

(51) International Patent Classification:
C12N 15/66 (2006.01)(21) International Application Number:
PCT/CN2013/001045(22) International Filing Date:
6 September 2013 (06.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201210336604.4
12 September 2012 (12.09.2012) CN(71) Applicant: SHANGHAI SIDANSI BIOTECHNO-
LOGY CO., LTD [CN/CN]; Room 313, Building 1,
No.720 Cailun Rd, Zhangjiang High-Tech Park, Shanghai
(CN).(72) Inventors: ZHAO, Jinlong; Room 313, Building 1,
No.720 Cailun Rd, Zhangjiang High-Tech Park, Shanghai
202103 (CN). WU, Zhao; Room 313, Building 1, No.720
Cailun Rd, Zhangjiang High-Tech Park, Shanghai 202103
(CN).(74) Agent: SHANGHAI BRIDGE INTELLECTUAL
PROPERTY AGENCY CO., LTD.; 22A7, Harvest
Building, No.585 West Longhua Road, Shanghai 200232
(CN).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,
TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,
EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).**Declarations under Rule 4.17:**

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- of inventorship (Rule 4.17(iv))

Published:

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

(54) Title: Transcription Activator-Like Effector Assembly

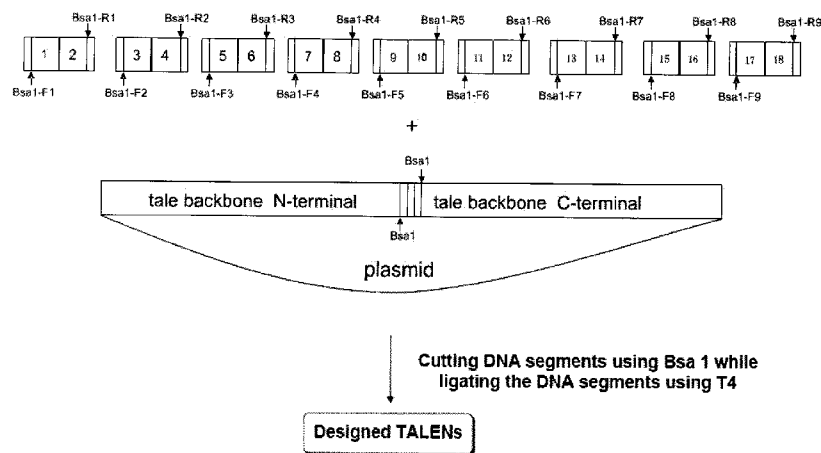


FIG. 4

(57) **Abstract:** Described herein are techniques for assembling a polynucleotide encoding a transcription activator-like effector nucleases (TALEN). The techniques ligate and digest necessary modules for a TALEN assembly in one reactor or system. Methods and Kits for generating a TALEN are also described.

SPECIFICATION

TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR ASSEMBLY

CROSS REFERENCE TO RELATED PATENT APPLICATIONS

This application claims priority to Chinese Patent Application No. No. 201210336604.4, filed on September 12, 2012, entitled "A DNA library and a method for transcription activator-like effector nuclease plasmid assembly," which is hereby incorporated by reference in its entirety.

SEQUENCE LISTING

The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification.

TECHNICAL FIELD

This disclosure relates to genome engineering. More specifically, the disclosure relates to designed transcription activator-like effector assemblies.

BACKGROUND

Target genome engineering is desirable for many scientists. By deleting or inserting a designed and specific nucleotide sequence in an endogenous genome, scientists can generate various animal models for performing fundamental biological research and studying mechanisms of disease. In addition, scientists can create transgenic animals to produce biological compositions and/or components, which may be difficult to obtain from other resources. However, it is challenging to perform targeted and specific genome modifications using traditional techniques. The traditional techniques rely on random fragment exchanges of homologous chromosomes in natural cellular processes. Therefore, the efficiency for the traditional techniques is low (e.g., 10^{-6} - 10^{-8} as a successfully rate). Because of this low

efficiency, these techniques are generally applied in mice rather than other animal models (e.g., large mammals).

In 2009, two research groups identified a transcription activator-like effector (TALE) in plant pathogen *Xanthomonas*, which modulates host gene functions by binding specific sequences within gene promoters. The TALE related techniques helped scientists develop an easier method for targeted genome engineering. This technique fuses TALE to FokI to generate a transcription activator-like effector nuclease (TALEN). In general, TALEs include tandem-like and nearly identical monomers (i.e., repeat domains), flanked by N-terminal and C-terminal sequences. Each monomer contains 34 amino acids, and the sequence of each monomer is highly conserved. Only two amino acids per repeat (i.e., residues 12th and 13th) are hypervariable, and are also known as repeat variable di-residues (RVDs). The RVDs determines the nucleotide-binding specificity of each TALE repeat domain.

TALE related techniques have increased the efficiency and usages of genome engineering, and make the genome engineering more convenient. However, assembling ten to twenty highly conserved DNA modules into a vector is a big challenge.

BRIEF DESCRIPTION OF THE DRAWINGS

The detailed description is described with reference to the accompanying figures.

FIG. 1A(1) and FIG. 1A(2) are diagrams showing an exemplary DNA library including dimer repeat modules.

FIG. 1B is a diagram showing an exemplary DNA library including monomer repeat modules.

FIG. 2 is a diagram showing an exemplary TALEN backbone plasmid.

FIG. 3 is a diagram showing another exemplary TALEN backbone plasmid.

FIG. 4 is an exemplary process showing 19 modules ligation.

FIG. 5 is an exemplary process showing a TALE assembly.

FIG. 6 is a diagram showing exemplary monomers and dimers for TALE assemblies.

FIG. 7 is a photograph of an agarose gel electrophoresis showing confirmation of assembly clones by restriction digestion analysis.

DETAILED DESCRIPTION

Various methods have been developed for assembling TALENs, such as chemical synthesis, two-step molecular cloning, and one-step molecular cloning. However, any of these methods has its own drawback. For example, although highly-repeated DNA sequences may be chemically synthesized, the cost is high and the outcome is hardly predictable. Also, two-step molecular cloning is also expensive, considering the cost of materials and sequencing, as well as time consuming. As for one-step molecular cloning, under current techniques, the maximum number of DNA modules encoding a TALEN is 14 using dimer modules. However, although natural TALEs may include 12-23 repeat modules, designed TALEs are generally more than 14 repeat modules. Therefore, to generate a TALEN including more than 14 repeat modules, current techniques require multiple steps for enzyme digestion, purifications, and ligation. This not only limits the scope of use of TALENs related genome engineering, but also affects TALENs specificity. In addition, it is a challenge to properly store intermediate products (e.g., digested DNA segments and a tail of single strand). In sum, assembling a polynucleotide encoding a TALE including more than 14 repeat domains in a single cloning reaction has not been accomplished.

Methods involving conventional molecular biology techniques are described herein. Such techniques are general known in the art unless otherwise specified in this disclosure. These techniques include PCR

amplification and detection, cell transfection, cell culture, and detection techniques.

Embodiments of this disclosure relate to a transcription activator-like effector nuclease (TALEN) assembly library, which can be used for ligation of multiple repeat DNA modules encoding TALENs. In certain embodiments, the number of the multiple repeat modules is greater than 14.

In certain related embodiments, the TAL assembly library may include 16 sets, and each set includes n dimers, wherein n is an integer. In some embodiments, the TAL assembly library may include 4 sets, and each set includes m monomers, wherein m is an integer and is not greater than n . As defined herein, a DNA module for TALE assemblies may encode a single nucleotide recognition domain, and is therefore referred as a monomer DNA module (i.e., monomer). The single nucleotide recognition domain includes two amino acids recognizing one of A, T, C, and G. In addition, a DNA module for TALE assemblies may encode a double nucleotide recognition domain, and therefore is referred as a dimer DNA module (i.e., dimer), which includes amino acids that recognize one of AA, AT, AC, AG, TT, TA, TC, TG, CC, CA, CT, CG, GG, GA, GT, and GC. In some embodiments, a set of monomers or dimers may recognize the same single nucleotide and the same pair of nucleotides respectively.

In some embodiments, each dimer or monomer may contain a 1st overhang and a 2nd overhang that are generated from digestion of type II restriction endonucleases, such as BsaI, BsmB1, BsmA1, and BbsI. In some instances, the digestion and later ligation are performed using only BsaI. In certain embodiments, a sequence of the 2nd overhang of a dimer (e.g., dimer i) may be complementary to a sequence of the 1st overhang of a dimer that is located after and adjacent to the dimer i , wherein i is an integer greater than 1 but less than n . In certain embodiments, a sequence 2nd overhang of a monomer (e.g., monomer j) may be complementary to a sequence 1st

overhang of a monomer that is located after and adjacent to the monomer j , wherein j is an integer greater than 1 but less than m .

For example, as illustrated in FIG. 5, there are four nucleotides at the overhang, from 5' to 3'. In this overhang, the 2nd to 4th nucleotides are a codon for Leu, and the first nucleotide in sense strand is the last nucleotide for a codon for Gly. In the antisense strand, the last two nucleotides are the first two nucleotides of a codon for Gly, while the first three nucleotides are complementary to a codon for Leu.

In some embodiments, dimers may be numbered as 1, ... l , ... and n , and monomers may be numbered as 1, ... j , ... and m . For example, when n is not less than 7, more than 14 modules are assembled; when n is 9 and m is 7, 19 modules are assembled.

In some embodiments, given that DNA modules are not easy for storage or self-amplification, DNA modules may be inserted into a plasmid in a circular structure for better storage and amplification.

Embodiments of this disclosure also relate to a DNA library including multiple DNA segments each corresponding to a repeat domain of a TALE. In some embodiments, each DNA segment may contain a module component and one or more fusion components fused to another DNA segment. Each DNA segment may also have cutting sites of type II restriction endonucleases. Therefore, DNA segments may be flanked by a type II restriction endonuclease to obtain DNA modules for TALE assemblies. In certain embodiments, the DNA segments may be PCR amplification products or recombinant plasmids, such as pMD18-T, TOPO® plasmids, pUC19, and pUC18.

Embodiments of this disclosure also relate to methods for transcription activator-like effector nuclease plasmid assembly. In certain embodiments, the method may include identifying target gene sequences, and designing corresponding TALENs, such as repeat domains of a TALE. Based on repeat domains, multiple DNA segments may be selected from a DNA library. In

these instances, in a single cloning reaction reactor, the multiple DNA segments, type II restriction endonucleases, DNA ligases, and TALE backbone vector (e.g., plasmids) may be mixed together to generate a polynucleotide encoding a TALEN. For example, the multiple DNA segments may be inserted into a backbone plasmid that contains a polynucleotide encoding a DNA restriction enzyme. The polynucleotides encoding TALENs may be purified by removing incomplete ligation products (e.g., linear DNA segments) using a plasmid-safe Deoxyribonuclease (DNase).

In some embodiments, individual DNA modules may be ligated to other DNA modules in an order. During ligations of a module to another module or a module to a plasmid, type II restriction endonucleases may not be able to cut additional nucleotides. In some embodiments, the multiple DNA segments (e.g., all DNA segments encoding a TALE), the backbone plasmids, the type II restriction endonucleases, and DNA ligases may be put in a single reactor to generate polynucleotides encoding TALENs, wherein digestion and ligation occur at substantially the same time. In certain embodiments, the type II restriction endonuclease may be BsaI, and the DNA ligase may be T4 ligase.

For example, a single ligation reactor or assembly reactor may include 40-200 ng plasmids, 20-200 DNA segments, 0.5-2 μ l type II restriction endonuclease, 0.5-2 μ l DNA ligase, 2 μ l DNA ligation buffer, and double-distilled water (ddH₂O) to be added to reach a final volume of 20 μ l. The ligation process may include 15 cycles: 37 °C for 5 min, 16 °C for 10 min, and followed by 80 °C for 10 min.

A polynucleotide sequence of a TALEN plasmid includes a DNA restriction enzyme, N-terminal and C-terminal may be set forth in any of SEQ ID NO. 41 and SEQ ID NO. 42. During the process, the TALEN backbone plasmid may be cut by type II restriction endonuclease to create a linear DNA segment with two overhangs. An overhang may be ligated to the 1st overhang of a

monomer *j* or dimer *i*; and the other overhang may be ligated to the 2nd overhang of the monomer *j* or dimer *i*.

In some embodiments, incomplete products may be removed using Plasmid-Safe nucleases. The incomplete linear or linearized DNA segments reduce the ligation efficiency by recombination. In some instances, before transformation of generated TALENs, Plasmid-Safe™ ATP-Dependent DNase (Epicentre, cat no: E3105K) may be used to digest linear or linearized DNA segments to increase the ligation efficiency.

In certain embodiments, a designed TALEN may include 20 repeat domains, and thus a polynucleotide encoding the designed TALEN may be generated using 20 DNA modules from a DNA library for TALEN assemblies. In certain embodiments, using appropriate primers, the DNA library for TALENs assembly may be obtained. The DNA library may contain multiple DNA modules (e.g., 172 modules). These DNA modules may be monomers each corresponding to a TALE recognition module recognizing one nucleotide, and/or dimers each corresponding to two TALE recognition modules recognizing two nucleotides. Each of the monomers and dimers contains type II restriction endonuclease cutting sites. By using this DNA library, enzyme digestion and ligation (e.g., 19-module ligation) may be performed in one reaction reactor or system, therefore avoiding purifications and additional ligation steps. This increases production efficiency, and thus improves TALE related techniques. In some embodiments, because DNA modules are plasmids or corresponding PCR products, certain risks (e.g., tail end damages and DNA degradations) are avoided. This simplifies TALEN generation procedures, and therefore reduces the cost.

In some embodiments, a polynucleotide encoding TALEN including 20 repeat domains may be assembled in a single reaction reactor or system. For example, an individual TALE repeat modules of these 20 repeat modules may identify each of 4 monomers (A, T, C, and G) or each of 16 dimers (AA, AT, AC,

AG, TA, TT, TC, TG, CA, CT, CC, CG, GA, GT, GC, and GG). Therefore, RVDs of the TALE repeat module may be NI, NG, HD, and NN if the TALE repeat module identifies one nucleotide, or NI-NI, NI-NG, NI-HD, NI-NN, NG-NI, NG-NG, NG-HD, NG-NN, HD-NI, HD-NG, HD-HD, HD-NN, NN-NI, NN-NG, NN-HD, NN-NN if the TALE repeat module identifies two nucleotides. Exemplary sequences of polynucleotides encoding the TALE repeat modules are listed in Table 1.

Table 1

Name	Sequence	SEQ ID
NI	CTGACCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTGTG CCAAGCGCACGGA	SEQ ID NO: 1
NG	CTGACCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTGT GCCAAGCGCACGGC	SEQ ID NO: 2
HD	TTGACCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAGG AAAGCAAGCCTTGGAAACAGTACAGAGGCTGTTGCCTGTGCTGT GCCAAGCGCACGGG	SEQ ID NO: 3
NN	CTTACCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG AAAACAGGCTTTGGAAACGGTGCAAGGCTCCTTCCAGTGCTGT GCCAAGCGCACGGG	SEQ ID NO: 4
NI-NI	CTGACCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCTC CAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTGC TCCCGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 5
NI-NG	CTGACCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG CTGCCGGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 6
NI-HD	CTGACCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA GCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGCT GTTGCCTGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 7
NI-NN	CTGACCCAGAGCAGGTCGTGGCAATCGCCTCCAACATTGGCGG GAAACAGGCACTCGAGACTGTCCAGCGCCTGCTTCCCGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA GCAATAACGGCGGAAACAGGCTTTGGAAACGGTGCAAGGCT CCTTCCAGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 8

NG-NI	CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCTC CAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTGC TTCCCGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 9
NG-NG	CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG CTGCCGGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 10
NG-HD	CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA GCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGCT GTTGCCTGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 11
NG-NN	CTGACCCCAGAGCAGGTCGTGGCCATTGCCTCGAATGGAGGGGG CAAACAGGCGTTGGAAACCGTACAACGATTGCTGCCGGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA GCAATAACGGCGGAAACAGGCTTTGGAAACGGTGCGAGAGGCT CCTTCCAGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 12
HD-NI	CTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAG GAAAGCAAGCCTTGGAAACAGTACAGAGGCTGTTGCCTGTGCTTT GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCT CCAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTG CTTCCCGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 13
HD-NG	CTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAG GAAAGCAAGCCTTGGAAACAGTACAGAGGCTGTTGCCTGTGCTTT GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCT CGAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATT GCTGCCGGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 14
HD-HD	CTGACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAG GAAAGCAAGCCTTGGAAACAGTACAGAGGCTGTTGCCTGTGCTTT GTCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCA AGCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGC TGTTGCCTGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 15
HD-NN	CTACCCCAGAGCAGGTCGTGGCGATCGCAAGCCACGACGGAGG AAAGCAAGCCTTGGAAACAGTACAGAGGCTGTTGCCTGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA GCAATAACGGCGGAAACAGGCTTTGGAAACGGTGCGAGAGGCT CCTTCCAGTGCTGTGCCAAGCGCACGGA	SEQ ID NO: 16
NN-NI	CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG AAAACAGGCTTTGGAAACGGTGCGAGAGGCTCCTTCCAGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCCTC CAACATTGGCGGGAAACAGGCACTCGAGACTGTCCAGCGCCTGC TTCCCGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 17
NN-NG	CTGACCCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG AAAACAGGCTTTGGAAACGGTGCGAGAGGCTCCTTCCAGTGCTTTG	SEQ ID NO: 18

	TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCCATTGCCTC GAATGGAGGGGGCAAACAGGCGTTGGAAACCGTACAACGATTG CTGCCGGTGCTGTGCCAAGCGCACGGT	
NN-HD	CTGACCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG AAAACAGGCTTTGGAAACGGTGCAGAGGCTCCTTCCAGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCGATCGCAA GCCACGACGGAGGAAAGCAAGCCTTGGAAACAGTACAGAGGCT GTTGCCTGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 19
NN-NN	CTGACCCAGAGCAGGTCGTGGCAATCGCGAGCAATAACGGCGG AAAACAGGCTTTGGAAACGGTGCAGAGGCTCCTTCCAGTGCTTTG TCAGGCACACGGCCTCACTCCGGAACAAGTGGTCGCAATCGCGA GCAATAACGGCGGAAAACAGGCTTTGGAAACGGTGCAGAGGCT CCTTCCAGTGCTGTGCCAAGCGCACGGT	SEQ ID NO: 20

Examples

A DNA library including 172 DNA segments was established by modifying the TALE repeat modules described above. PCR amplification was applied to add restriction enzyme cutting sites and adaptors. For dimers, PCR was performed using T-vectors containing 16 dimers and primer pairs including F1 and R1, F2 and R2, F3 and R3, F4 and R4, F5 and R5, F6 and R6, F7 and R7, F8 and R8, as well as F9 and R9. There were 144 (i.e., 16 X 9) PCR products. For monomers, PCR was performed using T-vectors containing 4 monomers and primer pairs including F1 and R1, F2 and R2, F3 and R3, F4 and R4, F5 and R5, F6 and R6, as well as F7 and R7. There were 28 (i.e., 4 X 7) PCR products. Thus, the DNA library includes 172 PCT products (i.e., 144 plus 28). Exemplary sequences of primer pairs F1 and R1, F2 and R2, F3 and R3, F4 and R4, F5 and R5, F6 and R6, F7 and R7, F8 and R8, as well as F9 and R9 may be listed in Table 2, and lower case letter indicates BsaI cutting sites.

Table 2

Name	Sequence	SEQ ID
TALE-F1	AATGGACGACCCGGCTTGATAggtctc CTG ACCCAGAGCAG GTCGTGG	SEQ ID NO: 21
TALE-R1	CATCACAGGTAGCTCGCTGGAgtctc TAA ACCGTGCCTTG GCAC	SEQ ID NO: 22
TALE-F2	ATCGATCGATCGCGATCGATCggtctc GTTT AACCCAGAGCA GGTCGTG	SEQ ID NO: 23

TALE-R2	GCAGCCACGGCTAGCTTAAGCggtctc TGAGG CCGTGCGCTTG GCAC	SEQ.ID NO: 24
TALE-F3	ATCGATCGATCGCGATCGATCggtctc GCTC ACCCAGAGCA GGTCGTG	SEQ.ID NO: 25
TALE-R3	GAACCGCCGTCTTACGTAGAGggtctc TAAT CCGTGCGCTTG GCAC	SEQ.ID NO: 26
TALE-F4	TTTAGCCCGTACCGTAGCCTAggtctc GATTA ACCCAGAGCA GGTCGTG	SEQ.ID NO: 27
TALE-R4	TTGCACCGGTATCGTCGAGGcgtctc TAAGT CCGTGCGCTTG GCAC	SEQ.ID NO: 28
TALE-F5	AAGCATGGATCGCAAGGGTTggtctc GACTT ACCCAGAGC AGGTCGTG	SEQ.ID NO: 29
TALE-R5	GGGTTGCGCTCGCAATTACCGggtctc TAAGG CCGTGCGCTTG GCAC	SEQ.ID NO: 30
TALE-F6	CGAAATCCGACCGGATGCCTAggtctc GCTT ACCCAGAGCA GGTCGTG	SEQ.ID NO: 31
TALE-R6	GCCATCGCGTCGCACGAAGCTggtctc TAGT CCGTGCGCTTG GCAC	SEQ.ID NO: 32
TALE-F7	ATAGCTGGTAGGGCTACGGGcgtctc GACTA ACCCAGAGC AGGTCGTG	SEQ.ID NO: 33
TALE-R7	GAACGACCCTGACAGTCGTTggtctc TGAGC CGTGCGCTTG GCAC	SEQ.ID NO: 34
TALE-F8	CGATATCGATCGCCTTACGCggtctc GGCTC ACCCAGAGCAG GTCGTG	SEQ.ID NO: 35
TALE-R8	CGCCACATATATAGCGCGTCCggtctc TAGC CGTGCGCTTGG CAC	SEQ.ID NO: 36
TALE-F9	GTGTGACGGCTAGCCTAGTAgtctc GGCTA ACCCAGAGCA GGTCGTG	SEQ.ID NO: 37
TALE-R9	GCTTGCGGATCGATAGCATGGggtctc TGAGT CCGTGCGCTTG GCAC	SEQ.ID NO: 38

Regarding the PCR, approximately 1 μ l Plasmid was mixed with a solution containing 0.2 μ l Primers (0.1 μ l for each of the primer pair), 1.5 μ l Buffer, 0.8 μ l dNTP, 0.35 μ l MgSO₄, 11.48 μ l ddH₂O, and 1 Unit DNA Polymerase. The following PCR reaction was used: 36 cycles 95 °C for 2 min, 95 °C for 15 sec, 55.8 °C for 30 sec, 68 °C for 30 sec, 68 °C for 2 sec, and followed by 68 °C for 1 min.

All 18 primers contain a BsaI cutting site: GGTCTCN'NNNN, wherein N represents a nucleotide. BsaI belongs to type II restriction endonuclease, and one cutting site can generate various overhangs. Using type II restriction

endonuclease, 24 fusion sites were generated with respect to 4 codons for Gly and 6 codons for Leu. In addition, 10 of those 24 were selected for primer designs. Except for F1 and R9, Fk can specifically ligate to Rk-1, but not other primers, wherein k is an integer between 3 and 9.

The 172 PCR products were purified by gel extraction, ligated and inserted into pMD18-T plasmids. The following ligation of 20 original modules into pMD18-T (from Takara) was used. First, 2.7 μ l PCR products was mixed with a solution containing 3 μ l solution 1 and 0.3 μ l pMD18-T. Then, the mixture was incubated at 16 °C for 2 hours, transformed into DH5a, and stroke onto LB plates containing kanamycin. Colonies were selected, and plasmids were isolated. The PCR products were verified by PCR and sequencing. Finally, a plasmid library containing 172 plasmids were established, as illustrated in FIG. 1.

A PCR product library was generated using assem-F and assem-R as primers (e.g., sequences in Table 3) and plasmids of the 172 plasmid library as PCR templates. The binding sites of primers are 400 bp upstream and downstream of polynucleotides encoding individual TALE repeat modules. In addition, the PCR products for dimers are about 1050 bp and for monomers are about 950 bp.

Table 3

Name	Sequence	SEQ ID
assem-F	TGTTGTGTGGAATTGTGAGCGGATAAC	SEQ ID NO:39
assem-R	TGCAAGGCGATTAAGTTGGGTAACG	SEQ ID NO:40

For PCR amplification (50 μ l), 0.5 μ l DNA template (about 50 ng) was mixed with a solution containing 0.3 μ l (50 μ M) for each primer, 0.25 μ l pfx polymerase (Invitrogen), 5 μ l 10 $\times$ buffer, 2.5 μ l dNTP (2.5 μ M), 1 μ l MgSO₄, 40.15 μ l ddH₂O. The following PCR amplification program was used: 36 cycles 95 °C for 2 min, 95 °C for 15 sec, 68 °C for 30 sec, 68 °C for 50 sec, and followed by 68 °C for 5 min.

The PCR products were purified using DNA purification kits (Taingen), and measured concentrations by agar gel electrophoresis. Enzyme digestion sites of two TALEN plasmids: pEF1a-NLS-TALE backbone-Fok1(R)-pA and pEF1a-NLS-TALE backbone-Fok1(L)-IRES-PURO-pA, were illustrated in FIGS. 2 and 3 respectively. The sequences are shown as SEQ ID NO: 41 and SEQ ID NO: 42. Sequences of N-terminal and C-terminal of transcription activator-like effectors are shown as SEQ ID NO: 43 and SEQ ID NO: 44. Before ligation, BsaI was added to digest TALEN vectors to obtain overhangs for repeats modules. Digested TALEN vectors were purified by gel extraction, and concentrations were determined by gel electrophoresis.

With respect to TALEN ligation, except? for F1 and R9 (F1 ligates to left end of TALEN vector, R9 ligates to right end of backbone vector), Fk can ligate to Rk-1 at overhangs, but not to others. After ligations, BsaI is not able to break modules and backbone vectors.

FIG. 4 illustrates a process for assembling a TALEN containing 19 repeat segments. As illustrated, the last half DNA segment encoding a module recognizing T is in the backbone vector already; thus the ligation of 18 modules is enough. 9 DNA segments may be selected based on target sequences, and mixed with a solution containing TALEN backbone vector, BsaI, and T4 ligase to digest and ligate in the same reactor or system.

The following assembly system was used: 150 ng vector, 50 ng each DNA segment, 1 μ l BsaI (NEB), 1 μ l T4 ligase (Fermentas), 2 μ l T4 Buffer (NEB), and double-distilled water (ddH₂O) to make to final 20 μ l. The following ligation program was used: 15 cycles 37 °C for 5 min, 16 °C for 10 min, and followed by 80 °C for 10 min.

If occasional incomplete ligation happens (e.g., only 1 to 8 modules are ligated), this incomplete ligation may slow down the ligation efficiency by recombination. Thus, before transformation, a Plasmid-Safe™ ATP-Dependent DNase (Epicentre, cat no: E3105K) may be used to digest the linear plasmids.

To remove the linear plasmids, 1 μ l plasmid-safe DNases and 0.5 μ l ATP were added into a 20 μ l ligation system for an additional incubation at 37 °C for 1 hour. 10 μ l of ligation products were taken to transform Trans-T1 competent cells. Colonies were selected to obtain isolated vectors. Restriction analysis was performed by using BamH1/Pst1. The expected size of smaller fragment is the length of ligated size plus 550 bp. The final precuts were sent for sequencing. Exemplary sequencing primers are listed in table 4.

Table 4

Name	Sequence	SEQ ID
Sequence-F	CTCCCCTTCAGCTGGACAC	SEQ ID NO:45
Sequence-R	AGCTGGGCCACGATTGAC	SEQ ID NO:46

Embodiments of this disclosure allow obtaining sequence-confirmed TALEN vectors within 3 days. For example, the ligation (4.5 hours), plasmid-safe DNase digestion (1 hour), and transformation (1 hour) may be performed in the first day. Colonies selection and bacterial inculcation may be performed in day 2. Finally, the sequence analysis results may be received in day 3. If the target sequence is 12-18 but not 19, the modules located in the front part can be changed from dimers into monomers, and thus the change of dimer to monomer can reduce a module. Exemplary options for different monomers or dimers specific to the targeting nucleotide(s) are shown in picture 6.

In some embodiments, polynucleotides encoding TALENs for targeting certain sequences may be assembled in a single reaction. Examples of the sequences may be found in table 5.

Table 5

Name	Sequence	SEQ ID
Sequence 1	CGCGCGCGCGCGCGCGT	SEQ ID NO:47
Sequence 2	CCCACTCCCATCCAGT	SEQ ID NO:48

In these instances, DNA segments encoding repeat modules were selected from the PCR library. For example, for sequence 1, DNA segments

corresponding to CG-1, CG-2, CG-3, CG-4, CG-5, CG-6, CG-7, CG-8, and CG-9 were chosen, and TALEN vectors containing pEF1a-NLS-TALE backbone-Fok1(R)-pA were used. For sequence 2, DNA segments corresponding to C-1, A-2, C-3, TC-4, CC-5, CA-6, TC-7, CA-8, and GT-9 were chosen, and TALEN vectors containing pEF1a-NLS-TALE backbone-Fok1 (L)-IRES-PURO-pA were used. The following assembly system was used: 150 ng vector, 50ng each modules, 1 μ l BsaI (NEB), 1 μ l T4 Ligase (fermentas), 2 μ l T4 Buffer (NEB), and H₂O to make the system solution to final 20 μ l. The following Ligation program was used for 15 cycles: 37 °C for 5 min, and 16 °C for 10 min, and followed by 80 °C for 10 min.

The ligation products were purified using plasmid-safe DNases for 1 hour. The products (plasmids) were then transformed into Trans-T1 chemically competent cells. The plasmids were isolated and analyzed by BamH1\EcoR1 restriction digestion and gel electrophoresis. FIG. 7 is a photograph of an agarose gel electrophoresis showing confirmation of assembly clones by restriction digestion analysis. As illustrated, enzyme digestion bands include: 1 kb DNA marker in the middle lane, ligation I indicating 3.1 kb and 2.2 kb on the right of the DNA marker, and ligation II indicating 4.2 kb and 3.7 kb on the left of the DNA marker. Cloning efficiency for assembling TALENs containing recognition domains to identify Sequence 1 and Sequence 2 in a single reaction is 70% and 80% respectively.

EXEMPLARY EMBODIMENTS

1. A transcription activator-like effector (TALE) assembly library comprising:

16 sets of dimer DNA modules, each set of the 16 sets including n number of dimer DNA modules and each dimer DNA modules encoding two recognition domains of a designed TALE; and

4 sets of monomer DNA modules, each set of the 4 sets including m number of monomer DNA modules and each monomer DNA modules encoding a recognition domain of the designed TALE, wherein:

each dimer DNA module or monomer DNA module includes a first overhang and a second overhang that are generated using a type II restriction endonuclease,

a sequence of a first overhang of a dimer DNA module of the TALE assembly library is complementary to a sequence of a second overhang of a particular dimer DNA module that is adjacent and before the dimer DNA module,

a sequence of a first overhang of a monomer DNA module of the TALE assembly library is complementary to a sequence of a second overhang of a particular monomer DNA module that is adjacent and before the monomer DNA module,

m and n are integers, and

m is not greater than n.

2. The TALE assembly library of embodiment 1, wherein the type II restriction endonuclease is Bsa I.

3. The TALE assembly library of embodiment 1, wherein the n is either 6 or 7.

4. The TALE assembly library of embodiment 1, wherein the n is not less than 7.

5. The TALE assembly library of embodiment 1, wherein the n is 9, and m is 7.

6. The TALE assembly library of embodiment 1, wherein the TALE assembly library is a DNA library, the dimer DNA module includes a DNA segment encoding two recognition domains of a designed TALE, the monomer DNA module includes a DNA segment encoding a recognition domain of a

designed TALE, and the DNA segment is flanked by a type II restriction endonuclease.

7. The TALE assembly library of embodiment 1, wherein the DNA segment is a overlap extension PCR product or a plasmid.

8. The TALE assembly library of embodiment 1, wherein the plasmid is pMD18-T.

9. A method for assembling a transcription activator-like effector nuclease (TALEN), the method comprising:

selecting, from a DNA library, multiple DNA fragments corresponding to recognition domain of the TALEN;

in a same reactor or system, mixing, at a substantially same time, the multiple DNA fragments, type II restriction endonucleases, DNA ligases, and a TALEN backbone plasmid including sequences encoding a N-terminal and a C-terminal of the TALEN to generate a polynucleotide encoding the TALEN; and

purifying the TALEN by removing linear DNA segments.

10. The method of embodiment 9, wherein a sequence of the plasmid include SEQ ID NO. 41 or 42.

11. The method of embodiment 9, wherein the reactor or system includes a 20 μ l solution including:

40-200 ng plasmids;

20-200 units for each module;

0.5-2 μ l type II restriction endonuclease;

0.5-2 μ l DNA ligase;

2 μ l DNA ligation buffer; and

a certain amount of ddH₂O.

CLAIMS

What is claimed is:

1. A method for assembling a polynucleotide encoding a transcription activator-like effector (TALE), the method comprising:

generating multiple Deoxyribonucleic acid (DNA) segments, an individual DNA segment of the multiple DNA segments corresponding to a repeat sequence of the TALE; and

mixing the multiple DNA segments with restriction enzymes, DNA ligases, and a vector to generate the polynucleotide.

2. The method of claim 1, wherein the multiple DNA segments are polymerase chain reaction (PCR) amplification products, and a number of the multiple DNA segments is greater than 14.

3. The method of claim 1, wherein the number of multiple DNA segments is greater than 18.

4. The method of claim 1, wherein the mixing the multiple DNA segments with restriction enzymes, the DNA ligases, and the vector comprises mixing the multiple DNA segments with the restriction enzymes, the DNA ligases, and the vector at a substantially same time.

5. The method of claim 1, further comprising purifying the polypeptide using a plasmid-safe Deoxyribonuclease (DNase).

6. The method of claim 1, wherein the mixing the multiple DNA segments with the restriction enzymes, DNA ligases, and the vector comprises mixing the multiple DNA segments with the restriction enzymes, DNA ligases, and the vector under a condition including multiple temperature cyclings

7. The method of claim 5, wherein the multiple temperature cyclings is followed by inactivation of the restriction enzymes.

8. The method of claim 1, wherein the mixing the multiple DNA segments with the restriction enzymes, DNA ligases, and the vector comprises mixing the multiple DNA segments with the restriction enzymes that belongs

to a type of type II restriction enzymes, DNA ligases, and the vector without other types of the type II restriction enzymes.

9. The method of claim 1, wherein the individual DNA segment of the multiple DNA segments has two cleavage sites that are generated using type II restriction enzymes, and the restriction enzymes are the type II restriction enzymes.

10. The method of claim 1, wherein a repeat sequence of the multiple repeat sequences identifies a single nucleotide of a particular polynucleotide or two contiguous nucleotides of a particular polynucleotide.

11. The method of claim 1, wherein the multiple DNA segments are assembled in one cloning reactor.

12. The method of claim 1, wherein the multiple DNA segments are assembled in a single cloning reaction.

13. A kit for generating polynucleotides encoding transcription activator-like effector nucleases (TALENs), the kit comprising:

multiple sets of dimer DNA segments, an individual set of dimer DNA segments including multiple dimer DNA segments, an individual dimer DNA segment encoding two recognition domains of a designed TALE; and

multiple sets of monomer DNA segments, an individual set of monomer DNA segments including multiple monomer DNA modules, an individual monomer DNA module encoding a recognition domain of the designed TALE,

wherein the individual dimer DNA segments and the individual monomer DNA segments include a first overhang and a second overhang that are generated using a type II restriction endonuclease.

14. The Kit of claim 13, wherein a sequence of a first overhang of a dimer DNA segment of multiple dimer DNA segments is complementary to a sequence of a second overhang of a particular dimer DNA segment that is ligated to the dimer DNA segment.

15. The Kit of claim 13, wherein a sequence of a first overhang of a monomer DNA segment of multiple monomer DNA segments is complementary to a sequence of a second overhang of a particular monomer DNA segment that is ligated to the monomer DNA segment.

16. The Kit of claim 13, wherein the type II restriction endonuclease is Bsa I.

17. The Kit of claim 13, wherein a number of the multiple sets of dimer DNA segments is from 8 to 10, and a number of the multiple sets of monomer DNA segments is from 6 to 8.

18. The Kit of claim 13, wherein an individual set of dimer DNA segments includes 15 to 17 dimer DNA segments, and an individual set of monomer DNA segments include 3 to 5 monomer DNA segments.

19. The Kit of claim 13, wherein the multiple dimer DNA segments and the multiple monomer DNA segments include an overlap extension polymerase chain reaction (PCR) product or a plasmid.

20. The Kit of claim 13, wherein the multiple dimer DNA segments and the multiple monomer DNA segments include includes pMD18-T.

DRAWINGS

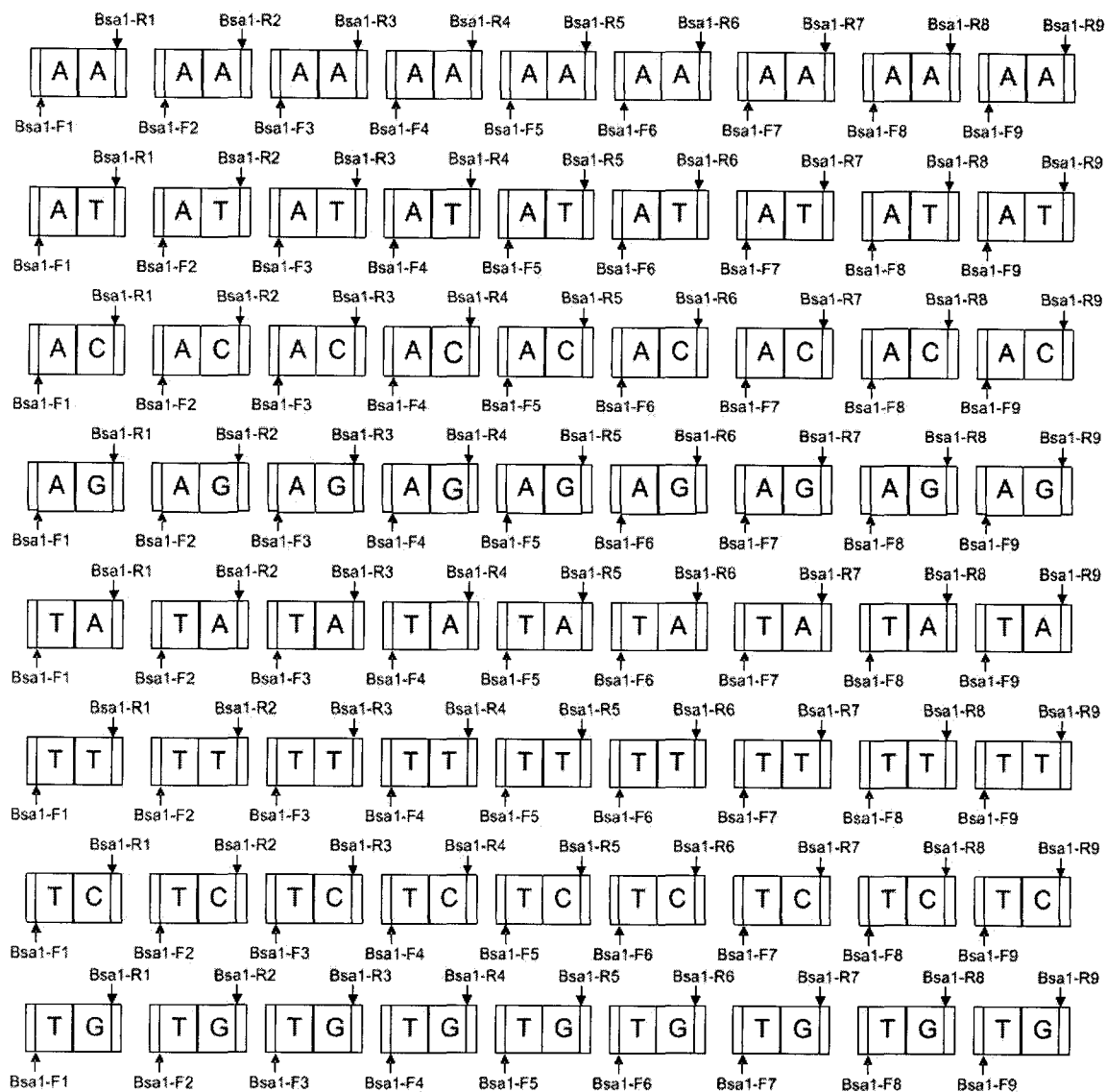


FIG. 1 A (1)

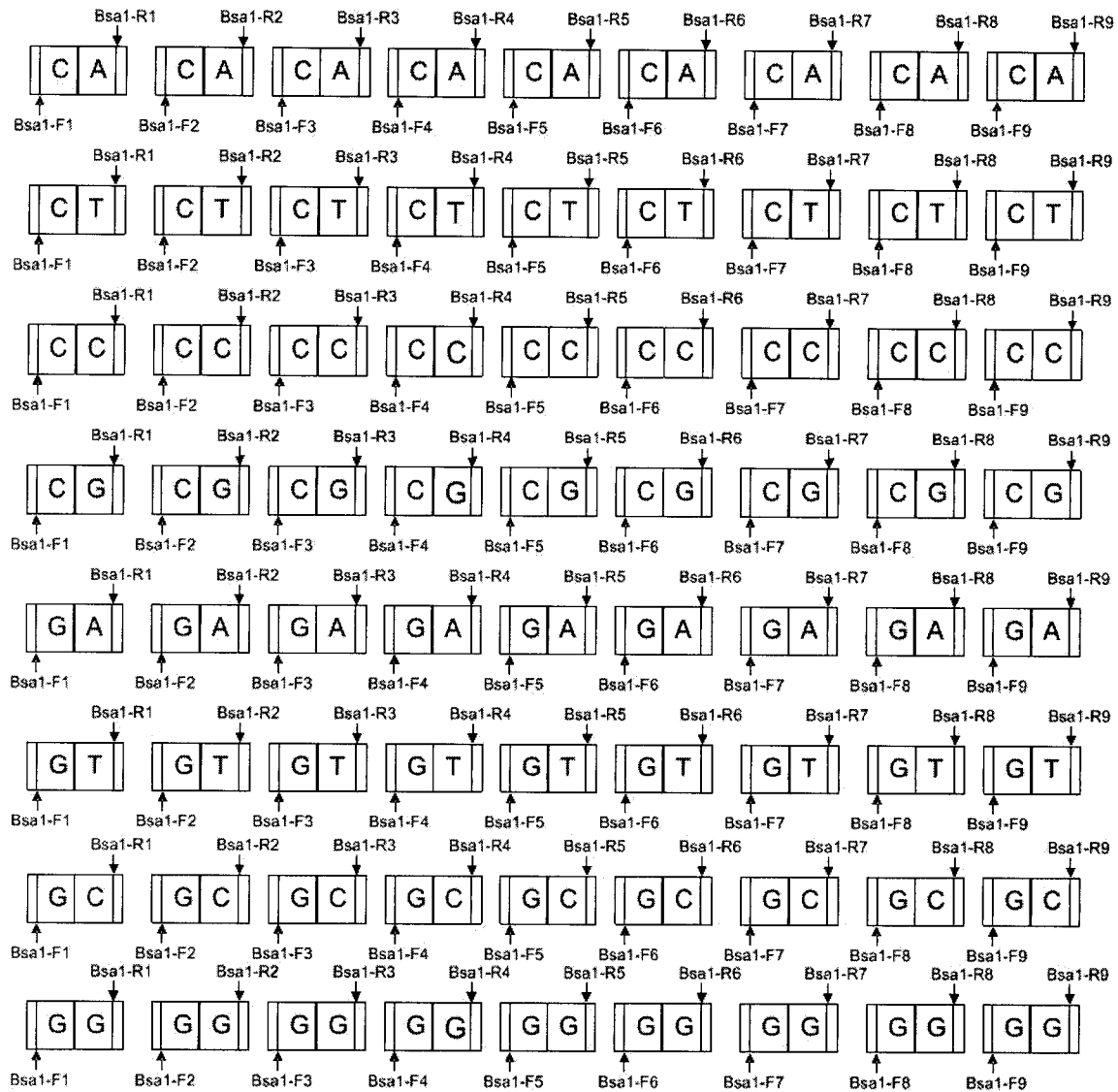


FIG. 1 A(2)

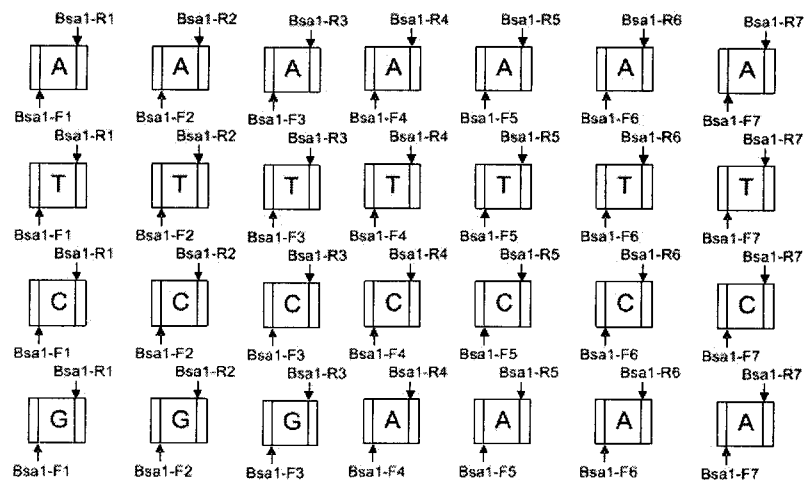


FIG. 1 (B)

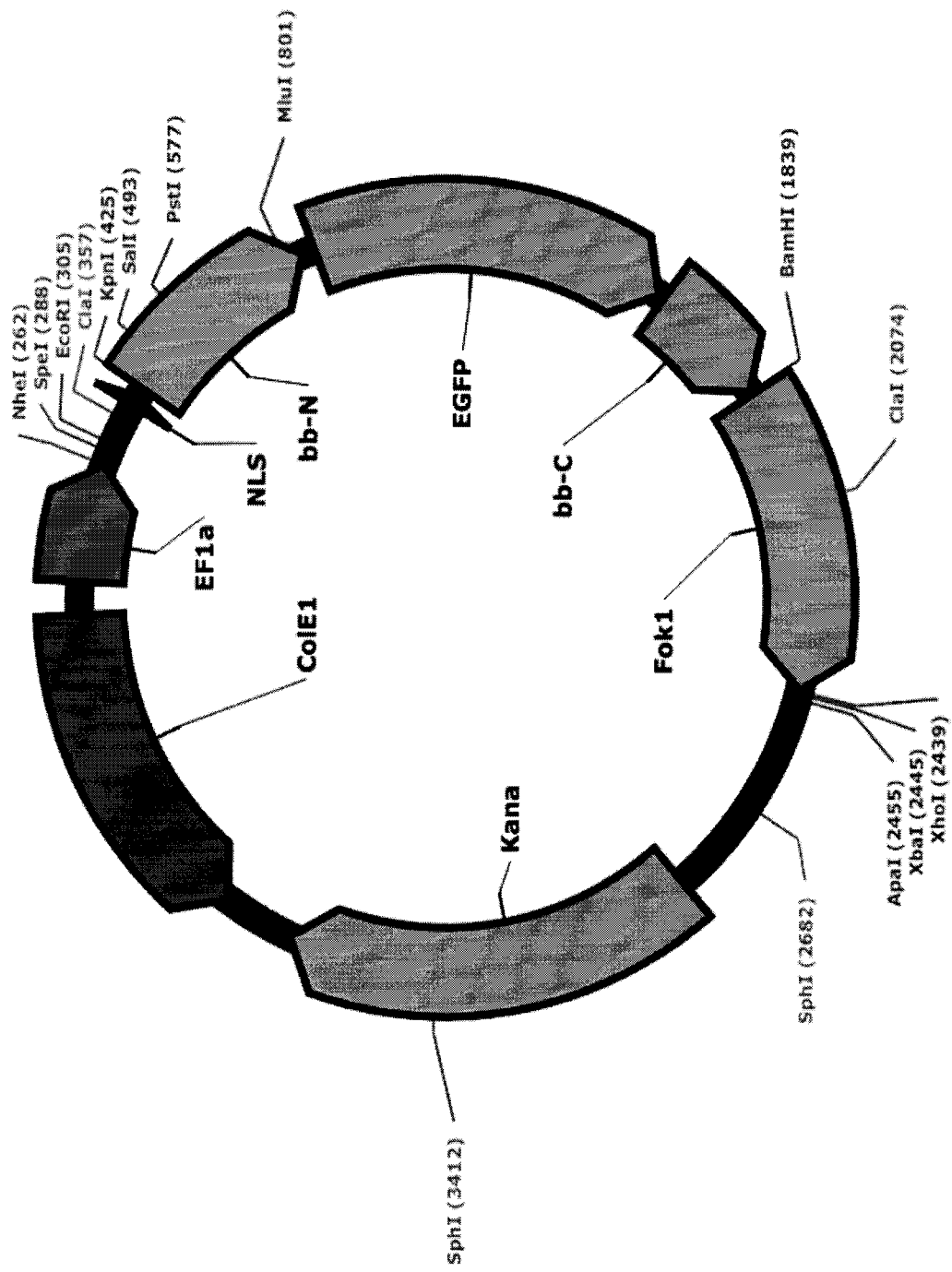


FIG.2

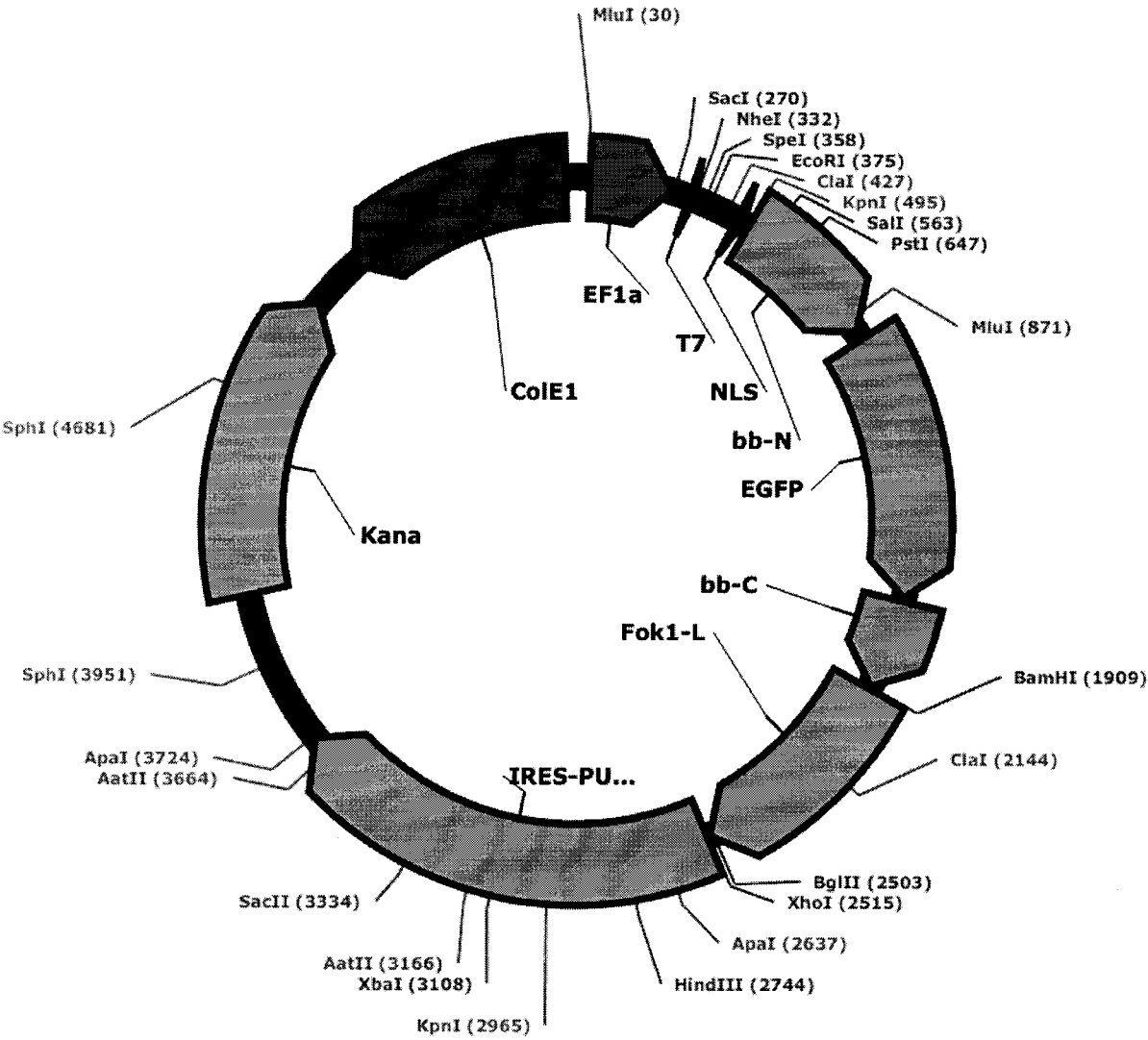


FIG. 3

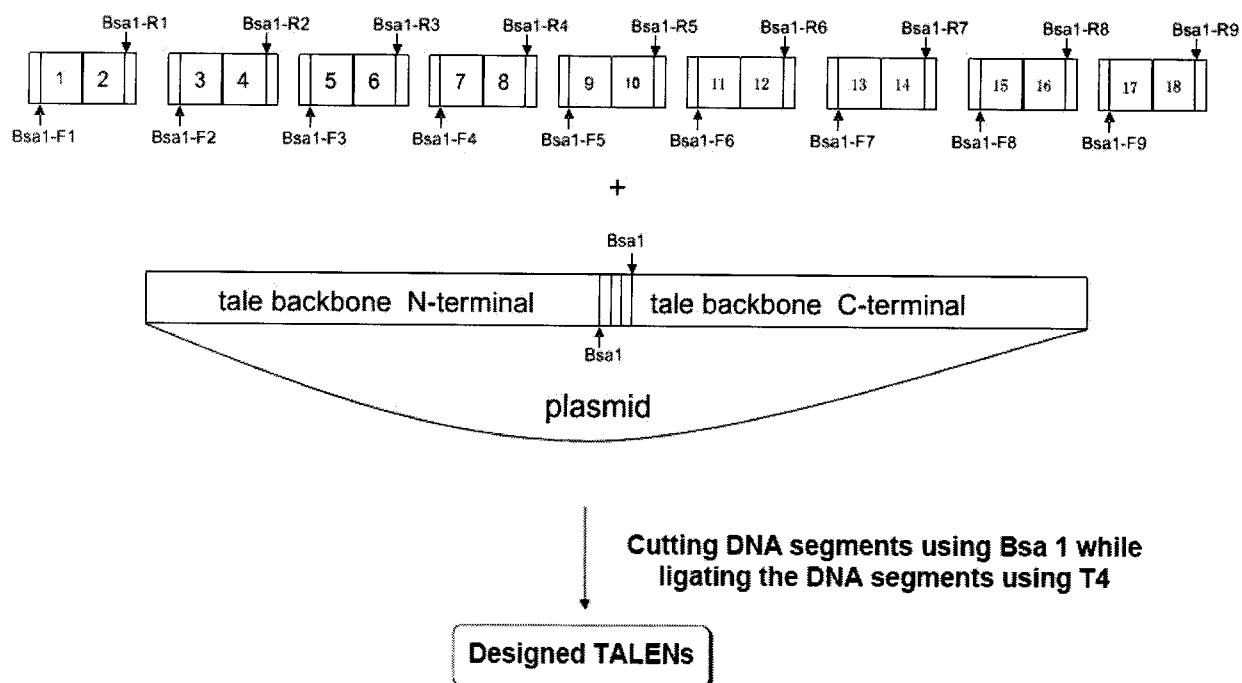


FIG. 4

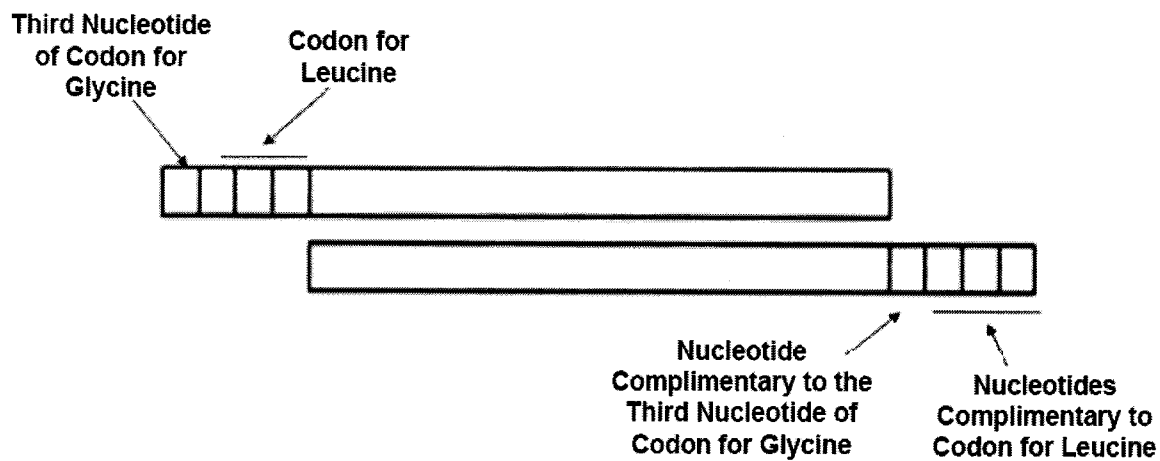


FIG. 5

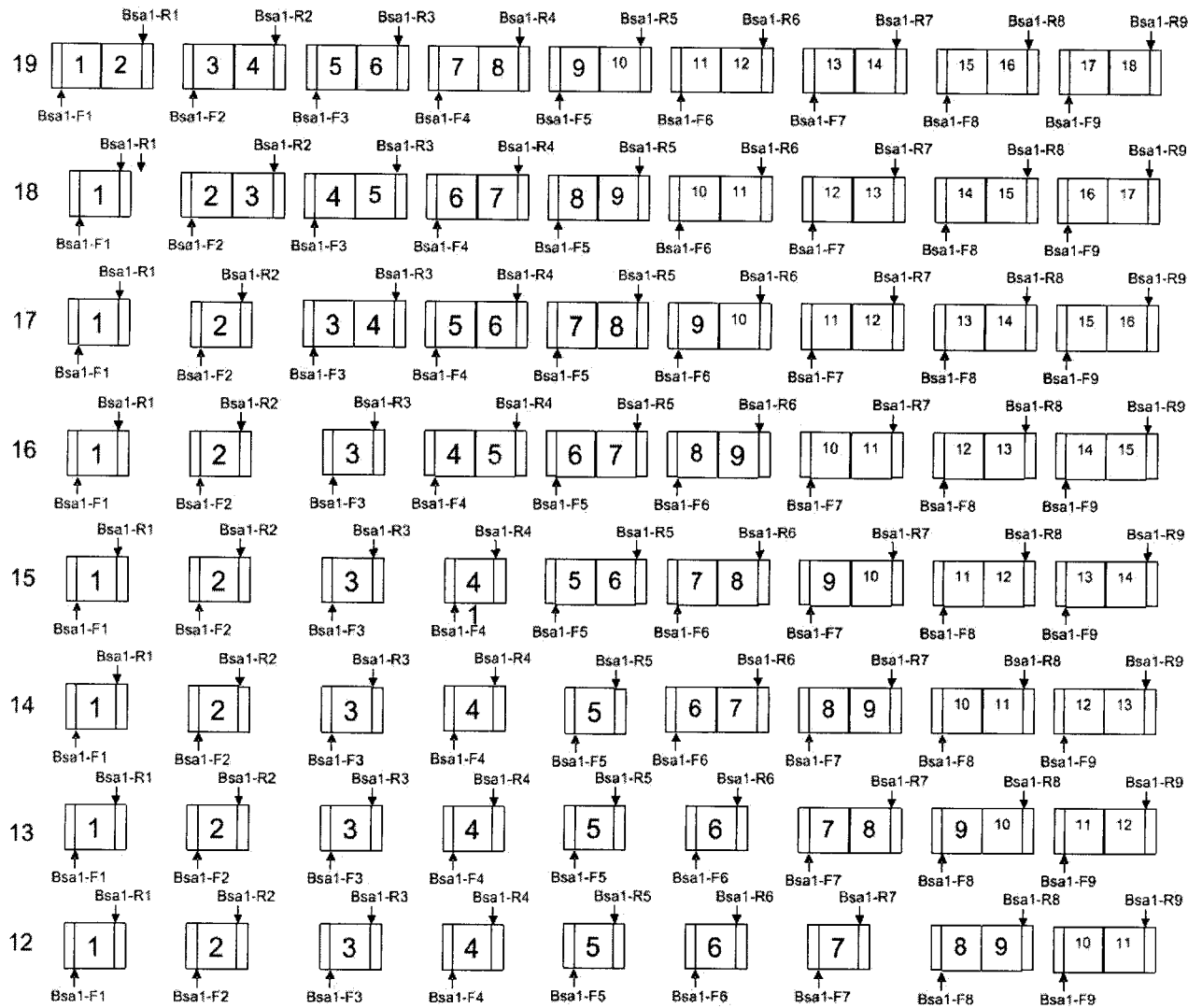


FIG. 6

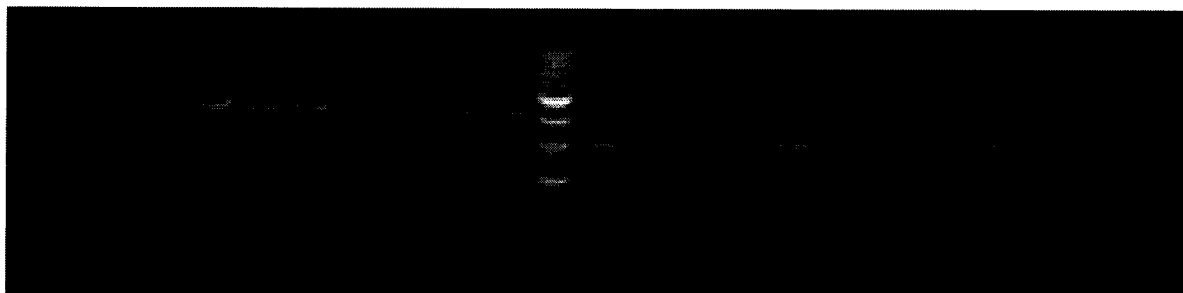


FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2013/001045

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/66 (2006.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C12N

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

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

DWPI, SIPOABS, CNABS, CNKI, Google scholar, Pubmed, TALE, transcription activator-like effector, assembly, construct+, cloning, type II restriction enzyme, fragment?, segment?, one step, golden gate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Ernst Weber et al., "Assembly of Designer TAL Effectors by Golden Gate Cloning", Plos One, May 2011, Volume 6, Issue 5	1-2, 4, 8-18, 20
Y	see the the 1st and 3rd paragraph of "Results - dTALE assembly strategy", figure 1.	3, 5-7, 19
Y	Carola Engler et al., "Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes", Plos One, May 2009, Volume 4, Issue 5	3,6-7,19
Y	see the 1st paragraph of "Results – DNA shuffling strategy", Figure 2.	
Y	Neville E Sanjana et al., "A transcription activator-like effector toolbox for genome engineering", Nature Protocols, Jan 2012, Volume 7, No.1, p171-p192	5
	see abstract, Figure 3., materials, procedure 16	

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

☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim (S) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 29 November 2013 (29.11.2013)	Date of mailing of the international search report 19 Dec. 2013 (19.12.2013)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer LI,Kangqi Telephone No. (86-10)62411034