DESCRIPTION

Title of the Invention: GENOME SEQUENCE MODIFICATION METHOD FOR SPECIFICALLY CONVERTING NUCLEIC ACID BASES OF TARGETED DNA SEQUENCE, AND MOLECULAR COMPLEX FOR USE IN SAME

[Technical Field]

[0001]

The present invention relates to a modification method of a genome sequence, which enables modification of a nucleic acid base in a particular region of a genome, without cleaving double-stranded DNA (no cleavage or single strand cleavage), or inserting a foreign DNA fragment, and a complex of a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme to be used therefor.

[Background Art]

[0002]

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

archaebacterium, 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.

[Document List]

[Patent Documents]

[0003]

patent document 1: JP-B-4968498

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

[non-patent document]

[0004]

non-patent document 1: Kelvin M Esvelt, Harris H Wang (2013) Genome-scale engineering for systems and synthetic biology, Molecular Systems Biology 9: 641

[SUMMARY OF THE INVENTION]

[Problems to be Solved by the Invention]

[0005]

The genome editing techniques heretofore been proposed 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.

Therefore, an object of the present invention is to provide a novel method of genome editing for modifying a nucleic acid base of a particular sequence of a gene without DSB or insertion of foreign DNA fragment, i.e., by non-cleavage of a double stranded DNA or single strand cleavage, and a complex of a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme therefor.

[Means of Solving the Problems]

[0006]

The present inventors have conducted intensive studies in an attempt to solve the above-mentioned problems and took note of adopting base conversion by a conversion reaction of DNA base, without accompanying DSB. The base conversion reaction by a deamination reaction of DNA base is already known; however, targeting any site by recognizing a particular sequence of DNA, and specifically modifying the targeted DNA by base conversion of DNA bases has not been realized yet.

Therefore, deaminase that catalyzes a deamination reaction was used as an enzyme for such conversion of nucleic acid bases, and a genome sequence was modified by nucleic acid base conversion in a region containing a particular DNA sequence, by linking the enzyme and a molecule having a DNA sequence recognition ability.

To be specific, CRISPR-Cas system (CRISPR-mutant Cas) was used. That is, a DNA encoding an RNA molecule wherein genome specific CRISPR-RNA:crRNA (gRNA) containing a sequence complementary to a target sequence of a gene to be modified is linked to an RNA (trans-activating crRNA: tracrRNA) for recruiting Cas protein was produced, a DNA wherein a DNA encoding a mutant Cas protein (dCas) wherein cleavage ability of one or both chains of a double stranded DNA is inactivated and a deaminase gene are linked was produced, and these DNAs were introduced into a host yeast cell containing a gene to be modified. As a result, mutation could be introduced randomly within the range of several hundred nucleotides of the object

gene including the target sequence. When compared to the use of a double mutant Cas protein free of cleavage of any DNA chain in the double stranded DNA, the mutation introduction efficiency increased when a mutant Cas protein with cleavage of one of the chains. In addition, it was clarified that the area of mutation region and variety of mutation vary depending on which of the DNA double strand is cleaved. Furthermore, mutation could be introduced extremely efficiently by targeting multiple regions in the object gene. That is, a host cell introduced with DNA was seeded in a nonselective medium, and the sequence of the object gene was examined in randomly selected colonies. As a result, introduction of mutation was confirmed in almost all colonies. It was also confirmed that genome editing can be simultaneously performed at multiple sites by targeting certain region in two or more object genes. It was further demonstrated that the method can simultaneously introduce mutation into alleles of diploid or polyploid genomes, can introduce mutation into not only eukaryotic cells but also prokaryotic cells such as Escherichia coli, and is widely applicable irrespective of species. It was also found that editing of essential gene, which showed low efficiency heretofore, can be efficiently performed by transiently performing a nucleic acid base conversion reaction at a desired stage.

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

[0007]

Accordingly, the present invention is as described below.

[1] A method of modifying a targeted site of a double stranded DNA, comprising a step of contacting a complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a selected double stranded DNA and a nucleic acid base converting enzyme are bonded, with said double stranded DNA, 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 at least one chain of said double stranded DNA in the targeted site.

[2] The method of the above-mentioned [1], wherein the aforementioned nucleic acid sequence-recognizing module is selected from the group consisting of a CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is inactivated, a zinc finger motif, a TAL effector and a PPR motif.

[3] The method of the above-mentioned [1], wherein the aforementioned nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is inactivated.

[4] The method of any of the above-mentioned [1] - [3], which uses two or more kinds of nucleic acid sequence-recognizing modules each specifically binding to a different target nucleotide sequence.

[5] The method of the above-mentioned [4], wherein the aforementioned different target nucleotide sequence is present in a different gene.

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

[7] The method of the above-mentioned [6], wherein the aforementioned deaminase is AID (AICDA).

[8] The method of any of the above-mentioned [1] - [7], wherein the double stranded DNA is contacted with the complex by introducing a nucleic acid encoding the complex into a cell having the double stranded DNA.

[9] The method of the above-mentioned [8], wherein the aforementioned cell is a prokaryotic cell.

[10] The method of the above-mentioned [8], wherein the aforementioned cell is an eucaryote.

[11] The method of the above-mentioned [8], wherein the aforementioned cell is a cell of a microorganism.

[12] The method of the above-mentioned [8], wherein the aforementioned cell is a plant cell.

[13] The method of the above-mentioned [8], wherein the aforementioned cell is an insect cell.

[14] The method of the above-mentioned [8], wherein the aforementioned cell is an animal cell.

[15] The method of the above-mentioned [8], wherein the aforementioned cell is a cell of a vertebrata.

[16] The method of the above-mentioned [8], wherein the aforementioned cell is a mammalian cell.

[17] The method of any of the above-mentioned [9] - [16], wherein the aforementioned cell is a polyploid cell, and a site in any targeted allele on a homologous chromosome is modified.

[18] The method of any of the above-mentioned [8] - [17], comprising a step of introducing an expression vector comprising a nucleic acid encoding the aforementioned complex in a form permitting control of an expression period into the aforementioned cell, and a step of inducing expression of the nucleic acid for a period necessary for fixing the modification of the targeted site in the double stranded DNA.

[19] The method of the above-mentioned [18], wherein the target nucleotide sequence in the double stranded DNA is present in a gene essential for the aforementioned cell.

[20] A nucleic acid-modifying enzyme complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a selected double stranded DNA and a nucleic acid base converting enzyme are bonded, which converts one or more nucleotides in the targeted site to other one or more nucleotides or deletes one or more nucleotides, or inserts one or more nucleotides into said targeted site, without cleaving at least one chain of said double stranded DNA in the targeted site.

[21] A nucleic acid encoding the nucleic acid-modifying enzyme complex of the above-mentioned [20].

[Effect of the Invention]

[0008]

According to the genome editing technique of the present invention, since it does not accompany insertion of a foreign DNA or double-stranded DNA breaks, the technique is superior in safety, and has no small possibility of affording a solution to cases causing biological or legal disputes on conventional methods as relating to gene recombination. It is also theoretically possible to set a wide range of mutation introduction from a pin point of one base to several hundred bases, and the technique can also be applied to topical evolution induction by introduction of random mutation into a particular restricted region, which has been almost impossible heretofore.

[Brief Description of the Drawings]

[0009]

Fig. 1 is a schematic showing of a mechanism of the genetic modification method of the present invention using the CRISPR-Cas system.

Fig. 2 shows the results of verification, by using a budding yeast, of the effect of the genetic modification method of the present invention comprising a CRISPR-Cas system and PmCDA1 deaminase derived from Petromyzon marinus in combination.

Fig. 3 shows changes in the number of surviving cells after expression induction when a CRISPR-Cas9 system using a D10A mutant of Cas9 having a nickase activity and a deaminase, PmCDA1, are used in combination (nCas9 D10A-PmCDA1), and when conventional Cas9 having a DNA double strand cleavage ability is used.

Fig. 4 shows the results when an expression construct constructed such that human-derived AID deaminase and dCas9 are bound via SH3 domain and a binding ligand thereof is introduced into a budding yeast together with two kinds of gRNA (targeting sequences of target 4 and target 5).

Fig. 5 shows that the mutation introduction efficiency increased by the use of Cas9 that cleaves any DNA single strand.

Fig. 6 shows that the area of mutation introduction region and frequency thereof change depending on whether a double stranded DNA is cleaved or any DNA single strand is cleaved.

Fig. 7 shows that extremely high mutation introduction efficiency can be realized by targeting two adjacent regions.

Fig. 8 shows that the genetic modification method of the present invention does not require selection by marker. It was found that mutation was introduced into all colonies sequenced.

Fig. 9 shows that multiple sites in a genome can be simultaneously edited by the genetic modification method of the present invention. The upper panel shows the nucleotide sequence and amino acid sequence of the target site of each gene, and an arrow on the nucleotide sequence shows the target nucleotide sequence. The number at the arrow end or arrow head indicates the position of the target nucleotide sequence terminus on ORF. The lower panel shows the results of sequencing of the target site in each 5 clones of red (R) and white (W) colonies. In the sequences, the nucleotides indicated with white characters show occurrence of base conversion. As for responsiveness to canavanine (Can^R), R shows resistance, and S shows sensitivity.

Fig. 10 shows that a mutation can be simultaneously introduced into both alleles on the homologous chromosome of diploid genome by the genetic modification method of the present invention. Fig. 10A shows homomutation introduction efficiency of ade1 gene (upper panel) and can1 gene, and Fig. 10B shows that homomutation was actually introduced into red colony (lower panel). Also, occurrence of heteromutation in white colony is shown (upper panel).

Fig. 11 shows that genome editing of prokaryotic cell, Escherichia coli, is possible by the genetic modification method of the present invention. Fig. 11A is a schematic showing of the plasmid used. Fig. 11B shows that a mutation (CAA→TAA) could be efficiently introduced by targeting a

region in the galK gene. Fig. 11C shows the results of sequence analysis of each two clones of the colony selected from a nonselective medium (none), or a medium containing 25 μg/ml rifampicin (Rif25) and 50 μg/ml rifampicin (Rif50). Introduction of a mutation imparting rifampicin resistance was confirmed (upper panel). The emergence frequency of rifampicin resistance strain was estimated to be about 10% (lower panel).

Fig. 12 shows control of the edited base sites by the length of guide RNA. Fig. 12A is a conceptual Figure of editing base site when the length of the target nucleotide sequence is 20 bases or 24 bases. Fig. 12B shows the results of editing by targeting gsiA gene and changing the length of the target nucleotide sequence. The mutated sites are shown with bold letters, "T" and "A" show introduction of complete mutation (C→T or G→A) into the clone, "t" shows that not less than 50% of mutation (C→T) is introduced into the clone (incomplete cloning), and "c" shows that the introduction efficiency of the mutation (C→T) into the clone is less than 50%.

Fig. 13 is a schematic showing of a temperature sensitive plasmid for mutation introduction, which was used in Example 11.

Fig. 14 shows the protocol of mutation introduction in Example 11.

Fig. 15 shows the results of introduction of mutation into the rpoB gene in Example 11.

Fig. 16 shows the results of introduction of mutation into the galK gene in Example 11.

[Description of Embodiments]

[0010]

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 at least one chain of the double stranded DNA to be modified. The method characteristically contains a step of contacting a

complex wherein a nucleic acid sequence-recognizing module that specifically binds to the target nucleotide sequence in the double stranded DNA 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.

[0011]

In the present invention, the "modification" of a double stranded DNA means that a nucleotide (e.g., dC) on a DNA chain is converted to other nucleotide (e.g., dT, dA or dG), or deleted, or a nucleotide or a nucleotide sequence is inserted between certain nucleotides on a DNA chain. While the double stranded DNA to be modified is not particularly limited, it is preferably a genomic DNA. The "targeted site" of a double stranded DNA means the whole or partial "target nucleotide 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), and the range thereof can be appropriately adjusted between 1 base and several hundred bases according to the object.

[0012]

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 chain. Binding of the nucleic acid sequence-recognizing 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.

[0013]

In the present invention, the "nucleic acid base 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 chain.

[0014]

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

[0015]

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.

[0016]

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 mammal (e.g., human, swine, bovine, horse, monkey etc.) can be used. The base sequence and amino acid sequence of CDS of PmCDA1 are shown in SEQ ID NOs: 1 and 2, and the base sequence and amino acid sequence of CDS of human AID are shown in SEQ ID NOs: 3 and 4.

[0017]

A target nucleotide sequence in a double stranded DNA to be recognized by the nucleic acid sequence-recognizing module in the nucleic acid-modifying enzyme complex of the present invention is not particularly limited as long as the module specifically binds to, and may be any sequence in the double stranded DNA. The length of the target nucleotide sequence only needs to be sufficient for specific binding of the nucleic acid sequence-recognizing module. For example, when mutation is introduced into a particular site in the genomic DNA of a mammal, it is not less than 12 nucleotides, preferably not less than 15 nucleotides, more preferably not less than 17 nucleotides, according to the genome size thereof. While the upper limit of the length is not particularly limited, it is preferably not more than 25 nucleotides, more preferably not more than 22 nucleotides.

[0018]

As the nucleic acid sequence-recognizing module in the nucleic acid-modifying enzyme complex of the present invention, CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is inactivated (CRISPR-mutant Cas), zinc finger motif, TAL effector and PPR motif and the like, as well as a fragment containing a DNA binding domain of a protein that specifically binds to DNA, such as restriction enzyme, transcription factor, RNA polymerase and the like, and free of a DNA double strand cleavage ability and the like can be used, but the module is not limited thereto. Preferably, CRISPR-mutant Cas, zinc finger motif, TAL effector, PPR motif and the like can be mentioned.

[0019]

A zinc finger motif is constituted by linkage of 3 - 6 different Cys2His2 type zinc finger units (1 finger recognizes about 3 bases), and can recognize a target nucleotide sequence of 9 - 18 bases. A zinc finger motif can be produced by a known method such as Modular assembly method (Nat Biotechnol (2002) 20: 135-141), OPEN method (Mol Cell (2008) 31: 294-301), CoDA method (Nat Methods (2011) 8: 67-69), Escherichia coli one-hybrid method (Nat Biotechnol (2008) 26:695-701) and the like. The above-mentioned patent document 1 can be referred to as for the detail of the zinc finger motif production.

[0020]

A TAL effector has a module repeat structure with about 34 amino acids as a unit, and the 12th and 13th amino acid residues (called RVD) of one module determine the binding stability and base specificity. Since each module is highly independent, TAL effector specific to a target nucleotide sequence can be produced by simply connecting the module. For TAL effector, a production method utilizing an open resource (REAL method (Curr Protoc Mol Biol (2012) Chapter 12: Unit 12.15), FLASH method (Nat Biotechnol (2012) 30: 460-465), and Golden Gate method (Nucleic Acids Res (2011) 39: e82) etc.) have been established, and a TAL effector for a target nucleotide sequence can be designed comparatively conveniently. The above-mentioned patent document 2 can be referred to as for the detail of the production of TAL effector.

[0021]

PPR motif is constituted such that a particular nucleotide sequence is recognized by a continuation of PPR motifs each consisting of 35 amino acids and recognizing one nucleic acid base, and recognizes a target base only by 1, 4 and ii(-2) amino acids of each motif. Motif constitution has no dependency, and is free of interference of motifs on both sides. Therefore, like TAL effector, a PPR protein specific to the target nucleotide sequence can be produced by simply connecting PPR motifs. The above-mentioned patent document 4

can be referred to as for the detail of the production of PPR motif.

[0022]

When a fragment of restriction enzyme, transcription factor, RNA polymerase and the like is used, since the DNA binding domains of these proteins are well known, a fragment containing the domain and free of a DNA double strand cleavage ability can be easily designed and constructed.

[0023]

Any of the above-mentioned nucleic acid sequence-recognizing module can be provided as a fusion protein with the above-mentioned nucleic acid base converting enzyme, or a protein binding domain such as SH3 domain, PDZ domain, GK domain, GB domain and the like and a binding partner thereof may be fused with a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme, respectively, and provided as a protein complex via an interaction of the domain and a binding partner thereof. Alternatively, a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme may be each fused with intein, and they can be linked by ligation after protein synthesis.

[0024]

The nucleic acid-modifying enzyme complex of the present invention containing a complex (including fusion protein) wherein a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme are bonded may be contacted with a double stranded DNA as an enzyme reaction in a cell-free system. In view of the main object of the present invention, a nucleic acid encoding the complex is desirably introduced into a cell having the object double stranded DNA (e.g., genomic DNA).

Therefore, the nucleic acid sequence-recognizing module and the nucleic acid base converting enzyme are preferably prepared as a nucleic acid encoding a fusion protein thereof, or in a form capable of forming a complex in a host cell after

translation into a protein by utilizing a binding domain, intein and the like, or as a nucleic acid encoding each of them. The nucleic acid here may be a DNA or an RNA. When it is a DNA, it is preferably a double stranded DNA, and provided in the form of an expression vector disposed under regulation of a functional promoter in a host cell. When it is an RNA, it is preferably a single strand RNA.

Since the complex of the present invention wherein a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme are bonded 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 to be introduced with nucleic acid encoding nucleic acid sequence-recognizing module and/or nucleic acid base converting enzyme can encompass cells of any species, from bacterium of Escherichia 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.

[0025]

A DNA encoding a nucleic acid sequence-recognizing module such as zinc finger motif, TAL effector, PPR motif and the like can be obtained by any method mentioned above for each module. A DNA encoding a sequence-recognizing module of restriction enzyme, transcription factor, RNA polymerase and the like can be cloned by, for example, synthesizing an oligoDNA primer covering a region encoding a desired part of the protein (part containing DNA binding domain) based on the cDNA sequence information thereof, and amplifying by the RT-PCR method using, as a template, the total RNA or mRNA fraction prepared from the protein-producing cells.

A DNA encoding a nucleic acid base converting enzyme can also be cloned similarly by synthesizing an oligoDNA primer

based on the cDNA sequence information thereof, and amplifying by the RT-PCR method using, as a template, the total RNA or mRNA fraction prepared 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.

The cloned DNA 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 oraganelle transfer signal when the object double stranded DNA is mitochondria or chloroplast DNA), ligated with a DNA encoding a nucleic acid sequence-recognizing module to prepare a DNA encoding a fusion protein. Alternatively, a DNA encoding a nucleic acid sequence-recognizing module, and a DNA encoding a nucleic acid base converting enzyme may be each fused with a DNA encoding a binding domain or a binding partner thereof, or both DNAs may be fused with a DNA encoding a separation intein, whereby the nucleic acid sequence-recognizing conversion module and the nucleic acid base converting enzyme are translated in a host cell to form a complex. 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.

[0026]

A DNA encoding a nucleic acid sequence-recognizing module and a DNA encoding a nucleic acid base converting enzyme can be obtained by chemically synthesizing the DNA chain, or by connecting synthesized partly overlapping oligoDNA short chains 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]

An expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme can be produced, for example, by linking the DNA to the downstream of a promoter in a suitable expression vector.

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 (e.g., pA1-11, pXT1, pRc/CMV, pRc/RSV, pcDNAI/Neo); bacteriophages such as λphage 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.

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 constitution promoter can also be used without limitation.

For example, when the host is an animal cell, SRα 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, SRα promoter and the like are preferable.

When the host is Escherichia coli, trp promoter, lac promoter, recA promoter, λP_L 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.

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.

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

[0028]

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

[0029]

An RNA encoding a nucleic acid sequence-recognizing

module and/or a nucleic acid base converting enzyme can be prepared by, for example, transcription to mRNA in a vitro transcription system known per se by using a vector encoding DNA encoding the above-mentioned nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme as a template.

[0030]

A complex of a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme can be intracellularly expressed by introducing an expression vector containing a DNA encoding a nucleic acid sequence-recognizing module and/or 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 the yeast, Saccharomyces cerevisiae AH22, AH22R$^-$, NA87-11A, DKD-5D, 20B-12, Schizosaccharomyces pombe NCYC1913, NCYC2036, Pichia pastoris KM71 and the like are used.

[0031]

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 Five™ 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)].

[0032]

As the animal cell, cell lines such as monkey COS-7 cell, monkey Vero cell, Chinese hamster ovary (CHO) cell, dhfr gene-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, Xenopus oocyte and the like can also be used.

[0033]

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 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.

[0034]

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, 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 can be expressed in a single generation even in the case of recessive mutation (Fig. 10), which is extremely useful since the problem of the conventional method can be solved.

[0035]

An expression vector can be introduced by a known method (e.g., lysozyme method, competent method, PEG method, $CaCl_2$ 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 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.

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 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, 263-267 (1995) (published by Shujunsha), and Virology, 52, 456 (1973).

[0036]

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

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 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 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 3β-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 and a nucleic acid base converting enzyme, i.e., nucleic acid-modifying enzyme complex, can be expressed intracellularly.

[0037]

An RNA encoding a nucleic acid sequence-recognizing module and/or a nucleic acid base converting enzyme can be introduced into a host cell by microinjection method, lipofection method and the like. RNA introduction can be performed once or repeated multiple times (e.g., 2 - 5 times) at suitable intervals.

[0038]

When a complex of a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme is expressed by an expression vector or RNA molecule introduced into the cell, the nucleic acid sequence-recognizing module specifically recognizes and binds to a target nucleotide sequence in the double stranded DNA (e.g., genomic DNA) of interest and, due to the action of the nucleic acid base converting enzyme linked to

the nucleic acid sequence-recognizing module, base conversion occurs in the sense strand or antisense strand of the targeted site (whole or partial target nucleotide sequence or appropriately adjusted within several hundred bases including the 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 chain forms a pair with a base of the converted chain (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.

[0039]

As for zinc finger motif, production of many actually functionable zinc finger motifs is not easy, since production efficiency of a zinc finger that specifically binds to a target nucleotide sequence is not high and selection of a zinc finger having high binding specificity is complicated. While TAL effector and PPR motif have a high degree of freedom of target nucleic acid sequence recognition as compared to zinc finger motif, a problem remains in the efficiency since a large protein needs to be designed and constructed every time according to the target nucleotide sequence.

In contrast, since the CRISPR-Cas system recognizes the object double stranded DNA sequence by a guide RNA complementary to the target nucleotide sequence, any sequence can be targeted by simply synthesizing an oligoDNA capable of specifically forming a hybrid with the target nucleotide sequence.

Therefore, in a more preferable embodiment of the present invention, a CRISPR-Cas system wherein at least one DNA

cleavage ability of Cas is inactivated (CRISPR-mutant Cas), a zinc finger motif, a TAL effector and a PPR motif is used as a nucleic acid sequence-recognizing module.

[0040]

Fig. 1 is a schematic showing of the double stranded DNA modification method of the present invention using CRISPR-mutant Cas as a nucleic acid sequence-recognizing module.

The nucleic acid sequence-recognizing module of the present invention using CRISPR-mutant Cas is provided as a complex of an RNA molecule consisting of a guide RNA complementary to the target nucleotide sequence and tracrRNA necessary for recruiting mutant Cas protein, and a mutant Cas protein.

[0041]

The Cas protein to be used in the present invention is not particularly limited as long as it belongs to the CRISPR system, and preferred is Cas9. Examples of Cas9 include, but are not limited to, Streptococcus pyogenes-derived Cas9 (SpCas9), Streptococcus thermophilus-derived Cas9 (StCas9) and the like. Preferred is SpCas9. AS a mutant Cas to be used in the present invention, any of Cas wherein the cleavage ability of the both chains of the double stranded DNA is inactivated and one having nickase activity wherein at least one cleavage ability of one chain alone is inactivated can be used. For example, in the case of SpCas9, a D10A mutant wherein the 10th Asp residue is converted to an Ala residue and lacking cleavage ability of a chain opposite to the chain forming a complementary chain with a guide RNA, or H840A mutant wherein the 840th His residue is converted to an Ala residue and lacking cleavage ability of chain complementary to guide RNA, or a double mutant thereof can be used, and other mutant Cas can be used similarly.

A nucleic acid base converting enzyme is provided as a complex with mutant Cas by a method similar to the coupling scheme with the above-mentioned zinc finger and the like.

Alternatively, a nucleic acid base converting enzyme and mutant Cas can also be bound by utilizing RNA aptamers MS2F6, PP7 and the like and RNA scaffold by binding proteins thereto. Guide RNA forms a complementary chain with the target nucleotide sequence, mutant Cas is recruited by the tracrRNA attached and mutant Cas recognizes DNA cleavage site recognition sequence PAM (protospacer adjacent motif) (when SpCas9 is used, PAM is 3 bases of NGG (N is any base), and, theoretically, can target any position on the genome). One or both DNAs cannot be cleaved, and, due to the action of the nucleic acid base converting enzyme linked to the mutant Cas, base conversion occurs in the targeted site (appropriately adjusted within several hundred bases including whole or partial target nucleotide sequence) and a mismatch occurs in the double stranded DNA. When the mismatch is not correctly repaired, and when repaired such that a base of the opposite chain forms a pair with a base of the converted chain, or when other nucleotide is further converted or when one to several dozen bases are deleted or inserted during repair, various mutations are introduced (e.g., Fig. 2).

[0042]

Even when CRISPR-mutant Cas is used as a nucleic acid sequence-recognizing module, a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme are desirably introduced, in the form of a nucleic acid encoding same, into a cell having a double stranded DNA of interest, similar to when zinc finger and the like are used as a nucleic acid sequence-recognizing module.

A DNA encoding Cas can be cloned by a method similar to the above-mentioned method for a DNA encoding a nucleic acid base converting enzyme, from a cell producing the enzyme. A mutant Cas can be obtained by introducing a mutation to convert an amino acid residue of the part important for the DNA cleavage activity (e.g., 10th Asp residue and 840th His residue for Cas9, though not limited thereto) to other amino acid, into

a DNA encoding cloned Cas, by a site specific mutation induction method known per se.

Alternatively, a DNA encoding mutant Cas can also be constructed as a DNA showing codon usage suitable for expression in a host cell to be used, by a method similar to those mentioned above for a DNA encoding a nucleic acid sequence-recognizing module and a DNA encoding a nucleic acid base converting enzyme, and by a combination of chemical synthesis or PCR method or Gibson Assembly method. For example, CDS sequence and amino acid sequence optimized for the expression of SpCas9 in eukaryotic cells are shown in SEQ ID NOs: 5 and 6. In the sequence shown in SEQ ID NO: 5, when "A" is converted to "C" in base No. 29, a DNA encoding a D10A mutant can be obtained, and when "CA" is converted to "GC" in base No. 2518-2519, a DNA encoding an H840A mutant can be obtained.

A DNA encoding a mutant Cas and a DNA encoding a nucleic acid base converting enzyme may be linked to allow for expression as a fusion protein, or designed to be separately expressed using a binding domain, intein and the like, and form a complex in a host cell via protein-protein interaction and protein ligation.

The obtained DNA encoding a mutant Cas and/or a nucleic acid base converting enzyme can be inserted into the downstream of a promoter of an expression vector similar to the one mentioned above, according to the host.

On the other hand, a DNA encoding guide RNA and tracrRNA can be obtained by designing an oligoDNA sequence linking guide RNA sequence complementary to the target nucleotide sequence and known tracrRNA sequence (e.g., gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggc accgagtcggtggtgctttt; SEQ ID NO: 7) and chemically synthesizing using a DNA/RNA synthesizer.

While the length of the guide RNA sequence is not particularly limited as long as it can specifically bind to a

target nucleotide sequence, for example, it is 15 - 30 nucleotides, preferably 18 - 24 nucleotides.

While a DNA encoding guide RNA and tracrRNA can also be inserted into an expression vector similar to the one mentioned above, according to the host. As the promoter, pol III system promoter (e.g., SNR6, SNR52, SCR1, RPR1, U6, H1 promoter etc.) and terminator (e.g., T_6 sequence) are preferably used.

[0043]

An RNA encoding mutant Cas and/or a nucleic acid base converting enzyme can be prepared by, for example, transcription to mRNA in a vitro transcription system known per se by using a vector encoding the above-mentioned mutant Cas and/or DNA encoding a nucleic acid base converting enzyme as a template.

Guide RNA-tracrRNA can be obtained by designing an oligoDNA sequence linking a sequence complementary to the target nucleotide sequence and known tracrRNA sequence and chemically synthesizing using a DNA/RNA synthesizer.

[0044]

A DNA or RNA encoding mutant Cas and/or a nucleic acid base converting enzyme, guide RNA-tracrRNA or a DNA encoding same can be introduced into a host cell by a method similar to the above, according to the host.

[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-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 conversion reaction of the substituent on the DNA base (particularly deamination reaction), and therefore, drastic reduction of toxicity can be realized. In fact, as shown in the comparison tests using a budding yeast as a host

in the below-mentioned Examples, when Cas9 having a conventional type of DSB activity is used, the number of surviving cells decreases by induction of expression, whereas it was confirmed that the cells continued to grow and the number of surviving cells increased by the technique of the present invention using mutant Cas and a nucleic acid base converting enzyme in combination (Fig. 3).

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 (appropriately adjusted within several hundred bases including whole or partial target nucleotide sequence). 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 cleavage of DNA chain not only in a targeted site of a selected double stranded DNA but in a site other than same.

[0046]

As shown in the below-mentioned Examples, when Cas having a nickase activity capable of cleaving only one of the chains of the double stranded DNA is used as a mutant Cas (Fig. 5), the mutation introduction efficiency increases as compared to mutant Cas incapable of cleaving both chains. Therefore, for example, when a protein having a nickase activity is further linked in addition to a nucleic acid sequence-recognizing module and a nucleic acid base converting enzyme and only a DNA single strand is cleaved in the vicinity of the target nucleotide sequence, the mutation introduction efficiency can be improved while avoiding the strong toxicity of DSB.

Furthermore, a comparison of the effects of mutant Cas having two kinds of nickase activity of cleaving different chain reveals mutated sites gathering near the center of the target nucleotide sequence in one of them and various mutations

randomly introduced into region of several hundred bases from the target nucleotide sequence in the other (Fig. 6). Therefore, by selecting a chain to be cleaved by the nickase, a mutation can be introduced into a particular nucleotide or nucleotide region at a pinpoint, or various mutations can be randomly introduced into a comparatively wide range, which can be property adopted according to the object. For example, when the former technique is applied to genetically diseased iPS cell, a cell transplantation therapeutic agent with a more reduced risk of rejection can be produced by repairing mutation of the pathogenic gene in an iPS cell produced from the patients' own cell, and differentiating the cell into the object somatic cell.

[0047]

Example 7 ff. mentioned below show that a mutation can be introduced into a particular nucleotide almost at a pinpoint. For pinpoint introduction of a mutation into a desired nucleotide, the target nucleotide sequence should be set to show certain regularity of the positional relationship between a nucleotide desired to be introduced with a mutation and the target nucleotide sequence. For example, when CRIPR-Cas system is used as a nucleic acid sequence-recognizing module and AID is used as a nucleic acid base converting enzyme, the target nucleotide sequence can be designed such that C (or G in the opposite chain) into which a mutation is desired to be introduced is at 2 - 5 nucleotides from the 5'-end of the target nucleotide sequence. As mentioned above, the length of the guide RNA sequence can be appropriately determined to fall between 15 - 30 nucleotides, preferably 18 - 24 nucleotides. Since the guide RNA sequence is a sequence complementary to the target nucleotide sequence, the length of the target nucleotide sequence changes when the length of the guide RNA 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 (Fig.

12). Therefore, by appropriately determining the length of the target nucleotide sequence (guide RNA as a complementary chain thereof), the site of a base into which a mutation can be introduced can be shifted. As a result, restriction by DNA cleavage site recognition sequence PAM (NGG) can also be removed, and the degree of freedom of mutation introduction becomes higher.

[0048]

As shown in the below-mentioned Examples, when sequence-recognizing modules are produced corresponding to the adjacent multiple target nucleotide sequences, and simultaneously used, the mutation introduction efficiency drastically increases than using a single nucleotide sequence as a target (Fig. 7). As the effect thereof, similarly mutation induction is realized even when both target nucleotide sequences partly overlap or when the both are apart by about 600 bp. It can occur both when the target nucleotide sequences are in the same direction (target nucleotide sequences are present on the same chain) (Fig. 7), and when they are opposed (target nucleotide sequence is present on each chain of double stranded DNA) (Fig. 4).

[0049]

As shown in the below-mentioned Examples, the genome sequence modification method of the present invention can introduce mutation into almost all cells in which the nucleic acid-modifying enzyme complex of the present invention has been expressed, by selecting a suitable target nucleotide sequence (Fig. 8). Thus, insertion and selection of a selection marker gene, which are essential in the conventional genome editing, are not necessary. This dramatically facilitates and simplifies gene manipulation and enlarges the applicability to crop breeding and the like since a recombinant organism with foreign DNA is not produced.

Since the genome sequence modification method of the present invention shows extremely high mutation introduction efficiency, and does not require selection by markers,

modification of multiple DNA regions at completely different positions as targets can be performed (Fig. 9). Therefore, in one preferable embodiment of the present invention, two or more kinds of nucleic acid sequence-recognizing modules that specifically bind to different target nucleotide sequences (which may be present in one object gene, or two or more different object genes, which object genes may be present on the same chromosome or different chromosomes) can be used. In this case, each one of these nucleic acid sequence-recognizing modules and nucleic acid base editing enzyme form a nucleic acid-modifying enzyme complex. Here, a common nucleic acid base editing enzyme can be used. For example, when CRISPR-Cas system is used as a nucleic acid sequence-recognizing module, a common complex of a Cas protein and a nucleic acid base converting enzyme (including fusion protein) is used, and two or more kinds of chimeric RNAs of tracrRNA and each of two or more guide RNAs that respectively form a complementary chain with a different target nucleotide sequence are produced and used as guide RNA-tracrRNAs. On the other hand, when zinc finger motif, TAL effector and the like are used as nucleic acid sequence-recognizing modules, for example, a nucleic acid base converting enzyme can be fused with a nucleic acid sequence-recognizing module that specifically binds to a different target nucleotide.

[0050]

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, or an RNA encoding the nucleic acid-modifying enzyme complex is introduced into a host cell. For efficient introduction of mutation, it is desirable to maintain 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 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 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 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) or RNA.

[0051]

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 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 efficiently by causing a nucleic acid base conversion reaction in 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 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 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. For example, when a budding yeast is subjected to liquid culture in a 0.02% galactose inducer medium, the expression induction period is, for example, 20 - 40 hr. The expression induction period of the a nucleic acid encoding the nucleic acid-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-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 nucleic acid (a DNA encoding a guide RNA-tracrRNA and a DNA encoding a mutant Cas and nucleic acid base substitution enzyme in the CRISPR-Cas system) encoding the nucleic acid-modifying enzyme complex, in a form capable of controlling the expression period, introducing the construct into a host cell can be mentioned. The "form capable of controlling the expression period" is specifically, for example, a nucleic acid 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 not particularly limited, it is, for example, in microorganism 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. Examples of the ts mutation repressor include, but are not limited to, ts mutation of λphage-derived cI repressor. In the case of λphage 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 induction of gene expression (Figs. 13 and 14). Therefore, the period when the expression of the target gene is suppressed can be minimized by culturing a host cell introduced with a nucleic acid 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 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 efficiently edited while suppressing the side effects (Fig. 15).

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 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. 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 λphage simultaneously enables transient expression of the nucleic acid-modifying enzyme complex of the present invention, and removal of the plasmid.

[0052]

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.

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.

[0053]

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, 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 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

vector).

[0054]

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

[Examples]

[0055]

In the below-mentioned Examples 1 - 6, experiments were performed as follows.

<cell line, culture, transformation, and expression induction>

Budding yeast Saccharomyces cerevisiae BY4741 strain (requiring leucine and uracil) was cultured in a standard YPDA medium or SD medium with a Dropout composition meeting the auxotrophicity. The culture performed was stand culture in an agar plate or shaking culture in a liquid medium between 25°C and 30°C. Transformation was performed by an acetic acid lithium method, and selection was made in SD medium showing appropriate auxotrophicity. For expression induction by galactose, after preculture overnight in an appropriate SD medium, culture in SR medium with carbon source changed from 2% glucose to 2% raffinose overnight, and further culture in SGal medium with carbon source changed to 0.2 - 2% galactose for 3 hr to about two nights were conducted for expression induction.

For the measurement of the number of surviving cells and Can1 mutation rate, a cell suspension was appropriately diluted, applied on SD plate medium or SD-Arg+60 mg/l Canavanine plate medium or SD+300 mg/l Canavanine plate medium, applied, and the number of colonies that emerge 3 days later was counted as the number of surviving cells. Using number of surviving colonies in SD plate as the total number of cells, and the number of surviving colonies in Canavanine plate as resistant mutant strain number, the mutation rate was calculated and evaluated. The site of mutation introduction was identified by amplifying DNA fragments containing the target gene region of each strain by a colony PCR method, performing DNA sequencing, and

performing an alignment analysis based on the sequence of Saccharomyces Genome Database (http://www.yeastgenome.org/).

[0056]

<nucleic acid operation>

DNA was processed or constructed by any of PCR method, restriction enzyme treatment, ligation, Gibson Assembly method, and artificial chemical synthesis. For plasmid, as a yeast · Escherichia coli shuttle vector, pRS315 for leucine selection and pRS426 for uracil selection were used as the backbone. Plasmid was amplified by Escherichia coli line XL-10 gold or DH5α, and introduced into yeast by the acetic acid lithium method.

[0057]

<construct>

For inducible expression, budding yeast pGal1/10 (SEQ ID NO: 8) which is a bidirectional promoter induced by galactose was used. At the downstream of a promoter, a nuclear localization signal (ccc aag aag aag agg aag gtg; SEQ ID NO: 9(PKKKRV; encoding SEQ ID NO: 10)) was added to Streptococcus pyogenes-derived Cas9 gene ORF having a codon optimized for eucaryon expression (SEQ ID NO: 5) and ORF (SEQ ID NO: 1 or 3) of deaminase gene (Petromyzon marinus Petromyzon marinus-derived PmCDA1 or human-derived hAID) was ligated via a linker sequence, and expressed as a fusion protein. As a linker sequence, GS linker (ggt gga gga ggt tct; repeat of SEQ ID NO: 11 (GGGGS; encoding SEQ ID NO: 12)), Flag tag (gac tat aag gac cacgac gga gac tac aag gat cat gat att gat tac aaa gac gat gac gat aag; SEQ ID NO: 13 (DYKDHDGDYKDHDIDYKDDDDK; encoding SEQ ID NO: 14)), Strep-tag (tgg agc cac ccg cag ttc gaa aaa; SEQ ID NO: 15 (WSHPQFEK; encoding SEQ ID NO: 16)), and other domains are selected and used in combination. Here, particularly, 2xGS, SH3 domain (SEQ ID NO: 17 and 18), and Flag tag were ligated and used. As a terminator, budding yeast-derived ADH1 terminator (SEQ ID NO: 19) and Top2 terminator (SEQ ID NO: 20) were ligated. In the domain integration method, Cas9 gene ORF

was ligated to SH3 domain via 2xGS linker to give one protein, deaminase added with SH3 ligand sequence (SEQ ID NOs: 21 and 22) as another protein and they were ligated to Gal1/10 promoter on both directions and simultaneously expressed. These were incorporated into pRS315 plasmid.

In Cas9, mutation to convert the 10th aspartic acid to alanine (D10A, corresponding DNA sequence mutation a29c) and mutation to convert the 840th histidine to alanine (H840A, corresponding DNA sequence mutation ca2518gc) were introduced to remove cleavage ability of each side of DNA chain.

gRNA as a chimeric structure with tracrRNA (derived from Streptococcus pyogenes; SEQ ID NO: 7) was disposed between SNR52 promoter (SEQ ID NO: 23) and Sup4 terminator (SEQ ID NO: 24), and incorporated into pRS426 plasmid. As gRNA target base sequence, CAN1 gene ORF, 187-206 (gatacgttctctatggagga; SEQ ID NO: 25) (target 1), 786-805 (ttggagaaacccaggtgcct; SEQ ID NO: 26) (target 3), 793-812 (aacccaggtgcctggggtcc; SEQ ID NO: 27) (target 4), 563-582 (ttggccaagtcattcaattt; SEQ ID NO: 28) (target 2) and chain sequence complementary to 767-786 (ataacggaatccaactgggc; SEQ ID NO: 29) (target 5r) were used. For simultaneous expression of multiple targets, a sequence from a promoter to a terminator as one set and a plurality thereof were incorporated into the same plasmid. They were introduced into cells along with Cas9-deaminase expression plasmid, intracellularly expressed, and a complex of gRNA-tracrRNA and Cas9-deaminase was formed.

[0058]

Example 1: Modification of genome sequence by linking DNA sequence recognition ability of CRISPR-Cas to deaminase PmCDA1

To test the effect of genome sequence modification technique of the present invention by utilizing deaminase and CRISPR-Cas nucleic acid sequence recognition ability, introduction of mutation into CAN1 gene encoding canavanine transporter that acquire canavanine-resistance due to gene deficiency was tried. As gRNA, a sequence complementary to

187-206 (target 1) of CAN1 gene ORF was used, a chimeric RNA expression vector obtained by linking thereto Streptococcus pyogenes-derived tracrRNA, and a vector expressing a protein obtained by fusing dCas9 with impaired nuclease activity by introducing mutations (D10A and H840A) into Streptococcus pyogenes-derived Cas9 (SpCas9) with Petromyzon marinus-derived PmCDA1 as deaminase were constructed, introduced into the budding yeast by the acetic acid lithium method, and coexpressed. The results are shown in Fig. 2. When cultured on a canavanine-containing SD plate, only the cells subjected to introduction and expression of gRNA-tracrRNA and dCas9-PmCDA1 formed colony. The resistant colony was picked up and the sequence of CAN1 gene region was determined. As a result, it was confirmed that a mutation was introduced into the target nucleotide sequence (target 1) and the vicinity thereof.

[0059]

Example 2: Drastic reduction of side effects·toxicity

In conventional Cas9 and other artificial nucleases (ZFN, TALEN), inhibition of growth and cell death assumedly caused by disordered cleavage of chromosome occur by targeting a sequence in the genome. The effect thereof is particularly fatal for many microorganisms and prokaryotes, and prevents applicability.

Therefore, to verify the safety of the genome sequence modification technique of the present invention and cell toxicity, a comparative test with conventional CRISPR-Cas9 was performed. Using sequences (targets 3, 4) in the CAN1 gene as gRNA target, the surviving cells were counted immediately after the start of expression induction by galactose and at 6 hr after the induction based on the colony forming ability on an SD plate. The results are shown in Fig. 3. In conventional Cas9, the growth was inhibited and cell death was induced, which decreased the number of surviving cells. In contrast, by the present technique (nCas9 D10A-PmCDA1), the cells could continue to grow, and the number of surviving cells drastically increased.

[0060]

Example 3: Use of different coupling scheme

Whether mutation can be introduced into a targeted gene even when Cas9 and deaminase are not used as a fusion protein but when a nucleic acid-modifying enzyme complex is formed via a binding domain and a ligand thereof was examined. As Cas9, dCas9 used in Example 1 was used and human AID instead of PmCDA1 was used as deaminase, SH3 domain was fused with the former, and a binding ligand thereof was fused with the latter to produce various constructs shown in Fig. 4. In addition, sequences (target 4,5r) in the CAN1 gene were used as gRNA targets. These constructs were introduced into a budding yeast. As a result, even when dCas9 and deaminase were bound via the binding domain, mutation was efficiently introduced into the targeted site of the CAN1 gene (Fig. 4). The mutation introduction efficiency was remarkably improved by introducing multiple binding domains into dCas9. The main site of mutation introduction was 782nd (g782c) of ORF.

[0061]

Example 4: High efficiency and changes in mutation pattern by use of Nickase

In the same manner as in Example 1 except that D10A mutant nCas9 (D10A) that cleaves only a chain complementary to gRNA, or H840A mutant nCas9 (H840A) that cleaves only a chain opposite to a chain complementary to gRNA was used instead of dCas9, mutation was introduced into the CAN1 gene, and the sequence in the CAN1 gene region of the colony generated on a canavanine-containing SD plate was examined. It was found that the efficiency increases in the former (nCas9 (D10A)) as compared to dCas9 (Fig. 5), and the mutation gathers in the center of the target sequence (Fig. 6). Therefore, this method enables pinpoint introduction of mutation. On the other hand, it was found in the latter (nCas9 (H840A)) that multiple random mutations were introduced into a region of several hundred bases from the targeted nucleotide (Fig. 6) along with an

increase in the efficiency as compared to dCas9 (Fig. 5).

Similar remarkable introduction of mutation could be confirmed even when the target nucleotide sequence was changed. In this genome editing system using CRISPR-Cas9 system and cytidine deaminase, it was confirmed as shown in Table 1 that cytosine present within the range of about 2 - 5 bp from the 5'-side of the target nucleotide sequence (20 bp) were preferentially deaminated. Therefore, by setting the target nucleotide sequence based on this regularity and further combining with nCas9 (D10A), precise genome editing of 1 nucleotide unit is possible. On the other hand, multiple mutations can be simultaneously inserted within the range of about several hundred bp in the vicinity of the target nucleotide sequence by using nCas9 (H840A). Furthermore, site specificity may possibly be further varied by changing the coupling scheme of deaminase.

These results show that the kind of Cas9 protein can be changed property according to the object.

[0062]

Table 1

position of CAN1 gene ORF	sequence (SEQ ID NO:)	site of main mutation introduction
187 - 206 (target 1)	Gatacgttctctatggagga (25)	c191a, g226t
563 - 582 (target 2)	Ttggccaagtcattcaattt (28)	cc567at, c567del
786 - 805 (target 3) and 793 - 812 (target 4)	Ttggagaaacccaggtgcct (26) Aacccaggtgcctggggtcc (27)	cc795tt, cc796tt
767 - 786 (complementary chain) (target 5r)	Ataacggaatccaactgggc (29)	g782c

[0063]

Example 5: Efficiency increases synergistically by targeting multiple adjacent DNA sequences

Efficiency drastically increased by simultaneously using

multiple adjacent targets rather than a single target (Fig. 7). In fact, 10 - 20% of cells had canavanine-resistant mutation (targets 3, 4). In the Figure, gRNA1 and gRNA2 target target 3 and target 4, respectively. As deaminase, PmCDA1 was used. The effect thereof was confirmed to occur not only when the sequences partly overlapped (targets 3, 4), but also when they were apart by about 600 bp (targets 1, 3). The effect was found both when the DNA sequences were in the same direction (targets 3, 4) and opposing (targets 4, 5) (Fig. 4).

[0064]

Example 6: Genetic modification not requiring selection marker

As for the cells (Targets 3, 4) targeting target 3 and target 4 in Example 5, 10 colonies were randomly selected from those grown on a non-selected (canavanine-free) plate (SD plate not containing Leu and Ura) and the sequences of the CAN1 gene region were determined. As a result, mutation was introduced into the targeted site of the CAN1 gene in all examined colonies (Fig. 8). That is, editing can be expected in almost all expressed cells by selecting a suitable target sequence according to the present invention. Therefore, insertion and selection of a marker gene, which are essential for the conventional gene manipulation, are not necessary. This dramatically facilitates and simplifies gene manipulation and enlarges the applicability to crop breeding and the like since a recombinant organism with foreign DNA is not produced.

[0065]

In the following Examples, experiment techniques common in Examples 1 - 6 were performed in the same manner as above.

[0066]

Example 7: Simultaneous editing of multiple sites (different gene)

In a general gene manipulation method, mutation of one site is generally achieved by one operation due to various restrictions. Thus, whether a simultaneous mutation operation of multiple sites is possible by the method of the present

invention was tested.

Using the 3-22-position of Ade1 gene ORF of budding yeast BY4741 strain as the first target nucleotide sequence (Ade1 target 5: GTCAATTACGAAGACTGAAC; SEQ ID NO: 30), and 767-786-position (complementary chain) of Can1 gene ORF as the second target nucleotide sequence (Can1 target8 (786-767; ATAACGGAATCCAACTGGGC; SEQ ID NO: 29), both DNAs encoding chimeric RNAs of two kinds of gRNAs each containing a nucleotide sequence complementary thereto and tracrRNA (SEQ ID NO: 7) were placed on the same plasmid (pRS426), and introduced into BY4741 strain together with plasmid nCas9 D10A-PmCDA1 containing a nucleic acid encoding a fusion protein of mutant Cas9 and PmCDA1, and expressed, and introduction of mutation into the both genes was verified. The cells were cultured in an SD drop-out medium (uracil and leucine deficient; SD-UL) as a base, which maintains plasmid. The cells were appropriately diluted, applied on SD-UL and canavaine addition medium and allowed to form a colony. After 2 days of culture at 28°C, colonies were observed, and the incidence of red colony due to ade1 mutation, and the survival rate in a canavanine medium were respectively counted. The results are shown in Table 2.

[0067]

Table 2

medium	incidence of red colony	survival rate in Canavanine medium (Can)	red colony and Can survival rate
SD-UL	0.54±0.04		
+canavanine	0.64±0.14	0.51±0.15	0.31±0.04

[0068]

As a phenotype, the proportion of introduction of mutation into both Ade1 gene and Can1 gene was high and about 31%.

Then, a colony on an SD-UL medium was subjected to PCR amplification followed by sequencing. Regions containing ORF of each of Ade1, Can1 were amplified, and sequence information

of about 500 b sequences surrounding the target sequence was obtained. To be specific, 5 red colonies and 5 white colonies were analyzed to find conversion of the 5th C of Ade1 gene ORF to G in all red colonies and the 5th C to T in all white colonies (Fig. 9). While the mutation rate of the target is 100%, as the mutation rate in light of the object of gene destruction, the desired mutation rate is 50% since the 5th C needs to be changed to G to be a stop codon. Similarly, as for the Can1 gene, mutation was confirmed in the 782nd G of ORF in all clones (Fig. 9); however, since only the mutation to C affords canavanine-resistance, the desired mutation rate is 70%. In both genes, desired mutations were simultaneously obtained in 40% (4 clones out of 10 clones) by investigation, and practically high efficiency was obtained.

[0069]

Example 8: Editing of polyploid genome

Many organisms have diploid or more genome polyploidy. In the conventional mutation introduction methods, mutation is, in principle, introduced into only one homologous chromosome to produce a hetero gene type. Therefore, desired phenotype is not obtained unless the mutation is dominant mutation, and homozygousness requires labor and time. Thus, whether the technique of the present invention can introduce mutation into all target alleles on the homologous chromosome in the genome was tested.

That is, simultaneous editing of Ade1 and Can1 genes was performed in budding yeast YPH501 strain as a diploid strain. The phenotype of these gene mutations (red colony and canavanine-resistant) is a recessive phenotype, and therefore, these phenotypes do not show unless both mutations of homologous gene (homomutation) is introduced.

Using the 1173-1154-position (complementary chain) of Ade1 gene ORF (Ade1 target 1: GTCAATAGGATCCCCTTTT; SEQ ID NO: 31) or 3-22-position (Ade1 target 5: GTCAATTACGAAGACTGAAC; SEQ ID NO: 30) as the first target nucleotide sequence, and the

767-786-position (complementary chain) of Can1 gene ORF as the second target nucleotide sequence (Can1 target8: ATAACGGAATCCAACTGGGC; SEQ ID NO: 29), both DNAs encoding chimeric RNAs of two kinds of gRNAs each containing a nucleotide sequence complementary thereto and tracrRNA (SEQ ID NO: 7) were placed on the same plasmid (pRS426), and introduced into BY4741 strain together with plasmid nCas9 D10A-PmCDA1 containing a nucleic acid encoding a fusion protein of mutant Cas9 and PmCDA1, and expressed, and introduction of mutation into each gene was verified.

As a result of colony count, it was found that each characteristic of phenotype could be obtained at a high probability (40% - 70%) (Fig. 10A).

To confirm mutation, Ade1 target region of each of white colony and red colony was sequenced to confirm overlapping of sequence signals indicating heteromutation in the target site of white colony (Fig. 10B, upper panel, G and T signals overlap at ↓). Being a heteromutation, phenotype was confirmed to be absent. On the other hand, in red colony, homomutation free of overlapping signal was confirmed (Fig. 10B, lower panel, T signal at ↓).

[0070]

Example 9: Genome editing in Escherichia coli

In this Example, it is demonstrated that this technique effectively functions in Escherichia coli which is a representative bacterium model organism. Particularly, conventional nuclease type genome editing technique is fatal for bacteria, and application is difficult. Thus, the superiority of this technique is emphasized. It is shown that this technique is widely applicable to any species irrespective of prokaryon and eukaryon, besides yeast which is an eucaryote model cell.

Amino acid mutation of D10A and H840A was introduced (dCas9) into Streptococcus pyogenes Cas9 gene containing bidirectional promoter region, and a construct to be expressed

as a fusion protein with PmCDA1 via a linker sequence was constructed, and chimeric gRNA encoding a sequence complementary to the target nucleotide sequence was simultaneously mounted on a plasmid (full-length nucleotide sequence is shown in SEQ ID NO: 32, in which sequence, a sequence complementary to a target sequence is introduced into the site of n_{20}) (Fig. 11A).

First, the 426-445-position (T CAA TGG GCT AAC TAC GTT C; SEQ ID NO: 33) of Escherichia coli galK gene ORF was introduced as a target nucleotide sequence into a plasmid, various Escherichia coli strains (XL10-gold, DH5a, MG1655, BW25113) were transformed with the plasmid by calcium method or electroporation method, SOC medium was added, recovery culture was performed overnight, plasmid-maintenance cells were selected from ampicillin-containing LB medium, and colony was formed. Introduction of mutation was verified by direct-sequence from colony PCR. The results are shown in Fig. 11B.

Independent colony (1 - 3) was selected randomly, and sequence was analyzed. As a result, the 427-position C of ORF was converted to T (clones 2, 3) at a probability of not less than 60%, and occurrence of gene destruction generating a stop codon (TAA) was confirmed.

[0071]

Then, with a complementary sequence (5'-GGTCCATAAACTGAGACAGC-3'; SEQ ID NO: 34) of 1530-1549 base region of rpoB gene ORF, which is an essential gene, as a target, particular point mutation was introduced by a method similar to the above-mentioned method to try to impart rifampicin-resistant function to Escherichia coli. The sequences of colonies selected from a nonselective medium (none), a 25 μg/ml rifampicin (Rif25) and 50 μg/ml rifampicin (Rif50)-containing medium were analyzed. As a result, it was confirmed that conversion of the 1546-position G of ORF to A introduced amino acid mutation from Asp(GAC) to Asn(AAC), and rifampicin-resistance was imparted (Fig. 11C, upper panel). A

10-fold dilution series of the cell suspension after transformation treatment was spotted on a nonselective medium (none), a 25 μg/ml rifampicin (Rif25) and 50 μg/ml rifampicin (Rif50)-containing medium and cultured. As a result, it is estimated that rifampicin-resistant strain was obtained at about 10% frequency (Fig. 11C, lower panel).

As shown above, by this technique, a new function can be added by particular point mutation, rather than simple gene destruction. This technique is superior since essential gene is directly edited.

[0072]

Example 10: Adjustment of editing base site by gRNA length

Conventionally, the gRNA length relative to a target nucleotide sequence was 20b as basic, and cytosine (or guanine in opposite chain) in a site of 2 - 5b from the 5'-terminus thereof (15 - 19b upstream of PAM sequence) is used as a mutation target. Whether expression of different gRNA length can shift the site of the base to be the target was examined (Fig. 12A).

Experimental Example performed on Escherichia coli is shown in Fig. 12B. A site containing many cytosines on Escherichia coli genome was searched for, and the experiment was performed using gsiA gene which is a putative ABC-transporter. Substituted cytosine was examined by changing the length of the target to 24 bp, 22 bp, 20 bp, 18 bp to find that the 898th, 899th cytosine was substituted by thymine in the case of 20 bp (standard length). When the target site is longer than 20 bp, the 896th, 897th cytosine was also substituted, and when the target site was shorter, the 900th, 901st cytosine was also substituted. In fact, the target site could be shifted by changing the length of the gRNA.

[0073]

Example 11: Development of temperature dependent genome editing plasmid

A plasmid that induces expression of the nucleic acid-

modifying enzyme complex of the present invention under high temperature conditions was designed. While optimizing efficiency by limitatively controlling the expression state, reduction of side effects (growth inhibition of host, unstable mutation introduction efficiency, mutation of site different from target and the like) was aimed. Simultaneously, by combining a mechanism for ceasing the replication of plasmid at a high temperature, simultaneous and easy removal of plasmid after editing was intended. The detail of the experiment is shown below.

With temperature sensitive plasmid pSC101-Rep101 system (sequence of pSC101 ori is shown in SEQ ID NO: 35, and sequence of temperature sensitive Rep101 is shown in SEQ ID NO: 36) as a backbone, temperature sensitive λ repressor (cI857) system was used for expression induction. For genome editing, G113E mutation imparting RecA resistance was introduced into λ repressor, to ensure normal function even under SOS response (SEQ ID NO: 37). dCas9-PmCDA1 (SEQ ID NO: 38) was ligated to Right Operator (SEQ ID NO: 39), and gRNA (SEQ ID NO: 40) was ligated to the downstream of Left Operator (SEQ ID NO: 41) to regulate the expression (full-length nucleotide sequence of the constructed expression vector is shown in SEQ ID NO: 42). During culture at not more than 30°C, transcription and expression of gRNA, dCas9-PmCDA1 is suppressed, and the cells can grow normally. When cultured at not less than 37°C, transcription of gRNA and expression of dCas9-PmCDA1 are induced, and replication of plasmid is suppressed simultaneously. Therefore, a nucleic acid-modifying enzyme complex necessary for genome editing is transiently supplied, and plasmid can be removed easily after editing (Fig. 13).

[0074]

Specific protocol of the base substitution is shown in Fig. 14.

The culture temperature for plasmid construction is set to around 28°C, and an Escherichia coli colony retaining the

desired plasmid is first established. Then, the colony is directly used, or after plasmid extraction when the strain is changed, transformation with the object strain is performed again, and the obtained colony is subjected to liquid culture at 28°C overnight. Thereafter, the colony is diluted with the medium, induction culture is performed at 42°C for 1 hr to overnight, the cell suspension is appropriately diluted and spread or spotted on a plate to acquire a single colony.

[0075]

As a verification experiment, point mutation introduction into essential gene rpoB was performed. When rpoB which is one of the RNA polymerase-constituting factors is deleted or its function is lost, Escherichia coli dies. On the other hand, it is known that resistance to antibiotic rifampicin (Rif) is acquired when point mutation is entered at a particular site. Therefore, aiming at such introduction of point mutation, a target site is selected and assay was performed.

The results are shown in Fig. 15. In the upper left panel, the left shows an LB (chloramphenicol addition) plate, and the right shows a rifampicin-added LB (chloramphenicol addition) plate, and samples with or without chloramphenicol were prepared and cultured at 28°C or 42°C. When cultured at 28°C, the rate of Rif resistance is low; however, when cultured at 42°C, rifampicin resistance was obtained with extremely high efficiency. When the colonies (non-selection) obtained on LB were sequenced by 8 colonies, the 1546th guanine (G) was substituted by adenine (A) in not less than 60% of the strain cultured at 42°C (lower and upper left panels). It is clear that the base is also completely substituted in actual sequence spectrum (lower right panel).

[0076]

Similarly, base substitution of galK, which is one of the factors involved in the galactose metabolism, was performed. Since metabolism of 2-deoxy-galactose (2DOG), which is an analogue of galactose, by galK is fatal to Escherichia coli,

this was used as a selection method. Target site was set such that missense mutation is induced in target 8, and stop codon is entered in target 12 (Fig. 16 lower right).

The results are shown in Fig. 16. In the upper left and lower left panels, the left shows an LB (chloramphenicol addition) plate, and the right shows a 2-DOG-added LB (chloramphenicol addition) plate, and samples with or without chloramphenicol were prepared and cultured at 28°C or 42°C. In target 8, colony was produced only slightly on a 2-DOG addition plate (upper left panel), the colonies on LB (red frame) were sequenced by 3 colonies to find that the 61st cytosine (C) was substituted by thymine (T) in all colonies (upper right). This mutation is assumed to be insufficient to lose function of galK. On the other hand, in target 12, colony was obtained on 2-DOG addition plate by culture at 28°C and 42°C (lower left panel). 3 colonies on LB were sequenced to find that the 271st cytosine was substituted by thymine in all colonies (lower right). It was shown that mutation can be also introduced stably and highly efficiently in such different targets.

[0077]

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

[0078]

This application is based on patent application Nos. 2014-43348 and 2014-201859 filed in Japan (filing dates: March 5, 2014 and September 30, 2014), the contents of which are incorporated in full herein.

[Industrial Applicability]

[0079]

The present invention makes it possible to safely introduce site specific mutation into any species without accompanying insertion of a foreign DNA or double-stranded DNA breaks. It is also possible to set a wide range of mutation

introduction from a pin point of one base to several hundred bases, and the technique can also be applied to topical evolution induction by introduction of random mutation into a particular restricted region, which has been almost impossible heretofore, and is extremely useful.

Claims

1. A method of modifying a targeted site of a double stranded DNA, comprising a step of contacting a complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a given double stranded DNA and a nucleic acid base converting enzyme are bonded, with said double stranded DNA, 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 at least one strand of said double stranded DNA in the targeted site, wherein the nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is inactivated.

2. The method according to claim 1, which uses two or more kinds of nucleic acid sequence-recognizing modules each specifically binding to a different target nucleotide sequence.

3. The method according to claim 2, wherein the aforementioned different target nucleotide sequences are present in different genes.

4. The method according to any one of claims 1, 2 and 3, wherein the aforementioned nucleic acid base converting enzyme is a deaminase.

5. The method according to claim 4, wherein the aforementioned deaminase is a cytidine deaminase.

6. The method according to any one of claims 1 and 2 to 5, wherein the double stranded DNA is contacted with the complex by introducing a nucleic acid encoding the complex into a cell

having the double stranded DNA.

7. The method according to claim 6, wherein the aforementioned cell is a prokaryotic cell.

8. The method according to claim 6, wherein the aforementioned cell is an eukaryotic cell.

9. The method according to claim 6, wherein the aforementioned cell is a microbial cell.

10. The method according to claim 6, wherein the aforementioned cell is a plant cell.

11. The method according to claim 6, wherein the aforementioned cell is an insect cell.

12. The method according to claim 6, wherein the aforementioned cell is an animal cell.

13. The method according to claim 6, wherein the aforementioned cell is a vertebrate cell.

14. The method according to claim 6, wherein the aforementioned cell is a mammalian cell.

15. The method according to any one of claims 7 to 14, wherein the aforementioned cell is a polyploid cell, and all of the targeted sites in alleles on homologous chromosomes are modified.

16. The method according to any one of claims 6 to 15, comprising a step of introducing an expression vector comprising a nucleic acid encoding the aforementioned complex

in a form permitting control of an expression period into the aforementioned cell, and a step of inducing expression of the nucleic acid for a period necessary for fixing the modification of the targeted site in the double stranded DNA.

17. The method according to claim 16, wherein the target nucleotide sequence in the double stranded DNA is present in a gene essential for the aforementioned cell.

18. A nucleic acid-modifying enzyme complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a given double stranded DNA and a nucleic acid base converting enzyme are bonded and the nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is inactivated, which complex converts one or more nucleotides in the targeted site to other one or more nucleotides or deletes one or more nucleotides, or inserts one or more nucleotides into said targeted site, without cleaving at least one strand of said double stranded DNA in the targeted site.

19. A nucleic acid encoding the nucleic acid-modifying enzyme complex according to claim 18.

20. A method of modifying a targeted site of a double stranded DNA, comprising a step of contacting a complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a given double stranded DNA and a nucleic acid base converting enzyme are bonded, with said double stranded DNA, 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 one of the strands of said double stranded DNA in the targeted site,

and the nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein at least one DNA cleavage ability of Cas is deactivated, wherein the nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein only one of the two DNA cleavage abilities of Cas is inactivated.

21. The method according to claim 20, wherein the Cas is a Cas9 wherein the 10th Asp residue is converted to an Ala residue.

22. The method according to claim 20, wherein the Cas is a Cas9 wherein the 840th His residue is converted to an Ala residue.

23. The method according to any one of claims 20 to 22, wherein the aforementioned nucleic acid base converting enzyme is a deaminase.

24. The method according to claim 23, wherein the aforementioned deaminase is a cytidine deaminase.

25. A nucleic acid-modifying enzyme complex wherein a nucleic acid sequence-recognizing module that specifically binds to a target nucleotide sequence in a given double stranded DNA and a nucleic acid base converting enzyme are bonded and the nucleic acid sequence-recognizing module is a CRISPR-Cas system wherein only one of the two DNA cleavage abilities of Cas is inactivated, which complex converts one or more nucleotides in the targeted site to other one or more nucleotides or deletes one or more nucleotides, or inserts one or more nucleotides into said targeted site, without cleaving one of the strands of said double stranded DNA in the targeted site.

26. A nucleic acid encoding the nucleic acid-modifying enzyme

complex according to claim 25.