

DESCRIPTION

Title of the Invention: METHOD FOR MODIFYING GENOME SEQUENCE
THAT SPECIFICALLY CONVERTS NUCLEOBASE OF TARGETED DNA SEQUENCE,
AND MOLECULAR COMPLEX USED IN SAID METHOD

5 [Technical Field]

[0001]

The present invention relates to a method for modifying a genomic sequence without cleaving a double stranded DNA or inserting a foreign DNA fragment and enabling modification of a nucleic acid base within a particular region of the genome, and to a complex of a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme to be used therefor.

[Background Art]

[0002]

15 In recent years, genome editing is attracting attention as a technique for modifying the object gene and genome region in various species. Conventionally, as a method of genome editing, a method utilizing an artificial nuclease comprising a molecule having a sequence-independent DNA cleavage ability and a molecule having a sequence recognition ability in combination has been proposed (non-patent document 1).

For example, a method of performing recombination at a target gene locus in DNA in a plant cell or insect cell as a host, by using a zinc finger nuclease (ZFN) wherein a zinc finger DNA binding domain and a non-specific DNA cleavage domain are linked (patent document 1), a method of cleaving or modifying a target gene in a particular nucleotide sequence or a site adjacent thereto by using TALEN wherein a transcription activator-like (TAL) effector which is a DNA binding module that the plant pathogenic bacteria *Xanthomonas* has, and a DNA endonuclease are linked (patent document 2), a method utilizing CRISPR-Cas9 system wherein DNA sequence CRISPR (Clustered Regularly interspaced short palindromic repeats) that functions in an acquired immune system possessed by eubacterium and archaeobacterium, and nuclease Cas (CRISPR-associated) protein

family having an important function along with CRISPR are combined (patent document 3) and the like have been reported. Furthermore, a method of cleaving a target gene in the vicinity of a particular sequence, by using artificial nuclease wherein a PPR protein constituted to recognize a particular nucleotide sequence by a continuation of PPR motifs each consisting of 35 amino acids and recognizing one nucleic acid base, and nuclease are linked (patent document 4) has also been reported.

[0003]

10 The genome editing techniques proposed heretofore basically presuppose double stranded DNA breaks (DSB). However, since they include unexpected genome modifications, side effects such as strong cytotoxicity, chromosomal rearrangement and the like occur, and they have common problems of impaired reliability in gene therapy, extremely small number of surviving cells by nucleotide modification, and difficulty in genetic modification itself in primate ovum and unicellular microorganisms.

[0004]

20 As a genome editing technique unaccompanied by DSB, an artificial enzyme obtained by combining a molecule having a sequence recognition ability such as zinc finger (ZF) motif and the like and deaminase that converts an amino group of a nucleic acid base to a carbonyl group has been proposed (patent document 5). However, experimental proof does not exist, and it is not clear whether genetic modification is at all possible, not to mention mutation introduction efficiency thereof. Indeed, as a system similarly unaccompanied by genome cleavage, a method using DNA glycosylase that catalyzes deamination reaction on the DNA strand shows an extremely low mutation introduction efficiency even when a mutant yeast with an attenuated DNA repair mechanism is used as a host (non-patent document 2), and practicalization for gene therapy, molecular breeding of useful organisms and the like is further away as the situation stands.

[Document List]

[Patent documents]

[0005]

patent document 1: JP-B-4968498

5 patent document 2: National Publication of International Patent
Application No. 2013-513389

patent document 3: National Publication of International Patent
Application No. 2010-519929

patent document 4: JP-A-2013-128413

10 patent document 5: US-A-2011/0104787

[non-patent document]

[0006]

non-patent document 1: Kelvin M Esvelt, Harris H Wang (2013)
Genome-scale engineering for systems and synthetic biology,

15 Molecular Systems Biology 9: 641

non-patent document 2: Prashant Mali, Kevin M Esvelt, George M
Church (2013) Nucleic Acids Res. 41: e99

[SUMMARY OF THE INVENTION]

[Problems to be Solved by the Invention]

20 [0007]

Therefore, an object of the present invention is to
provide a novel method of genome editing to modify a nucleic
acid base in a particular sequence of a gene without DSB or
insertion of a foreign DNA fragment, and a complex of a nucleic
25 acid sequence-recognizing module and a nucleic acid base
converting enzyme therefor.

[Means of Solving the Problems]

[0008]

The present inventors have already succeeded in
30 efficiently modifying, without accompanying DSB, genomic
sequence in a region containing a particular DNA sequence, by
using a Type II CRISPR-Cas system having mutant Cas9, in which
cleavage ability of one or both strands of a double stranded
DNA has been inactivated, as a nucleic acid sequence-
35 recognizing module and deaminase as a nucleic acid base

converting enzyme (WO 2015/133554). However, since Cas protein recognizes and binds to a sequence called protospacer adjacent motif (PAM) on a DNA strand, the mutation introduction site is limited by the presence or otherwise of PAM. Cas9 (SpCas9) derived from *Streptococcus pyogenes* in the Type II CRISPR-Cas system currently used frequently for genome editing recognizes NGG (N is any base) as PAM. It has also been clarified by the research of the present inventors that the nucleotide sequence to be targeted in Type II CRISPR (i.e., nucleotide sequence on DNA strand complementary to guide RNA) is 5' upstream 18 - 25 nucleotides of PAM in length and, despite its length, base conversion most frequently occurs at the positions of 2 - 5 nucleotides from the 5'-end of the target nucleotide sequence. [0009]

On the other hand, in the Type I-E CRISPR-Cas system of *Escherichia coli* and the like, it is known that ATG, AAG, AGG or GAG adjacent to the 5'-side of the target nucleotide sequence with 32 - 33 nucleotide length is recognized as PAM sequence (see Fig. 1; PAM sequence shows the case of AAG in Fig. 1D and E), and other Type I CRISPR-Cas systems also recognize specific PAM sequences of 2 or 3 bases. Therefore, if Type I CRISPR-Cas system can be used as a nucleic acid sequence-recognizing module, base conversion can be performed at a site where introduction of mutation is difficult for Cas9 due to the restriction of PAM sequence and introduction site of mutation.

In the Type II CRISPR-Cas system, a nucleic acid sequence-recognizing module could be constituted only from a Chimeric RNA of crRNA complementary to the target nucleotide sequence and trans-acting crRNA (tracrRNA) for recruiting Cas 9, and Cas9. However, for example, in the Type I-E CRISPR-Cas system of *Escherichia coli*, a complicated ribonucleoprotein complex composed of 5 kinds of Cas proteins (CasA, CasB, CasC, CasD and CasE) presenting crRNA, which is called CRISPR-associated complex for antiviral defense (Cascade), recognizes the target nucleotide sequence and the PAM sequence. Combined

with Cas3 having nuclease and helicase activities, the complex exhibits a DNA recognition and cleavage function similar to that of Cas9 in the Type II CRISPR-Cas system (see Fig. 1). Due to such complicated constitution, the Type I CRISPR-Cas
5 system has hardly been utilized even for genome editing technique as an artificial nuclease.

[0010]

The present inventors prepared a DNA encoding genomic specific CRISPR-RNA in which 5' handle and 3' handle necessary
10 for correct display in Cascade are linked at both ends of a sequence complementary to the target strand of the target nucleotide sequence of the rpoB gene which is an essential gene of Escherichia coli (crRNA). On the other hand, they isolated casA - casE gene group necessary for the constitution of
15 Cascade from Cas operon of Escherichia coli, linked a deaminase gene thereto to produce a DNA, and introduced these DNAs into a host Escherichia coli containing a gene to be modified. As a result, they successfully introduced mutation into the target nucleotide sequence of the gene of interest and the vicinity
20 thereof, without accompanying cleavage of the genome DNA. In addition, it was found that cytosine at 32 - 44 bases downstream from the PAM sequence is mainly edited when the length of the target nucleotide sequence is set to 32 nucleotides, unlike the use of Cas9.

25 The present inventors have conducted further studies based on these findings and completed the present invention.

[0011]

That is, the present invention is as described below.

[1] A method of modifying a targeted site of a double stranded
30 DNA in a host cell, the method comprising introducing
(a) a DNA encoding a crRNA comprising a sequence complementary to a target strand of a target nucleotide sequence in the given double stranded DNA, and
(b) a DNA encoding a protein group constituting Cascade and a
35 nucleic acid base converting enzyme, in which the nucleic acid

base converting enzyme is constituted in a form capable of forming a complex with any protein in the protein group into the host cell to convert one or more nucleotides in the targeted site to other one or more nucleotides, or delete one or more nucleotides, or insert one or more nucleotides into said targeted site, without cleaving the double stranded DNA in the targeted site.

[2] The method of the above-mentioned [1], wherein the aforementioned protein group constituting Cascade comprises CasA, CasB, CasC, CasD and CasE.

[3] The method of the above-mentioned [2], wherein the aforementioned Cascade is derived from *Escherichia coli*.

[4] The method of the above-mentioned [2] or [3], wherein the protein that forms the complex with the aforementioned nucleic acid base converting enzyme is CasE.

[5] The method of any of the above-mentioned [1] to [4], wherein the aforementioned nucleic acid base converting enzyme is deaminase.

[6] The method of the above-mentioned [5], wherein the aforementioned deaminase is cytidine deaminase.

[7] The method of any of the above-mentioned [1] to [6], wherein the aforementioned host cell is a prokaryotic cell.

[8] The method of any of the above-mentioned [1] to [7], comprising a step of introducing an expression vector comprising the DNAs of the aforementioned (a) and (b) in a form capable of controlling an expression period into the aforementioned host cell, and inducing expression of the DNAs for a period necessary for fixing the modification of the targeted site in the double stranded DNA.

[9] The method of the above-mentioned [8], wherein the target nucleotide sequence in the double stranded DNA is present in a gene essential for the aforementioned host cell.

[10] A nucleic acid-modifying enzyme complex for modifying a targeted site of a double stranded DNA in a host cell, the complex comprising

(a) a crRNA comprising a sequence complementary to a target strand of a target nucleotide sequence in the given double stranded DNA, and

(b) a nucleic acid-modifying enzyme complex comprising a protein group constituting Cascade, and a nucleic acid base converting enzyme that has formed a complex with any protein in the protein group.

[11] A DNA encoding the nucleic acid-modifying enzyme complex of the above-mentioned [10].

10 [Effect of the Invention]

[0012]

The genome editing technique of the present invention does not accompany insertion of a foreign DNA or cleavage of double stranded DNA. Therefore, it is superior in safety and has no small possibility of becoming a solution to the cases associated with biological or legal disputes of gene recombination in conventional methods. In addition, using Cascade as a nucleic acid sequence-recognizing module, PAM sequence, which is different from Cas9, can be utilized and a mutation is introduced highly frequently into a different site, which expands the choice of the site into which a mutation can be introduced in the target gene.

[Brief Description of the Drawings]

[0013]

25 Fig. 1 schematically shows cas operon on the Escherichia coli genome and CRISPR gene locus (A), structure of crRNA (B), structure of Cascade (C), R-loop structure produced when Cascade recognizes and binds to the target nucleotide sequence (D), and comparison of the target nucleotide sequences and PAM sequences of Cas9 and Cascade (E).

Fig. 2 shows the constitution of respective subtypes (I-A to I-F) of the cas gene group in the Type I CRISPR-Cas system.

Fig. 3 is a schematic showing of the structure (top) of the major part of the representative vector used in the present invention, and genetic modification (bottom) by the nucleic

acid-modifying enzyme complex of the present invention produced from the vector.

Fig. 4 relates to a physical map and the detailed sequence information of the vector used in the Examples.

5 Fig. 5-1 shows the sequencing results of the target nucleotide sequence and the vicinity thereof in the rifampicin resistant colony obtained by the modification of rpoB gene by using a Cascade-deaminase complex. Target 3r and target 5r were used.

10 Fig. 5-2 shows the sequencing results of the target nucleotide sequence and the vicinity thereof in the rifampicin resistant colony obtained by the modification of rpoB gene by using a Cascade-deaminase complex. Target 4r was used.

[Description of Embodiments]

15 [0014]

The present invention provides a method of modifying a targeted site of a double stranded DNA by converting the target nucleotide sequence and nucleotides in the vicinity thereof in the double stranded DNA to other nucleotides, without cleaving
20 the double stranded DNA to be modified in the host cell (hereinafter to be also referred to as "the method of the present invention"). The method characteristically contains a step of contacting a complex wherein a nucleic acid sequence-recognizing module that specifically binds to the target
25 nucleotide sequence in the double stranded DNA in the host cell and a nucleic acid base converting enzyme are bonded with the double stranded DNA to convert the targeted site, i.e., the target nucleotide sequence and nucleotides in the vicinity thereof, to other nucleotides.

30 [0015]

In the present invention, the "modification" of a double stranded DNA means that a nucleotide (e.g., dC) on a DNA strand is converted to other nucleotide (e.g., dT, dA or dG), or deleted, or a nucleotide or a nucleotide sequence is inserted
35 between certain nucleotides on a DNA strand. While the double

stranded DNA to be modified is not particularly limited as long as it is a double stranded DNA present in the host cell, it is preferably a genomic DNA. The "targeted site" of a double stranded DNA means the whole or partial "target nucleotide
5 sequence", which a nucleic acid sequence-recognizing module specifically recognizes and binds to, or the vicinity of the target nucleotide sequence (one or both of 5' upstream and 3' downstream).

[0016]

10 In the present invention, the "nucleic acid sequence-recognizing module" means a molecule or molecule complex having an ability to specifically recognize and bind to a particular nucleotide sequence (i.e., target nucleotide sequence) on a DNA strand. Binding of the nucleic acid sequence-recognizing
15 module to a target nucleotide sequence enables a nucleic acid base converting enzyme linked to the module to specifically act on a targeted site of a double stranded DNA.

[0017]

In the present invention, the "nucleic acid base
20 converting enzyme" means an enzyme capable of converting a target nucleotide to other nucleotide by catalyzing a reaction for converting a substituent on a purine or pyrimidine ring on a DNA base to other group or atom, without cleaving the DNA strand.

25 [0018]

In the present invention, the "nucleic acid-modifying enzyme complex" means a molecule complex comprising a complex wherein the above-mentioned nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme are linked,
30 which molecule complex is imparted with a particular nucleotide sequence recognition ability, and nucleic acid base converting enzyme activity. The "complex" here encompasses not only one constituted of multiple molecules, but also one having a nucleic acid sequence-recognizing module and nucleic acid base
35 converting enzyme in a single molecule, like a fusion protein.

[0019]

The nucleic acid base converting enzyme to be used in the present invention is not particularly limited as long as it can catalyze the above-mentioned reaction, and examples thereof include deaminase belonging to the nucleic acid/nucleotide deaminase superfamily, which catalyzes a deamination reaction that converts an amino group to a carbonyl group. Preferable examples thereof include cytidine deaminase capable of converting cytosine or 5-methylcytosine to uracil or thymine, respectively, adenosine deaminase capable of converting adenine to hypoxanthine, guanosine deaminase capable of converting guanine to xanthine and the like. As cytidine deaminase, more preferred is activation-induced cytidine deaminase (hereinafter to be also referred to as AID) which is an enzyme that introduces a mutation into an immunoglobulin gene in the acquired immunity of vertebrata or the like.

[0020]

While the derivation of nucleic acid base converting enzyme is not particularly limited, for example, PmCDA1 derived from *Petromyzon marinus* (*Petromyzon marinus* cytosine deaminase 1) or AID (Activation-induced cytidine deaminase; AICDA) derived from vertebrata (e.g., mammal such as human, swine, bovine, dog, chimpanzee and the like, birds such as chicken and the like, amphibian such as *xenopus* and the like, fish such as zebrafish, sweetfish, channel catfish and the like) can be used. The base sequence and amino acid sequence of CDS of PmCDA1 are shown in SEQ ID.NOs: 1 and 2.

[0021]

As the nucleic acid sequence-recognizing module in the nucleic acid-modifying enzyme complex of the present invention, a Type I CRISPR-Cas system is specifically used.

As a molecule having a nucleic acid sequence recognition ability, zinc finger (ZF) motif, TAL effector, PPR motif and the like are known. However, the production efficiency of ZF of a ZF motif that specifically binds to the target nucleotide

sequence is not high and selection of ZF having high binding specificity is complicated, and therefore, production of a large number of ZF motifs that actually function is not easy. On the other hand, while TAL effector and PPR motif have a high
5 degree of freedom in the target nucleic acid sequence recognition as compared to ZF, since a huge protein according to the target nucleotide sequence needs to be designed and constructed each time, they have a problem in terms of efficiency. In contrast, the CRISPR-Cas system, which is an
10 acquired immunity mechanism widely distributed in eubacteria and archaeobacteria, can target any sequence merely by synthesizing an oligoDNA capable of specifically forming a hybrid with the target nucleotide sequence since CRISPR-RNA (crRNA) complementary to the target nucleotide sequence
15 recognize the sequence of the object double stranded DNA.
[0022]

In the present specification, the "target nucleotide sequence" means a double stranded DNA sequence to which a nucleic acid sequence-recognizing module (specifically, Type I
20 CRISPR-Cas) binds, and a strand that hybridizes to crRNA is referred to as a "targeted strand", and an opposite strand thereof which becomes single-stranded by hybridization to a target strand and crRNA is referred to as a "non-targeted strand". Since a nucleic acid base conversion reaction
25 generally frequently occurs on a single-stranded non-target strand, when the target nucleotide sequence is to be expressed by a single strand (e.g., when PAM sequence is indicated, when positional relationship of target nucleotide sequence and PAM is shown etc.), it is represented by a sequence of a non-target
30 strand.

[0023]

CRISPR-Cas system is largely divided into three types. One currently frequently used for genome editing is the Type II CRISPR-Cas system in which Cas9 has both a DNA sequence
35 recognition ability and a nuclease activity. On the other hand,

in the Type I CRISPR-Cas system, the function is shared between Cas protein complex Cascade, which presents crRNA containing a sequence complementary to the target nucleotide sequence and recognizes the target nucleotide sequence, and Cas3 having
5 nuclease and helicase activities. Therefore, DSB with strong cytotoxicity can be avoided by simply removing Cas3 from the Cas operon even without using a mutant having an inactivated cleavage ability of at least one strand of a double-stranded DNA like Cas9. That is, (a) a DNA encoding a chimeric crRNA in
10 which 5' handle and 3' handle of crRNA are linked to a sequence complementary to a target strand of a target nucleotide sequence (targeting sequence), and (b) a DNA encoding a protein group constituting a Cascade complex and a nucleic acid base converting enzyme are introduced into a host cell having a
15 double stranded DNA to be modified and expressed to form a Cascade complex composed of the crRNA containing the targeting sequence and the protein group in the host cell, whereby the target nucleotide sequence on the double stranded DNA can be identified. The DNA encoding the nucleic acid base converting
20 enzyme is disposed in the DNA of the above-mentioned (b) such that the nucleic acid base converting enzyme here is expressed in a form capable of forming a complex with any protein in the protein group constituting the Cascade. In this way, the nucleic acid base converting enzyme can convert the target
25 nucleotide sequence and the nucleic acid bases in the vicinity thereof recognized and bound by the Cascade complex to other bases.

[0024]

Type I CRISPR-Cas system is known to contain six subtypes
30 of A - F, and the Cascade of any subtype can be used as the nucleic acid sequence-recognizing module of the present invention. Fig. 2 schematically shows the Cas operons of subtypes I-A to I-F. In the operon of each subtype, the protein group excluding Cas3 which is nuclease/helicase, Cas1
35 involved in foreign gene cleavage during immunity acquisition,

Cas2 and Cas4 constitutes the Cascade complex. Cse1, Cse2, Cas7, Cas5 and Cas6e in the I-E subtypes are also respectively called CasA, CasB, CasC, CasD and CasE (see Fig. 1; the latter names are used in the present specification). In the I-E subtype, a Cascade complex is formed by CasA, CasB, CasC, CasD and CasE respectively at 1:2:6:1:1 molecule crRNA (see Fig. 1). In the I-E subtype, CasA is considered to play an important role in the identification of PAM sequence (ATG, AAG, AGG or GAG).

10 Relatively many I-A, I-B and I-D subtypes are distributed in archaeobacterium, and relatively many I-C, I-E and I-F subtypes are distributed in eubacterium. The I-A subtype is analyzed in *S. solfataricus*, *T. tenax* and the like, I-B subtype is analyzed in *Haloferax volcanii* and the like, I-C subtype is analyzed in *B. halodurans* and the like, I-E subtype is analyzed in *Escherichia coli* and the like, and I-F subtype is analyzed in *P. aeruginosa*, *Escherichia coli*, *P. atosepticum* and the like. [0025]

 Therefore, a DNA encoding a protein group constituting a Cascade complex can be obtained by, for example, isolating a region containing ORF of cas5 from csa5 in the case of I-A subtype, a region containing ORF of cas5 from csa6 in the case of I-B subtype, a region containing ORF of cas7 from csa5 in the case of I-C subtype, a region containing ORF of casE from casA in the case of I-E subtype, a region containing ORF of cas6f from csy1 in the case of I-F subtype, from cas operon by genome PCR using genome DNA derived from the above-mentioned bacterial strain as a template. Preferably, the protein groups constituting the Cascade complexes are CasA, CasB, CasC, CasD and CasE constituting Cascade complexes of I-E subtypes, more preferably CasA, CasB, CasC, CasD and CasE derived from *Escherichia coli*. The ORF sequences of casA, casB, casC, casD and casE derived from *Escherichia coli* are respectively the 5933rd-7441st, 5458th-5940th, 4341st-5432nd, 3664th-4338th and 3081st-3677th (all Opposite strands) sequences of the

nucleotide sequence shown in SEQ ID NO: 4.

[0026]

A protein group constituting a DNA encoding a Cascade complex can be obtained by chemically synthesizing the DNA strand, or by connecting synthesized partly overlapping oligoDNA short strands by utilizing the PCR method and the Gibson Assembly method to construct a DNA encoding the full length thereof. The advantage of constructing a full-length DNA by chemical synthesis or a combination of PCR method or Gibson Assembly method is that the codon to be used can be designed in CDS full-length according to the host into which the DNA is introduced. In the expression of a heterologous DNA, the protein expression level is expected to increase by converting the DNA sequence thereof to a codon highly frequently used in the host organism. As the data of codon use frequency in host to be used, for example, the genetic code use frequency database (<http://www.kazusa.or.jp/codon/index.html>) disclosed in the home page of Kazusa DNA Research Institute can be used, or documents showing the codon use frequency in each host may be referred to. By reference to the obtained data and the DNA sequence to be introduced, codons showing low use frequency in the host from among those used for the DNA sequence may be converted to a codon coding the same amino acid and showing high use frequency.

[0027]

A DNA encoding a nucleic acid base converting enzyme can also be cloned similarly from the enzyme-producing cells. For example, a DNA encoding PmCDA1 of *Petromyzon marinus* can be cloned by designing suitable primers for the upstream and downstream of CDS based on the cDNA sequence (accession No. EF094822) registered in the NCBI database, and cloning from *Petromyzon marinus*-derived mRNA by the RT-PCR method. A DNA encoding human AID can be cloned by designing suitable primers for the upstream and downstream of CDS based on the cDNA sequence (accession No. AB040431) registered in the NCBI

database, and cloning from, for example, human lymph node-derived mRNA by the RT-PCR method. AID homologues derived from other vertebrates can also be cloned in the same manner as in the above based on known cDNA sequence information (e.g., swine
5 (accession No. CU582981), bovine (accession No. NM_110138682), dog (accession No. NM_001003380), chimpanzee (accession No. NM_001071809), chicken (accession No. NM_001243222), xenopus (accession No. NM_001095712), zebrafish (accession No. AAI62573), sweetfish (accession No. AB619797) and channel
10 catfish (accession No. NM_001200185) etc.).

Alternatively, similar to the above, it can also be constructed as a DNA showing codon usage suitable for expression in a host cell to be used, by a combination of chemical synthesis or PCR method or Gibson Assembly method.

15 [0028]

A DNA encoding a cloned or synthesized nucleic acid base converting enzyme may be directly, or after digestion with a restriction enzyme when desired, or after addition of a suitable linker and/or a nuclear localization signal (each
20 organelle transfer signal when the object double stranded DNA is mitochondria or chloroplast DNA), linked with a DNA encoding one or more proteins selected from a group of proteins constituting the Cascade complex. While the linker to be used is not particularly limited, for example, Flag-tag, GS linker,
25 Strep-tag and the like can be mentioned, and tandem repeats of these are also encompassed therein.

Alternatively, a DNA encoding a protein constituting the Cascade complex, and a DNA encoding a nucleic acid base
30 converting enzyme may be each fused with a DNA encoding a binding domain (e.g., SH3 domain, PDZ domain, GK domain, GB domain etc.) or a binding partner thereof, or both DNAs may be fused with a DNA encoding a separation intein, whereby the protein constituting the Cascade complex and the nucleic acid
base converting enzyme are translated in a host cell to form a
35 complex. Furthermore, a nucleic acid base converting enzyme

and a Cascade constituent protein may be bound using an RNA scaffold by RNA aptamer (MS2F6, PP7 etc.) and their binding proteins. Also in these cases, a linker and/or a nuclear localization signal can be linked to a suitable position of one of or both DNAs when desired.

[0029]

While the Cascade constituent protein that forms a complex with the nucleic acid base converting enzyme is not particularly limited, when it is expressed as a fusion protein, it is easy for gene manipulation to link the protein to the terminal of the operon (downstream of CasE on the C-terminal side in the case of I-E subtype (see Fig. 3), upstream of CasA on the N terminal side). However, as long as the Cas protein-nucleic acid base converting enzyme complex can form a correct Cascade complex (Casprotein group is associated in such a manner that crRNA can hybridize to a target strand of a target nucleotide sequence, and crRNA can be presented), DNA encoding the nucleic acid base converting enzyme may be inserted in the inside of operon and linked to a DNA encoding any of the Cas proteins (CasB, CasC or CasD in I-E subtype). For example, in the I-E subtype, 2 molecules of CasB and 6 molecules of CasC per 1 molecule of crRNA form a Cascade complex. Therefore, as long as the nucleic acid base converting enzymes do not produce a steric disorder, introduction of multiple molecules of nucleic acid base converting enzymes has a possibility of improving the mutation introduction efficiency.

A complex of a nucleic acid base converting enzyme and a Cascade constituent protein can also be formed using a binding domain and the like by, similarly, linking a DNA encoding a binding domain or a binding partner thereof to the terminal or inside of the operon and linking the binding partner or binding domain to a DNA encoding the nucleic acid base converting enzyme.

[0030]

The DNA encoding a chimeric crRNA in which 5' handle and

3' handle of crRNA are linked to a sequence complementary to a target strand of a target nucleotide sequence (targeting sequence) can be chemically synthesized by designing an oligo DNA sequence in which known 5' handle and 3' handle (e.g.,
5 ataaaccg as 5' handle and gagttccccgcgccagcgggg (SEQ ID NO: 3) as 3' handle when I-E subtype derived from *Escherichia coli* is used as Cascade complex) are respectively linked to the 5' upstream and 3' downstream of the targeting sequence and using a DNA/RNA synthesizer. The 5' handle and 3' handle described
10 in J. Biol. Chem. 286(24): 21643-21656 (2011) can be used when I-A subtype derived from *S. solfataricus* is used, the 5' handle and 3' handle described in J. Biol. Chem. 287(40): 33351-33365 (2012) can be used when I-B subtype derived from *H. volcanii* is used, the 5' handle and 3' handle described in Structure 20:
15 1574-1584 (2012) can be used when I-C subtype derived from *B. halodurans* is used, and the 5' handle and 3' handle described in PNAS 108(25): 10092-10097 (2011) can be used when I-F subtype derived from *P. aeruginosa* is used.

[0031]

20 The length of the targeting sequence is not particularly limited as long as it is correctly presented to a Cascade complex and can specifically bind to a target strand of a target nucleotide sequence. It is, for example, 30 - 45 nucleotides, preferably 32 - 33 nucleotides for I-C, I-E and I-
25 F subtypes, and preferably 34-44 nucleotides for I-A and I-B subtypes.

The target nucleotide sequence in the Type I CRISPR-Cas system is subject to restriction by PAM sequence specific to the subtype. In the I-E subtype, any of AAG, ATG, AGG, GAG
30 needs to be adjacent to the non-target strand on the 5' upstream side of the target nucleotide sequence in the 5' → 3' direction. For example, in the I-E subtype, when a targeting sequence with 32 nucleotide length is used in combination with deaminase, as shown in the below-mentioned Examples, cytosine
35 present at 32 - 44 bases downstream of the PAM sequence causes

base conversion, and a mutation is introduced highly frequently.
[0032]

Since the nucleic acid-modifying enzyme complex of the present invention, in which the Type I CRISPR-Cas system and a
5 nucleic acid base converting enzyme are combined, does not accompany double stranded DNA breaks (DSB), genome editing with low toxicity is possible, and the genetic modification method of the present invention can be applied to a wide range of biological materials. Therefore, the cells into which (a) a
10 DNA encoding a chimeric crRNA in which 5' handle and 3' handle of crRNA are linked to the targeting sequence and (b) a DNA encoding a protein group constituting a Cascade complex and a nucleic acid base converting enzyme are introduced can encompass cells of any species, from bacterium of *Escherichia*
15 *coli* and the like which are prokaryotes, cells of microorganism such as yeast and the like which are lower eucaryotes, to cells of vertebrata including mammals such as human and the like, and cells of higher eukaryote such as insect, plant and the like.
[0033]

20 An expression vector containing the DNAs of the above-mentioned (a) and (b) can be produced, for example, by linking the DNA to the downstream of a promoter in a suitable expression vector. The DNAs of (a) and (b) may be combined on the same vector or separate vectors.

25 As the expression vector, *Escherichia coli*-derived plasmids (e.g., pBR322, pBR325, pUC12, pUC13); *Bacillus subtilis*-derived plasmids (e.g., pUB110, pTP5, pC194); yeast-derived plasmids (e.g., pSH19, pSH15); insect cell expression plasmids (e.g., pFast-Bac); animal cell expression plasmids
30 (e.g., pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNA1/Neo); bacteriophages such as λ phage and the like; insect virus vectors such as baculovirus and the like (e.g., BmNPV, AcNPV); animal virus vectors such as retrovirus, vaccinia virus, adenovirus and the like, and the like are used.

35 As the promoter, any promoter appropriate for a host to

be used for gene expression can be used. In a conventional method using DSB, since the survival rate of the host cell sometimes decreases markedly due to the toxicity, it is desirable to increase the number of cells by the start of the induction by using an inductive promoter. However, since sufficient cell proliferation can also be afforded by expressing the nucleic acid-modifying enzyme complex of the present invention, a constituent promoter can also be used without limitation.

10 When the host is *Escherichia coli*, trp promoter, lac promoter, recA promoter, λP_L promoter, λP_R promoter, lpp promoter, T7 promoter and the like are preferable.

 When the host is genus *Bacillus*, SPO1 promoter, SPO2 promoter, penP promoter and the like are preferable.

15 When the host is a yeast, Gal1/10 promoter, PHO5 promoter, PGK promoter, GAP promoter, ADH promoter and the like are preferable.

 When the host is an insect cell, polyhedrin promoter, P10 promoter and the like are preferable.

20 For example, when the host is an animal cell, SR α promoter, SV40 promoter, LTR promoter, CMV (cytomegalovirus) promoter, RSV (Rous sarcoma virus) promoter, MoMuLV (Moloney mouse leukemia virus) LTR, HSV-TK (simple herpes virus thymidine kinase) promoter and the like are used. Of these, CMV promoter, SR α promoter and the like are preferable.

 When the host is a plant cell, CaMV35S promoter, CaMV19S promoter, NOS promoter and the like are preferable.

 A DNA encoding the chimeric crRNA can also use a pol III system promoter (e.g., SNR6, SNR52, SCR1, RPR1, U6, H1 promoter etc.) as the promoter and a terminator (e.g., T₆ sequence).

[0034]

 When prokaryotic cell or lower eukaryotic cell such as yeast and the like is used as a host, a DNA encoding a Cascade constituent protein group can be polycistronically expressed as operon (see Fig. 3). When a higher eukaryotic cell such as

animal cell and the like is used as a host, an intervening sequence (e.g., IRES sequence, 2A sequence derived from foot-and-mouth disease virus and the like) that enables polycistronic expression can be inserted between respective
5 DNAs encoding Cas proteins. Alternatively, a promoter may be inserted into the 5' upstream of each DNA to cause monocistronic expression.

[0035]

DNAs encoding the chimeric crRNA can also be inserted in
10 tandem in the downstream of one promoter (see Fig. 3). In this case, the targeting sequences mounted on respective chimeric crRNAs may be the same or different. When the sequences are different, a different region in the same gene may be used as the target nucleotide sequence or a region in a different gene
15 may be used as the target nucleotide sequence. The linking site between respective chimeric crRNA units is a 3' handle-5' handle crRNA repeat sequence having a partial secondary structure of a characteristic hairpin structure. It is therefore considered to be cleaved by a non-specific RNase
20 endogenous to the host cell. The sequence can also be intracellularly cleaved into each chimeric crRNA unit even in a host cell without an endogenous Type I CRISPR-Cas system. It is of course possible to express each chimeric crRNA individually by inserting a promoter in the upstream of the
25 chimeric crRNA.

[0036]

As the expression vector, besides those mentioned above, one containing enhancer, splicing signal, terminator, polyA addition signal, a selection marker such as drug resistance
30 gene, auxotrophic complementary gene and the like, replication origin and the like on demand can be used.

[0037]

A nucleic acid-modifying enzyme complex of the present invention can be intracellularly expressed by introducing an
35 expression vector containing (a) a DNA encoding chimeric crRNA

and (b) a DNA encoding a Cascade constituent protein group and a nucleic acid base converting enzyme into a host cell, and culturing the host cell.

As the host, genus *Escherichia*, genus *Bacillus*, yeast, insect cell, insect, animal cell and the like are used.

As the genus *Escherichia*, *Escherichia coli* K12·DH1 [Proc. Natl. Acad. Sci. USA, 60, 160 (1968)], *Escherichia coli* JM103 [Nucleic Acids Research, 9, 309 (1981)], *Escherichia coli* JA221 [Journal of Molecular Biology, 120, 517 (1978)], *Escherichia coli* HB101 [Journal of Molecular Biology, 41, 459 (1969)], *Escherichia coli* C600 [Genetics, 39, 440 (1954)] and the like are used.

As the genus *Bacillus*, *Bacillus subtilis* MI114 [Gene, 24, 255 (1983)], *Bacillus subtilis* 207-21 [Journal of Biochemistry, 95, 87 (1984)] and the like are used.

As a bacterium having an endogenous Type I CRISPR-Cas system, a mutant strain deficient in endogenous Cas3 is desirably used.

As the yeast, *Saccharomyces cerevisiae* AH22, AH22R⁻, NA87-11A, DKD-5D, 20B-12, *Schizosaccharomyces pombe* NCYC1913, NCYC2036, *Pichia pastoris* KM71 and the like are used.
[0038]

As the insect cell when the virus is AcNPV, cells of cabbage armyworm larva-derived established line (*Spodoptera frugiperda* cell; Sf cell), MG1 cells derived from the mid-intestine of *Trichoplusia ni*, High FiveTM cells derived from an egg of *Trichoplusia ni*, *Mamestra brassicae*-derived cells, *Estigmena acrea*-derived cells and the like are used. When the virus is BmNPV, cells of *Bombyx mori*-derived established line (*Bombyx mori* N cell; BmN cell) and the like are used as insect cells. As the Sf cell, for example, Sf9 cell (ATCC CRL1711), Sf21 cell [all above, In Vivo, 13, 213-217 (1977)] and the like are used.

As the insect, for example, larva of *Bombyx mori*, *Drosophila*, cricket and the like are used [Nature, 315, 592

(1985)].

[0039]

As the animal cell, cell lines such as monkey COS-7 cell, monkey Vero cell, Chinese hamster ovary (CHO) cell, dhfr gene-
5 deficient CHO cell, mouse L cell, mouse AtT-20 cell, mouse myeloma cell, rat GH3 cell, human FL cell and the like, pluripotent stem cells such as iPS cell, ES cell and the like of human and other mammals, and primary cultured cells prepared from various tissues are used. Furthermore, zebrafish embryo,
10 *Xenopus* oocyte and the like can also be used.

[0040]

As the plant cell, suspend cultured cells, callus, protoplast, leaf segment, root segment and the like prepared from various plants (e.g., grain such as rice, wheat, corn and
15 the like, product crops such as tomato, cucumber, egg plant and the like, garden plants such as carnation, *Eustoma russellianum* and the like, experiment plants such as tobacco, *arabidopsis thaliana* and the like, and the like) are used.

[0041]

20 All the above-mentioned host cells may be haploid (monoploid), or polyploid (e.g., diploid, triploid, tetraploid and the like). In the conventional mutation introduction methods, mutation is, in principle, introduced into only one homologous chromosome to produce a hetero gene type. Therefore,
25 desired phenotype is not expressed unless dominant mutation occurs, and homozygousness inconveniently requires labor and time. In contrast, according to the present invention, since mutation can be introduced into any allele on the homologous chromosome in the genome, desired phenotype may be expressed in
30 a single generation even in the case of recessive mutation, and the problems of the conventional methods can be solved.

[0042]

An expression vector can be introduced by a known method (e.g., lysozyme method, competent method, PEG method, CaCl_2
35 coprecipitation method, electroporation method, the

microinjection method, the particle gun method, lipofection method, Agrobacterium method and the like) according to the kind of the host.

Escherichia coli can be transformed according to the
5 methods described in, for example, Proc. Natl. Acad. Sci. USA, 69, 2110 (1972), Gene, 17, 107 (1982) and the like.

The genus Bacillus can be introduced into a vector according to the methods described in, for example, Molecular & General Genetics, 168, 111 (1979) and the like.

10 A yeast can be introduced into a vector according to the methods described in, for example, Methods in Enzymology, 194, 182-187 (1991), Proc. Natl. Acad. Sci. USA, 75, 1929 (1978) and the like.

An insect cell and an insect can be introduced into a
15 vector according to the methods described in, for example, Bio/Technology, 6, 47-55 (1988) and the like.

An animal cell can be introduced into a vector according to the methods described in, for example, Cell Engineering additional volume 8, New Cell Engineering Experiment Protocol,
20 263-267 (1995) (published by Shujunsha), and Virology, 52, 456 (1973).

[0043]

A cell introduced with a vector can be cultured according to a known method according to the kind of the host.

25 For example, when Escherichia coli or genus Bacillus is cultured, a liquid medium is preferable as a medium to be used for the culture. The medium preferably contains a carbon source, nitrogen source, inorganic substance and the like necessary for the growth of the transformant. Examples of the
30 carbon source include glucose, dextrin, soluble starch, sucrose and the like; examples of the nitrogen source include inorganic or organic substances such as ammonium salts, nitrate salts, corn steep liquor, peptone, casein, meat extract, soybean cake, potato extract and the like; and examples of the inorganic
35 substance include calcium chloride, sodium dihydrogen phosphate,

magnesium chloride and the like. The medium may contain yeast extract, vitamins, growth promoting factor and the like. The pH of the medium is preferably about 5 - about 8.

As a medium for culturing *Escherichia coli*, for example, M9 medium containing glucose, casamino acid [Journal of Experiments in Molecular Genetics, 431-433, Cold Spring Harbor Laboratory, New York 1972] is preferable. Where necessary, for example, agents such as 3 β -indolylacrylic acid may be added to the medium to ensure an efficient function of a promoter. *Escherichia coli* is cultured at generally about 15 - about 43°C. Where necessary, aeration and stirring may be performed.

The genus *Bacillus* is cultured at generally about 30 - about 40°C. Where necessary, aeration and stirring may be performed.

Examples of the medium for culturing yeast include Burkholder minimum medium [Proc. Natl. Acad. Sci. USA, 77, 4505 (1980)], SD medium containing 0.5% casamino acid [Proc. Natl. Acad. Sci. USA, 81, 5330 (1984)] and the like. The pH of the medium is preferably about 5 - about 8. The culture is performed at generally about 20°C - about 35°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing an insect cell or insect, for example, Grace's Insect Medium [Nature, 195, 788 (1962)] containing an additive such as inactivated 10% bovine serum and the like as appropriate and the like are used. The pH of the medium is preferably about 6.2 - about 6.4. The culture is performed at generally about 27°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing an animal cell, for example, minimum essential medium (MEM) containing about 5 - about 20% of fetal bovine serum [Science, 122, 501 (1952)], Dulbecco's modified Eagle medium (DMEM) [Virology, 8, 396 (1959)], RPMI 1640 medium [The Journal of the American Medical Association, 199, 519 (1967)], 199 medium [Proceeding of the Society for the Biological Medicine, 73, 1 (1950)] and the like are used. The

pH of the medium is preferably about 6 - about 8. The culture is performed at generally about 30°C - about 40°C. Where necessary, aeration and stirring may be performed.

As a medium for culturing a plant cell, for example, MS medium, LS medium, B5 medium and the like are used. The pH of the medium is preferably about 5 - about 8. The culture is performed at generally about 20°C - about 30°C. Where necessary, aeration and stirring may be performed.

As mentioned above, a complex of a nucleic acid sequence-recognizing module (Type I CRISPR-Cas) and a nucleic acid base converting enzyme, i.e., nucleic acid-modifying enzyme complex, can be expressed intracellularly.

[0044]

When (a) a chimeric crRNA and (b) a complex of a Cascade constituent protein group and a nucleic acid base converting enzyme are expressed from an expression vector introduced into a cell, the protein group forms a Cascade complex and presents the chimeric crRNA. When the Cascade complex specifically recognizes and binds to a target nucleotide sequence in the double stranded DNA (e.g., genomic DNA) of interest, due to the action of the nucleic acid base converting enzyme linked to the Cascade constituent protein group, base conversion occurs in the sense strand or antisense strand of the targeted site (whole or partial target nucleotide sequence or vicinity thereof) and a mismatch occurs in the double stranded DNA (e.g., when cytidine deaminase such as PmCDA1, AID and the like is used as a nucleic acid base converting enzyme, cytosine on the sense strand or antisense strand at the targeted site is converted to uracil to cause U:G or G:U mismatch). When the mismatch is not correctly repaired, and when repaired such that a base of the opposite strand forms a pair with a base of the converted strand (T-A or A-T in the above-mentioned example), or when other nucleotide is further substituted (e.g., U→A, G) or when one to several dozen bases are deleted or inserted during repair, various mutations are introduced.

[0045]

Since conventional artificial nuclease accompanies double stranded DNA breaks (DSB), inhibition of growth and cell death assumedly caused by disordered cleavage of chromosome (off-
5 target cleavage) occur by targeting a sequence in the genome. The effect thereof is particularly fatal for many microorganisms and prokaryotes, and prevents applicability. In the present invention, mutation is introduced not by DNA cleavage but by a substituent conversion reaction (particularly
10 deamination reaction) on the DNA base, and therefore, drastic reduction of toxicity can be realized.

The modification of the double stranded DNA in the present invention does not prevent occurrence of cleavage of the double stranded DNA in a site other than the targeted site.
15 However, one of the greatest advantages of the present invention is avoidance of toxicity by off-target cleavage, which is generally applicable to any species. In preferable one embodiment, therefore, the modification of the double stranded DNA in the present invention does not accompany
20 cleavage of DNA strand not only in a targeted site of a given double stranded DNA but in a site other than same.

[0046]

The present inventors have already found that, when Type II CRIPR-Cas system is used as a nucleic acid sequence-
25 recognizing module and AID is used as a nucleic acid base converting enzyme, it is preferable to design the target nucleotide sequence such that C (or G in the opposite strand) into which a mutation is desired to be introduced is at 2 - 5 nucleotides from the 5'-end of the target nucleotide sequence.
30 In the Type II CRISPR-Cas system, the length of the targeting sequence can be appropriately determined to fall between 15 - 30 nucleotides, preferably 18 - 25 nucleotides. Since the targeting sequence is a sequence complementary to the target strand of the target nucleotide sequence, the length of the
35 target nucleotide sequence changes when the length of the

targeting sequence is changed; however, the regularity that a mutation is easily introduced into C or G at 2 - 5 nucleotides from the 5'-end irrespective of the length of the nucleotide is maintained. Therefore, by appropriately determining the length of the target nucleotide sequence (targeting sequence as a complementary strand thereof), the site of a base into which a mutation can be introduced can be shifted. As a result, restriction by PAM (NGG in the case of SpCas9) can be removed partially but not completely.

10 On the other hand, when the Type I-E CRISPR-Cas system (targeting sequence 32 nucleotides) was used as a nucleic acid sequence-recognizing module and AID was used as a nucleic acid base converting enzyme, a mutation was mainly introduced into the range of 32 - 44 bases downstream of PAM (AAG, ATG, AGG, 15 GAG in I-E subtype derived from *Escherichia coli*). As mentioned above, the PAM sequence varies according to the Type I subtype and has a high degree of freedom of 2 - 3 nucleotides (e.g., CCN in I-A subtype, CC in I-F subtype). In addition, it shows highly frequent mutation introduction in a wider range 20 than by the use of the Type II CRISPR-Cas system. While the tolerance of the length of the targeting sequence has a room for consideration, a longer target nucleotide sequence can be determined by using the I-A or I-B subtype.

By utilizing the Type I CRISPR-Cas system as a nucleic acid sequence-recognizing module in this manner, a mutation can be introduced even into a site difficult for mutation introduction in the Type II CRISPR-Cas system, and the systems can be utilized complementarily.

[0047]

30 As shown in Fig. 3, a chimeric crRNA containing multiple targeting sequences can be introduced into a host cell. The present inventors have already found that, when Type II CRISPR-Cas system is used and sequence-recognizing modules are produced corresponding to the adjacent multiple target 35 nucleotide sequences and simultaneously used, the mutation

introduction efficiency markedly increases than by using a single nucleotide sequence as a target. As the effect thereof, similarly mutation induction is realized even when both target nucleotide sequences partly overlap or when the both are apart
5 by about 600 bp. mutation introduction can occur both when the target nucleotide sequences are in the same direction (target strands are present on the same strand), and when they are opposed (target strand is present on each strand of double stranded DNA).

10 Therefore, further improvement of the mutation induction efficiency can be expected in the present invention using the Type I CRISPR-Cas system by targeting multiple nucleotide sequences.

[0048]

15 To express the nucleic acid-modifying enzyme complex of the present invention in a host cell, as mentioned above, an expression vector containing a DNA encoding the nucleic acid-modifying enzyme complex is introduced into a host cell. For efficient introduction of mutation, it is desirable to maintain
20 an expression of nucleic acid-modifying enzyme complex of a given level or above for not less than a given period. From such aspect, it is ensuring to introduce an expression vector (plasmid etc.) autonomously replicatable in a host cell. However, since the plasmid etc. are foreign DNAs, they are
25 preferably removed rapidly after successful introduction of mutation. Therefore, though subject to change depending on the kind of host cell and the like, for example, the introduced plasmid is desirably removed from the host cell after a lapse of 6 hr - 2 days from the introduction of an expression vector
30 by using various plasmid removal methods well known in the art.

Alternatively, as long as expression of a nucleic acid-modifying enzyme complex, which is sufficient for the introduction of mutation, is obtained, it is preferable to introduce mutation into the object double stranded DNA by
35 transient expression by using an expression vector without

autonomous replicatability in a host cell (e.g., vector etc. lacking replication origin that functions in host cell and/or gene encoding protein necessary for replication).

[0049]

5 Expression of target gene is suppressed while the nucleic acid-modifying enzyme complex of the present invention is expressed in a host cell to perform a nucleic acid base conversion reaction. Therefore, it was difficult to directly edit a gene essential for the survival of the host cell as a
10 target gene (due to side effects such as growth inhibition of host, unstable mutation introduction efficiency, mutation of site different from target and the like). In the present invention, direct editing of an essential gene is successfully realized by causing a nucleic acid base conversion reaction in
15 a desired stage, and transiently expressing the nucleic acid-modifying enzyme complex of the present invention in a host cell for a period necessary for fixing the modification of the targeted site. While a period necessary for a nucleic acid base conversion reaction and fixing the modification of the
20 targeted site varies depending on the kind of the host cell, culture conditions and the like, 2 - 20 generations are generally considered to be necessary. For example, when the host cell is yeast or bacterium (e.g., Escherichia coli), expression of a nucleic acid-modifying enzyme complex needs to
25 be induced for 5 - 10 generations. Those of ordinary skill in the art can appropriately determine a preferable expression induction period based on the doubling time of the host cell under culture conditions to be used. The expression induction period of the a nucleic acid encoding the nucleic acid-
30 modifying enzyme complex of the present invention may be extended beyond the above-mentioned "period necessary for fixing the modification of the targeted site" as long as the host cell is free of side effects.

As a means for transiently expressing the nucleic acid-
35 modifying enzyme complex of the present invention at a desired

stage for a desired period, a method including producing a construct (expression vector) containing a DNA encoding the nucleic acid-modifying enzyme complex, in a form capable of controlling the expression period, introducing the construct
5 into a host cell can be mentioned. The "form capable of controlling the expression period" is specifically, for example, a DNA encoding the nucleic acid-modifying enzyme complex of the present invention placed under regulation of an inducible regulatory region. While the "inducible regulatory region" is
10 not particularly limited, it is, for example, in microbial cells of bacterium (e.g., *Escherichia coli*), yeast and the like, an operon of a temperature sensitive (ts) mutation repressor and an operator regulated thereby (see Fig. 4). Examples of the ts mutation repressor include, but are not limited to, ts
15 mutation of λ phage-derived cI repressor. In the case of λ phage cI repressor (ts), it is bound to an operator to suppress expression of gene in the downstream at not more than 30°C (e.g., 28°C). At a high temperature of not less than 37°C (e.g., 42°C), it is dissociated from the operator to allow for
20 induction of gene expression. Therefore, the period when the expression of the target gene is suppressed can be minimized by culturing a host cell introduced with a DNA encoding nucleic acid-modifying enzyme complex generally at not more than 30°C, raising the temperature to not less than 37°C at an appropriate
25 stage, performing culture for a given period to carry out a nucleic acid base conversion reaction and, after introduction of mutation into the target gene, rapidly lowering the temperature to not more than 30°C. Thus, even when an essential gene for the host cell is targeted, it can be
30 efficiently edited while suppressing the side effects.

When temperature sensitive mutation is utilized, for example, a temperature sensitive mutant of a protein necessary for autonomous replication of a vector is mounted on a vector containing a DNA encoding the nucleic acid-modifying enzyme
35 complex of the present invention. As a result, autonomous

replication cannot occur rapidly after expression of the nucleic acid-modifying enzyme complex, and the vector naturally falls off along with the cell division. Examples of the temperature sensitive mutant protein include, but are not limited to, a temperature sensitive mutant of Rep101 ori necessary for the replication of pSC101 ori (see Fig. 4). Rep101 ori (ts) acts on pSC101 ori to enable autonomous replication of plasmid at not more than 30°C (e.g., 28°C), but loses function at not less than 37°C (e.g., 42°C), and plasmid cannot replicate autonomously. Therefore, a combined use with cI repressor (ts) of the above-mentioned λ phage simultaneously enables transient expression of the nucleic acid-modifying enzyme complex of the present invention, and removal of the plasmid.

15 [0050]

On the other hand, when a higher eukaryotic cell such as animal cell, insect cell, plant cell and the like is used as a host cell, a DNA encoding the nucleic acid-modifying enzyme complex of the present invention is introduced into a host cell under regulation of an inductive promoter (e.g., metallothionein promoter (induced by heavy metal ion), heat shock protein promoter (induced by heat shock), Tet-ON/Tet-OFF system promoter (induced by addition or removal of tetracycline or a derivative thereof), steroid-responsive promoter (induced by steroid hormone or a derivative thereof) etc.), the induction substance is added to the medium (or removed from the medium) at an appropriate stage to induce expression of the nucleic acid-modifying enzyme complex, culture is performed for a given period to carry out a nucleic acid base conversion reaction and, introduction of mutation into the target gene, transient expression of the nucleic acid-modifying enzyme complex can be realized by removing (or re-adding) the induction substances.

Prokaryotic cells such as *Escherichia coli* and the like can utilize an inductive promoter. Examples of the inductive

promoter include, but are not limited to, lac promoter (induced by IPTG), cspA promoter (induced by cold shock), araBAD promoter (induced by arabinose) and the like.

[0051]

5 Alternatively, the above-mentioned inductive promoter can also be utilized as a vector removal mechanism when higher eukaryotic cells such as animal cell, insect cell, plant cell and the like are used as a host cell. That is, a vector is mounted with a replication origin that functions in a host cell,
10 and a nucleic acid encoding a protein necessary for replication (e.g., SV40 ori and large T antigen, oriP and EBNA-1 etc. for animal cells), of the expression of the nucleic acid encoding the protein is regulated by the above-mentioned inductive promoter. As a result, while the vector is autonomously
15 replicatable in the presence of an induction substance, when the induction substance is removed, autonomous replication is not available, and the vector naturally falls off along with cell division (autonomous replication is not possible by the addition of tetracycline and doxycycline in Tet-OFF system
20 vector).

[0052]

The present invention is explained in the following by referring to Examples, which are not to be construed as limitative.

25 [Examples]

[0053]

(1) Production of CRISPR/Cascade-PmCDA1 expression vector

In the Cascade region derived from Escherichia coli DL21 (DE3) strain genome, an operon containing ORF of from CasA to
30 CasE was isolated by PCR, and a composite gene segment in which PmCDA1 gene was fused in the downstream of CasE with 3xFlag tag as a linker was introduced into the downstream of the pR promoter of a temperature induction type vector. As a crRNA region, pL as a promoter and 32bp targeting sequence combined
35 between 5' handle sequence and 3' handle sequence in the

downstream of the promoter (see Fig. 5) were introduced. The full-length DNA sequence of the obtained expression vector is shown in SEQ ID NO: 4, and a schematic drawing and the sequence information of the vector are shown in Fig. 4. The targeting
5 sequence is inserted into the portion of n_{32} in SEQ ID NO: 4.

[0054]

(2) Modification of *rpoB* gene using CRISPR/Cascade-PmCDA1

Nuclease *cas3* deficient *Escherichia coli* strain (JW2731) was used for a genome editing test. *Escherichia coli* competent
10 cells were produced by a conventional method and transformed with the expression vector produced in the above-mentioned (1). After recovery culture (about 2.5 hr) in SOC medium (500 μ l), the cells were diluted with a drug selection medium (LB+10 μ g/ml chloramphenicol (Cm)) (2.5 ml) and cultured at a non-
15 induction temperature of 28°C almost one night. Thereafter, the cells were diluted 20-fold with the same medium and shake cultured at 37°C for about 4 hr to induce expression. A 10-fold dilution series of the culture medium was produced, the cells were selected in a Cm-containing or not containing
20 mutation selection plate medium (LB+25 μ g/ml rifampicin (Rif)) and Rif resistant colony was obtained.

The target region in the *rpoB* gene in the obtained colony was amplified by PCR and mutation was identified by sanger sequence. The results are shown in Fig. 5-1 and 5-2. When the
25 target nucleotide sequence was set to 32 nucleotides, cytosine present in 32 - 44 bases downstream from PAM sequence was mainly converted.

[0055]

The contents disclosed in any publication cited herein,
30 including patents and patent applications, are hereby incorporated in their entireties by reference, to the extent that they have been disclosed herein.

[Industrial Applicability]

[0056]

35 According to the present invention, a site specific

mutation can be safely introduced into any species without accompanying insertion of foreign DNA or DNA double strand cleavage. Therefore, the present invention is useful for molecular breeding of useful microorganisms and application
5 thereof to gene therapy.

[0057]

This application is based on a patent application No. 2015-178023 filed in Japan (filing date: September 9, 2015), the contents of which are incorporated in full herein.

Claims

1. A method of modifying a targeted site of a double stranded DNA in a host cell, the method comprising introducing
5 (a) a DNA encoding a crRNA comprising a sequence complementary to a target strand of a target nucleotide sequence in the given double stranded DNA, and
(b) a DNA encoding a protein group constituting Cascade and a deaminase, in which the deaminase is constituted in a form
10 capable of forming a complex with Cas E in the protein group into the host cell to convert one or more nucleotides in the targeted site to other one or more nucleotides, or delete one or more nucleotides, or insert one or more nucleotides into said targeted site, without cleaving the double stranded DNA in
15 the targeted site

wherein said protein group constituting Cascade comprises CasA, CasB, CasC, CasD and CasE, and wherein the complex lacks Cas3, Cas1, Cas2 and Cas4, and wherein the method is not performed on
20 human reproductive cells, human embryos or human body.

2. The method according to claim 1, wherein said Cascade is derived from *Escherichia coli*.

25 3. The method according to claim 1 or 2, wherein said deaminase is cytidine deaminase.

4. The method according to any one of claims 1 to 3, wherein said host cell is a prokaryotic cell.

30

5. The method according to any one of claims 1 to 4, comprising a step of introducing an expression vector comprising the DNAs of said (a) and (b) in a form capable of controlling an expression period into said host cell, and inducing expression
35 of the DNAs for a period necessary for fixing the modification

of the targeted site in the double stranded DNA.

6. The method according to claim 5, wherein the target
nucleotide sequence in the double stranded DNA is present in a
5 gene essential for said host cell.

7. A nucleic acid-modifying enzyme complex for modifying a
targeted site of a double stranded DNA in a host cell, the
complex comprising

10 (a) a crRNA comprising a sequence complementary to a target
strand of a target nucleotide sequence in the given double
stranded DNA, and

(b) a nucleic acid-modifying enzyme complex comprising a
protein group constituting Cascade, and a deaminase that has
15 formed a complex with CasE in the protein group

wherein said protein group constituting Cascade comprises CasA,
CasB, CasC, CasD and CasE.

20 8. The nucleic acid-modifying enzyme complex according to claim
7, wherein the Cascade is derived from Escherichia coli.

9. The nucleic acid-modifying enzyme complex according to claim
25 7 or 8, wherein said deaminase is cytidine deaminase.

10. A DNA encoding the nucleic acid-modifying enzyme complex
according to any one of claims 7 to 9.