein parts (i) and (ii) are each incubated with part (iii) in a single or separate reactions; wherein the one or more restriction enzymes cleave the rare restriction enzyme recognition sites of first and second insert cTAGs and their corresponding distinct cTAGs to generate digested DNA ends, and
- b) incubating the replacement DNA insert part and modular CRISPR DNA construct with digested DNA ends generated in step (a) under conditions which allow for the covalent joining of the digested DNA ends; wherein the resulting edited modular CRISPR DNA

construct comprises the complete cTAG sequences of the original insert part that is covalently joined by the method.

44. The method for DNA sequence editing of claim 43, wherein the first and second insert cTAGs comprise the same rare restriction enzyme site.

45. A method for preparing a recombinant nucleic acid molecule, the method comprising:

a) incubating a mixture comprising:

i) a plurality of DNA insert parts, wherein each DNA insert part is flanked by two cloning tags (cTAGs), each cTAG comprising:

1) one or more validated CRISPR landing sites, each comprising a protospacer sequence operably linked to a protospacer adjacent motif (PAM); wherein at least two of the DNA insert parts share the same cTAG;

ii) a single stranded DNA (ssDNA) exonuclease;

under conditions which allow for digestion of the shared cTAG in the least two DNA insert parts, thereby generating compatible overhang DNA ends in the at least two DNA insert parts, and

b) incubating the DNA insert parts with digested cTAGs generated in (a) under conditions which allow for the hybridization and covalent joining of the compatible overhang DNA ends of the least two DNA insert parts; wherein the resulting recombinant nucleic acid molecule comprises the complete cTAG sequences of the shared cTAG before digestion.

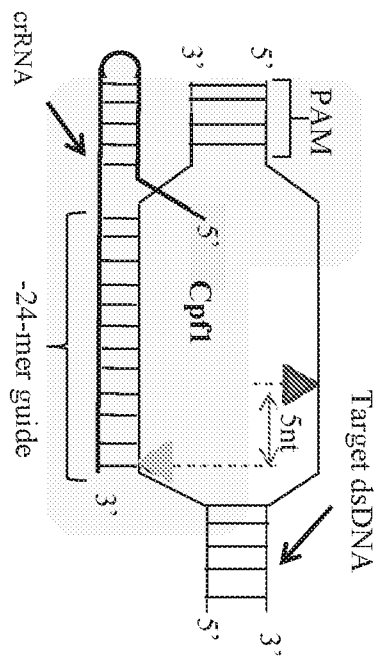
46. The method for preparing a recombinant nucleic acid molecule of claim 45, wherein the mixture of step (a) further comprises a polymerase capable of filling in any sequence gaps prior to the covalent joining of the compatible overhang DNA ends.

47. The method for preparing a recombinant nucleic acid molecule of claim 45, wherein the plurality of DNA insert parts are digested with a double stranded DNA (dsDNA) exonuclease prior before step (a).

48. The method for preparing a recombinant nucleic acid molecule of claim 45, wherein one or more of the plurality of DNA insert parts were cleaved with a CRISPR endonuclease prior to their incubation in step (a).

49. The insert parts of any one of claims 1-49, wherein one or more of the cTAGs is selected from the group consisting of SEQ ID NO: 65-74, 78-81, and combinations thereof.

FIGURE 1 (Continued)



CRISPR-Cpf1

FIGURE 1C

FIGURE 2

FIGURE 2A

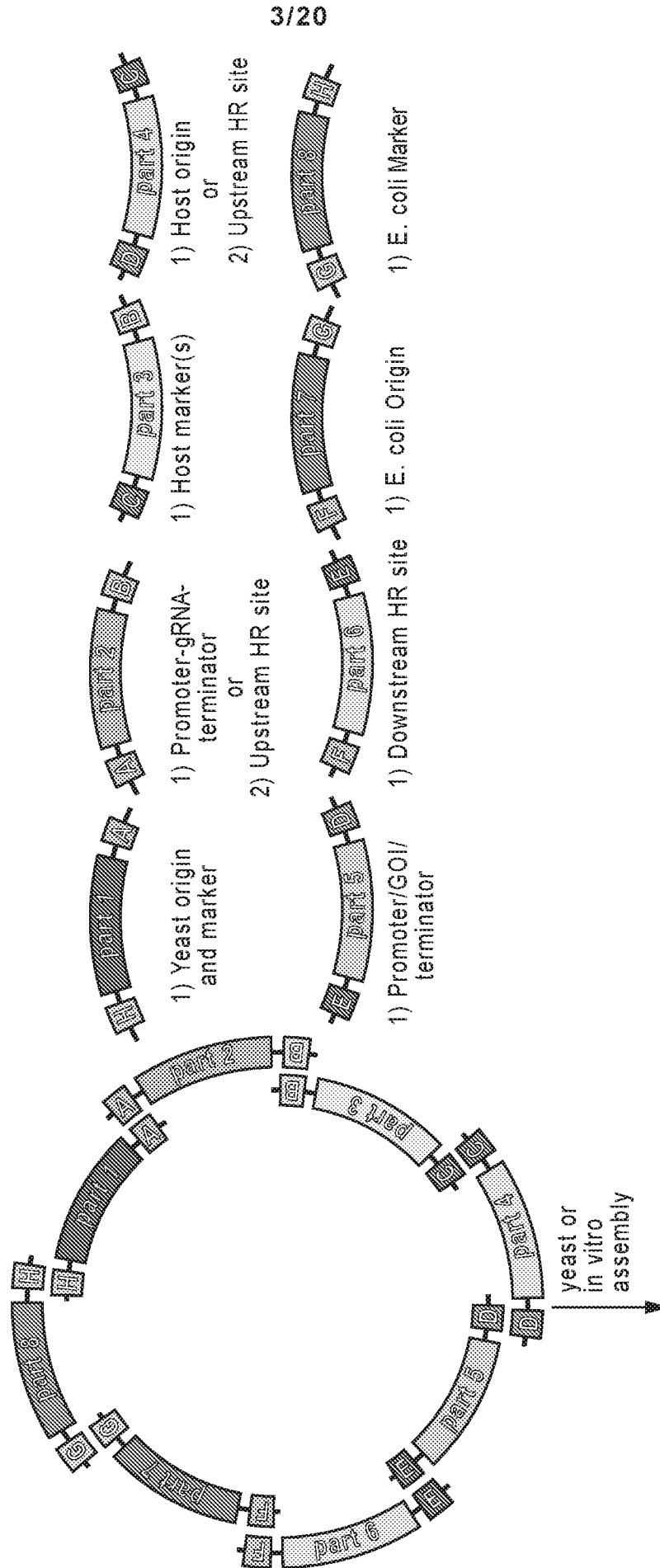


FIGURE 2 (Continued)

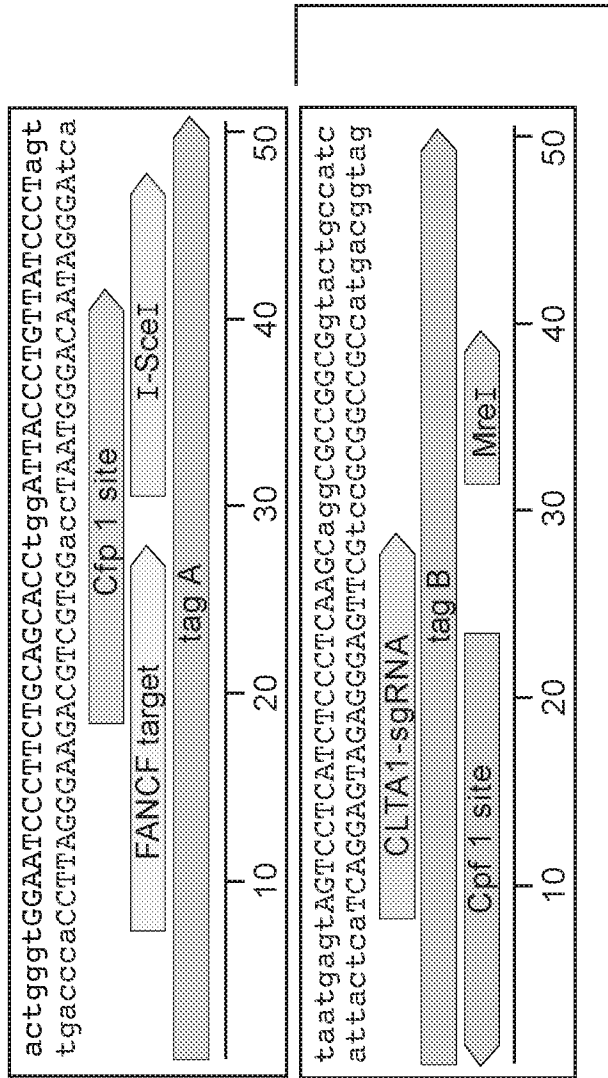


FIGURE 2B

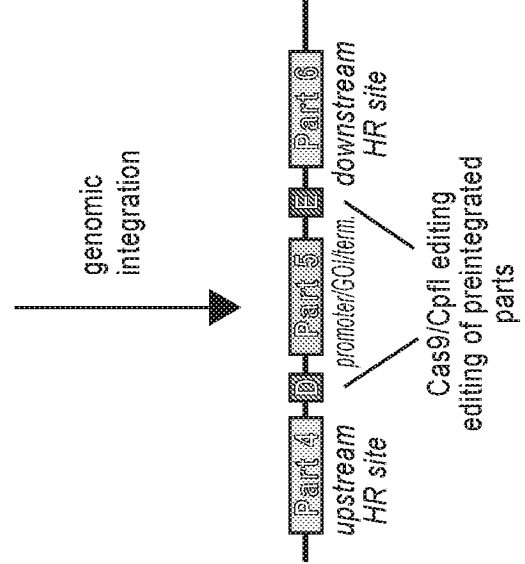
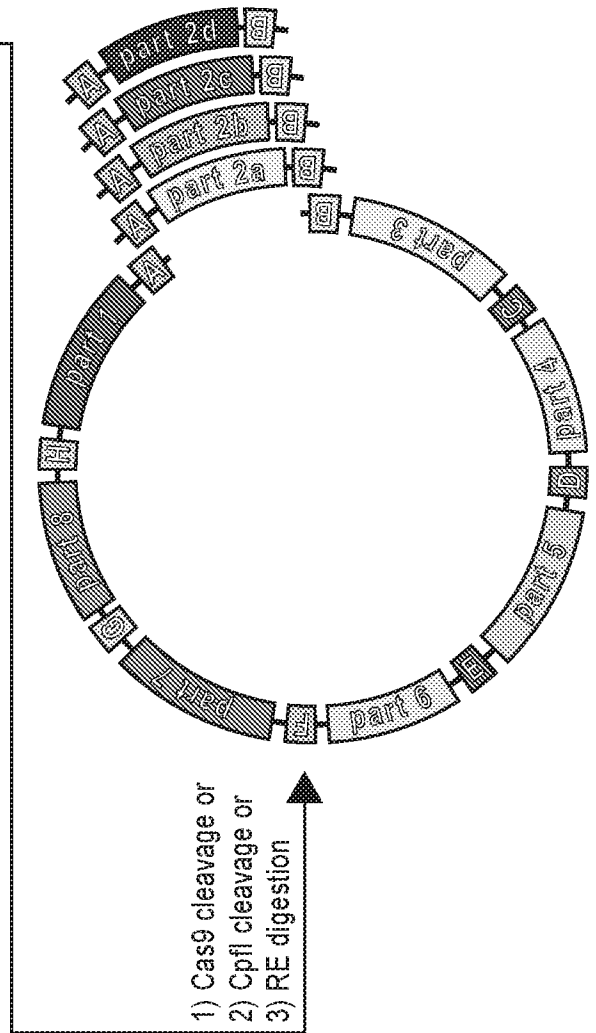
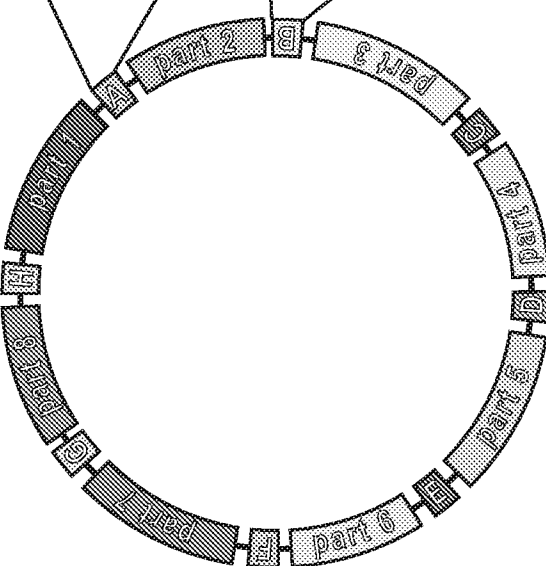


FIGURE 2C

FIGURE 3

FIGURE 3A

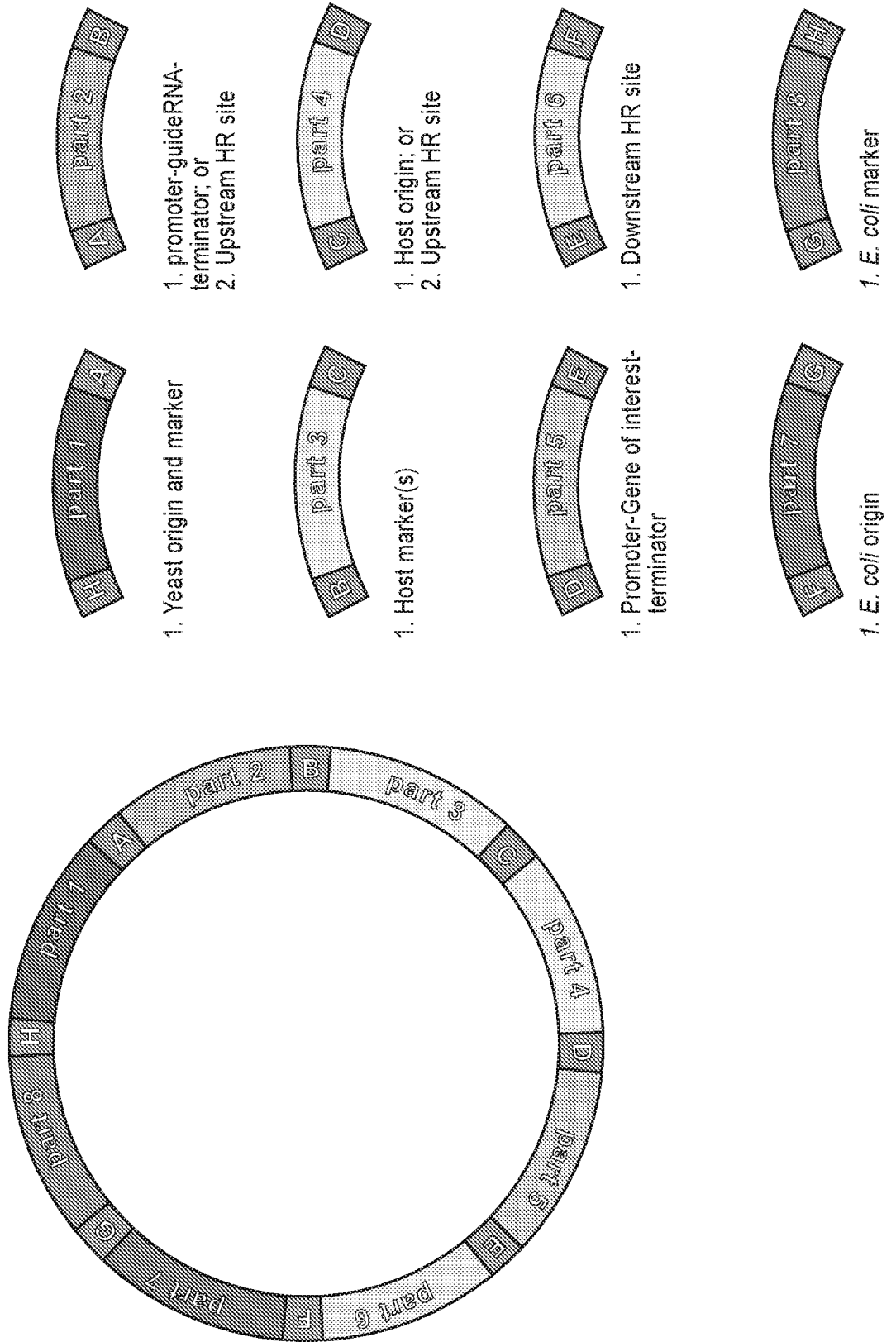


FIGURE 3 (Continued)

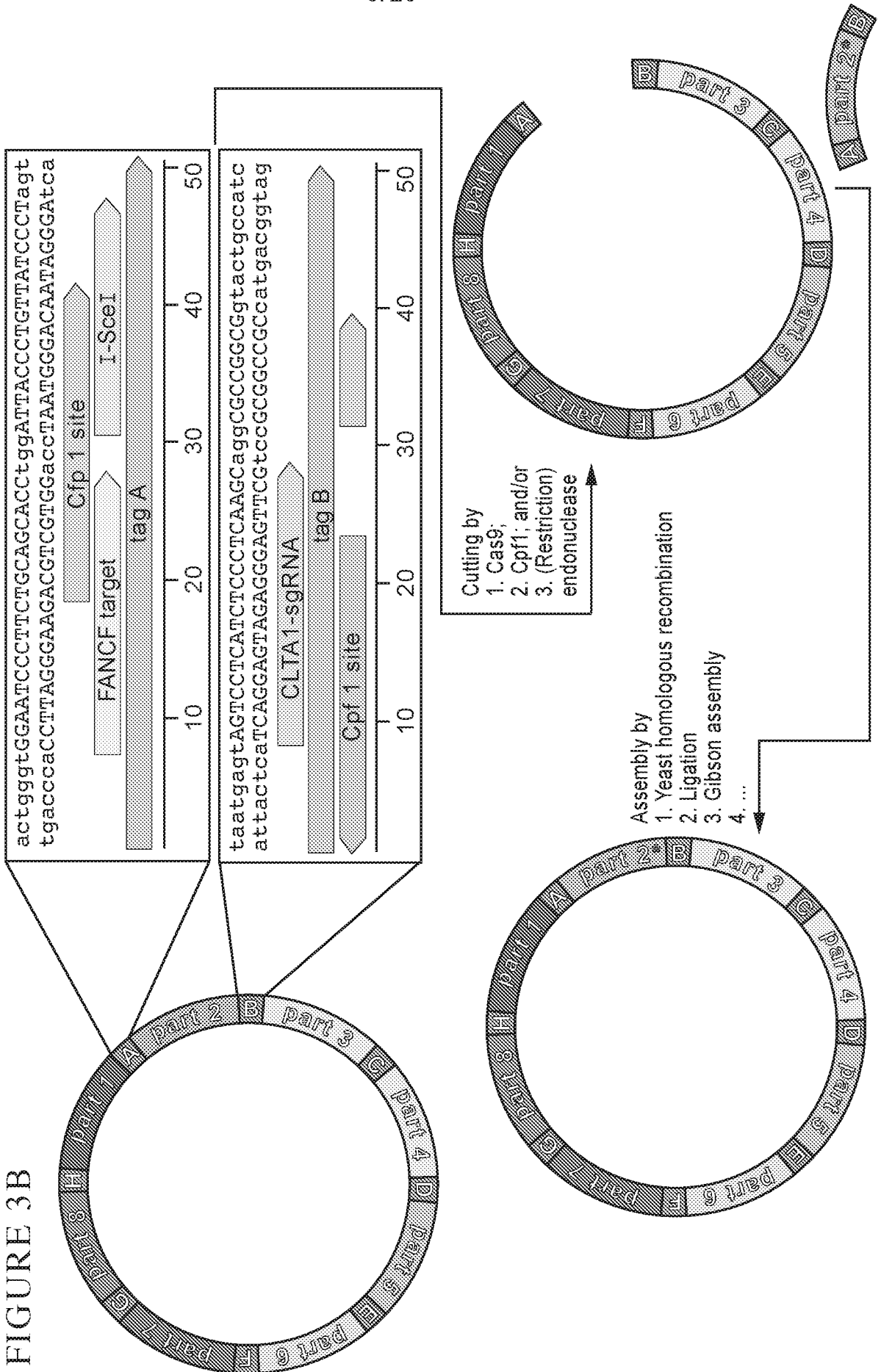


FIGURE 3B

FIGURE 3 (Continued)

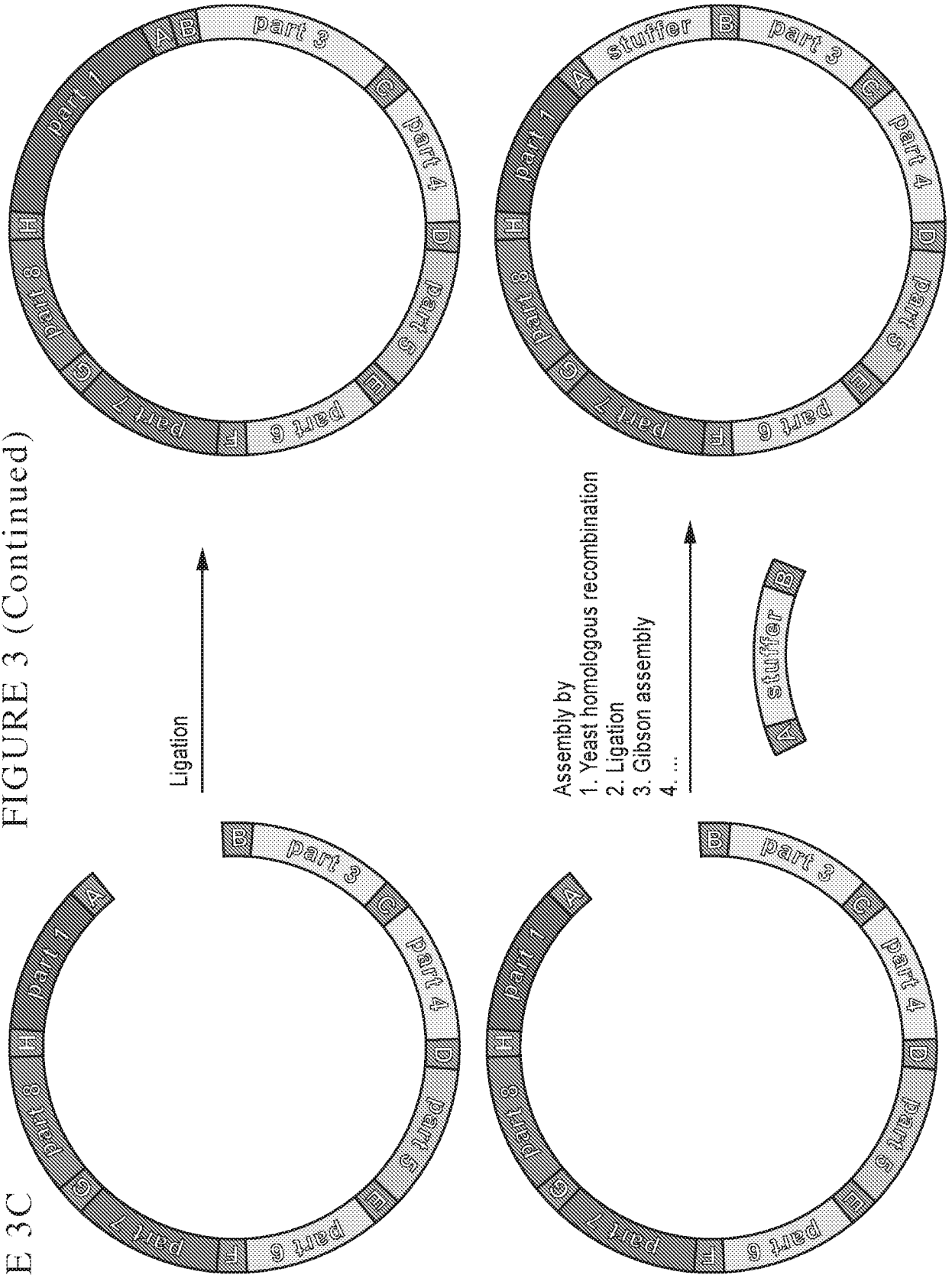


FIGURE 3C

FIGURE 3 (Continued)

FIGURE 3D

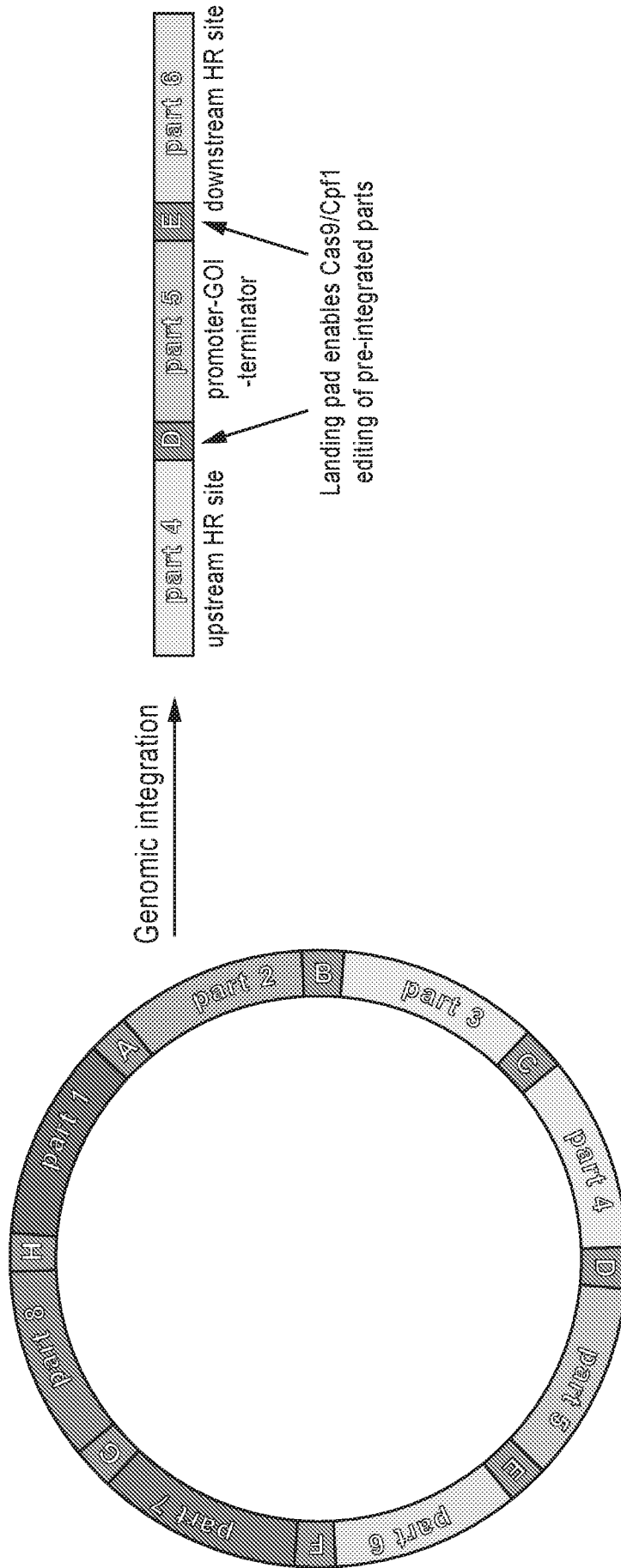


FIGURE 4

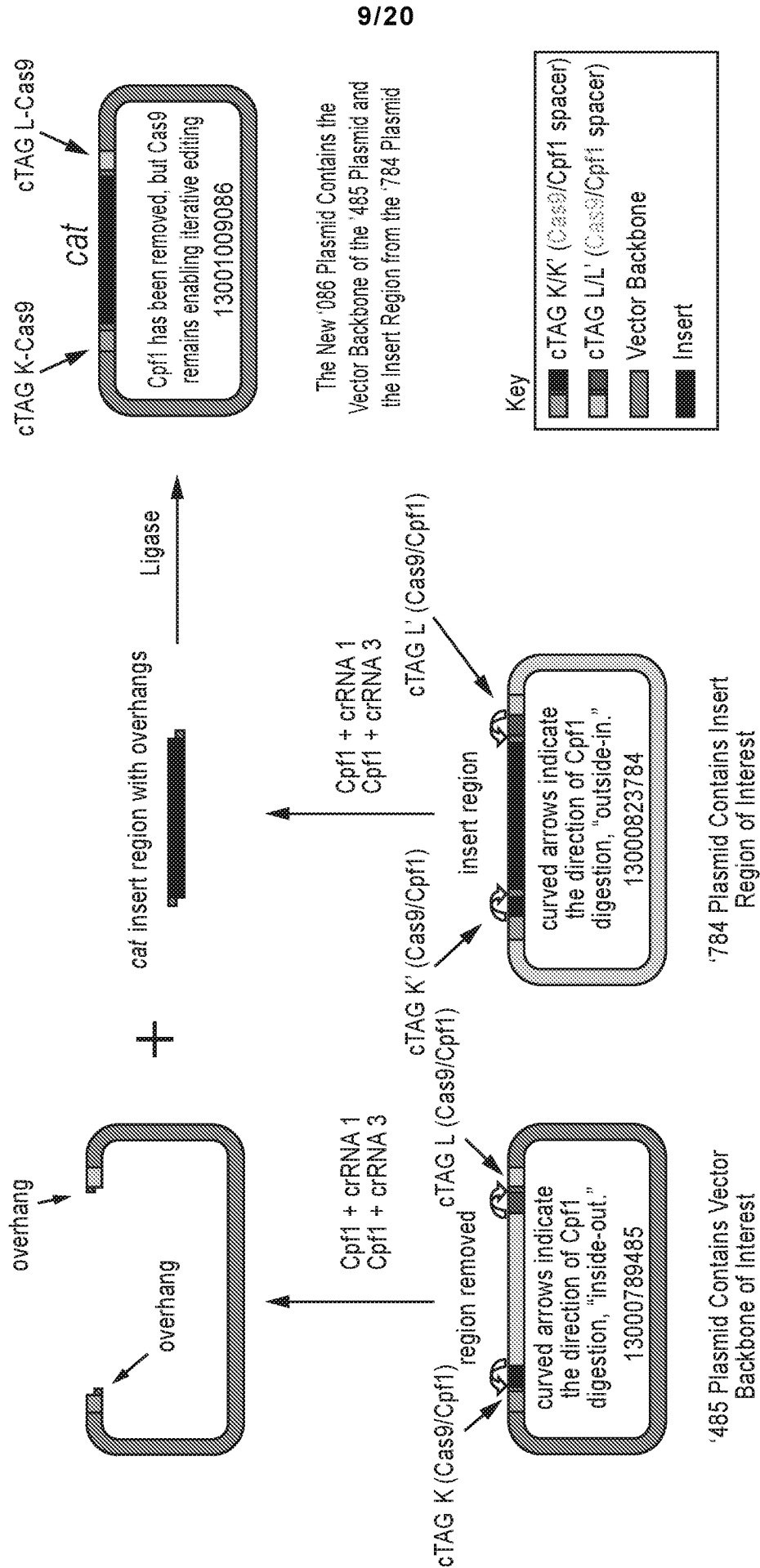


FIGURE 5

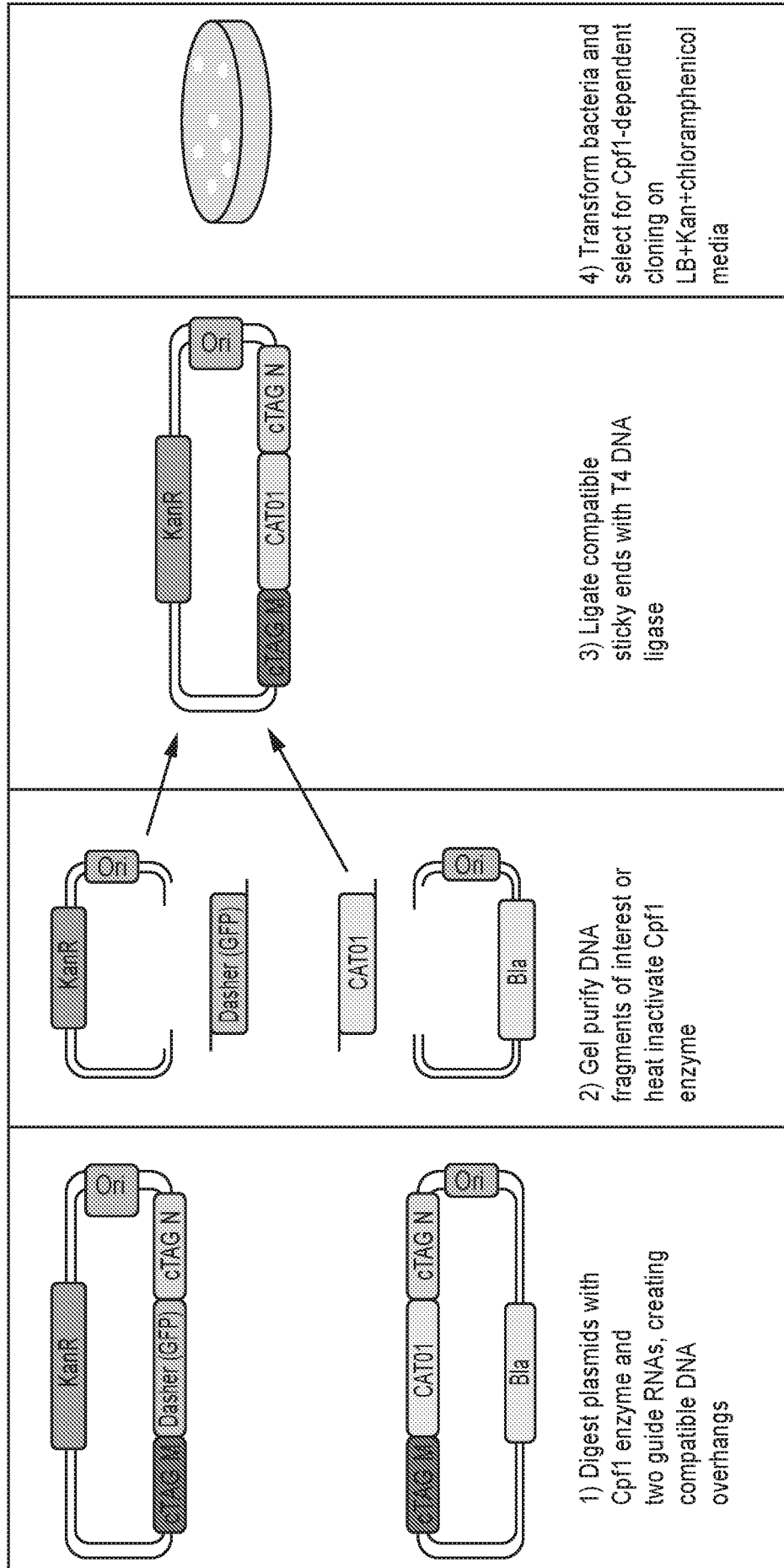
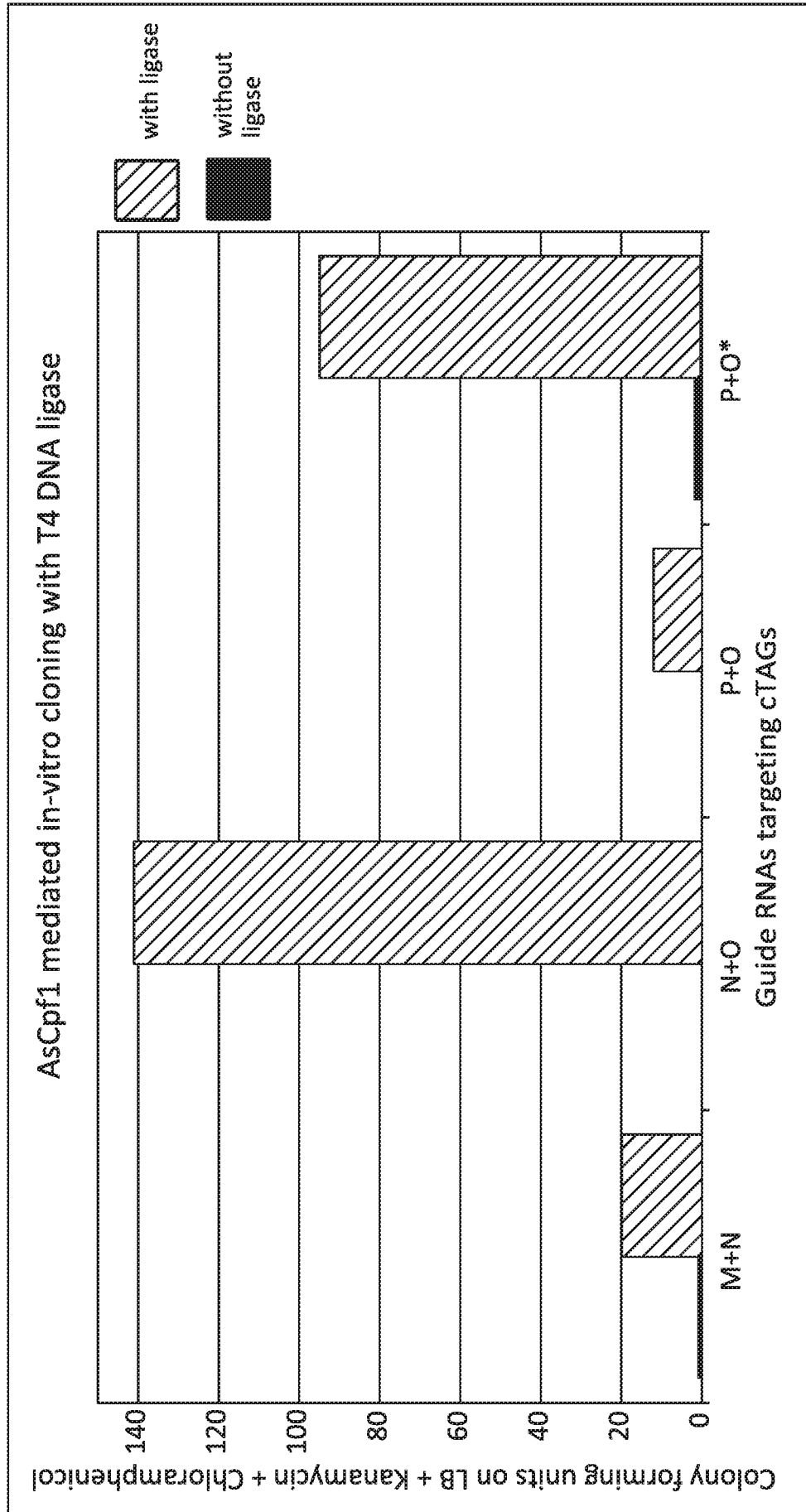


FIGURE 6



12/20

FIGURE 7



FIGURE 8



FIGURE 9

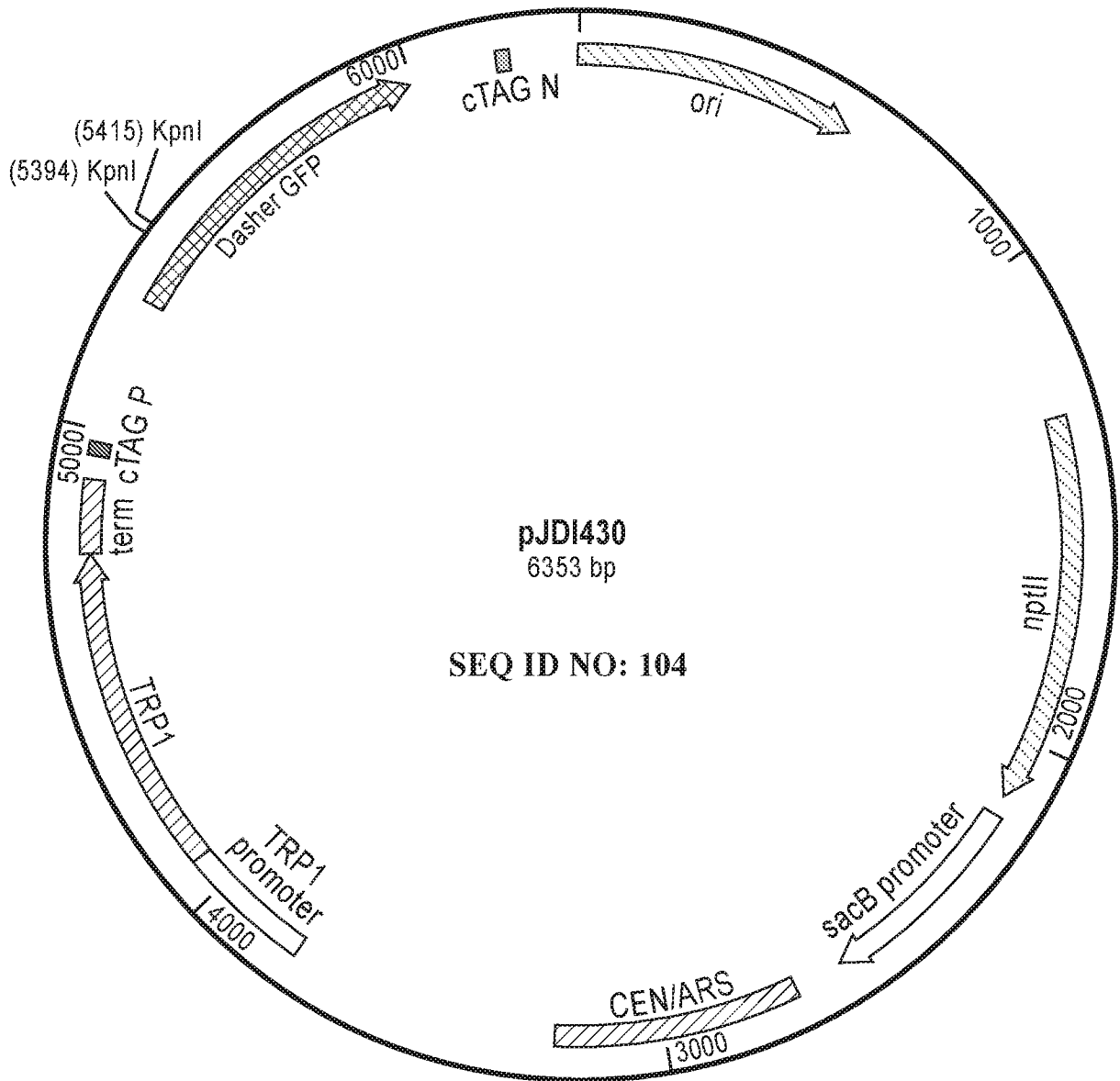


FIGURE 10

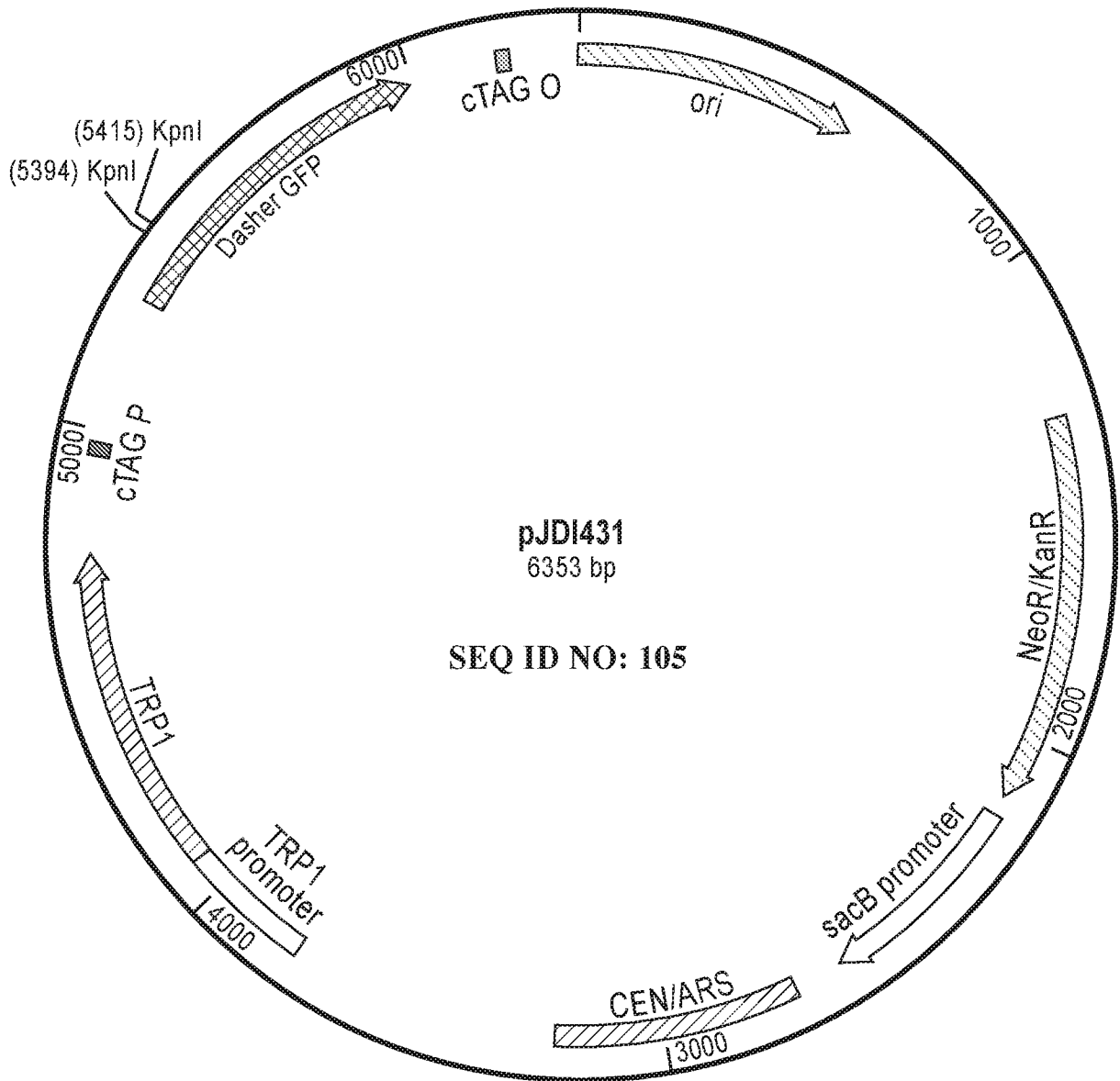


FIGURE 11

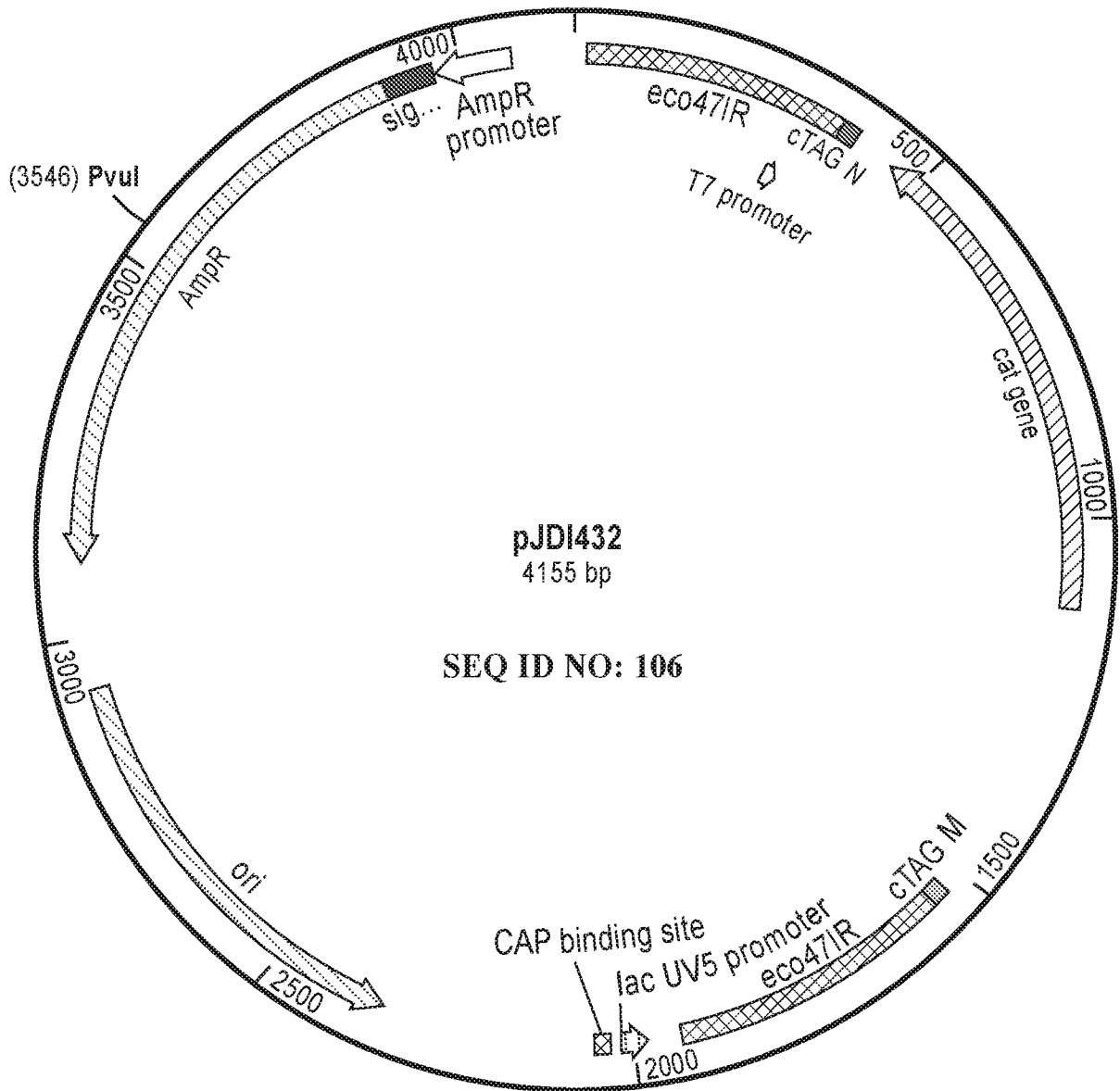


FIGURE 12

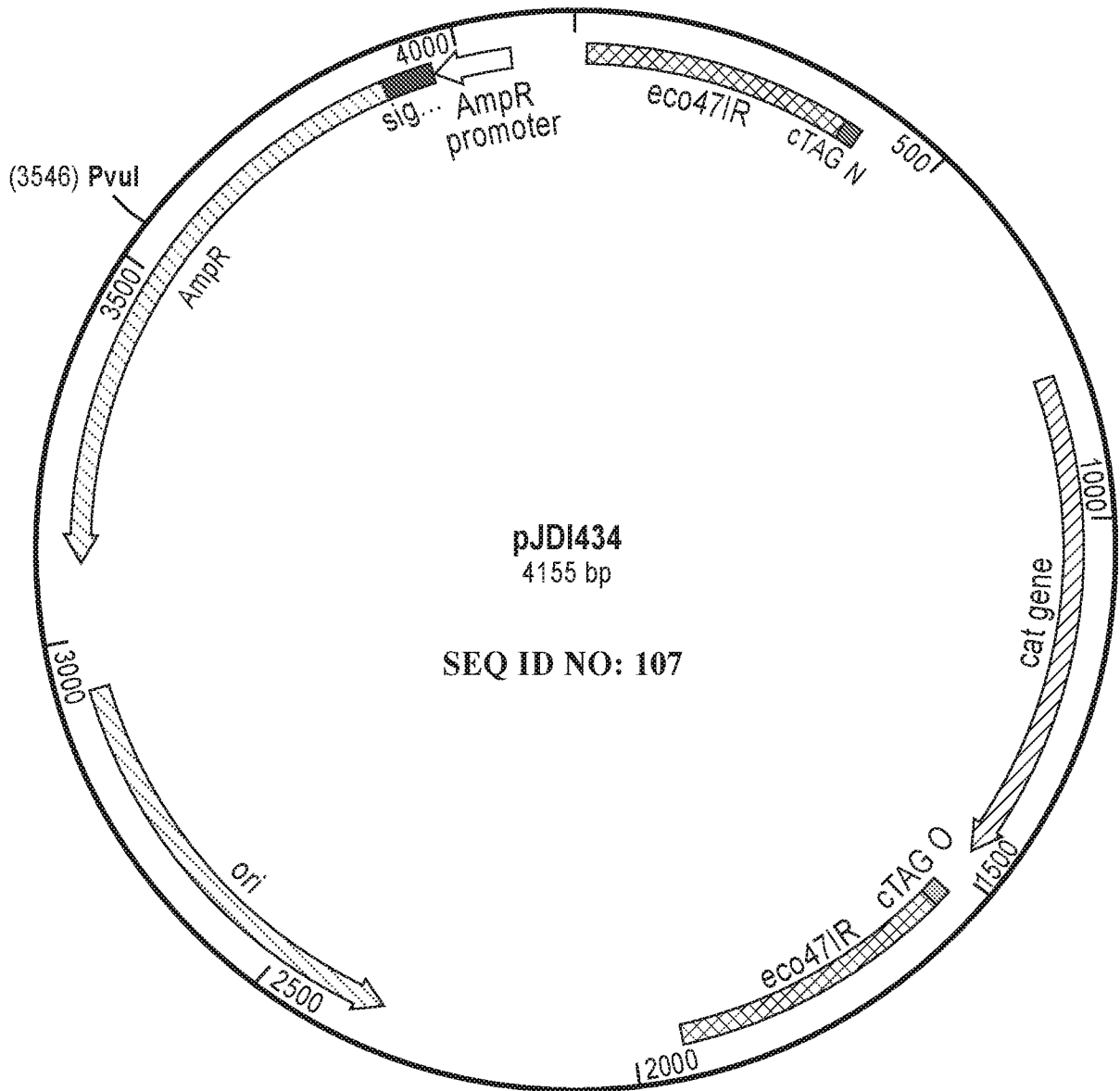


FIGURE 13

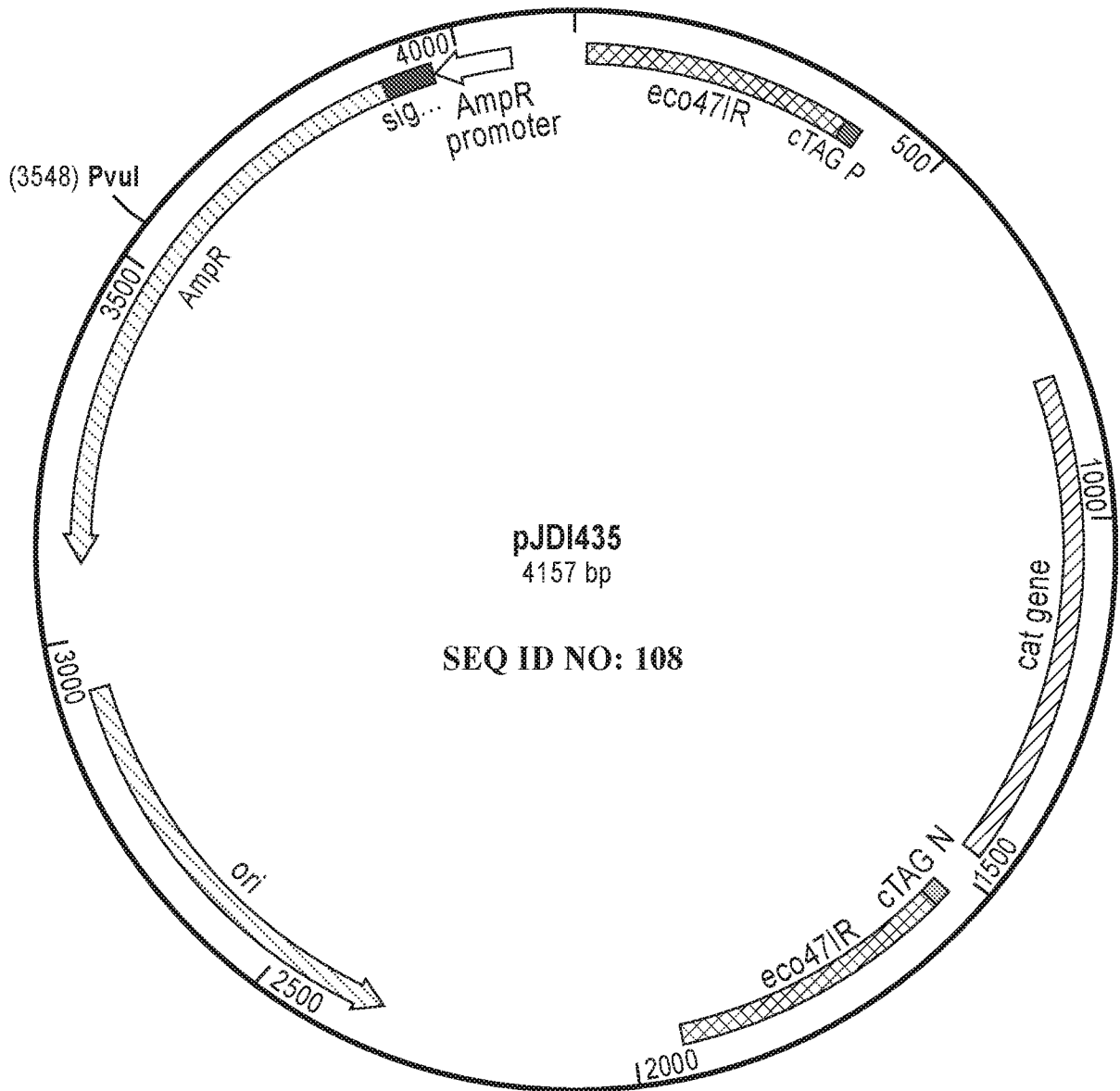


FIGURE 14

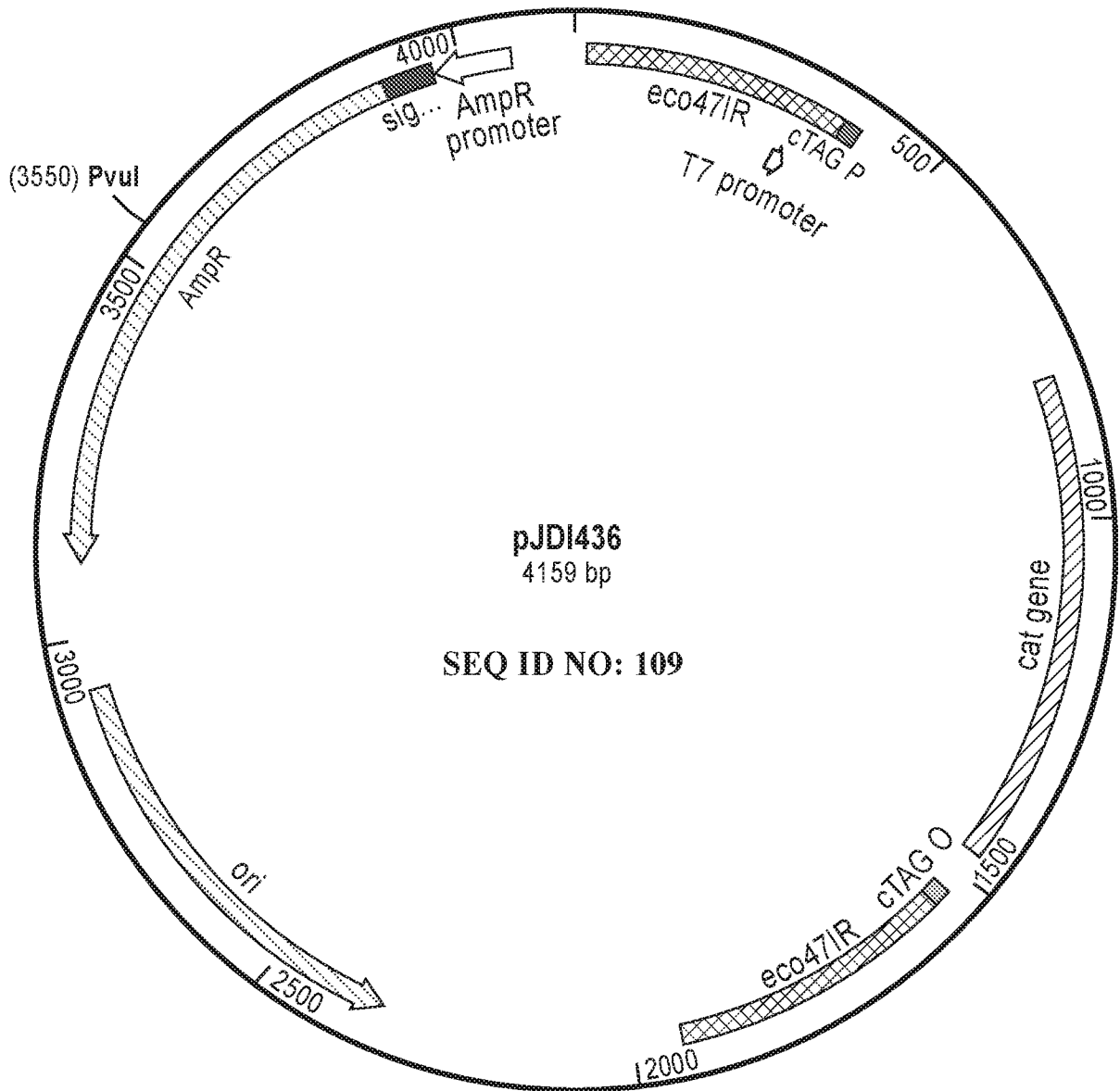
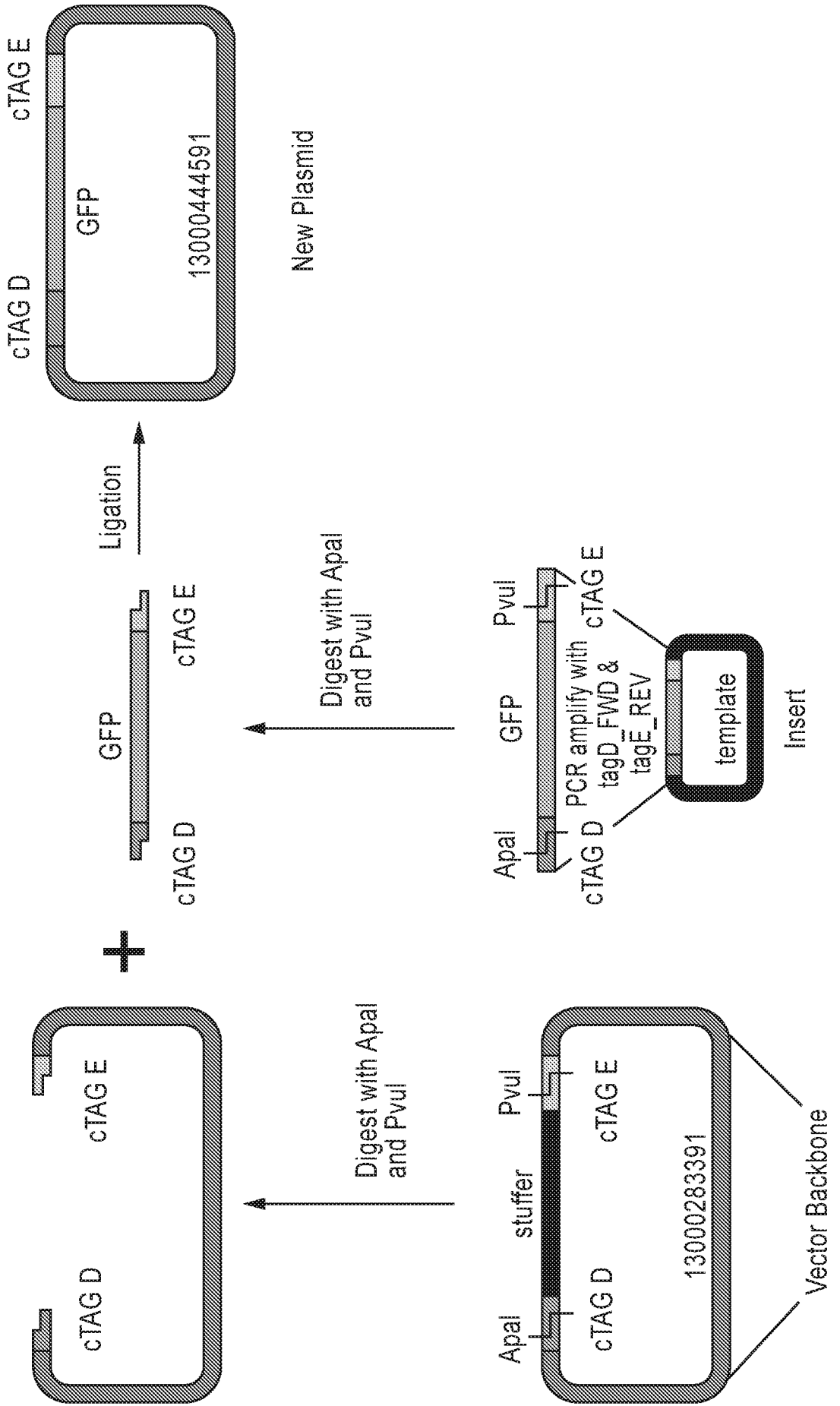


FIGURE 15



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/017573

A. CLASSIFICATION OF SUBJECT MATTER
 IPC(8) - C12N 15/09; C12N 15/10; C12N 15/113; C12N 15/63; C12N 15/64; C12N 15/66 (2018.01)
 CPC - C12N 15/09; C12N 15/10; C12N 15/1031; C12N 15/113; C12N 15/63; C12N 15/64; C12N 15/66 (2018.02)

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)

See Search History document

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

USPC - 435/440; 435/441; 435/462; 435/91.4; 435/91.52; 506/4; 506/16 (keyword delimited)

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2015/200334 A1 (REGENERON PHARMACEUTICALS, INC.) 30 December 2015 (30.12.2015) entire document	1-6, 8-10, 12-25, 27-31, 33-38, 41-48
Y		7, 11, 26, 32, 39, 40
Y	TOTH et al. "Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells." Biol Direct. 14 September 2016 (14.09.2016), Vol. 11, No. 46, Pgs. 1-14. entire document	7, 11, 26, 32
Y	WO 2010/143917 A2 (TOOLGEN INCORPORATION et al) 16 December 2010 (16.12.2010) entire document	39
Y	KLEINSTIVER et al. "Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition," Nat Biotechnol, 02 November 2015 (02.11.2015), Vol. 33, No. 12, Pgs. 1293-1298. entire document	40
A	WO 2011/161413 A2 (ITI SCOTLAND LTD) 29 December 2011 (29.12.2011) entire document	1-48
A	WO 2014/144761 A2 (THE GENERAL HOSPITAL CORPORATION) 18 September 2014 (18.09.2014) entire document	1-48
A	WO 2016/109255 A1 (UNIVERSITY OF SOUTH FLORIDA) 07 July 2016 (07.07.2016) entire document	1-48
P,A	WO 2017/037304 A2 (DSM IP ASSETS B.V.) 09 March 2017 (09.03.2017) entire document	1-48

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search
 11 April 2018

Date of mailing of the international search report
25 MAY 2018

Name and mailing address of the ISA/US
 Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
 P.O. Box 1450, Alexandria, VA 22313-1450
 Facsimile No. 571-273-8300

Authorized officer
 Blaine R. Copenheaver
 PCT Helpdesk: 571-272-4300
 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/017573

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.: 49
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.