

(19) **DANMARK**

(10) **DK/EP 2313498 T3**



(12)

Oversættelse af europæisk patentskrift

Patent- og
Varemærkestyrelsen

-
- (51) Int.Cl.: **C 12 N 9/22 (2006.01)** **C 12 N 5/10 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2017-05-22**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2017-03-15**
- (86) Europæisk ansøgning nr.: **09798667.3**
- (86) Europæisk indleveringsdag: **2009-07-14**
- (87) Den europæiske ansøgnings publiceringsdag: **2011-04-27**
- (86) International ansøgning nr.: **US2009050566**
- (87) Internationalt publikationsnr.: **WO2010009147**
- (30) Prioritet: **2008-07-14 US 80453 P**
- (84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK SM TR**
- (73) Patenthaver: **Precision Biosciences, Inc., 302 East Pettigrew Street , Dibrell Building, Suite A-100, Durham, NC 27701, USA**
- (72) Opfinder: **JANTZ, Derek, 215 Hillview Drive, Durham, NC 27703, USA**
SMITH, James, Jefferson, 1808 Carnation Drive, Durham, NC 27703, USA
- (74) Fuldmægtig i Danmark: **NORDIC PATENT SERVICE A/S, Bredgade 30, 1260 København K, Danmark**
- (54) Benævnelse: **GENKENDELSESSEKVENSER FOR I-CREI-AFLEDTE MEGANUKLEASER OG ANVENDELSER DERAFT**
- (56) Fremdragne publikationer:
WO-A2-2004/067736
WO-A2-2007/047859
WO-A2-2007/093918
WO-A2-2008/010093
US-A1- 2006 206 949
US-A1- 2007 117 128
ARNOULD ET AL.: 'Engineered I-Crel Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells.' J. MOL. BIOL. vol. 371, 03 August 2007, pages 49 - 65, XP022145891

DESCRIPTION

REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority to U.S. Provisional Patent Application No. 61/080,453, filed July 14, 2008.

FIELD OF THE INVENTION

[0002] The invention relates to the field of molecular biology and recombinant nucleic acid technology. In particular, the invention relates to DNA sequences that can be recognized and cleaved by a non-naturally-occurring, rationally-designed, I-CreI-derived homing endonuclease and methods of using same. The invention also relates to methods of producing recombinant nucleic acids, cells, and organisms using such meganucleases which cleave such DNA sites. The invention further relates to methods of conducting a custom-designed, I-CreI-derived meganuclease business.

BACKGROUND OF THE INVENTION

[0003] Genome engineering requires the ability to insert, delete, substitute and otherwise manipulate specific genetic sequences within a genome, and has numerous therapeutic and biotechnological applications. The development of effective means for genome modification remains a major goal in gene therapy, agrotechnology, and synthetic biology (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73; Tzfira et al. (2005), Trends Biotechnol. 23: 567-9; McDaniel et al. (2005), Curr. Opin. Biotechnol. 16: 476-83). A common method for inserting or modifying a DNA sequence involves introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target and selecting or screening for a successful homologous recombination event. Recombination with the transgenic DNA occurs rarely, but can be stimulated by a double-stranded break in the genomic DNA at the target site. Numerous methods have been employed to create DNA double-stranded breaks, including irradiation and chemical treatments. Although these methods efficiently stimulate recombination, the double-stranded breaks are randomly dispersed in the genome, which can be highly mutagenic and toxic. At present, the inability to target gene modifications to unique sites within a chromosomal background is a major impediment to successful genome engineering.

[0004] One approach to achieving this goal is stimulating homologous recombination at a double-stranded break in a target locus using a nuclease with specificity for a sequence that is sufficiently large to be present at only a single site within the genome (see, e.g., Porteus et al. (2005), Nat. Biotechnol. 23: 967-73). The effectiveness of this strategy has been demonstrated in a variety of organisms using chimeric fusions between an engineered zinc finger DNA-binding domain and the non-specific nuclease domain of the FokI restriction enzyme (Porteus (2006), Mol Ther 13: 438-46; Wright et al. (2005), Plant J. 44: 693-705; Urnov et al. (2005), Nature 435: 646-51). Although these artificial zinc finger nucleases stimulate site-specific recombination, they retain residual non-specific cleavage activity resulting from under-regulation of the nuclease domain and frequently cleave at unintended sites (Smith et al. (2000), Nucleic Acids Res. 28: 3361-9). Such unintended cleavage can cause mutations and toxicity in the treated organism (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73).

[0005] A group of naturally-occurring nucleases which recognize 15-40 base-pair cleavage sites commonly found in the genomes of plants and fungi may provide a less toxic genome engineering alternative. Such "meganucleases" or "homing endonucleases" are frequently associated with parasitic DNA elements, such as group 1 self-splicing introns and inteins. They naturally promote homologous recombination or gene insertion at specific locations in the host genome by producing a double-stranded break in the chromosome, which recruits the cellular DNA-repair machinery (Stoddard (2006), Q. Rev. Biophys. 38: 49-95). Meganucleases are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cys box family and the HNH family. These families are characterized by structural motifs, which affect catalytic activity and recognition sequence. For instance, members of the LAGLIDADG family are characterized by having either one or two copies of the conserved LAGLIDADG motif (see Chevalier et al. (2001), Nucleic Acids Res. 29(18): 3757-3774). The LAGLIDADG meganucleases with a single copy of the LAGLIDADG motif form homodimers, whereas members with two copies of the LAGLIDADG motif are found as monomers.

[0006] Natural meganucleases, primarily from the LAGLIDADG family, have been used to effectively promote site-specific genome modification in plants, yeast, *Drosophila*, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the meganuclease recognition sequence (Monnat et al. (1999), Biochem.

Biophys. Res. Commun. 255: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Rouet et al. (1994), Mol. Cell. Biol. 14: 8096-106; Chilton et al. (2003), Plant Physiol. 133: 956-65; Puchta et al. (1996), Proc. Natl Acad. Sci. USA 93: 5055-60; Rong et al. (2002), Genes Dev. 16: 1568-81; Gouble et al. (2006), J. Gene Med. 8(5):616-622).

[0007] Systematic implementation of nuclease-stimulated gene modification requires the use of engineered enzymes with customized specificities to target DNA breaks to existing sites in a genome and, therefore, there has been great interest in adapting meganucleases to promote gene modifications at medically or biotechnologically relevant sites (Porteus et al. (2005), Nat. Biotechnol. 23: 967-73; Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62).

[0008] I-CreI (SEQ ID NO: 1) is a member of the LAGLIDADG family which recognizes and cleaves a 22 base pair recognition sequence in the chloroplast chromosome, and which presents an attractive target for meganuclease redesign. Genetic selection techniques have been used to modify the wild-type I-CreI recognition site preference (Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Chames et al. (2005), Nucleic Acids Res. 33: e178; Seligman et al. (2002), Nucleic Acids Res. 30: 3870-9; Arnould et al. (2006), J. Mol. Biol. 355: 443-58). More recently, a method of rationally-designing mono-LAGLIDADG meganucleases was described which is capable of comprehensively redesigning I-CreI and other such meganucleases to target widely-divergent DNA sites, including sites in mammalian, yeast, plant, bacterial, and viral genomes (WO 2007/047859).

[0009] The DNA sequences recognized by I-CreI are 22 base pairs in length. One example of a naturally-occurring I-CreI recognition site is provided in SEQ ID NO: 2 and SEQ ID NO: 3, but the enzyme will bind to a variety of related sequences with varying affinity. The enzyme binds DNA as a homodimer in which each monomer makes direct contacts with a nine base pair "half-site" and the two half-sites are separated by four base pairs that are not directly contacted by the enzyme (Figure 1a). Like all LAGLIDADG family meganucleases, I-CreI produces a staggered double-strand break at the center of its recognition sequences which results in the production of a four base pair 3'-overhang (Figure 1a). The present invention concerns the central four base pairs in the I-CreI recognition sequences (*i.e.* the four base pairs that become the 3' overhang following I-CreI cleavage, or "center sequence", Figure 1b). In the case of the native I-CreI recognition sequence in the *Chlamydomonas reinhardtii* 23S rRNA gene, this four base pair sequence is 5'-GTGA-3'. In the interest of producing genetically-engineered meganucleases which recognize DNA sequences that deviate from the wild-type I-CreI recognition sequences, it is desirable to know the extent to which the four base pair center sequence can deviate from the wild-type sequences. A number of published studies concerning I-CreI or its derivatives evaluated the enzyme, either wild-type or genetically-engineered, using DNA substrates that employed either the native 5'-GTGA-3' central sequence or the palindromic sequence 5'-GTAC-3'. Recently, Arnould *et. al.* (Arnould et al. (2007), J. Mol. Biol. 371: 49-65) reported that a set of genetically-engineered meganucleases derived from I-CreI cleaved DNA substrates with varying efficiencies depending on whether the substrate sequences were centered around 5'-GTAC-3', 5'-TTGA-3', 5'-GAAA-3', or 5'-ACAC-3' (cleavage efficiency: GTAC > ACAC >> TTGA ≈ GAAA).

SUMMARY OF THE INVENTION

[0010] The present invention is based, in part, upon the identification and characterization of a subset of DNA recognition sequences that can act as efficient substrates for cleavage by the rationally-designed, I-CreI-derived meganucleases (hereinafter, "I-CreI-derived meganucleases"). The invention is defined by the appended claims.

[0011] Disclosed are methods of identifying sets of 22 base pair DNA sequences which can be cleaved by I-CreI-derived meganucleases and which have, at their center, one of a limited set of four base pair DNA center sequences that contribute to more efficient cleavage by the I-CreI-derived meganucleases. The invention also provides methods that use such DNA sequences to produce recombinant nucleic acids, cells and organisms by utilizing the recognition sequences as substrates for I-CreI-derived meganucleases, and products incorporating such DNA sequences.

[0012] Thus, in one aspect, the invention provides an *ex vivo* method for cleaving a double-stranded DNA comprising: (a) identifying in the DNA at least one recognition site for a rationally-designed I CreI-derived meganuclease with altered specificity relative to I-CreI, wherein the recognition site is not cleaved by a naturally-occurring I-CreI, wherein the recognition site is identified based on the presence of a four base pair central sequence selected from the group consisting of TTGT, TTAT, TTTC, TTCC, TTAG, TTAC, and TTGC; (b) providing the rationally-designed meganuclease; and (c) contacting the DNA with the rationally-designed meganuclease; whereby the rationally-designed meganuclease cleaves the DNA.

[0013] In another aspect, the invention provides a method for cleaving a double-stranded DNA comprising: (a) incorporating a four base pair central sequence into a recognition site for a rationally-designed I CreI derived meganuclease with altered specificity relative to I-CreI, wherein the recognition site is not cleaved by a naturally-occurring I-CreI, wherein the four base pair

central sequence is selected from the group consisting of TTGT, TTAT, TTTC, TTCC, TTAG, TTAC, and TTGC; and (b) introducing said recognition site into said DNA; and (c) providing the rationally-designed meganuclease; and (d) contacting the DNA with the rationally-designed meganuclease; whereby the rationally-designed meganuclease cleaves the DNA.

[0014] In some embodiments, the DNA cleavage is *in vitro*.

[0015] In some embodiments, the DNA is selected from the group consisting of a PCR product; an artificial chromosome; genomic DNA isolated from bacteria, fungi, plants, or animal cells; and viral DNA.

[0016] In some embodiments, the DNA is present in a cell selected from the group consisting of a bacterial, fungal, plant and animal cell.

[0017] In some embodiments, the DNA is present in a nucleic acid selected from the group consisting of a plasmid, a prophage and a chromosome.

[0018] In certain embodiments, the method further comprises rationally-designing the I Crel derived meganuclease to recognize the recognition site.

[0019] In some embodiments, the method further comprises producing the rationally-designed I-Crel-derived meganuclease.

[0020] Disclosed is a cell transformed with a nucleic acid comprising, in order: a) a first 9 base pair DNA sequence which can be bound by an I Crel derived meganuclease monomer or by a first domain from a single-chain I Crel derived meganuclease; b) a four base pair DNA sequence selected from the group consisting of GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA and GCAT; and c) a second 9 base pair DNA sequence which can be bound by an I Crel derived meganuclease monomer or by a second domain from the single-chain I Crel derived meganuclease, wherein the second 9 base pair DNA sequence is in the reverse orientation relative to the first.

[0021] Disclosed is also a cell containing an exogenous nucleic acid sequence integrated into its genome, comprising, in order: a) a first exogenous 9 base pair DNA sequence which can be bound by an I Crel derived meganuclease monomer or by a first domain from a single-chain I Crel derived meganuclease; b) an exogenous four base pair DNA sequence selected from the group consisting of GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA and GCAT; and a) a second exogenous 9 base pair DNA sequence which can be bound by an I Crel derived meganuclease monomer or by a second domain from the single-chain I Crel derived meganuclease, wherein the second 9 base pair DNA sequence is in the reverse orientation relative to the first.

[0022] The nucleic acid can be a plasmid, an artificial chromosome, or a viral nucleic acid.

[0023] The cell can be a non-human animal cell, a plant cell, a bacterial cell, or a fungal cell.

[0024] The four base pair DNA sequence can be TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG or GCAG.

[0025] The four base pair DNA sequence can be GTGT, GTAT, TTAG, GTAG, TTAC, TCTC, TCAC, GTCC, GTAC, TCGC, AAGC, GAGC, GCGC, GTGC, TAGC, TTGC, ATGC, ACAC, ATAC, CTAA, CTAC, GTAA, GAGA, GTGA, GGAC, GTAC, GCGA, GCTT, GCTC, GCGC, GCAC, GCTA, GCAA or GCAT.

[0026] Disclosed is a method of conducting a custom-designed, I-Crel-derived meganuclease business comprising: (a) receiving a DNA sequence into which a double-strand break is to be introduced by a rationally-designed I Crel-derived meganuclease; (b) identifying in the DNA sequence at least one recognition site for a rationally-designed I Crel-derived meganuclease with altered specificity relative to I-Crel, wherein the recognition site is not cleaved by a naturally-occurring I-Crel, wherein the recognition site has a four base pair central sequence selected from the group consisting of TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG and GCAG; and (c) providing the rationally-designed meganuclease.

[0027] The method can further comprise rationally-designing the I-CreI derived meganuclease to recognize the recognition site.

[0028] The method can further comprise producing the rationally-designed meganuclease.

[0029] The rationally-designed meganuclease can be provided to the same party from which the DNA sequence has been received.

[0030] These and other aspects and embodiments of the invention will be apparent to one of ordinary skill in the art from the following detailed description of the invention, figures and appended claims.

BRIEF DESCRIPTION OF THE FIGURES

[0031]

Figure 1. (A) Schematic illustration of the interactions between the naturally-occurring I-CreI homodimer and a double-stranded recognition sequence, based upon crystallographic data. This schematic representation depicts one recognition sequence (SEQ ID NO: 2 and SEQ ID NO: 3), shown as unwound for illustration purposes only, bound by the homodimer, shown as two ovals. The bases of each DNA half-site are numbered -1 through -9, and the amino acid residues of I-CreI which form the recognition surface are indicated by one-letter amino acid designations and numbers indicating residue position. The four base pairs that comprise the center sequence are numbered +1 to +4. Solid black lines: hydrogen bonds to DNA bases. (B) One wild-type I-CreI recognition sequence (SEQ ID NO: 2 and SEQ ID NO: 3) showing the locations of the inverted half-sites and center sequence.

Figure 2. (A) Schematic diagram of the plasmid substrates evaluated to determine center sequence preference. A set of pUC-19 plasmids were produced which harbored potential recognition sequences for the genetically-engineered meganuclease DJ1. These potential recognition sequences comprised a pair of inverted DJ1 half-sites separated by a variety of different four base pair center sequences (numbered +1 through +4), as described below. (B) Example of gel electrophoresis data showing DJ1 meganuclease cleavage of plasmid substrates described in (A). The "uncut" arrow indicates XmnI linearized plasmid substrate. The "cut" arrows indicate XmnI linearized plasmid substrates which have also been successfully cleaved by DJ1.

Figure 3. (A) Schematic diagram of a T-DNA that was stably integrated into the *Arabidopsis thaliana* genome as described in Example 1. In this T-DNA construct, a codon-optimized gene encoding the genetically-engineered BRP2 meganuclease (BRP2) (SEQ ID NO: 8) is under the control of a Hsp70 promoter (HSP) and a NOS terminator (TERM). A pair of potential BRP2 recognition sequences (Site1, Site2) are housed adjacent to the terminator separated by 7 base pairs containing a PstI restriction enzyme site (PstI). A kanamycin resistance marker (Kan) is also housed on the T-DNA to allow selection for stable transformants. (B) The expected product following BRP2 meganuclease cleavage of Site1 and Site2 showing loss of the intervening 7 base pair fragment and PstI restriction site. Arrows show the location of PCR primers used to screen for cleavage of the T-DNA. (C) Sequences of the BRP2 recognition sequences housed on either the GTAC construct (GTAC) or the TAGA construct (TAGA) with center sequences underlined. (D) Example of electrophoresis data from a plant transformed with the GTAC construct. Genomic DNA was isolated from the leaves of *Arabidopsis* seedlings stably transformed with either the GTAC T-DNA construct before and after a 2 hour "heat-shock" to induce BRP2 expression. DNA samples were then added to PCR reactions using the primers shown in (B). PCR reactions were digested with PstI and visualized by gel electrophoresis. C: control lane lacking PstI. 44, 45, and 46: PCR samples from three representative plants showing nearly complete digestion by PstI in samples taken prior to heat shock (-lanes) and very little digestion by PstI in samples taken after heat-shock (+lanes). These results indicate that the BRP2 meganuclease was able to cleave the BRP2 recognition sequence which incorporated a GTAC center sequence *in vivo*.

Figure 4. (A) Schematic diagram of a T-DNA that was stably integrated into the *Arabidopsis thaliana* genome as described in Example 2. In this T-DNA construct, a codon-optimized gene encoding the BRP12-SC meganuclease (BRP12-SC) (SEQ ID NO: 15) is under the control of a Hsp70 promoter (HSP) and a NOS terminator (TERM). A pair of potential BRP12-SC recognition sequences (Site1, Site2) are housed adjacent to the terminator separated by 7 base pairs containing a PstI restriction enzyme site (PstI). A kanamycin resistance marker (Kan) is also housed on the T-DNA to allow selection for stable transformants. (B) The expected product following BRP12-SC meganuclease cleavage of Site1 and Site2 showing loss of the intervening 7 base pair fragment and PstI restriction site. Arrows show the location of PCR primers used to screen for cleavage of the T-DNA. (C) Sequences of the BRP12-SC recognition sequences housed on either the GTAC construct (GTAC) or the TAGA construct (TAGA) with center sequences underlined.

Figure 5. Graphic representation of the effects of meganuclease concentration and center sequence on *in vitro* meganuclease cleavage. The BRP2 meganuclease (SEQ ID NO: 8, see Example 1) was added at the indicated concentration to a digest reaction

containing 0.25 picomoles of a plasmid substrate harboring either a BRP2 recognition sequence with the center sequence GTAC or a BRP2 recognition sequence with the center sequence TAGA. Reactions were 25 microliters in SA buffer (25 mM Tris-HCL, pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 5mM EDTA). Reactions were incubated at 37°C for 2 hours and were then visualized by gel electrophoresis and the percent of plasmid substrate cleaved by the meganuclease was plotted as a function of meganuclease concentration.

DETAILED DESCRIPTION OF THE INVENTION

1.1 Introduction

[0032] The present invention is based, in part, upon the identification and characterization of particular DNA sequences that are more efficiently cleaved by the rationally-designed, I-Crel-derived meganucleases. Specifically, the invention is based on the discovery that certain four-base pair DNA sequences, when incorporated as the central four-base pairs of a rationally-designed, I-Crel-derived meganuclease recognition sequence, can significantly impact cleavage by the corresponding meganuclease although the meganuclease does not, based upon analysis of crystal structures, appear to contact the central four base pairs. As there are four DNA bases (A, C, G, and T), there are 4⁴ or 256 possible DNA sequences that are four base pairs in length. All of these possible sequences were examined to determine the subsets of sequences that are more efficiently cleaved by I-Crel-derived meganucleases. The results of this analysis allow for more accurate prediction of whether or not a particular double-stranded DNA site 22 base pairs in length can be more efficiently cleaved by the I-Crel-derived meganuclease.

1.2 References and Definitions

[0033] The patent and scientific literature referred to herein establishes knowledge that is available to those of skill in the art. The issued U.S. patents, allowed applications, published foreign applications, and references, including GenBank database sequences, that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference.

[0034] As used herein, the term "I-Crel-derived meganuclease" refers to a rationally-designed (*i.e.*, genetically-engineered) meganuclease that is derived from I-Crel. The term genetically-engineered meganuclease, as used herein, refers to a recombinant variant of an I-Crel homing endonuclease that has been modified by one or more amino acid insertions, deletions or substitutions that affect one or more of DNA-binding specificity, DNA cleavage activity, DNA-binding affinity, and/or dimerization properties. Some genetically-engineered meganucleases are known in the art (see, *e.g.*, Porteus et al. (2005), Nat. Biotechnol. 23: 967-73; Sussman et al. (2004), J. Mol. Biol. 342: 31-41; Epinat et al. (2003), Nucleic Acids Res. 31: 2952-62) and a general method for rationally-designing such variants is disclosed in WO 2007/047859. Additional methods for genetically-engineering such variants are disclosed in WO 04/067736, WO 07/060495, WO 06/097853, WO 07/049095, WO 08/102198, WO 08/010093, WO 08/010009, WO 07/093918, WO 07/093836, WO 08/102274, WO 08/059317, WO 09/013622, WO 09/019614, WO 09/019528, WO 08/152523, WO 04/067753, WO 03/078619, WO 06/097784, WO 07/034262, WO 07/049156, WO 07/057781, WO 08/093152, WO 08/102199, WO 08/102274, WO 08/149176, WO 09/013559, WO 09/013622, and WO 09/019528.

[0035] A meganuclease may bind to double-stranded DNA as a homodimer, as is the case for wild-type I-Crel, or it may bind to DNA as a heterodimer. A meganuclease may also be a "single-chain heterodimer" in which a pair of DNA-binding domains derived from I-Crel are joined into a single polypeptide using a peptide linker. The term "homing endonuclease" is synonymous with the term "meganuclease."

[0036] As used herein, the term "rationally-designed" means non-naturally occurring and/or genetically engineered. The rationally-designed meganucleases of the invention differ from wild-type or naturally-occurring meganucleases in their amino acid sequence or primary structure, and may also differ in their secondary, tertiary or quaternary structure. In addition, the rationally-designed meganucleases of the invention also differ from wild-type or naturally-occurring meganucleases in recognition sequence-specificity and/or activity.

[0037] As used herein, with respect to a protein, the term "recombinant" means having an altered amino acid sequence as a result of the application of genetic engineering techniques to nucleic acids which encode the protein, and cells or organisms

which express the protein. With respect to a nucleic acid, the term "recombinant" means having an altered nucleic acid sequence as a result of the application of genetic engineering techniques. Genetic engineering techniques include, but are not limited to, PCR and DNA cloning technologies; transfection, transformation and other gene transfer technologies; homologous recombination; site-directed mutagenesis; and gene fusion. In accordance with this definition, a protein having an amino acid sequence identical to a naturally-occurring protein, but produced by cloning and expression in a heterologous host, is not considered recombinant.

[0038] As used herein, the term "genetically-modified" refers to a cell or organism in which, or in an ancestor of which, a genomic DNA sequence has been deliberately modified by recombinant technology. As used herein, the term "genetically-modified" encompasses the term "transgenic."

[0039] As used herein, the term "wild-type" refers to any naturally-occurring form of a meganuclease. The term "wild-type" is not intended to mean the most common allelic variant of the enzyme in nature but, rather, any allelic variant found in nature. Wild-type meganucleases are distinguished from recombinant or non-naturally-occurring meganucleases.

[0040] As used herein, the term "recognition sequence half-site" or simply "half site" means a 9 base pair DNA sequence which is recognized by a meganuclease monomer, in the case of a dimeric meganuclease, or by one domain of a single-chain meganuclease.

[0041] As used herein, the term "recognition sequence" refers to a pair of half-sites which is bound and cleaved by a meganuclease. A recognition sequence comprises a pair of inverted, 9 base pair half sites separated by four base-pairs. The recognition sequence is, therefore, 22 base-pairs in length. The base pairs of each half-site are designated -9 through -1, with the -9 position being most distal from the cleavage site and the -1 position being adjacent to the 4 base pair center sequence, the base pairs of which are designated +1 through +4. The strand of each half-site which is oriented 5' to 3' in the direction from -9 to -1 (*i.e.*, towards the cleavage site), is designated the "sense" strand, and the opposite strand is designated the "antisense strand", although neither strand may encode protein. Thus, the "sense" strand of one half-site is the antisense (opposite) strand of the other half-site. See, for example, Figure 1a.

[0042] As used herein, the term "center sequence" refers to the four base pairs separating half sites in the meganuclease recognition sequence. These bases are numbered +1 through +4 in Figure 1 a. The center sequence comprises the four bases that become the 3' single-strand overhangs following meganuclease cleavage. "Center sequence" can refer to the sequence of the sense strand or the antisense (opposite) strand.

[0043] As used herein, the term "specificity" refers to the ability of a meganuclease to recognize and cleave double-stranded DNA molecules only at a particular subset of all possible recognition sequences. The set of recognition sequences will share certain conserved positions or sequence motifs, but may be degenerate at one or more positions. A more specific meganuclease is capable of binding and cleaving a smaller subset of the possible recognition sequences, whereas a less specific meganuclease is capable of binding and cleaving a larger subset of the possible recognition sequences.

[0044] As used herein, the term "palindromic" refers to a recognition sequence consisting of inverted repeats of identical half-sites. In this case, however, the palindromic sequence need not be palindromic with respect to the center sequence, which is not contacted by the enzyme. In the case of dimeric meganucleases, palindromic DNA sequences are recognized by homodimers in which the two monomers make contacts with identical half-sites.

[0045] As used herein, the term "pseudo-palindromic" refers to a recognition sequence consisting of inverted repeats of non-identical or imperfectly palindromic half-sites. In this case, the pseudo-palindromic sequence need not be palindromic with respect to the center sequence, and also can deviate from a perfectly palindromic sequence between the two half-sites. Pseudo-palindromic DNA sequences are typical of the natural DNA sites recognized by wild-type homodimeric meganucleases in which two identical enzyme monomers make contacts with different half-sites.

[0046] As used herein, the term "non-palindromic" refers to a recognition sequence composed of two unrelated half-sites of a meganuclease. In this case, the non-palindromic sequence need not be palindromic with respect to either the center sequence or the two monomer half-sites. Non-palindromic DNA sequences are recognized by either heterodimeric meganucleases or single-chain meganucleases comprising a pair of domains that recognize non-identical half-sites.

[0047] As used herein, the term "activity" refers to the rate at which a meganuclease of the invention cleaves a particular recognition sequence. Such activity is a measurable enzymatic reaction, involving the hydrolysis of phosphodiester bonds of

double-stranded DNA. The activity of a meganuclease acting on a particular DNA substrate is affected by the affinity or avidity of the meganuclease for that particular DNA substrate which is, in turn, affected by both sequence-specific and non-sequence-specific interactions with the DNA.

[0048] As used herein, the term "homologous recombination" refers to the natural, cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template (see, *e.g.*, Cahill et al. (2006), *Front. Biosci.* 11:1958-1976). The homologous DNA sequence may be an endogenous chromosomal sequence or an exogenous nucleic acid that was delivered to the cell. Thus, for some applications, a meganuclease is used to cleave a recognition sequence within a target sequence in a genome and an exogenous nucleic acid with homology to or substantial sequence similarity with the target sequence is delivered into the cell and used as a template for repair by homologous recombination. The DNA sequence of the exogenous nucleic acid, which may differ significantly from the target sequence, is thereby incorporated into the chromosomal sequence. The process of homologous recombination occurs primarily in eukaryotic organisms. The term "homology" is used herein as equivalent to "sequence similarity" and is not intended to require identity by descent or phylogenetic relatedness.

[0049] As used herein, the term "non-homologous end-joining" refers to the natural, cellular process in which a double-stranded DNA-break is repaired by the direct joining of two non-homologous DNA segments (see, *e.g.*, Cahill et al. (2006), *Front. Biosci.* 11:1958-1976). DNA repair by non-homologous end-joining is error-prone and frequently results in the untemplated addition or deletion of DNA sequences at the site of repair. Thus, for some applications, a meganuclease can be used to produce a double-stranded break at a meganuclease recognition sequence within a target sequence in a genome to disrupt a gene (*e.g.*, by introducing base insertions, base deletions, or frameshift mutations) by non-homologous end-joining. For other applications, an exogenous nucleic acid lacking homology to or substantial sequence similarity with the target sequence may be captured at the site of a meganuclease-stimulated double-stranded DNA break by non-homologous end-joining (see, *e.g.*, Salomon et al. (1998), *EMBO J.* 17:6086-6095). The process of non-homologous end-joining occurs in both eukaryotes and prokaryotes such as bacteria.

[0050] As used herein, the term "sequence of interest" means any nucleic acid sequence, whether it codes for a protein, RNA, or regulatory element (*e.g.*, an enhancer, silencer, or promoter sequence), that can be inserted into a genome or used to replace a genomic DNA sequence using a meganuclease protein. Sequences of interest can have heterologous DNA sequences that allow for tagging a protein or RNA that is expressed from the sequence of interest. For instance, a protein can be tagged with tags including, but not limited to, an epitope (*e.g.*, c-myc, FLAG) or other ligand (*e.g.*, poly-His). Furthermore, a sequence of interest can encode a fusion protein, according to techniques known in the art (see, *e.g.*, Ausubel et al., *Current Protocols in Molecular Biology*, Wiley 1999). For some applications, the sequence of interest is flanked by a DNA sequence that is recognized by the meganuclease for cleavage. Thus, the flanking sequences are cleaved allowing for proper insertion of the sequence of interest into genomic recognition sequences cleaved by a meganuclease. For some applications, the entire sequence of interest is homologous to or has substantial sequence similarity with a target sequence in the genome such that homologous recombination effectively replaces the target sequence with the sequence of interest. For other applications, the sequence of interest is flanked by DNA sequences with homology to or substantial sequence similarity with the target sequence such that homologous recombination inserts the sequence of interest within the genome at the locus of the target sequence. For some applications, the sequence of interest is substantially identical to the target sequence except for mutations or other modifications in the meganuclease recognition sequence such that the meganuclease can not cleave the target sequence after it has been modified by the sequence of interest.

[0051] As used herein, the term "single-chain meganuclease" refers to a polypeptide comprising a pair of meganuclease subunits joined by a linker. A single-chain meganuclease has the organization: N-terminal subunit - Linker - C-terminal subunit. The two meganuclease subunits, each of which is derived from I-CreI, will generally be non-identical in amino acid sequence and will recognize non-identical half-sites. Thus, single-chain meganucleases typically cleave pseudo-palindromic or non-palindromic recognition sequences. A single chain meganuclease may be referred to as a "single-chain heterodimer" or "single-chain heterodimeric meganuclease" although it is not, in fact, dimeric.

[0052] As used herein, unless specifically indicated otherwise, the word "or" is used in the inclusive sense of "and/or" and not the exclusive sense of "either/or."

2.1 Preferred Center Sequences for I-CreI-Derived Meganucleases

[0053] The present invention is based, in part, in the identification of subsets of the possible four base pair center sequences that are preferred by I-CreI-derived meganucleases. As the wild-type enzyme does not make significant contacts with the bases in

the center sequence, the same center sequence preferences of the wild-type I-CreI homing nuclease apply to rationally-designed I-CreI-derived meganucleases which have been redesigned with respect to, for example, half-site preference, DNA-binding affinity, and/or heterodimerization ability. This invention provides, therefore, important criteria that can be considered in determining whether or not a particular 22 base pair DNA sequence is a suitable I-CreI-derived meganuclease recognition sequence.

[0054] The preferred set of center sequences was determined using a genetically-engineered meganuclease called "DJ1" (SEQ ID NO: 4). The production of this meganuclease is described in WO 2007/047859. DJ1 is a homodimeric I-CreI-derived meganuclease which was designed to recognize a palindromic meganuclease recognition sequence (SEQ ID NO: 5, SEQ ID NO: 6) that differs at 4 positions per half-site relative to wild-type I-CreI. This change in half-site specificity was achieved by the introduction of 6 amino acid substitutions to wild type I-CreI (K28D, N30R, S32N, Q38E, S40R, and T42R).

[0055] To test for cleavage activity with respect to various recognition sequences, DJ1 was expressed in *E. coli* and purified as described in Example 1 of WO 2007/047859. Then, 25 picomoles of purified meganuclease protein were added to a 10 nM solution of plasmid DNA substrate in SA buffer (25 mM Tris-HCL, pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 5mM EDTA) in a 25 microliter reaction. 1 microliter of XmnI restriction enzyme was added to linearize the plasmid substrates. Reactions were incubated at 37°C for 4 hours and were then visualized by gel electrophoresis to determine the extent to which each was cleaved by the DJ1 meganuclease.

[0056] The plasmid substrates used in these experiments comprised a pUC-19 plasmid in which a potential meganuclease recognition sequence was inserted into the polylinker site (SmaI site). Each potential meganuclease recognition site comprised a pair of inverted DJ1 half-sites (SEQ ID NO: 7) separated by a different center sequence. Thus, by evaluating DJ1 cleavage of multiple DNA substrates differing only by center sequence, it was possible to determine which center sequences are the most amenable to meganuclease cleavage (Figure 2).

[0057] Initially, only the influence of the N₊₂ and N₊₃ bases were evaluated. The X-ray crystal structure of I-CreI in complex with its natural DNA site shows that the DNA is distorted at these central two base pairs (Jurica et al. (1998), Mol. Cell. 2:469-76). Computer modeling suggests that a purine (G or A) at N₊₂ is incompatible with a pyrimidine (C or T) at N₊₃. This is because the distortion introduced by I-CreI binding causes a steric clash between a purine base at N₊₂ and a second purine base-paired to a pyrimidine at N₊₃. This expected incompatibility was verified experimentally by incubating DJ1 protein with plasmid substrates harboring meganuclease recognition sites with all possible center sequences of the form A₊₁X₊₂X₊₃T₊₄ in which X is any base. The results are summarized in Table 1. For Tables 1-5, "Activity" refers to the following:

- : no cleavage in 4 hours

+ : 1% - 25% cleavage in 4 hours

++ : 26%-75% cleavage in 4 hours

+++ : 75%-100% cleavage in 4 hours

Table 1. The effect of changes at N₊₂ and N₊₃

Seq. No.	N ₊₁	N ₊₂	N ₊₃	N ₊₄	Activity
1	A	A	A	T	+
2	A	A	C	T	-
3	A	A	G	T	+
4	A	A	T	T	-
5	A	C	A	T	+
6	A	C	C	T	+
7	A	C	G	T	+
8	A	C	T	T	+
9	A	G	A	T	+
10	A	G	C	T	-
11	A	G	G	T	+

Seq. No.	N ₊₁	N ₊₂	N ₊₃	N ₊₄	Activity
12	A	G	T	T	-
13	A	T	A	T	++
14	A	T	C	T	+
15	A	T	G	T	++
16	A	T	T	T	+

[0058] Consistent with the computer modeling, it was found that the four plasmid substrates with a purine base at N₊₂ and a pyrimidine base at N₊₃ (sequence numbers 2, 4, 10, and 12) were not cut efficiently by DJ1.

[0059] Next, a more comprehensive evaluation of center sequence preference was performed. There are 4⁴ or 256 possible center sequences. Of these, 25%, or 64, have a purine base at N₊₂ and pyrimidine at N₊₃ and, therefore, were eliminated as center sequences based on the experiment described above. Of the remaining 192, 92 are redundant because meganucleases are symmetric and recognize bases equally on both the sense and antisense strand. For example, the sequence A₊₁A₊₂A₊₃A₊₄ on the sense strand is recognized by the meganuclease as T₊₁T₊₂T₊₃T₊₄ on the antisense strand and, thus, A₊₁A₊₂A₊₃A₊₄ and T₊₁T₊₂T₊₃T₊₄ are functionally equivalent. Taking these redundancies into account, as well as the aforementioned N₊₂ / N₊₃ conflicts, there were 100 possible center sequences remaining. To determine which of these were preferred by meganucleases, we produced 100 plasmid substrates harboring these 100 center sequences flanked by inverted recognition half-sites for the DJ1 meganuclease. DJ1 was then incubated with each of the 100 plasmids and cleavage activity was evaluated as described above. These results are summarized in Table 2.

Table 2. Cleavable Center Sequences

Seq. No.	N ₊₁	N ₊₂	N ₊₃	N ₊₄	Activity
1	T	T	T	T	+
2	T	T	G	T	++
3	T	T	C	T	+
4	T	T	A	T	++
5	T	G	G	T	+
6	T	G	A	T	+
7	T	C	T	T	++
8	T	C	G	T	++
9	T	C	C	T	+
10	T	C	A	T	++
11	T	A	G	T	+
12	T	A	A	T	+
13	G	T	T	T	++
14	G	T	G	T	+++
15	G	T	C	T	++
16	G	T	A	T	+++
17	G	G	G	T	+
18	G	G	A	T	++
19	G	A	G	T	++
20	G	A	A	T	++
21	C	T	T	T	+
22	C	T	G	T	+
23	C	T	C	T	+
24	C	T	A	T	+

Seq. No.	N ₊₁	N ₊₂	N ₊₃	N ₊₄	Activity
25	C	G	G	T	+
26	C	G	A	T	+
27	C	C	T	T	+
28	C	C	G	T	+
29	C	C	C	T	+
30	C	C	A	T	+
31	C	A	G	T	+
32	C	A	A	T	+
33	A	T	T	T	+
34	A	T	G	T	++
35	A	T	C	T	+
36	A	G	G	T	+
37	A	C	T	T	+
38	T	T	T	G	+
39	T	T	G	G	+
40	T	T	C	G	+
41	T	T	A	G	+++
42	T	G	G	G	+
43	T	G	A	G	+
44	T	C	T	G	+
45	T	C	G	G	+
46	T	C	C	G	+
47	T	C	A	G	+
48	T	A	G	G	+
49	T	A	A	G	+
50	G	T	T	G	+
51	G	T	G	G	+
52	G	T	C	G	+
53	G	T	A	G	+++
54	G	G	G	G	+
55	G	G	A	G	+
56	G	A	G	G	+
57	G	A	A	G	+
58	C	T	T	G	+
59	C	T	G	G	+
60	C	T	C	G	+
61	C	G	G	G	+
62	C	C	T	G	+
63	T	T	T	C	++
64	T	T	C	C	++
65	T	T	A	C	+++
66	T	G	A	C	++
67	T	C	T	C	+++

Seq. No.	N ₊₁	N ₊₂	N ₊₃	N ₊₄	Activity
68	T	C	C	C	+
69	T	C	A	C	+++
70	T	A	A	C	++
71	G	T	T	C	++
72	G	T	C	C	+++
73	T	T	T	A	+
74	T	T	G	A	+
75	T	T	C	A	+
76	T	G	G	A	+
77	T	C	T	A	+
78	A	T	A	T	++
79	A	C	G	T	+
80	C	T	A	G	+
81	C	C	G	G	+
82	G	T	A	C	+++
83	T	C	G	A	++
84	T	T	A	A	++
85	T	C	G	C	+++
86	A	A	G	C	+++
87	G	A	G	C	+++
88	G	C	G	C	+++
89	G	G	G	C	++
90	G	T	G	C	+++
91	T	A	G	C	+++
92	T	G	G	C	+
93	T	T	G	C	+++
94	A	C	G	C	++
95	A	G	G	C	+
96	A	T	G	C	+++
97	C	A	G	C	+
98	C	C	G	C	++
99	C	G	G	C	+
100	C	T	G	C	++

[0060] For clarity, each of the center sequences listed in Table 2 is equivalent to its opposite strand sequence due to the fact that the I-CreI meganuclease binds its recognition sequence as a symmetric homodimer. Thus, sequence no. 100 in Table 2, C₊₁T₊₂G₊₃C₊₄, is equivalent to its opposite strand sequence, G₊₁C₊₂A₊₃G₊₄. From these data, a general set of center sequence preference rules emerge. These rules, which are not meant to supersede Table 1 or Table 2, include:

1. Center sequences with a purine base at N₊₂ and a pyrimidine base at N₊₃ cut very poorly, if at all.
2. G is preferred at N₊₁. This is equivalent to C at N₊₄. All of the most preferred center sequences have G at N₊₁ and/or C at N₊₄.
3. C is preferred at N₊₂. This is equivalent to G at N₊₃.
4. There is a preference for center sequences with a pyrimidine base at N₊₂ and a purine base at N₊₃.
5. There is a preference for sequences with at least 1 A-T base pair in the center sequence.

[0061] Thus, in general, preferred center sequences have the form $G_{+1}Y_{+2}R_{+3}X_{+4}$ where Y is a pyrimidine (C or T), R is a purine (A or G), and X is any base (A, C, G, or T).

2.2 *In Vitro* Applications Using Preferred Center Sequences.

[0062] Genetically-engineered meganucleases have numerous potential *in vitro* applications including restriction mapping and cloning. These applications are known in the art and are discussed in WO 2007/047859.

[0063] One advantage of using genetically-engineered meganucleases rather than conventional restriction enzymes for applications such as cloning is the possibility of cutting DNA to leave a wide range of different 3' overhangs ("sticky ends") that are compatible with, for example, the 3' overhangs produced by cleaving a particular vector of interest. Thus, there are occasions when it is desirable to cleave a meganuclease recognition sequence with a sub-optimal center sequence in order to create a desired overhang.

[0064] Because *in vitro* DNA cleavage conditions are, in general, less stringent than conditions *in vivo*, the use of sub-optimal center sequences may be acceptable for such applications. For example, relative to *in vivo* applications, *in vitro* digests using engineered meganucleases can be performed at a higher ratio of meganuclease to DNA, there is typically less non-specific (genomic) DNA competing for meganuclease, and solution conditions can be optimized to favor meganuclease cleavage (e.g., using SA buffer as described above). Thus, a larger number of center sequences are suitable for *in vitro* applications than for *in vivo* applications. All of the center sequences listed in Table 2 are suitable for *in vitro* applications, but preferred and most preferred center sequences for *in vitro* applications are listed in Table 3 and Table 4, respectively, with their opposite strand sequences.

Table 3. Preferred Center Sequences for *in vitro* Applications

Seq. No.	$N_{+1}N_{+2}N_{+3}N_{+4}$	Opposite Strand Sequence
1	TTGT	ACAA
2	TTAT	ATAA
3	TCTT	AAGA
4	TCGT	ACGA
5	TCAT	ATGA
6	GTTT	AAAC
7	GTCT	AGAC
8	GGAT	ATCC
9	GAGT	ACTC
10	GAAT	ATTC
11	ATGT	ACAT
12	TTTC	GAAA
13	TTCC	GGAA
14	TGAC	GTCA
15	TAAC	GTTA
16	GTTC	GAAC
17	ATAT	ATAT
18	TCGA	TCGA
19	TTAA	TTAA
20	GGGC	GCCC
21	ACGC	GCGT
22	CCGC	GCGG
23	CTGC	GCAG

Table 4. Most Preferred Center Sequences for *in vitro* Applications

Seq. No.	N ₁ N ₂ N ₃ N ₄	Opposite Strand Sequence
1	GTGT	ACAC
2	GTAT	ATAC
3	TTAG	CTAA
4	GTAG	CTAC
5	TTAC	GTAA
6	TCTC	GAGA
7	TCAC	GTGA
8	GTCC	GGAC
9	GTAC	GTAC
10	TCGC	GCGA
11	AAGC	GCTT
12	GAGC	GCTC
13	GCGC	GCGC
14	GTGC	GCAC
15	TAGC	GCTA
16	TTGC	GCAA
17	ATGC	GCAT

[0065] Obviously, not every 22 base pair DNA sequence having a preferred or most preferred center sequence is capable of being a meganuclease recognition sequence *in vitro*. The sequence of the half-sites flanking the center sequence must also be amenable to meganuclease recognition and cleavage. Methods for engineering a meganuclease including I-CreI, to recognize a pre-determined half-site are known in the art (see, e.g., WO 2007/047859). Thus, a preferred I-CreI-derived meganuclease recognition sequence for *in vitro* applications will comprise: (1) a first 9 base pair half-site amenable to recognition by a meganuclease monomer (or a first domain of a single-chain meganuclease); (2) a preferred or most preferred center sequence from Table 2 or Table 3; and (3) a second 9 base pair half-site amenable to recognition by a meganuclease monomer (or a second domain of a single-chain meganuclease) in the reverse orientation relative to the first half-site.

[0066] Thus, in one aspect, the invention provides methods for cleaving a double-stranded DNA *in vitro* by (a) identifying at least one potential recognition site for at least one I-CreI-derived meganuclease within the DNA, wherein the potential recognition site has a four base pair central sequence selected from the group of central sequences of Table 2; (b) identifying an I-CreI-derived meganuclease which recognizes that recognition site in the DNA; and (c) contacting the I-CreI-derived meganuclease with the DNA; whereby the I-CreI meganuclease cleaves the DNA.

[0067] In another aspect, the invention provides methods for cleaving a double-stranded DNA *in vitro* by (a) introducing into the DNA a recognition site for an I-CreI-derived meganuclease having a four base pair central sequence selected from the group consisting of central sequences of Table 2; and (b) contacting the I-CreI-derived meganuclease with the DNA; whereby the I-CreI-derived meganuclease cleaves the DNA.

[0068] In particular, in some embodiments, the DNA is selected from a PCR product; an artificial chromosome; genomic DNA isolated from bacteria, fungi, plants, or animal cells; and viral DNA.

[0069] In some embodiments, the DNA is present in a nucleic acid selected from a plasmid, a prophage and a chromosome.

[0070] In some of the foregoing embodiments, the four base pair DNA sequence is selected from Table 3. In other embodiments, the four base pair DNA sequence is selected from Table 4.

[0071] In some embodiments, the I-CreI-derived meganuclease can be specifically designed for use with the chosen recognition site in the method.

2.3 *In vivo* Applications Using Preferred Center Sequences.

[0072] Applications such as gene therapy, cell engineering, and plant engineering require meganuclease function inside of a living cell (for clarity, any intracellular application will be referred to as an "*in vivo*" application whether or not such cell is isolated or part of a multicellular organism). These applications are known in the art and are described in, e.g., WO 2007/047859. *In vivo* applications are significantly restricted relative to *in vitro* applications with regard to the center sequence. This is because intracellular conditions cannot be manipulated to any great extent to favor meganuclease activity and/or because vast amounts of genomic DNA compete for meganuclease binding. Thus, only meganuclease recognition sequences with optimal center sequences are preferred for *in vivo* applications. Such sequences are listed in Table 5 with their opposite strand sequences.

Table 5. Preferred Center Sequences for *in vivo* applications.

Seq. No.	N ₁ N ₂ N ₃ N ₄	Opposite Strand Sequence
	GTGT	ACAC
2	GTAT	ATAC
3	TTAG	CTAA
4	GTAG	CTAC
5	TTAC	GTAA
6	TCTC	GAGA
7	TCAC	GTGA
8	GTCC	GGAC
9	GTAC	GTAC
10	TCGC	GCGA
11	AAGC	GCTT
12	GAGC	GCTC
13	GCGC	GCGC
14	GTGC	GCAC
15	TAGC	GCTA
16	TTGC	GCAA
17	ATGC	GCAT

[0073] Obviously, not every 22 base pair DNA sequence having a preferred center sequence is capable of being a meganuclease recognition sequence *in vivo*. The sequence of the half-sites flanking the center sequence must also be amenable to meganuclease recognition and cleavage. Methods for engineering a meganuclease, including I-Crel, to recognize a pre-determined half-site are known in the art (see, e.g., WO 2007/047859). Thus, a preferred *in vivo* meganuclease recognition sequence will comprise: (1) a first 9 base pair half-site amenable to recognition by a meganuclease monomer (or a first domain of a single-chain meganuclease); (2) a preferred center sequence from Table 5; and (3) a second 9 base pair half-site amenable to recognition by a meganuclease monomer (or a second domain of a single-chain meganuclease) in the reverse orientation relative to the first half-site.

[0074] Thus, in one aspect, the invention provides methods for cleaving a double-stranded DNA *in vivo* by (a) identifying at least one potential recognition site for at least one I-Crel-derived meganuclease within the DNA, wherein the potential recognition site has a four base pair central sequence selected from the group of central sequences of Table 2; (b) identifying an I-Crel-derived meganuclease which recognizes that recognition site in the DNA; and (c) contacting the I-Crel-derived meganuclease with the DNA; whereby the I-Crel-derived meganuclease cleaves the DNA.

[0075] In another aspect, the invention provides methods for cleaving a double-stranded DNA *in vivo* by (a) introducing into the DNA a recognition site for an I-Crel-derived meganuclease having a four base pair central sequence selected from the group consisting of central sequences of Table 2; and (b) contacting the I-Crel-derived meganuclease with the DNA; whereby the I-Crel-derived meganuclease cleaves the DNA.

[0076] In some embodiments, the DNA is present in a cell selected from a bacterial, fungal, plant and animal cell.

[0077] In some embodiments, the DNA is present in a nucleic acid selected from a plasmid, a prophage and a chromosome.

[0078] In some of the foregoing embodiments, the four base pair DNA sequence is selected from Table 3. In other embodiments, the four base pair DNA sequence is selected from Table 4.

[0079] In some embodiments, the I-CreI-derived meganuclease is specifically designed for use with the chosen recognition site in the methods of the invention.

[0080] In some of the foregoing embodiments, the method includes the additional step of rationally-designing the I-CreI-derived meganuclease to recognize the chosen recognition site. In some embodiments, the method further comprises producing the I-CreI-derived meganuclease.

[0081] In another aspect, the invention provides cells transformed with a nucleic acid including (a) a first 9 base pair DNA sequence which can be bound by an I-CreI-derived meganuclease monomer or by a first domain from a single-chain I-CreI-derived meganuclease; (b) a four base pair DNA sequence selected from Table 2; and (c) a second 9 base pair DNA sequence which can be bound by an I-CreI-derived meganuclease monomer or by a second domain from a single-chain I-CreI-derived meganuclease; wherein the second 9 base pair DNA sequence is in the reverse orientation relative to the first.

[0082] In another aspect, the invention provides a cell containing an exogenous nucleic acid sequence integrated into its genome, including, in order: (a) a first exogenous 9 base pair DNA sequence which can be bound by an I-CreI-derived meganuclease monomer or by a first domain from a single-chain I-CreI-derived meganuclease; (b) an exogenous four base pair DNA sequence selected from Table 2; and (c) a second exogenous 9 base pair DNA sequence which can be bound by an I-CreI-derived meganuclease monomer or by a second domain from a single-chain I-CreI-derived meganuclease; wherein the second 9 base pair DNA sequence is in the reverse orientation relative to the first.

[0083] In another aspect, the invention provides a cell containing an exogenous nucleic acid sequence integrated into its genome, including, in order: (a) a first exogenous 9 base pair DNA sequence which can be bound by an I-CreI-derived meganuclease monomer or by a first domain from a single-chain I-CreI-derived meganuclease; (b) an exogenous two base pair DNA sequence, wherein the two base pairs correspond to bases N₊₁ and N₊₂ of a four base pair DNA sequence selected from Table 2; (c) an exogenous DNA sequence comprising a coding sequence which is expressed in the cell; (d) an exogenous two base pair DNA sequence, wherein the two base pairs correspond to bases N₊₃ and N₊₄ of a four base pair DNA sequence selected from Table 2; and (e) a second exogenous 9 base pair DNA sequence which can be bound by the I-CreI-derived meganuclease monomer or by a second domain from the single-chain I-CreI-derived meganuclease; wherein the second 9 base pair DNA sequence is in the reverse orientation relative to the first.

[0084] In some embodiments, the nucleic acid is a plasmid. In other embodiments, the nucleic acid is an artificial chromosome. In other embodiments, the nucleic acid is integrated into the genomic DNA of the cell. In other embodiments, the nucleic acid is a viral nucleic acid.

[0085] In some embodiments, the cell is selected from the group a human cell, a non-human animal cell, a plant cell, a bacterial cell, and a fungal cell.

[0086] In some of the foregoing embodiments, the four base pair DNA sequence is selected from Table 3. In other embodiments, the four base pair DNA sequence is selected from Table 4.

[0087] In some embodiments, the I-CreI meganuclease is specifically designed for use with the chosen recognition site in the methods of the invention.

2.4 Methods of conducting a custom-designed, I-CreI-derived meganuclease business

[0088] A meganuclease business can be conducted based on I-CreI-derived meganucleases. For example, such business can operate as following. The business received a DNA sequence into which a double-strand break is to be introduced by a rationally-designed I CreI-derived meganuclease. The business identifies in the DNA sequence at least one recognition site for a rationally-designed I CreI-derived meganuclease with altered specificity relative to I-CreI, wherein the recognition site is not cleaved by a

naturally-occurring I-CreI, wherein the recognition site has a four base pair central sequence selected from the group consisting of TTGT, TTAT, TCTT, TCGT, TCAT, GTTT, GTCT, GGAT, GAGT, GAAT, ATGT, TTTC, TTCC, TGAC, TAAC, GTTC, ATAT, TCGA, TTAA, GGGC, ACGC, CCGC, CTGC, ACAA, ATAA, AAGA, ACGA, ATGA, AAAC, AGAC, ATCC, ACTC, ATTC, ACAT, GAAA, GGAA, GTCA, GTTA, GAAC, ATAT, TCGA, TTAA, GCCC, GCGT, GCGG and GCAG. The business then provides a rationally-designed meganuclease that cleaves the recognition site in the DNA.

[0089] Optionally, the business rationally-designs an I-CreI-derived meganuclease that cleaves the recognition site in the DNA. Optionally, the business produces the rationally-designed I-CreI-derived meganuclease.

2.2 Specifically Excluded Center Sequences.

[0090] The center sequences GTAC, ACAC, and GTGA have previously been shown to be effective center sequences for *in vitro* and *in vivo* applications. These center sequences are specifically excluded from some aspects of the present invention. In addition, the center sequences TTGA and GAAA have previously been shown to be poor center sequences for *in vivo* applications (Arnould, et al. (2007). J. Mol. Biol. 371: 49-65).

EXAMPLES

[0091] This invention is further illustrated by the following examples, which should not be construed as limiting. Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific substances and procedures described herein. Such equivalents are intended to be encompassed in the scope of the claims that follow the examples below. Examples 1 and 2 refer to engineered meganucleases cleaving optimized meganuclease recognition sites *in vivo* in a model plant system. Example 3 refers to an engineered meganuclease cleaving optimized meganuclease recognition sites *in vitro*. Examples not falling within the scope of the claims are for illustrative purposes only.

EXAMPLE 1

Cleavage of an optimized meganuclease recognition site by a rationally-designed, I-CreI-derived meganuclease homodimer *in vivo*

[0092] An engineered meganuclease called BRP2 (SEQ ID NO: 8) was produced using the method disclosed in WO 2007/047859. This meganuclease is derived from I-CreI and was engineered to recognize DNA sites that are not recognized by wild-type I-CreI (e.g., BRP2 recognition sequences include SEQ ID NO: 9 and SEQ ID NO: 10, or SEQ ID NO: 11 and SEQ ID NO: 12). To facilitate nuclear localization of the engineered meganuclease, an SV40 nuclear localization signal (NLS, SEQ ID NO: 13) was added to the N-terminus of the protein. Conventional *Agrobacterium*-mediated transformation procedures were used to transform *Arabidopsis thaliana* with a T-DNA containing a codon-optimized BRP2 coding sequence (SEQ ID NO: 14). Expression of BRP2 meganuclease was under the control of a Hsp70 promoter and a NOS terminator. A pair of BRP2 recognition sequences were housed on the same T-DNA separated by 7 base pairs containing a PstI restriction enzyme site (Figure 3a). BRP2 cutting of the pair of BRP2 recognition sequences in this construct was expected to excise the region between the recognition sequences and thereby remove the PstI restriction site (Figure 3b). Two such T-DNA constructs were produced which varied the center sequence of the meganuclease recognition sequences flanking the PstI restriction enzyme site (Figure 3c). In the first construct (the "GTAC construct"), the meganuclease recognition sites had the center sequence GTAC (a preferred *in vivo* center sequence, Table 5, sequence 9; SEQ ID NO: 9 and SEQ ID NO 10). The second construct (the "TAGA construct") had the center sequence TAGA (a non-preferred center sequence, opposite strand sequence to Table 2, sequence 77; SEQ ID NO: 11 and SEQ ID NO 12).

[0093] Stably transformed *Arabidopsis* plants carrying each construct were produced by selection for a kanamycin resistance marker housed on the T-DNA. Genomic DNA was then isolated from the transformed plants (by leaf punch) before and after heat-shock to induce BRP2 meganuclease expression. Genomic DNA samples were added to PCR reactions using primers to amplify the region of the T-DNA housing the meganuclease recognition sequences. PCR products were then digested with PstI and visualized by gel electrophoresis (Fig. 3d). Results are summarized in Table 6. Any PCR sample in which a significant percentage (>25%) of product was found to be resistant to PstI was considered to be indicative of *in vivo* meganuclease cleavage in that

particular plant and was scored as "cut" in Table 6. It was found that, prior to heat-shock, the vast majority of PCR samples from plants carrying either construct retained the PstI site. After heat-shock, however, a large percentage of samples taken from plants transformed with the GTAC construct, but not the TAGA construct, had lost the PstI site. PCR products from the GTAC construct-transformed plants lacking a PstI site were cloned into a pUC-19 plasmid and sequenced. 100% of sequenced clones had a precise deletion of the region between the two BRP2 cut sites (as diagrammed in Fig. 3b). These results indicate that an engineered meganuclease is able to cleave a meganuclease recognition site *in vivo* provided it has an optimized center sequence.

Table 6. *In vivo* cleavage of optimized meganuclease recognition sequences by an engineered meganuclease homodimer.

Construct	Before heat-shock		After heat-shock	
	Cut	Uncut	Cut	Uncut
GTAC	0	4	3	1
TAGA	0	22	0	22

EXAMPLE 2

Cleavage of an optimized meganuclease recognition site by a rationally-designed, I-CreI-derived meganuclease single-chain heterodimer *in vivo*

[0094] The engineered meganuclease BRP12-SC (SEQ ID NO: 15) was produced in accordance with WO 2007/047859, except that this meganuclease is a single-chain heterodimer. As discussed in WO 2007/047859, wild-type I-CreI binds to and cleaves DNA as a homodimer. As a consequence, the natural recognition sequence for I-CreI is pseudo-palindromic. The BRP12-SC recognition sequences, however, are non-palindromic (e.g., SEQ ID NO: 16 and SEQ ID NO: 17, or SEQ ID NO: 18 and SEQ ID NO: 19). This necessitates the use of an engineered meganuclease heterodimer comprising a pair of subunits each of which recognizes one half-site within the full-length recognition sequence. In the case of BRP12-SC, the two engineered meganuclease monomers are physically linked to one another using an amino acid linker to produce a single-chain heterodimer. This linker comprises amino acids 166-204 (SEQ ID NO: 20) of BRP12-SC. The linker sequence joins an N-terminal meganuclease subunit terminated at L165 (corresponding to L155 of wild-type I-CreI) with a C-terminal meganuclease subunit starting at K204 (corresponding to K7 of wild-type I-CreI). The benefits of physically linking the two meganuclease monomers using this novel linker is twofold: First, it ensures that the meganuclease monomers can only associate with one another (heterodimerize) to cut the non-palindromic BRP12-SC recognition sequence rather than also forming homodimers which can recognize palindromic or pseudopalindromic DNA sites that differ from the BRP12-SC recognition sequence. Second, the physical linking of meganuclease monomers obviates the need to express two monomers simultaneously in the same cell to obtain the desired heterodimer. This significantly simplifies vector construction in that it only requires a single gene expression cassette. As was the case with the BRP2 meganuclease discussed in Example 1, the BRP12-SC meganuclease has an SV40 nuclear localization signal (SEQ ID NO: 13) at its N-terminus.

[0095] Conventional *Agrobacterium*-mediated transformation procedures were used to transform *Arabidopsis thaliana* with a T-DNA containing a codon-optimized BRP12-SC coding sequence (SEQ ID NO: 21). Expression of BRP12-SC meganuclease was under the control of a Hsp70 promoter and a NOS terminator. A pair of BRP12-SC recognition sequences were housed on the same T-DNA separated by 7 base pairs containing a PstI restriction enzyme site (Figure 4a). BRP12-SC cutting of the pair of BRP12-SC recognition sequences in this construct was expected to excise the region between the recognition sequences and thereby remove the PstI restriction site (Figure 4b). Two such T-DNA constructs were produced which varied only in the center sequences of the meganuclease recognition sequences flanking the PstI restriction enzyme site (Figure 4c). In the first construct (the "GTAC construct"), the meganuclease recognition sites had the center sequence GTAC (a preferred *in vivo* center sequence, Table 5, sequence 9; SEQ ID NO: 16 and SEQ ID NO 17). The second construct (the "TAGA construct") had the center sequence TAGA (a non-preferred center sequence, opposite strand sequence to Table 2, sequence 77; SEQ ID NO: 18 and SEQ ID NO 19).

[0096] Stably transformed *Arabidopsis* plants carrying each construct were produced by selection for a kanamycin resistance marker housed on the T-DNA. Genomic DNA was then isolated from the transformed plants (by leaf punch) before and after heat-shock to induce BRP12-SC meganuclease expression. Genomic DNA samples were added to PCR reactions using primers to amplify the region of the T-DNA housing the meganuclease recognition sequences. PCR products were then digested with PstI

and visualized by gel electrophoresis. The results of this analysis are presented in Table 7. Any PCR sample in which a significant percentage (>25%) of product was found to be resistant to PstI was considered to be indicative of *in vivo* meganuclease cleavage and was scored as "cut" in Table 7. It was found that, prior to heat-shock, the vast majority of PCR samples from plants carrying either construct retained the PstI site. After heat-shock, however, a large percentage of samples taken from plants transformed with the GTAC construct, but not the TAGA construct, had lost the PstI site. PCR products from the GTAC construct-transformed plants lacking a PstI site were cloned into a pUC-19 plasmid and sequenced. 100% of sequenced clones had a precise deletion of the region between the two BRP 12-SC cut sites (as diagrammed in Fig. 4b). These results indicate that an engineered single chain meganuclease is able to cleave a meganuclease recognition site *in vivo* provided it has an optimized center sequence.

Table 7. *In vivo* cleavage of optimized meganuclease recognition sequences by an engineered meganuclease homodimer.

Construct	Before heat-shock		After heat-shock	
	Cut	Uncut	Cut	Uncut
GTAC	0	23	8	15
TAGA	0	59	1	58

EXAMPLE 3

Cleavage of an optimized meganuclease recognition site by a rationally-designed, I-CreI-derived meganuclease homodimer *in vitro*

[0097] The BRP2 meganuclease described in Example 1 (SEQ ID NO: 8) was expressed in *E. coli* and purified as in Example 1 of WO 2007/047859. The purified meganuclease was then added at varying concentrations to reactions containing plasmids harboring BRP2 recognition sequences with either a GTAC or TAGA center sequence (0.25 picomoles of plasmid substrate in 25 microliters of SA buffer: 25 mM Tris-HCL, pH 8.0, 100 mM NaCl, 5 mM MgCl₂, 5mM EDTA). Reactions were incubated at 37°C for 2 hours and were then visualized by gel electrophoresis and the percentage of each plasmid substrate cleaved by the meganuclease was plotted as a function of meganuclease concentration (Figure 5). It was found that the plasmid substrate with the TAGA center sequence was cleaved by the meganuclease *in vitro*, but that cleavage of this substrate required a far higher concentration of BRP2 meganuclease than did cleavage of the GTAC substrate.

SEQUENCE LISTING

[0098]

SEQ ID NO: 1 (wild-type I-CreI, Genbank Accession # PO5725)

```

1 MNTKYNKEFL LYLAFVVDGD GSIIAQIKPN QSYKFKHQLS LAFQVTQKTQ RRWFLDKLVD
61 EIGVGIVRDR GSVSDYILSE IKPLHNFLTQ LQPFLLKQK QANLVLLKIIW RLPSAKESPD
121 KFLEVCTWVD QIAALNDSKT RKTTSQTVRA VLDSLSEKKK SSP

```

SEQ ID NO: 2 (wild-type I-CreI recognition sequence)

1 GAAACTGTCT CACGACGTTT TG

SEQ ID NO: 3 (wild-type I-CreI recognition sequence)

1 GAAAACGTCTG TGAGACAGTT TC

SEQ ID NO: 4 (DJ1 amino acid sequence)

```

1 MNTKYNKEFL LYLAFVVDGD GSIIAQIDPR QNYKFKHELRL RLFQVTQKTQ RRWFLDKLVD
61 EIGVGIVRDR GSVSDYILSE IKPLHNFLTQ LQPFLLKQK QANLVLLKIIW QLPSAKESPD
121 KFLEVCTWVD QIAALNDSKT RKTTSQTVRA VLDSLSEKKK SSP

```

SEQ ID NO: 5 (DJ1 recognition sequence - GTGA center sequence)

1 AACGGTGTCTG TGAGACACCG TT

SEQ ID NO: 6 (DJ1 recognition sequence - GTGA center sequence)

1 AACGGTGTCT CACGACACCG TT

SEQ ID NO: 7 (DJ1 half-site)

1 AACGGTGTCT

SEQ ID NO: 8 (BRP2 amino acid sequence)

1 MGPKKKRKVI MNTKYNKEFL LYLAGEFVDGD GSIIASIRPR QSKFKHELE
51 LRFQVTQKTQ RRWFLDKLVD EIGVGKYRDR GSVSDYRLSQ IKPLHNFLTQ
101 LQPFLLKQK QANLVLKIIIE QLPSAKESPD KFLEVCTWVD QIAALNDSKT
151 RKTSETVRA VLDLSEKKK SSP

SEQ ID NO: 9 (BRP2 recognition sequence - GTAC center sequence)

1 CTCGGGGTCG TACGACCCGG AG

SEQ ID NO: 10 (BRP2 recognition sequence - GTAC center sequence)

1 CTCGGGGTCG TACGACCCGG AG

SEQ ID NO: 11 (BRP2 recognition sequence - TAGA center sequence)

1 CTCGGGGTCT AGAGACCCGG AG

SEQ ID NO: 12 (BRP2 recognition sequence - TAGA center sequence)

1 CTCGGGGTCT CTAGACCCGG AG

SEQ ID NO: 13 (SV40 nuclear localization signal amino acid sequence)

1 MAPKKRKRV

SEQ ID NO: 14 (BRP2 codon-optimized DNA sequence)

1 ATGGGCCCCGA AGAAGAAGCG CAAGGTCATC ATGAACACCA AGTACAACAA
51 GGAGTTCTCTG CTCTACCTGG CGGGCTTCGT GGACGGGGAC GGCTCCATCA
101 TCGCCTCCAT CCGCCCGCGT CAGTCCTGCA AGTTCAAGCA TGAGCTGGAA
151 CTCGGTTTCC AGGTCACGCA GAAGACACAG CGCGTTGGT TCCTCGACAA
201 GCTGGTGGAC GAGATCGGGG TGGGCTACGT GCGCGACCGC GGCAGCGTCT
251 CCGACTACCG CCTGAGCCAG ATCAAGCCTC TGCACAACCT CCTGACCCAG
301 CTCAGCCCT TCCTGAAGCT CAAGCAGAAG CAGGCCAACC TCCTGCTGAA
351 GATCATCGAG CAGCTGCCCT CCGCCAAGGA ATCCCGGAC AAGTTCCTGG
401 AGGTGTGCAC CTGGGTGGAC CAGATCGCGC CTCTGAACGA CTCCAAGACC
451 CGCAAGACCA CTTCCGAGAC CGTCCGCGCC GTGCTGGACA GTCTCTCCGA
501 GAAGAAGAAG TCGTCCCCT AG

SEQ ID NO: 15 (BRP12-SC amino acid sequence)

1 MGPKKKRKVI MNTKYNKEFL LYLAGEFVDGD GSIKAQIRPR QSRKFKHELE
51 LTFQVTQKTQ RRWFLDKLVD EIGVGKYRDR GSVSDYELSQ IKPLHNFLTQ
101 LQPFLLKQK QANLVLKIIIE QLPSAKESPD KFLEVCTWVD QIAALNDSKT
151 RKTSETVRA VLDLPGSVG GLSPSQASSA ASSASSSPGS GISEALRAGA
201 TKSKEFLLYL AGFVDGDSI IASIRPRQSC KFKHELELRF QVTQKTQRRW
251 FLDKLVDLIG VGYVRDRGSV SDYRLSRIK LHNFLTQLP FLKLKQKQAN
301 LVLKIIIEQLP SAKESPDKFL EVCTWVDQIA ALNDSKTRKT TSETVRAVLD
351 SLSEKKKSSP

SEQ ID NO: 16 (BRP12-SC recognition sequence - GTAC center sequence)

1 TGCTCCTCG TACGACCCGG AG

SEQ ID NO: 17 (BRP12-SC recognition sequence - GTAC center sequence)

1 CTCGGGGTCG TACGAGGAGG CA

SEQ ID NO: 18 (BRP12-SC recognition sequence - TAGA center sequence)

1 TGCTCCTCT AGAGACCCGG AG

SEQ ID NO: 19 (BRP12-SC recognition sequence - TAGA center sequence)

1 CTCGGGGTCT CTAGAGGAGG CA

SEQ ID NO: 20 (BRP12-SC linker amino acid sequence)

1 PGSVGGLSPS QASSAASSAS SSPGSGISEA LRAGATKS

SEQ ID NO: 21 (BRP12-SC codon-optimized DNA sequence)

```

1  ATGGGCCCCGA AGAAGAAGCG CAAGGTCATC ATGAACACCA AGTACAACAA
51  GGAGTTCCCTG CTCTACCTGG CCGGCTTCGT GGACGGCGAC GGCTCCATCA
101 AGGCGCAGAT CCGTCCGCGG CAGAGCCGGA AGTTCAAGCA CGAGCTCGAG
151 CTGACCTTCC AGGTGACCCA GAAGACGCAG AGGCGCTGGT TCCTCGACAA
201 GCTGGTGGAC GAGATCGGGG TGGGCAAGGT CTACGACCGC GGCTCGGTGT
251 CCGACTACGA GCTCTCCCAG ATCAAGCCCC TGCACAACCT CCTCACCAG
301 CTCCAGCCGT TCCTGAAGCT CAAGCAGAAG CAGGCCAACC TCGTGCTGAA
351 GATCATCGAG CAGCTGCCCT CCGCCAAGGA ATCCCCGGAC AAGTTCTCTGG
401 AGGTGTGCAC GTGGGTGGAC CAGATCGCGG CCCTCAACGA CAGCAAGACC
451 CGCAAGACGA CCTCGGAGAC GGTGCGGGCG GTCTGGACT CCTCCCAGG
501 ATCCGTGGGA GGTCTATCGC CATCTCAGGC ATCCAGCGCC GCATCCTCGG
551 CTTCCCTCAAG CCCGGGTTCA GGGATCTCCG AAGCACTCAG AGCTGGAGCA
601 ACTAAGTCCA AGGAATTCTT GCTCTACCTG GCGGGCTTCG TGGACGGGGA
651 CGGCTCCATC ATCGCCTCCA TCCGCCCGCG TCAGTCTTGC AAGTTCAAGC
701 ATGAGCTGGA ACTCCGGTTC CAGGTCACGC AGAAGACACA GCGCCGTTCG
751 TTCTCGACA AGCTGGTGA CGAGATCGGG GTGGGCTACG TCGCGGACCG
801 CCGCAGCGTC TCCGACTACC GCCTGAGCCA GATCAAGCCT CTGCACAACT
851 TCCTGACCCA GCTCCAGCCC TTCCTGAAGC TCAAGCAGAA GCAGGCCAAC
901 TCTGTGCTGA AGATCATCGA GCAGCTGCC TCCGCCAAGG AATCCCCGGA
951 CAAGTTCCCTG GAGGTGTGCA CCTGGGTGGA CCAGATCGCC GCTCTGAACG
1001 ACTCCAAGAC CCGCAAGACC ACTTCCGAGA CCGTCCGCGC CGTGCTGGAC
1051 AGTCTCTCCG AGAAGAAGAA GTCGTCCCC TAG

```

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [US61080453A \[0001\]](#)
- [WO2007047859A \[0008\] \[0034\] \[0054\] \[0055\] \[0062\] \[0065\] \[0072\] \[0073\] \[0092\] \[0094\] \[0097\]](#)
- [WO04067736A \[0034\]](#)
- [WO07060495A \[0034\]](#)
- [WO06097853A \[0034\]](#)
- [WO07049095A \[0034\]](#)
- [WO08102198A \[0034\]](#)
- [WO08010093A \[0034\]](#)
- [WO08010009A \[0034\]](#)
- [WO07093918A \[0034\]](#)
- [WO07093836A \[0034\]](#)
- [WO08102274A \[0034\] \[0034\]](#)
- [WO08059317A \[0034\]](#)
- [WO09013622A \[0034\] \[0034\]](#)
- [WO09019614A \[0034\]](#)
- [WO09019528A \[0034\] \[0034\]](#)
- [WO08152523A \[0034\]](#)
- [WO04067753A \[0034\]](#)
- [WO03078619A \[0034\]](#)
- [WO06097784A \[0034\]](#)
- [WO07034262A \[0034\]](#)
- [WO07049156A \[0034\]](#)
- [WO07057781A \[0034\]](#)
- [WO08093152A \[0034\]](#)
- [WO08102199A \[0034\]](#)
- [WO08149176A \[0034\]](#)
- [WO09013559A \[0034\]](#)

Non-patent literature cited in the description

- PORTEUS et al. Nat. Biotechnol., 2005, vol. 23, 967-73 [0003] [0004] [0004] [0007] [0034]
- TZFIRA et al. Trends Biotechnol., 2005, vol. 23, 567-9 [0002]
- MCDANIEL et al. Curr. Opin. Biotechnol., 2005, vol. 16, 476-83 [0003]
- PORTEUS Mol Ther, 2006, vol. 13, 438-46 [0004]
- WRIGHT et al. Plant J., 2005, vol. 44, 693-705 [0004]
- URNOV et al. Nature, 2005, vol. 435, 646-51 [0004]
- SMITH et al. Nucleic Acids Res., 2000, vol. 28, 3361-9 [0004]
- STODDARD Q. Rev. Biophys., 2006, vol. 38, 49-95 [0005]
- CHEVALIER et al. Nucleic Acids Res., 2001, vol. 29, 183757-3774 [0005]
- MONNAT et al. Biochem. Biophys. Res. Commun., 1999, vol. 255, 88-93 [0006]
- ROUET et al. Mol. Cell. Biol., 1994, vol. 14, 8096-106 [0006]
- CHILTON et al. Plant Physiol., 2003, vol. 133, 956-65 [0006]
- PUCHTA et al. Proc. Natl Acad. Sci. USA, 1996, vol. 93, 5055-60 [0006]
- RONG et al. Genes Dev., 2002, vol. 16, 1568-81 [0006]
- GOUBLE et al. J. Gene Med., 2006, vol. 8, 5616-622 [0006]
- SUSSMAN et al. J. Mol. Biol., 2004, vol. 342, 31-41 [0007] [0008] [0034]
- EPINAT et al. Nucleic Acids Res., 2003, vol. 31, 2952-62 [0007] [0034]
- CHAMES et al. Nucleic Acids Res., 2005, vol. 33, e178- [0008]
- SELIGMAN et al. Nucleic Acids Res., 2002, vol. 30, 3870-9 [0008]
- ARNOULD et al. J. Mol. Biol., 2006, vol. 355, 443-58 [0008]
- ARNOULD et al. J. Mol. Biol., 2007, vol. 371, 49-65 [0009] [0090]
- CAHILL et al. Front. Biosci., 2006, vol. 11, 1958-1976 [0048] [0049]
- SALOMON et al. EMBO J., 1998, vol. 17, 6086-6095 [0049]
- AUSUBEL et al. Current Protocols in Molecular Biology Wiley 1999 0000 [0050]
- JURICA et al. Mol Cell., 1998, vol. 2, 469-76 [0057]

Patentkrav

1. Fremgangsmåde *ex vivo* til spaltning af et dobbeltstrenget DNA, hvilken fremgangsmåde omfatter:
 - 5 (a) identificering i DNA'et af mindst ét genkendelsessted for en rationelt designet ICrel-afledt meganuklease med ændret specificitet i forhold til I-CreI, hvor genkendelsesstedet ikke spaltes af en naturligt forekommende I-CreI, hvor genkendelsesstedet identificeres baseret på tilstedeværelse af en central sekvens med fire basepar udvalgt fra gruppen bestående af TTGT, TTAT, TTTC, TTCC, TTAG, TTAC og TTGC;
 - 10 (b) tilvejebringelse af den rationelt designede meganuklease og
 - (c) etablering af kontakt mellem DNA'et og den rationelt designede meganuklease; hvorved den rationelt designede meganuklease spalter DNA'et.
2. Fremgangsmåde *ex vivo* til spaltning af et dobbeltstrenget DNA, hvilken fremgangsmåde omfatter:
 - 15 (a) inkorporering af en central sekvens med fire basepar udvalgt fra gruppen bestående af TTGT, TTAT, TTTC, TTCC, TTAG, TTAC og TTGC i et genkendelsessted for en rationelt designet I-CreI-afledt meganuklease med ændret specificitet i forhold til I-CreI, hvor genkendelsesstedet ikke spaltes af en naturligt forekommende I-CreI,
 - 20 (b) indføring af genkendelsesstedet i DNA'et,
 - (c) tilvejebringelse af den rationelt designede meganuklease og
 - (d) etablering af kontakt mellem DNA'et og den rationelt designede meganuklease; hvorved den rationelt designede meganuklease spalter DNA'et.
- 25 3. Fremgangsmåde *ex vivo* ifølge et hvilket som helst af kravene 1-2, hvor DNA-spaltningen sker *in vitro*, eventuelt hvor DNA'et er udvalgt fra gruppen bestående af et PCR-produkt; et kunstigt kromosom; genomisk DNA isoleret fra bakterier, svampe, planter eller dyreceller og viralt DNA.
- 30 4. Fremgangsmåde *ex vivo* ifølge et hvilket som helst af kravene 1-3, der endvidere omfatter rationelt design af den I-CreI-afledte meganuklease for at genkende genkendelsesstedet.
5. Fremgangsmåde *ex vivo* ifølge et hvilket som helst af kravene 1-4, der endvidere omfatter frembringelse af den rationelt designede I-CreI-afledte meganuklease.

DRAWINGS

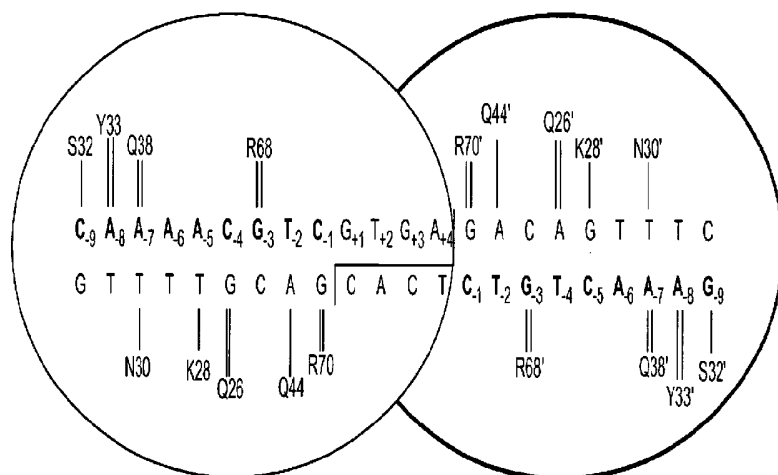


FIG. 1A

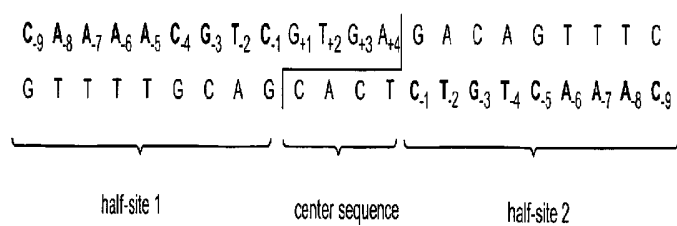


FIG. 1B

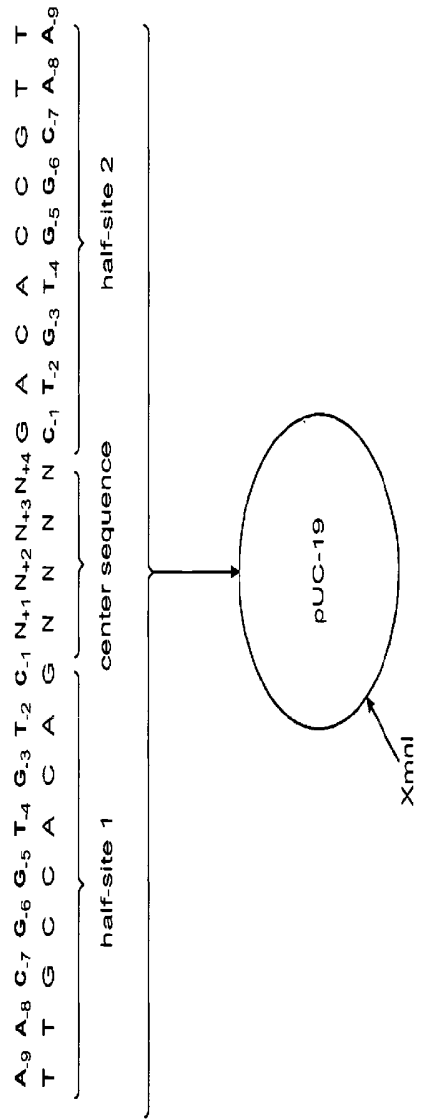


FIG. 2A

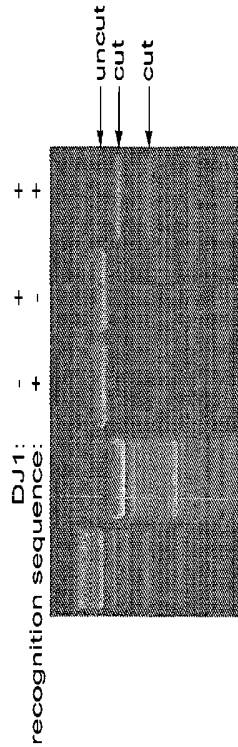
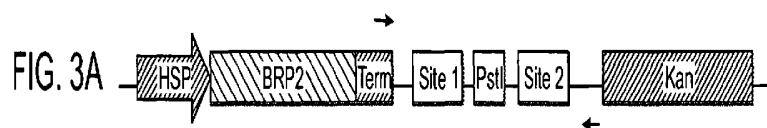


FIG. 2B

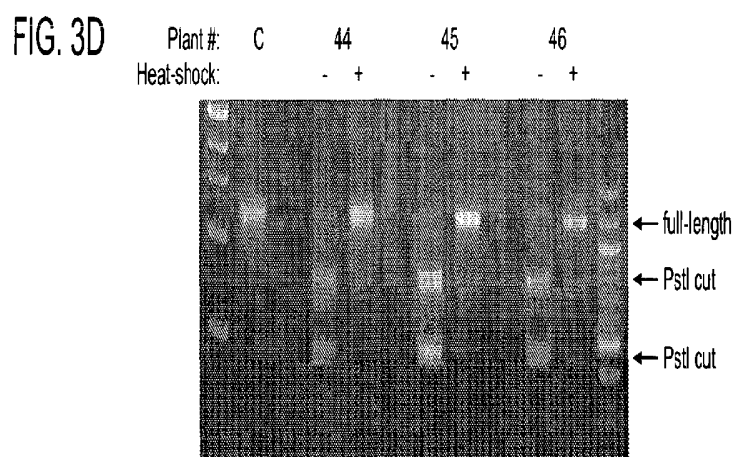


Meganuclease cleavage



FIG. 3C

GTAC: CTCCGGGTCGTACGACCCGGAG
 TAGA: CTCCGGGCTAGAGACCCGGAG



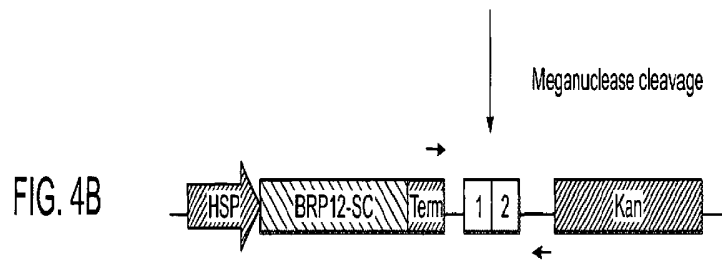
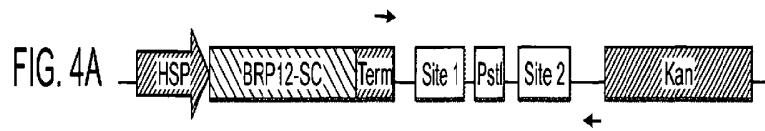


FIG. 4C

GTAC: TGCCTCCTCGTACGACCCGGAG
 TAGA: TGCCTCCTCTAGAGACCCGGAG

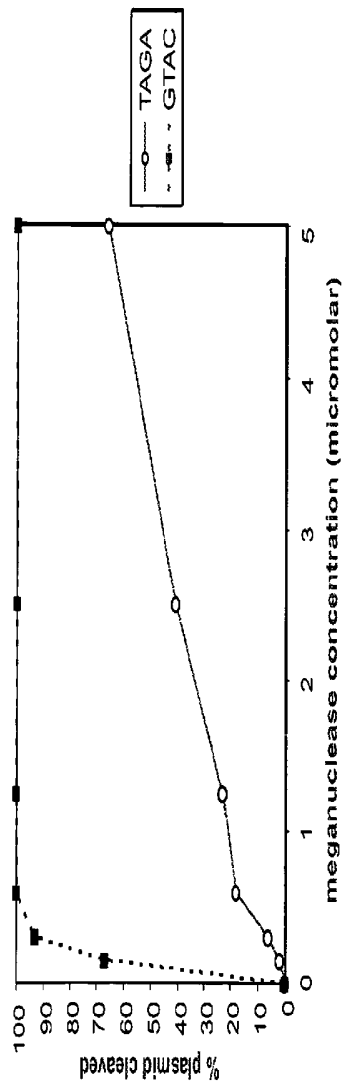


FIG. 5