



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

(13) **A1**

(86) **Date de dépôt PCT/PCT Filing Date:** 2022/09/07
 (87) **Date publication PCT/PCT Publication Date:** 2023/03/16
 (85) **Entrée phase nationale/National Entry:** 2024/01/31
 (86) **N° demande PCT/PCT Application No.:** US 2022/076059
 (87) **N° publication PCT/PCT Publication No.:** 2023/039436
 (30) **Priorité/Priority:** 2021/09/08 (US63/241,934)

(51) **Cl.Int./Int.Cl. C12N 5/10** (2006.01),
C12N 15/62 (2006.01), **C12N 15/70** (2006.01),
C12N 15/90 (2006.01), **C12N 9/12** (2006.01)
 (71) **Demandeur/Applicant:**
 METAGENOMI, INC., US
 (72) **Inventeurs/Inventors:**
 THOMAS, BRIAN C., US;
 BROWN, CHRISTOPHER, US;
 GOLTSMAN, DANIELA S.A., US;
 ALEXANDER, LISA, US;
 LAPERRIERE, SARAH, US
 (74) **Agent:** GOWLING WLG (CANADA) LLP

(54) **Titre : SYSTEMES ET PROCEDES DE TRANSPOSITION DE SEQUENCES NUCLEOTIDIQUES DE CHARGE**
 (54) **Title: SYSTEMS AND METHODS FOR TRANSPOSING CARGO NUCLEOTIDE SEQUENCES**

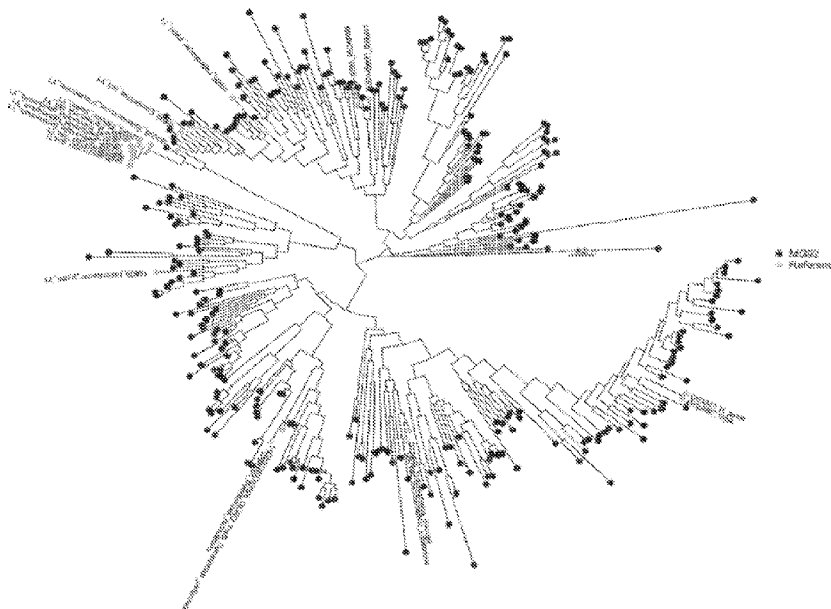


FIG. 2

(57) **Abrégé/Abstract:**

The present disclosure provides systems and methods for transposing a cargo nucleotide sequence to a target nucleic acid site. These systems and methods may comprise a first double-stranded nucleic acid comprising the cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase, and the transposase, wherein said transposase is configured to transpose the cargo nucleotide sequence to the target nucleic acid site.

Date Submitted: 2024/01/31

CA App. No.: 3227683

Abstract:

The present disclosure provides systems and methods for transposing a cargo nucleotide sequence to a target nucleic acid site. These systems and methods may comprise a first double-stranded nucleic acid comprising the cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase, and the transposase, wherein said transposase is configured to transpose the cargo nucleotide sequence to the target nucleic acid site.

SYSTEMS AND METHODS FOR TRANSPOSING CARGO NUCLEOTIDE SEQUENCES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 63/241,934, entitled “SYSTEMS AND METHODS FOR TRANSPOSING CARGO NUCLEOTIDE SEQUENCES”, filed September 8, 2021, which is incorporated herein by this reference in its entirety.

BACKGROUND

[0002] Transposable elements are movable DNA sequences which play a crucial role in gene function and evolution. While transposable elements are found in nearly all forms of life, their prevalence varies among organisms, with a large proportion of the eukaryotic genome encoding for transposable elements (at least 45% in humans). While the foundational research on transposable elements was conducted in the 1940s, their potential utility in DNA manipulation and gene editing applications has only been recognized in recent years.

SEQUENCE LISTING

[0003] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on September 7, 2022, is named 55921-733601.xml and is 452,421 bytes in size.

SUMMARY

[0004] In some aspects, the present disclosure provides for an engineered transposase system, comprising: a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and a transposase, wherein: the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; and the transposase is derived from an uncultivated microorganism.

[0005] In some embodiments, the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about

98%, or at least about 99% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase comprises one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, the NLS comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NO: 455-470. In some embodiments, the sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm. In some embodiments, the sequence identity is determined by the BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

[0006] In some aspects, the present disclosure provides for an engineered transposase system, comprising: a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and a transposase, wherein: the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; and the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.

[0007] In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-

stranded deoxyribonucleic acid polynucleotide. In some embodiments, the sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm. In some embodiments, the sequence identity is determined by the BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

[0008] In some aspects, the present disclosure provides for a deoxyribonucleic acid polynucleotide encoding any engineered transposase system disclosed herein.

[0009] In some aspects, the present disclosure provides for a nucleic acid comprising an engineered nucleic acid sequence optimized for expression in an organism, wherein the nucleic acid encodes a transposase, and wherein the transposase is derived from an uncultivated microorganism, wherein the organism is not the uncultivated microorganism.

[0010] In some embodiments, the transposase comprises a variant having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase comprises a sequence encoding one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, the NLS comprises a sequence selected from SEQ ID NOs: 455-470. In some embodiments, the NLS comprises SEQ ID NO: 456. In some embodiments, the NLS is proximal to the N-terminus of the transposase. In some embodiments, the NLS comprises SEQ ID NO: 455. In some embodiments, the NLS is proximal to the C-terminus of the transposase. In some embodiments, the organism is prokaryotic, bacterial, eukaryotic, fungal, plant, mammalian, rodent, or human.

[0011] In some aspects, the present disclosure provides for a vector comprising any nucleic acid disclosed herein. In some embodiments, the nucleic acid further comprises a nucleic acid encoding a cargo nucleotide sequence configured to form a complex with the transposase. In some embodiments, the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, or a lentivirus.

[0012] In some aspects, the present disclosure provides for a cell comprising any vector disclosed herein.

[0013] In some aspects, the present disclosure provides for a method of manufacturing a transposase, comprising cultivating any cell disclosed herein.

[0014] In some aspects, the present disclosure provides for a method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide comprising a cargo sequence, comprising: contacting the double-stranded deoxyribonucleic acid polynucleotide with a transposase configured to transpose the cargo

nucleotide sequence to a target nucleic acid locus; and wherein the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.

[0015] In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.

[0016] In some aspects, the present disclosure provides for a method of modifying a target nucleic acid locus, the method comprising delivering to the target nucleic acid locus an engineered transposase system disclosed herein, wherein the transposase is configured to transpose the cargo nucleotide sequence to the target nucleic acid locus, and wherein the complex is configured such that upon binding of the complex to the target nucleic acid locus, the complex modifies the target nucleic acid locus.

[0017] In some embodiments, modifying the target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing the target nucleic acid locus. In some embodiments, the target nucleic acid locus comprises deoxyribonucleic acid (DNA). In some embodiments, the target nucleic acid locus comprises genomic DNA, viral DNA, or bacterial DNA. In some embodiments, the target nucleic acid locus is *in vitro*. In some embodiments, the target nucleic acid locus is within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell. In some embodiments, the cell is a primary cell. In some embodiments, the primary cell is a T cell. In some embodiments, the primary cell is a hematopoietic stem cell (HSC). In some embodiments, delivering the engineered

transposase system to the target nucleic acid locus comprises delivering an nucleic acid disclosed herein or any vector disclosed herein. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding the transposase. In some embodiments, the nucleic acid comprises a promoter to which the open reading frame encoding the transposase is operably linked. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a capped mRNA containing the open reading frame encoding the transposase. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a translated polypeptide. In some embodiments, the transposase induces a single-stranded break or a double-stranded break at or proximal to the target nucleic acid locus. In some embodiments, the transposase induces a staggered single stranded break within or 5' to the target locus.

[0018] In some aspects, the present disclosure provides for a host cell comprising an open reading frame encoding a heterologous transposase having at least 75% sequence identity to any one of SEQ ID NOs: 1-349 or a variant thereof. In some embodiments, the transposase has at least 75% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 18-19. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 18-19. In some embodiments, the transposase has at least 75% sequence identity to any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17. In some embodiments, the host cell is an *E. coli* cell. In some embodiments, the *E. coli* cell is a λ DE3 lysogen or the *E. coli* cell is a BL21(DE3) strain. In some embodiments, the *E. coli* cell has an *ompT lon* genotype. In some embodiments, the open reading frame is operably linked to a T7 promoter sequence, a T7-lac promoter sequence, a lac promoter sequence, a tac promoter sequence, a trc promoter sequence, a ParaBAD promoter sequence, a PrhaBAD promoter sequence, a T5 promoter sequence, a *cspA* promoter sequence, an *araP*_{BAD} promoter, a strong leftward promoter from phage lambda (pL promoter), or any combination thereof. In some embodiments, the open reading frame comprises a sequence encoding an affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the affinity tag is an immobilized metal affinity chromatography (IMAC) tag. In some embodiments, the IMAC tag is a polyhistidine tag. In some embodiments, the affinity tag is a myc tag, a human influenza hemagglutinin (HA) tag, a maltose binding protein (MBP) tag, a glutathione S-transferase (GST) tag, a streptavidin tag, a FLAG tag, or any combination thereof. In some embodiments, the

affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding a protease cleavage site. In some embodiments, the protease cleavage site is a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof. In some embodiments, the open reading frame is codon-optimized for expression in the host cell. In some embodiments, the open reading frame is provided on a vector. In some embodiments, the open reading frame is integrated into a genome of the host cell.

[0019] In some aspects, the present disclosure provides for a culture comprising any host cell disclosed herein in compatible liquid medium.

[0020] In some aspects, the present disclosure provides for a method of producing a transposase, comprising cultivating any host cell disclosed herein in compatible growth medium.

[0021] In some embodiments, the method further comprises inducing expression of the transposase by addition of an additional chemical agent or an increased amount of a nutrient. In some embodiments, the additional chemical agent or increased amount of a nutrient comprises **Isopropyl β-D-1-thiogalactopyranoside (IPTG) or additional amounts of lactose.** In some embodiments, the method further comprises isolating the host cell after the cultivation and lysing the host cell to produce a protein extract. In some embodiments, the method further comprises subjecting the protein extract to IMAC, or ion-affinity chromatography. In some embodiments, the open reading frame comprises a sequence encoding an IMAC affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the IMAC affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding protease cleavage site. In some embodiments, the protease cleavage site comprises a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof. In some embodiments, the method further comprises cleaving the IMAC affinity tag by contacting a protease corresponding to the protease cleavage site to the transposase. In some embodiments, the method further comprises performing subtractive IMAC affinity chromatography to remove the affinity tag from a composition comprising the transposase.

[0022] In some aspects, the present disclosure provides for a method of disrupting a locus in a cell, comprising contacting to the cell a composition comprising: a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and a transposase, wherein: the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349; and the transposase has at least equivalent transposition activity to TnpA transposase in a cell.

[0023] In some embodiments, the transposition activity is measured *in vitro* by introducing the transposase to cells comprising the target nucleic acid locus and detecting transposition of the target nucleic acid locus in the cells. In some embodiments, the composition comprises 20 picomoles (pmol) or less of the transposase. In some embodiments, the composition comprises 1 pmol or less of the transposase.

[0024] In some aspects, the present disclosure provides for an engineered transposase system, comprising: a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and a transposase, wherein the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; and the double-stranded nucleic acid comprises a flanking sequence flanking the cargo sequence, wherein the flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.

[0025] In some embodiments, the transposase is derived from an uncultivated organism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase comprises one or more nuclear localization signals (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, a NLS of the one or more NLSs comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NOs: 455-470. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least

about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350, 352, 355, 356, 359, 361, 362, and 367. In some embodiments, the double-stranded nucleic acid comprises another flanking sequence flanking the cargo sequence, wherein the another flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454. In some embodiments, the another flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 351, 353, 354, 357, 358, 360, 363, and 366. In some embodiments, the flanking sequence flanks a left end of the cargo nucleic acid sequence and wherein the another flanking sequence flanks a right end of the cargo nucleic acid sequence. In some embodiments, the transposase is configured to recognize an insertion motif adjacent to the target nucleic acid locus. In some embodiments, the insertion motif comprises at least three, four, five, or six consecutive nucleotides of the sequence AATGAC.

[0026] In some aspects, the present disclosure provides for a deoxyribonucleic acid polynucleotide encoding any engineered transposase system disclosed herein.

[0027] In some aspects, the present disclosure provides for a method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide comprising a cargo sequence, the method comprising: contacting the double-stranded deoxyribonucleic acid polynucleotide with a transposase configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; wherein the double-stranded deoxyribonucleic acid polynucleotide comprises a flanking sequence flanking the cargo sequence, wherein the flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.

[0028] In some embodiments, the transposase is derived from an uncultivated organism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least

about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase comprises one or more nuclear localization signals (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, a NLS of the one or more NLSs comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NOs: 455-470. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350, 352, 355, 356, 359, 361, 362, and 367. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide comprises another flanking sequence flanking the cargo sequence, wherein the another flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454. In some embodiments, the another flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 351, 353, 354, 357, 358, 360, 363, and 366. In some embodiments, the flanking sequence flanks a left end of the cargo nucleic acid sequence and wherein the another flanking sequence flanks a right end of the cargo nucleic acid sequence. In some embodiments, the transposase is configured to recognize an insertion motif adjacent to the target nucleic acid locus. In some embodiments, the insertion motif comprises at least three, four, five, or six consecutive nucleotides of the sequence AATGAC.

[0029] In some aspects, the present disclosure provides for a method of modifying a target nucleic acid locus, the method comprising delivering to the target nucleic acid locus an engineered transposase system disclosed herein, wherein the transposase is configured to

transpose the cargo nucleotide sequence to the target nucleic acid locus, and wherein the complex is configured such that upon binding of the complex to the target nucleic acid locus, the complex modifies the target nucleic acid locus.

[0030] In some embodiments, modifying the target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing the target nucleic acid locus. In some embodiments, the target nucleic acid locus comprises deoxyribonucleic acid (DNA). In some embodiments, the target nucleic acid locus comprises genomic DNA, viral DNA, or bacterial DNA. In some embodiments, the target nucleic acid locus is *in vitro*. In some embodiments, the target nucleic acid locus is within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell. In some embodiments, the cell is a primary cell. In some embodiments, the primary cell is a T cell. In some embodiments, the primary cell is a hematopoietic stem cell (HSC). In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding the transposase. In some embodiments, the nucleic acid comprises a promoter to which the open reading frame encoding the transposase is operably linked. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a capped mRNA containing the open reading frame encoding the transposase. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a translated polypeptide. In some embodiments, the transposase induces a single-stranded break or a double-stranded break at or proximal to the target nucleic acid locus. In some embodiments, the transposase induces a staggered single stranded break within or 5' to the target locus.

[0031] In some aspects, the present disclosure provides for an engineered transposase system, comprising: (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and (b) a transposase, wherein: (i) the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; and (ii) the transposase is derived from an uncultivated microorganism. In some embodiments, the cargo nucleotide sequence is a heterologous sequence. In some embodiments, the cargo nucleotide sequence is an engineered sequence. In some embodiments, the cargo nucleotide sequence is not a wild-type genome sequence present in an organism. In some embodiments, the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than

80% sequence identity to a TnpB transposase. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase comprises one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, the NLS comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NO: 455-470. In some embodiments, the sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm. In some embodiments, the sequence identity is determined by the BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

[0032] In some aspects, the present disclosure provides for an engineered transposase system, comprising: (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and (b) a transposase, wherein: (i) the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; and (ii) the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm. In some embodiments, the sequence identity is determined by the BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

[0033] In some aspects, the present disclosure provides for a deoxyribonucleic acid

polynucleotide encoding the engineered transposase system of any one of the aspects or embodiments described herein

[0034] In some aspects, the present disclosure provides for a nucleic acid comprising an engineered nucleic acid sequence optimized for expression in an organism, wherein the nucleic acid encodes a transposase, and wherein the transposase is derived from an uncultivated microorganism, wherein the organism is not the uncultivated microorganism. In some embodiments, the transposase comprises a variant having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase comprises a sequence encoding one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of the transposase. In some embodiments, the NLS comprises a sequence selected from SEQ ID NOs: 455-470. In some embodiments, the NLS comprises SEQ ID NO: 456. In some embodiments, the NLS is proximal to the N-terminus of the transposase. In some embodiments, the NLS comprises SEQ ID NO: 455. In some embodiments, the NLS is proximal to the C-terminus of the transposase. In some embodiments, the organism is prokaryotic, bacterial, eukaryotic, fungal, plant, mammalian, rodent, or human.

[0035] In some aspects, the present disclosure provides for a vector comprising the nucleic acid of any one of the aspects or embodiments described herein. In some embodiments, the vector further comprises a nucleic acid encoding a cargo nucleotide sequence configured to form a complex with the transposase. In some embodiments, the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, or a lentivirus.

[0036] In some aspects, the present disclosure provides for a cell comprising the vector of any one of any one of the aspects or embodiments described herein.

[0037] In some aspects, the present disclosure provides for a method of manufacturing a transposase, comprising cultivating the cell of any one of the aspects or embodiments described herein.

[0038] In some aspects, the present disclosure provides for a method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide, comprising: (a) contacting the double-stranded deoxyribonucleic acid polynucleotide with a transposase configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; wherein the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than 80% sequence identity to a TnpB transposase. In some embodiments, the transposase

comprises a catalytic tyrosine residue. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.

[0039] In some aspects, the present disclosure provides for a method of modifying a target nucleic acid locus, the method comprising delivering to the target nucleic acid locus the engineered transposase system of any one of the aspects or embodiments described herein, wherein the transposase is configured to transpose the cargo nucleotide sequence to the target nucleic acid locus, and wherein the complex is configured such that upon binding of the complex to the target nucleic acid locus, the complex modifies the target nucleic acid locus. In some embodiments, modifying the target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing the target nucleic acid locus. In some embodiments, the target nucleic acid locus comprises deoxyribonucleic acid (DNA). In some embodiments, the target nucleic acid locus comprises genomic DNA, viral DNA, or bacterial DNA. In some embodiments, the target nucleic acid locus is *in vitro*. In some embodiments, the target nucleic acid locus is within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell. In some embodiments, the cell is a primary cell. In some embodiments, the primary cell is a T cell. In some embodiments, the primary cell is a hematopoietic stem cell (HSC). In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering the nucleic acid of any one of the aspects or embodiments described herein or the vector of any of the aspects or embodiments described herein. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding the transposase. In some embodiments, the nucleic acid comprises a promoter to which the open reading frame encoding the transposase is operably linked. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a capped mRNA containing the open reading frame encoding the transposase. In some embodiments, delivering the engineered transposase system to the target nucleic acid locus comprises delivering a translated polypeptide. In some embodiments, the transposase induces a single-stranded break or a double-stranded break at or proximal to the target nucleic acid locus. In some embodiments, the transposase induces a staggered single stranded break within or 5' to

the target locus.

[0040] In some aspects, the present disclosure provides for a host cell comprising an open reading frame encoding a heterologous transposase having at least 75% sequence identity to any one of SEQ ID NOs: 1-349 or a variant thereof. In some embodiments, the transposase has at least 75% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16. In some embodiments, the transposase has at least 75% sequence identity to any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17. In some embodiments, the host cell is an *E. coli* cell. In some embodiments, the *E. coli* cell is a λ DE3 lysogen or the *E. coli* cell is a BL21(DE3) strain. In some embodiments, the *E. coli* cell has an *ompT lon* genotype. In some embodiments, the open reading frame is operably linked to a T7 promoter sequence, a T7-lac promoter sequence, a lac promoter sequence, a tac promoter sequence, a trc promoter sequence, a ParaBAD promoter sequence, a PrhaBAD promoter sequence, a T5 promoter sequence, a *cspA* promoter sequence, an *araP_{BAD}* promoter, a strong leftward promoter from phage lambda (pL promoter), or any combination thereof. In some embodiments, the open reading frame comprises a sequence encoding an affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the affinity tag is an immobilized metal affinity chromatography (IMAC) tag. In some embodiments, the IMAC tag is a polyhistidine tag. In some embodiments, the affinity tag is a myc tag, a human influenza hemagglutinin (HA) tag, a maltose binding protein (MBP) tag, a glutathione S-transferase (GST) tag, a streptavidin tag, a FLAG tag, or any combination thereof. In some embodiments, the affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding a protease cleavage site. In some embodiments, the protease cleavage site is a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof. In some embodiments, the open reading frame is codon-optimized for expression in the host cell. In some embodiments, the open reading frame is provided on a vector. In some embodiments, the open reading frame is integrated into a genome of the host cell.

[0041] In some aspects, the present disclosure provides for a culture comprising the host cell of any one of the aspects or embodiments described herein in compatible liquid medium.

[0042] In some aspects, the present disclosure provides for a method of producing a transposase, comprising cultivating the host cell of any one of the aspects or embodiments described herein in compatible growth medium. In some embodiments, the method further comprises inducing expression of the transposase by addition of an additional chemical agent or an increased amount of a nutrient. In some embodiments, the additional chemical agent or increased amount of a nutrient comprises Isopropyl β -D-1-thiogalactopyranoside (IPTG) or additional amounts of lactose. In some embodiments, the method further comprises isolating the host cell after the

cultivation and lysing the host cell to produce a protein extract. In some embodiments, the method further comprises subjecting the protein extract to IMAC, or ion-affinity chromatography. In some embodiments, the open reading frame comprises a sequence encoding an IMAC affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the IMAC affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding protease cleavage site. In some embodiments, the protease cleavage site comprises a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof. In some embodiments, the method further comprises cleaving the IMAC affinity tag by contacting a protease corresponding to the protease cleavage site to the transposase. In some embodiments, the method further comprises performing subtractive IMAC affinity chromatography to remove the affinity tag from a composition comprising the transposase.

[0043] In some aspects, the present disclosure provides for a method of disrupting a locus in a cell, comprising contacting to the cell a composition comprising: (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein the cargo nucleotide sequence is configured to interact with a transposase; and (b) a transposase, wherein: (i) the transposase is configured to transpose the cargo nucleotide sequence to a target nucleic acid locus; (ii) the transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349; and (iii) the transposase has at least equivalent transposition activity to TnpA transposase in a cell. In some embodiments, the transposition activity is measured *in vitro* by introducing the transposase to cells comprising the target nucleic acid locus and detecting transposition of the target nucleic acid locus in the cells. In some embodiments, the composition comprises 20 pmoles or less of the transposase. In some embodiments, the composition comprises 1 pmol or less of the transposase.

[0044] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure.

Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0045] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or

patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0046] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0047] **FIGS. 1A** and **1B** depict MG transposases. **FIG. 1A** depicts the organization of a transposon comprising the tyrosine (Y1) transposase MG92-1 locus. MG92-1 is encoded at the 5' end of the transposon, followed by the accessory transposition protein TnpB and other cargo. The transposon ends contain direct repeats of 16-17 bp, and they exhibit secondary structure likely involved in transposition activity. **FIG. 1B** depicts multiple sequence alignment of MG Y1 transposase homologs. Catalytic residues HUH and Y are highlighted on the consensus sequence and on the MSA (boxes).

[0048] **FIG. 2** depicts a phylogenetic tree of TnpA protein sequences. The tree was built from a multiple sequence alignment of 414 novel TnpA sequences recovered here (black dots) and 19 reference TnpA sequences (grey dots). Labels for reference sequences were included.

[0049] **FIG. 3** depicts an example insertion sequence IS200/IS605 MG92-28. Top panel: Genomic context of the MG92-28 insertion sequence encoding the TnpA-like transposase and its associated TnpB-like gene. Both genes are flanked by LE and RE (boxes) predicted from covariance models. Bottom panel: LE (top left) and RE (bottom right) delineate the boundaries of the insertion sequence. Region predicted by the covariance models is annotated as arrows below the sequence. LE and RE secondary structures are shown for each end.

[0050] **FIG. 4** depicts a Western blot of TnpA-like proteins expressed in PureExpress. Lanes are: ladder, 1: HpTnpA, 2: HhTnpA, 3: 92-2, 4: 92-3, 5: 92-4, 6: 92-5, 7: 92-6, 8: 92-7, 9: 92-8, 10: 92-10, 11: 92-11. HpTnpA and HhTnpA are positive controls from *H. pylori* and *H. Heilmannii*, respectively. Molecular weights range from 17-23 kilodaltons (kDa).

[0051] **FIG. 5A** depicts the PCR product for the LE of the transposition reaction. All reactions have the protein and its paired specific cargo, except the control lane where the cargo is specified. Lanes are: 1: Ladder, 2: negative control NTC with HpTnpA cargo, 3: 92-1, 4: 92-2, 5: 92-3, 6: 92-4, 7: 92-5, 8: 92-6, 9: 92-7, 10: 92-8, 11: 92-10, 12: 92-11, 13: HpTnpA, 14: HhTnpA. Expected transposition product can range from 200 to 300 bp depending on LE size and is marked with an arrow. The band at <200 bp in 92-5 is related to non-specific primer interactions. **FIG. 5B** depicts the PCR product for the RE of the transposition reaction. All reactions have the protein and its paired specific cargo, except the control lane where the cargo is specified. Lanes are: 1: NTC with HpTnpA cargo, 2: 92-1, 3: 92-2, 4: 92-3, 5: 92-4, 6: 92-5, 7:

92-6, 8; 92-7, 9; 92-8, 10; 92-10, 11; 92-11, 12; HpTnpA, 13; HhTnpA, and 14: ladder. Expected transposition product can range from 300 to 500 bp depending on RE size and is marked with an arrow. Transposition that occurs into the 8N region will have a much weaker band than transposition into flanking sequence, so the faint bands are expected.

[0052] FIG. 6 depicts Sanger sequencing data confirming transposition for MG92-3. The chromatogram trace is shown mapped to the cargo sequence, where shaded letters match the cargo. At the cleavage point (arrow) the trace instead maps onto the target sequence (boxed). Analysis of the target reveals the insertion motif, which is shared sequence between the LE and the target. Downstream hairpins with flanking non-canonical base interactions can be identified.

[0053] FIG. 7 depicts Sanger sequencing data confirming transposition for MG92-3. The chromatogram trace is shown mapped to the cargo, and shaded letters match the cargo. At the cleavage point (arrow) the trace instead maps onto the target sequence (boxed). Analysis of the target reveals the insertion motif. The cleavage position in the putative RE defines the boundary of the RE, which folds into a canonical hairpin to allow TnpA recognition and strand cleavage (inset of dotted box).

[0054] FIG. 8 depicts analysis of chimeric NGS reads showing cargo and target sequence joints which were analyzed to determine the breakpoint. The x-axis is the position along the cargo sequence and the y-axis is the count of reads which transition at that position. The identified peak in the breakpoint at 2030 nt on the cargo matches the breakpoint identified in Sanger sequencing, confirming the position of LE cleavage.

[0055] FIG. 9 depicts NGS sequencing data confirming transposition for MG92-4. The NGS reads are shown mapped to the target, and light-shaded letters match the cargo. At the cleavage point (arrow) the trace instead maps onto the cargo sequence (boxed). The cleavage position in the putative RE defines the boundary of the RE, which folds into a canonical hairpin to allow TnpA recognition and strand cleavage (inset of dotted box). The NGS read histogram shows the frequency of reads corresponding to this breakpoint on the cargo.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

[0056] The Sequence Listing filed herewith provides exemplary polynucleotide and polypeptide sequences for use in methods, compositions, and systems according to the disclosure. Below are exemplary descriptions of sequences therein.

MG92

[0057] SEQ ID NOs: 1-349 show the full-length peptide sequences of MG92 transposition proteins.

[0058] SEQ ID NOs: 350-454 show the full-length peptide sequences of MG92 transposon ends.

Nuclear Localization Sequences

[0059] SEQ ID NOs: 455-470 show the full-length peptide sequences of nuclear localization sequences (NLSs) suitable for use with MG92 transposition proteins described herein.

DETAILED DESCRIPTION

[0060] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[0061] The practice of some methods disclosed herein employ, unless otherwise indicated, techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics, and recombinant DNA. See for example Sambrook and Green, *Molecular Cloning: A Laboratory Manual*, 4th Edition (2012); the series *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. eds.); the series *Methods In Enzymology* (Academic Press, Inc.), *PCR 2: A Practical Approach* (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) *Antibodies, A Laboratory Manual*, and *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications*, 6th Edition (R.I. Freshney, ed. (2010)) (which is entirely incorporated by reference herein).

[0062] As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms “including”, “includes”, “having”, “has”, “with”, or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term “comprising”.

[0063] The term “about” or “approximately” means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, “about” can mean within one or more than one standard deviation, per the practice in the art. Alternatively, “about” can mean a range of up to 20%, up to 15%, up to 10%, up to 5%, or up to 1% of a given value.

[0064] As used herein, a “cell” generally refers to a biological cell. A cell may be the basic structural, functional and/or biological unit of a living organism. A cell may originate from any organism having one or more cells. Some non-limiting examples include: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a protozoa cell, a cell from a plant (e.g., cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatoes, rice, cassava, sugarcane, pumpkin, hay, potatoes, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, ferns, clubmosses, hornworts,

liverworts, mosses), an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens* C. Agardh, and the like), seaweeds (e.g., kelp), a fungal cell (e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., a pig, a cow, a goat, a sheep, a rodent, a rat, a mouse, a non-human primate, a human, etc.), and etcetera. Sometimes a cell is not originating from a natural organism (e.g., a cell can be a synthetically made, sometimes termed an artificial cell).

[0065] The term “nucleotide,” as used herein, generally refers to a base-sugar-phosphate combination. A nucleotide may comprise a synthetic nucleotide. A nucleotide may comprise a synthetic nucleotide analog. Nucleotides may be monomeric units of a nucleic acid sequence (e.g., deoxyribonucleic acid (DNA) and ribonucleic acid (RNA)). The term nucleotide may include ribonucleoside triphosphates adenosine triphosphate (ATP), uridine triphosphate (UTP), cytosine triphosphate (CTP), guanosine triphosphate (GTP) and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such **derivatives may include, for example, [α S]dATP, 7-deaza-dGTP and 7-deaza-dATP, and** nucleotide derivatives that confer nuclease resistance on the nucleic acid molecule containing them. The term nucleotide as used herein may refer to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrative examples of dideoxyribonucleoside triphosphates may include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP. A nucleotide may be unlabeled or detectably labeled, such as using moieties comprising optically detectable moieties (e.g., fluorophores). Labeling may also be carried out with quantum dots. Detectable labels may include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels, and enzyme labels. Fluorescent labels of nucleotides may include **but are not limited fluorescein, 5-carboxyfluorescein (FAM), 2'7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE), rhodamine, 6-carboxyrhodamine (R6G), N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), 6-carboxy-X-rhodamine (ROX), 4-(4'dimethylaminophenylazo) benzoic acid (DABCYL), Cascade Blue, Oregon Green, Texas Red, Cyanine and 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS).** Specific examples of fluorescently labeled nucleotides can include [R6G]dUTP, [TAMRA]dUTP, [R110]dCTP, [R6G]dCTP, [TAMRA]dCTP, [JOE]ddATP, [R6G]ddATP, [FAM]ddCTP, [R110]ddCTP, [TAMRA]ddGTP, [ROX]ddTTP, [dR6G]ddATP, [dR110]ddCTP, [dTAMRA]ddGTP, and [dROX]ddTTP available from Perkin Elmer, Foster City, Calif; FluoroLink DeoxyNucleotides, FluoroLink Cy3-dCTP, FluoroLink Cy5-dCTP, FluoroLink Fluor X-dCTP, FluoroLink Cy3-dUTP, and FluoroLink Cy5-dUTP available from Amersham, Arlington Heights, Il.; Fluorescein-15-dATP, Fluorescein-12-

dUTP, Tetramethyl-rhodamine-6-dUTP, IR770-9-dATP, Fluorescein-12-ddUTP, Fluorescein-12-UTP, and Fluorescein-15-2'-dATP available from Boehringer Mannheim, Indianapolis, Ind.; and Chromosome Labeled Nucleotides, BODIPY-FL-14-UTP, BODIPY-FL-4-UTP, BODIPY-TMR-14-UTP, BODIPY-TMR-14-dUTP, BODIPY-TR-14-UTP, BODIPY-TR-14-dUTP, Cascade Blue-7-UTP, Cascade Blue-7-dUTP, fluorescein-12-UTP, fluorescein-12-dUTP, Oregon Green 488-5-dUTP, Rhodamine Green-5-UTP, Rhodamine Green-5-dUTP, tetramethylrhodamine-6-UTP, tetramethylrhodamine-6-dUTP, Texas Red-5-UTP, Texas Red-5-dUTP, and Texas Red-12-dUTP available from Molecular Probes, Eugene, Oreg. Nucleotides can also be labeled or marked by chemical modification. A chemically-modified single nucleotide can be biotin-dNTP. Some non-limiting examples of biotinylated dNTPs can include, biotin-dATP (e.g., bio-N6-ddATP, biotin-14-dATP), biotin-dCTP (e.g., biotin-11-dCTP, biotin-14-dCTP), and biotin-dUTP (e.g., biotin-11-dUTP, biotin-16-dUTP, biotin-20-dUTP).

[0066] The terms “polynucleotide,” “oligonucleotide,” and “nucleic acid” are used interchangeably to generally refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof, either in single-, double-, or multi-stranded form. A polynucleotide may be exogenous or endogenous to a cell. A polynucleotide may exist in a cell-free environment. A polynucleotide may be a gene or fragment thereof. A polynucleotide may be DNA. A polynucleotide may be RNA. A polynucleotide may have any three-dimensional structure and may perform any function. A polynucleotide may comprise one or more analogs (e.g., altered backbone, sugar, or nucleobase). If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. Some non-limiting examples of analogs include: 5-bromouracil, peptide nucleic acid, xeno nucleic acid, morpholinos, locked nucleic acids, glycol nucleic acids, threose nucleic acids, dideoxynucleotides, cordycepin, 7-deaza-GTP, fluorophores (e.g., rhodamine or fluorescein linked to the sugar), thiol-containing nucleotides, biotin-linked nucleotides, fluorescent base analogs, CpG islands, methyl-7-guanosine, methylated nucleotides, inosine, thiouridine, pseudouridine, dihydrouridine, queuosine, and wyosine. Non-limiting examples of polynucleotides include coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), short interfering RNA (siRNA), short-hairpin RNA (shRNA), microRNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, cell-free polynucleotides including cell-free DNA (cfDNA) and cell-free RNA (cfRNA), nucleic acid probes, and primers. The sequence of nucleotides may be interrupted by non-nucleotide components.

[0067] The terms “transfection” or “transfected” generally refer to introduction of a nucleic acid into a cell by non-viral or viral-based methods. The nucleic acid molecules may be gene sequences encoding complete proteins or functional portions thereof. See, e.g., Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual*, 18.1-18.88 (which is entirely incorporated by reference herein).

[0068] The terms “peptide,” “polypeptide,” and “protein” are used interchangeably herein to generally refer to a polymer of at least two amino acid residues joined by peptide bond(s). This term does not connote a specific length of polymer, nor is it intended to imply or distinguish whether the peptide is produced using recombinant techniques, chemical or enzymatic synthesis, or is naturally occurring. The terms apply to naturally occurring amino acid polymers as well as amino acid polymers comprising at least one modified amino acid. In some embodiments, the polymer may be interrupted by non-amino acids. The terms include amino acid chains of any length, including full length proteins, and proteins with or without secondary and/or tertiary structure (e.g., domains). The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, oxidation, and any other manipulation such as conjugation with a labeling component. The terms “amino acid” and “amino acids,” as used herein, generally refer to natural and non-natural amino acids, including, but not limited to, modified amino acids and amino acid analogues. Modified amino acids may include natural amino acids and non-natural amino acids, which have been chemically modified to include a group or a chemical moiety not naturally present on the amino acid. Amino acid analogues may refer to amino acid derivatives. The term “amino acid” includes both D-amino acids and L-amino acids.

[0069] As used herein, the “non-native” can generally refer to a nucleic acid or polypeptide sequence that is not found in a native nucleic acid or protein. Non-native may refer to affinity tags. Non-native may refer to fusions. Non-native may refer to a naturally occurring nucleic acid or polypeptide sequence that comprises mutations, insertions and/or deletions. A non-native sequence may exhibit and/or encode for an activity (e.g., enzymatic activity, methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.) that may also be exhibited by the nucleic acid and/or polypeptide sequence to which the non-native sequence is fused. A non-native nucleic acid or polypeptide sequence may be linked to a naturally-occurring nucleic acid or polypeptide sequence (or a variant thereof) by genetic engineering to generate a chimeric nucleic acid and/or polypeptide sequence encoding a chimeric nucleic acid and/or polypeptide.

[0070] The term “promoter”, as used herein, generally refers to the regulatory DNA region which controls transcription or expression of a gene and which may be located adjacent to or

overlapping a nucleotide or region of nucleotides at which RNA transcription is initiated. A promoter may contain specific DNA sequences which bind protein factors, often referred to as transcription factors, which facilitate binding of RNA polymerase to the DNA leading to gene transcription. A ‘basal promoter’, also referred to as a ‘core promoter’, may generally refer to a promoter that contains all the basic elements to promote transcriptional expression of an operably linked polynucleotide. In some embodiments eukaryotic basal promoters contain a TATA-box and/or a CAAT box.

[0071] The term “expression”, as used herein, generally refers to the process by which a nucleic acid sequence or a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as “gene product.” If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell.

[0072] As used herein, “operably linked”, “operable linkage”, “operatively linked”, or grammatical equivalents thereof generally refer to juxtaposition of genetic elements, e.g., a promoter, an enhancer, a polyadenylation sequence, etc., wherein the elements are in a relationship permitting them to operate in the expected manner. For instance, a regulatory element, which may comprise promoter and/or enhancer sequences, is operatively linked to a coding region if the regulatory element helps initiate transcription of the coding sequence. There may be intervening residues between the regulatory element and coding region so long as this functional relationship is maintained.

[0073] A “vector” as used herein, generally refers to a macromolecule or association of macromolecules that comprises or associates with a polynucleotide and which may be used to mediate delivery of the polynucleotide to a cell. Examples of vectors include plasmids, viral vectors, liposomes, and other gene delivery vehicles. The vector generally comprises genetic elements, e.g., regulatory elements, operatively linked to a gene to facilitate expression of the gene in a target.

[0074] As used herein, “an expression cassette” and “a nucleic acid cassette” are used interchangeably generally to refer to a combination of nucleic acid sequences or elements that are expressed together or are operably linked for expression. In some embodiments, an expression cassette refers to the combination of regulatory elements and a gene or genes to which they are operably linked for expression.

[0075] A “functional fragment” of a DNA or protein sequence generally refers to a fragment that retains a biological activity (either functional or structural) that is substantially similar to a biological activity of the full-length DNA or protein sequence. A biological activity of a DNA

sequence may be its ability to influence expression in a manner attributed to the full-length sequence.

[0076] As used herein, an “engineered” object generally indicates that the object has been modified by human intervention. According to non-limiting examples: a nucleic acid may be modified by changing its sequence to a sequence that does not occur in nature; a nucleic acid may be modified by ligating it to a nucleic acid that it does not associate with in nature such that the ligated product possesses a function not present in the original nucleic acid; an engineered nucleic acid may be synthesized in vitro with a sequence that does not exist in nature; a protein may be modified by changing its amino acid sequence to a sequence that does not exist in nature; an engineered protein may acquire a new function or property. An “engineered” system comprises at least one engineered component.

[0077] As used herein, “synthetic” and “artificial” can generally be used interchangeably to refer to a protein or a domain thereof that has low sequence identity (e.g., less than 50% sequence identity, less than 25% sequence identity, less than 10% sequence identity, less than 5% sequence identity, less than 1% sequence identity) to a naturally occurring human protein. For example, VPR and VP64 domains are synthetic transactivation domains.

[0078] As used herein, the term “transposable element” refers to a DNA sequence that can move from one location in the genome to another (i.e., they can be “transposed”). Transposable elements can be generally divided into two classes. Class I transposable elements, or “retrotransposons”, are transposed via transcription and translation of an RNA intermediate which is subsequently reincorporated into its new location into the genome via reverse transcription (a process mediated by a reverse transcriptase). Class II transposable elements, or “DNA transposons”, are transposed via a complex of single- or double-stranded DNA flanked on either side by a transposase. Further features of this family of enzymes can be found, e.g. in *Nature Education* **2008**, *1* (1), 204; and *Genome Biology* **2018**, *19* (199), 1-12; each of which is incorporated herein by reference.

[0079] As used herein, the term “TnpA” generally refers to the transposase found in members of the IS200/IS605 bacterial insertion sequence (“IS”) family. Unlike other documented IS transposases, which carry out DNA transposition via double-stranded DNA intermediates, TnpA proceeds via a single-stranded DNA intermediate. TnpA also differs from other documented IS transposases in that it contains flanking subterminal palindromic sequences rather than terminal inverted repeats. Further, TnpA inserts 3’ to specific AT-rich tetra- or pentanucleotides without duplication of the target site. Finally, TnpA belongs to the His-hydrophobic-His (“HuH”) superfamily of enzymes rather than the “DDE” superfamily of other IS transposases. As used herein, “TnpB” generally refers to an enzyme of undocumented function (though speculated to

play a regulatory role in transposition) found alongside TnpA in IS200/IS605 bacteria.

IS200/IS605 transposases are “Y1 transposases”, meaning that they are single-domain proteins comprising a single catalytic tyrosine residue. As used herein, the term “TnpA-like” generally refers to a protein which exhibits one or more functional, structural, biochemical, biophysical, or other properties or characteristics in common with a TnpA protein. As used herein, the term “TnpB-like” generally refers to a protein which exhibits one or more function, structural, biochemical, biophysical, or other properties or characteristics in common with a TnpB protein.

[0080] The term “sequence identity” or “percent identity” in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a local or global comparison window, as measured using a sequence comparison algorithm. Suitable sequence comparison algorithms for polypeptide sequences include, e.g., BLASTP using parameters of a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment for polypeptide sequences longer than 30 residues; BLASTP using parameters of a wordlength (W) of 2, an expectation (E) of 1000000, and the PAM30 scoring matrix setting gap costs at 9 to open gaps and 1 to extend gaps for sequences of less than 30 residues (these are the default parameters for BLASTP in the BLAST suite available at <https://blast.ncbi.nlm.nih.gov>); CLUSTALW with the Smith-Waterman homology search algorithm parameters with a match of 2, a mismatch of -1, and a gap of -1; MUSCLE with default parameters; MAFFT with parameters of a retree of 2 and max iterations of 1000; Novafold with default parameters; HMMER hmalign with default parameters.

[0081] The term “optimally aligned” in the context of two or more nucleic acids or polypeptide sequences, generally refers to two (e.g., in a pairwise alignment) or more (e.g., in a multiple sequence alignment) sequences that have been aligned to maximal correspondence of amino acids residues or nucleotides, for example, as determined by the alignment producing a highest or “optimized” percent identity score.

[0082] Included in the current disclosure are variants of any of the enzymes described herein with one or more conservative amino acid substitutions. Such conservative substitutions can be made in the amino acid sequence of a polypeptide without disrupting the three-dimensional structure or function of the polypeptide. Conservative substitutions can be accomplished by substituting amino acids with similar hydrophobicity, polarity, and R chain length for one another. Additionally, or alternatively, by comparing aligned sequences of homologous proteins from different species, conservative substitutions can be identified by locating amino acid

residues that have been mutated between species (e.g., non-conserved residues) without altering the basic functions of the encoded proteins. Such conservatively substituted variants may include variants with at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identity to any one of the transposase protein sequences described herein (e.g. MG92 family transposases described herein, or any other family transposase described herein). In some embodiments, such conservatively substituted variants are functional variants. Such functional variants can encompass sequences with substitutions such that the activity of one or more critical active site residues of the transposase are not disrupted. In some embodiments, a functional variant of any of the proteins described herein lacks substitution of at least one of the conserved or functional residues called out in **FIG. 1B**. In some embodiments, a functional variant of any of the proteins described herein lacks substitution of all of the conserved or functional residues called out in **FIG. 1B**.

[0083] Also included in the current disclosure are variants of any of the enzymes described herein with substitution of one or more catalytic residues to decrease or eliminate activity of the enzyme (e.g. decreased-activity variants). In some embodiments, a decreased activity variant as a protein described herein comprises a disrupting substitution of at least one, at least two, or all three catalytic residues called out in **FIG. 1B**.

[0084] Conservative substitution tables providing functionally similar amino acids are available from a variety of references (see, for e.g., Creighton, *Proteins: Structures and Molecular Properties* (W H Freeman & Co.; 2nd edition (December 1993)). The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

Overview

[0085] The discovery of new transposable elements with unique functionality and structure may offer the potential to further disrupt deoxyribonucleic acid (DNA) editing technologies, improving speed, specificity, functionality, and ease of use. Relative to the predicted prevalence of transposable elements in microbes and the sheer diversity of microbial species, relatively few functionally characterized transposable elements exist in the literature. This is partly because a huge number of microbial species may not be readily cultivated in laboratory conditions. Metagenomic sequencing from natural environmental niches containing large numbers of microbial species may offer the potential to drastically increase the number of new transposable elements documented and speed the discovery of new oligonucleotide editing functionalities.

[0086] Transposable elements are deoxyribonucleic acid sequences that can change position within a genome, often resulting in the generation or amelioration of mutations. In eukaryotes, a great proportion of the genome, and a large share of the mass of cellular DNA, is attributable to transposable elements. Although transposable elements are “selfish genes” which propagate themselves at the expense of other genes, they have been found to serve various important functions and to be crucial to genome evolution. Based on their mechanism, transposable elements are classified as either Class I “retrotransposons” or Class II “DNA transposons”.

[0087] Class I transposable elements, also referred to as retrotransposons, function according to a two-part “copy and paste” mechanism involving an RNA intermediate. First, the retrotransposon is transcribed. The resulting RNA is subsequently converted back to DNA by reverse transcriptase (generally encoded by the retrotransposon itself), and the reverse transcribed retrotransposon is finally integrated into its new position in the genome by integrase.

Retrotransposons are further classified into three orders. Retrotransposons with long terminal repeats (“LTRs”) encode reverse transcriptase and are flanked by long strands of repeating DNA. Retrotransposons with long interspersed nuclear elements (“LINEs”) encode reverse transcriptase, lack LTRs, and are transcribed by RNA polymerase II. Retrotransposons with short interspersed nuclear elements (“SINEs”) are transcribed by RNA polymerase III but lack reverse transcriptase, instead relying on the reverse transcription machinery of other transposable elements (e.g. LINEs).

[0088] Class II transposable elements, also referred to as DNA transposons, function according to mechanisms that do not involve an RNA intermediate. Many DNA transposons display a “cut and paste” mechanism in which transposase binds terminal inverted repeats (“TIRs”) flanking the transposon, cleaves the transposon from the donor region, and inserts it into the target region of the genome. Others, referred to as “helitrons”, display a “rolling circle” mechanism involving a single-stranded DNA intermediate and mediated by an undocumented protein believed to possess HUH endonuclease function and 5’ to 3’ helicase activity. First, a circular strand of DNA is

nicked to create two single DNA strands. The protein remains attached to the 5' phosphate of the nicked strand, leaving the 3' hydroxyl end of the complementary strand exposed and thus allowing a polymerase to replicate the non-nicked strand. Once replication is complete, the new strand disassociates and is itself replicated along with the original template strand. Still other DNA transposons, "Polintons", are theorized to undergo a "self-synthesis" mechanism. The transposition is initiated by an integrase's excision of a single-stranded extra-chromosomal Polinton element, which forms a racket-like structure. The Polinton undergoes replication with DNA polymerase B, and the double stranded Polinton is inserted into the genome by the integrase. Finally, some DNA transposons, such as those in the IS200/IS605 family, proceed via a "peel and paste" mechanism in which TnpA excises a piece of single-stranded DNA (as a circular "transposon joint") from the lagging strand template of the donor gene and reinserts it into the replication fork of the target gene.

[0089] While transposable elements have found some use as biological tools, documented transposable elements do not encompass the full range of possible biodiversity and targetability, and may not represent all possible activities. Here, thousands of genomic fragments were mined from numerous metagenomes for transposable elements. The documented diversity of transposable elements may have been expanded and novel systems may have been developed into highly targetable, compact, and precise gene editing agents.

MG Enzymes

[0090] In some aspects, the present disclosure provides for novel transposases. These candidates may represent one or more novel subtypes and some sub-families may have been identified. These transposases are less than about 500 amino acids in length. These transposases may simplify delivery and may extend therapeutic applications.

[0091] In some aspects, the present disclosure provides for a novel transposase. Such a transposase may be MG92 as described herein (see **FIGS. 1A** and **1B**).

[0092] In one aspect, the present disclosure provides for an engineered transposase system discovered through metagenomic sequencing. In some embodiments, the metagenomic sequencing is conducted on samples. In some embodiments, the samples may be collected from a variety of environments. Such environments may be a human microbiome, an animal microbiome, environments with high temperatures, environments with low temperatures. Such environments may include sediment.

[0093] In one aspect, the present disclosure provides for an engineered transposase system comprising a transposase. In some embodiments, the transposase is derived from an uncultivated microorganism. The transposase may be configured to bind a left-hand region comprising a

subterminal palindromic sequence. The transposase may bind a right-hand region comprising a subterminal palindromic sequence.

[0094] In one aspect, the present disclosure provides for an engineered transposase system comprising a transposase. In some embodiments, the transposase has at least about 70% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349.

[0095] In some embodiments, the transposase comprises a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase may be substantially identical to any one of SEQ ID NOs: 1-349.

[0096] In some embodiments, the transposase is not a TnpA or TnpB transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpB transposase.

[0097] In some embodiments, the transposase comprises a catalytic tyrosine residue.

[0098] In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic

sequence and a right-hand region comprising a subterminal palindromic sequence.

[0099] In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.

[00100] In some embodiments, the transposase comprises a sequence complementary to a eukaryotic, fungal, plant, mammalian, or human genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a eukaryotic genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a fungal genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a plant genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a mammalian genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a human genomic polynucleotide sequence.

[00101] In some embodiments, the transposase may comprise a variant having one or more nuclear localization sequences (NLSs). The NLS may be proximal to the N- or C-terminus of the transposase. The NLS may be appended N-terminal or C-terminal to any one of SEQ ID NOs: 455-470, or to a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 455. In some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 456.

Table 1: Example NLS Sequences that may be used with transposases according to the disclosure

| Source | NLS amino acid sequence | SEQ ID NO: |
|-----------------------------|--|------------|
| SV40 | PKKKRKV | 455 |
| nucleoplasmin bipartite NLS | KRPAATKKAGQAKKKK | 456 |
| c-myc NLS | PAAKRVKLD | 457 |
| c-myc NLS | RQRRNELKRSP | 458 |
| hRNPA1 M9 NLS | NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY | 459 |

| Source | NLS amino acid sequence | SEQ ID NO: |
|--|--|------------|
| Importin-alpha IBB domain | RMRI ZFKNKGKDTAELRRRRRVEVSVELRKAKKDEQILKRRNV | 460 |
| Myoma T protein | VSRKRPRP | 461 |
| Myoma T protein | PPKKARED | 462 |
| p53 | PQPKKKPL | 463 |
| mouse c-abl IV | SALIKKKKKMAP | 464 |
| influenza virus NS1 | DRLRR | 465 |
| influenza virus NS1 | PKQKKRK | 466 |
| Hepatitis virus delta antigen | RKLKKKIKKL | 467 |
| mouse Mx1 protein | REKKKFLKRR | 468 |
| human poly(ADP-ribose) polymerase | KRKGDEV DGVDEVAKKKSKK | 469 |
| steroid hormone receptors (human) glucocorticoid | RKCLQAGMNLEARKTKK | 470 |

[00102] In some embodiments, the transposase comprises a sequence at least 70% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 75% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 80% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 85% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 90% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 95% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof.

[00103] In some embodiments, the transposase comprises a sequence at least 70% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 75% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 80% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 85% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 90% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 95% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof.

[00104] In some embodiments, sequence may be determined by a BLASTP, CLUSTALW, MUSCLE, or MAFFT algorithm, or a CLUSTALW algorithm with the Smith-Waterman homology search algorithm parameters. The sequence identity may be determined by the BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.

[00105] In one aspect, the present disclosure provides a deoxyribonucleic acid polynucleotide encoding the engineered transposase system described herein.

[00106] In one aspect, the present disclosure provides a nucleic acid comprising an engineered nucleic acid sequence. In some embodiments, the engineered nucleic acid sequence is optimized for expression in an organism. In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the organism is not the uncultivated organism.

[00107] In some embodiments, the transposase has at least about 70% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349.

[00108] In some embodiments, the transposase comprises a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase may be substantially identical to any one of SEQ ID NOs: 1-349.

[00109] In some embodiments, the transposase is not a TnpA or TnpB transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpA transposase. In some

embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpB transposase.

[00110] In some embodiments, the transposase comprises a catalytic tyrosine residue.

[00111] In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.

[00112] In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.

[00113] In some embodiments, the transposase comprises a sequence complementary to a eukaryotic, fungal, plant, mammalian, or human genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a eukaryotic genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a fungal genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a plant genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a mammalian genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a human genomic polynucleotide sequence.

[00114] In some embodiments, the transposase may comprise a variant having one or more nuclear localization sequences (NLSs). The NLS may be proximal to the N- or C-terminus of the transposase. The NLS may be appended N-terminal or C-terminal to any one of SEQ ID NOs: 455-470, or to a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 455. In

some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 456.

[00115] In some embodiments, the organism is prokaryotic. In some embodiments, the organism is bacterial. In some embodiments, the organism is eukaryotic. In some embodiments, the organism is fungal. In some embodiments, the organism is a plant. In some embodiments, the organism is mammalian. In some embodiments, the organism is a rodent. In some embodiments, the organism is human.

[00116] In one aspect, the present disclosure provides an engineered vector. In some embodiments, the engineered vector comprises a nucleic acid sequence encoding a transposase. In some embodiments, the transposase is derived from an uncultivated microorganism.

[00117] In some embodiments, the engineered vector comprises a nucleic acid described herein. In some embodiments, the nucleic acid described herein is a deoxyribonucleic acid polynucleotide described herein. In some embodiments, the vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, or a lentivirus.

[00118] In one aspect, the present disclosure provides a cell comprising a vector described herein.

[00119] In one aspect, the present disclosure provides a method of manufacturing a transposase. In some embodiments, the method comprises cultivating the cell.

[00120] In one aspect, the present disclosure provides a method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide. The method may comprise contacting the double-stranded deoxyribonucleic acid polynucleotide with a transposase. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.

[00121] In some embodiments, the transposase is not a TnpA transposase or a TnpB transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about

35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpB transposase.

[00122] In some embodiments, the transposase comprises a catalytic tyrosine residue.

[00123] In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.

[00124] In some embodiments, the transposase is derived from an uncultivated microorganism. In some embodiments, the double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.

[00125] In one aspect, the present disclosure provides a method of modifying a target nucleic acid locus. The method may comprise delivering to the target nucleic acid locus the engineered transposase system described herein. In some embodiments, the complex is configured such that upon binding of the complex to the target nucleic acid locus, the complex modifies the target nucleic acid locus.

[00126] In some embodiments, modifying the target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing the target nucleic acid locus. In some embodiments, the target nucleic acid locus comprises deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). In some embodiments, the target nucleic acid comprises genomic DNA, viral DNA, viral RNA, or bacterial DNA. In some embodiments, the target nucleic acid locus is *in vitro*. In some embodiments, the target nucleic acid locus is within a cell. In some embodiments, the cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, or a human cell. In some embodiments, the cell is a primary cell. In some embodiments, the primary cell is a T cell. In some embodiments, the primary cell is a hematopoietic stem cell (HSC).

[00127] In some embodiments, delivery of the engineered transposase system to the target nucleic acid locus comprises delivering the nucleic acid described herein or the vector described herein. In some embodiments, delivery of engineered transposase system to the target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding the transposase. In some embodiments, the nucleic acid comprises a promoter. In some embodiments, the open reading frame encoding the transposase is operably linked to the promoter.

[00128] In some embodiments, delivery of the engineered transposase system to the target nucleic acid locus comprises delivering a capped mRNA containing the open reading frame

encoding the transposase. In some embodiments, delivery of the engineered transposase system to the target nucleic acid locus comprises delivering a translated polypeptide. In some embodiments, delivery of the engineered transposase system to the target nucleic acid locus comprises delivering a deoxyribonucleic acid (DNA) encoding the engineered guide RNA operably linked to a ribonucleic acid (RNA) pol III promoter.

[00129] In some embodiments, the transposase induces a single-stranded break or a double-stranded break at or proximal to the target locus. In some embodiments, the transposase induces a staggered single stranded break within or 5' to the target locus.

[00130] In one aspect, the present disclosure provides a host cell comprising an open reading frame encoding a heterologous transposase. In some embodiments, the transposase has at least about 70% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349.

[00131] In some embodiments, the transposase comprises a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase may be substantially identical to any one of SEQ ID NOs: 1-349.

[00132] In some embodiments, the transposase is not a TnpA or TnpB transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less

than about 10%, or less than about 5% sequence identity to a TnpB transposase.

[00133] In some embodiments, the transposase comprises a catalytic tyrosine residue.

[00134] In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.

[00135] In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.

[00136] In some embodiments, the transposase comprises a sequence at least 70% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 75% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 80% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 85% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 90% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 95% identical to a variant of any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 16, or a variant thereof.

[00137] In some embodiments, the transposase comprises a sequence at least 70% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 75% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 80% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 85% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 90% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof. In some embodiments, the transposase comprises a sequence at least 95% identical to a variant of any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17, or a variant thereof.

[00138] In some embodiments, the host cell is an *E. coli* cell. In some embodiments, the *E. coli*

cell is a λ DE3 lysogen or the *E. coli* cell is a BL21(DE3) strain. In some embodiments, the *E. coli* cell has an *ompT lon* genotype.

[00139] In some embodiments, the open reading frame is operably linked to a T7 promoter sequence, a T7-lac promoter sequence, a lac promoter sequence, a tac promoter sequence, a trc promoter sequence, a ParaBAD promoter sequence, a PrhaBAD promoter sequence, a T5 promoter sequence, a *cspA* promoter sequence, an *araP*_{BAD} promoter, a strong leftward promoter from phage lambda (pL promoter), or any combination thereof.

[00140] In some embodiments, the open reading frame comprises a sequence encoding an affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the affinity tag is an immobilized metal affinity chromatography (IMAC) tag. In some embodiments, the IMAC tag is a polyhistidine tag. In some embodiments, the affinity tag is a myc tag, a human influenza hemagglutinin (HA) tag, a maltose binding protein (MBP) tag, a glutathione S-transferase (GST) tag, a streptavidin tag, a FLAG tag, or any combination thereof. In some embodiments, the affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding a protease cleavage site. In some embodiments, the protease cleavage site is a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof.

[00141] In some embodiments, the open reading frame is codon-optimized for expression in the host cell. In some embodiments, the open reading frame is provided on a vector. In some embodiments, the open reading frame is integrated into a genome of the host cell.

[00142] In one aspect, the present disclosure provides a culture comprising a host cell described herein in compatible liquid medium.

[00143] In one aspect, the present disclosure provides a method of producing a transposase, comprising cultivating a host cell described herein in compatible growth medium. In some embodiments, the method further comprises inducing expression of the transposase by addition of an additional chemical agent or an increased amount of a nutrient. In some embodiments, the **additional chemical agent or increased amount of a nutrient comprises Isopropyl β -D-1-thiogalactopyranoside (IPTG) or additional amounts of lactose.** In some embodiments, the method further comprises isolating the host cell after the cultivation and lysing the host cell to produce a protein extract. In some embodiments, the method further comprises subjecting the protein extract to IMAC, or ion-affinity chromatography. In some embodiments, the open reading frame comprises a sequence encoding an IMAC affinity tag linked in-frame to a sequence encoding the transposase. In some embodiments, the IMAC affinity tag is linked in-frame to the sequence encoding the transposase via a linker sequence encoding protease cleavage

site. In some embodiments, the protease cleavage site comprises a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof. In some embodiments, the method further comprises cleaving the IMAC affinity tag by contacting a protease corresponding to the protease cleavage site to the transposase. In some embodiments, the method further comprises performing subtractive IMAC affinity chromatography to remove the affinity tag from a composition comprising the transposase.

[00144] In one aspect, the present disclosure provides a method of disrupting a locus in a cell. In some embodiments, the method comprises contacting to the cell a composition comprising a transposase. In some embodiments, the transposase has at least equivalent transposition activity to TnpA transposase in a cell. In some embodiments, the transposase has at least about 70% sequence identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase has at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349.

[00145] In some embodiments, the transposase comprises a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 1-349. In some embodiments, the transposase may be substantially identical to any one of SEQ ID NOs: 1-349.

[00146] In some embodiments, the transposase is not a TnpA or TnpB transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about 35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpA transposase. In some embodiments, the transposase has less than about 90%, less than about 85%, less than about 80%, less than about 75%, less than about 70%, less than about 65%, less than about 60%, less than about 55%, less than about 50%, less than about 45%, less than about 40%, less than about

35%, less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, or less than about 5% sequence identity to a TnpB transposase.

[00147] In some embodiments, the transposase comprises a catalytic tyrosine residue.

[00148] In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a right-hand region comprising a subterminal palindromic sequence. In some embodiments, the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.

[00149] In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as double-stranded deoxyribonucleic acid polynucleotide. In some embodiments, the transposase is configured to transpose the cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.

[00150] In some embodiments, the transposase comprises a sequence complementary to a eukaryotic, fungal, plant, mammalian, or human genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a eukaryotic genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a fungal genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a plant genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a mammalian genomic polynucleotide sequence. In some embodiments, the transposase comprises a sequence complementary to a human genomic polynucleotide sequence.

[00151] In some embodiments, the transposase may comprise a variant having one or more nuclear localization sequences (NLSs). The NLS may be proximal to the N- or C-terminus of the transposase. The NLS may be appended N-terminal or C-terminal to any one of SEQ ID NOs: 455-470, or to a variant having at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% identity to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to any one of SEQ ID NOs: 455-470. In some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 455. In some embodiments, the NLS may comprise a sequence substantially identical to SEQ ID NO: 456.

[00152] In some embodiments, the transposition activity is measured *in vitro* by introducing the

transposase to cells comprising the target nucleic acid locus and detecting transposition of the target nucleic acid locus in the cells. In some embodiments, the composition comprises 20 picomoles (pmol) or less of the transposase. In some embodiments, the composition comprises 1 pmol or less of the transposase.

[00153] Systems of the present disclosure may be used for various applications, such as, for example, nucleic acid editing (e.g., gene editing), binding to a nucleic acid molecule (e.g., sequence-specific binding). Such systems may be used, for example, for addressing (e.g., removing or replacing) a genetically inherited mutation that may cause a disease in a subject, inactivating a gene in order to ascertain its function in a cell, as a diagnostic tool to detect disease-causing genetic elements (e.g. via cleavage of reverse-transcribed viral RNA or an amplified DNA sequence encoding a disease-causing mutation), as deactivated enzymes in combination with a probe to target and detect a specific nucleotide sequence (e.g. sequence encoding antibiotic resistance in bacteria), to render viruses inactive or incapable of infecting host cells by targeting viral genomes, to add genes or amend metabolic pathways to engineer organisms to produce valuable small molecules, macromolecules, or secondary metabolites, to establish a gene drive element for evolutionary selection, to detect cell perturbations by foreign small molecules and nucleotides as a biosensor.

EXAMPLES

[00154] In accordance with IUPAC conventions, the following abbreviations are used throughout the examples:

A = adenine
C = cytosine
G = guanine
T = thymine
R = adenine or guanine
Y = cytosine or thymine
S = guanine or cytosine
W = adenine or thymine
K = guanine or thymine
M = adenine or cytosine
B = C, G, or T
D = A, G, or T
H = A, C, or T
V = A, C, or G

Example 1 – A method of metagenomic analysis for new proteins

[00155] Metagenomic samples were collected from sediment, soil, and animals.

Deoxyribonucleic acid (DNA) was extracted with a Zymobiomics DNA mini-prep kit and sequenced on an Illumina HiSeq[®] 2500. Samples were collected with consent of property

owners. Additional raw sequence data from public sources included animal microbiomes, sediment, soil, hot springs, hydrothermal vents, marine, peat bogs, permafrost, and sewage sequences. Metagenomic sequence data was searched using Hidden Markov Models generated based on documented transposase protein sequences to identify new transposases. Novel transposase proteins identified by the search were aligned to documented proteins to identify potential active sites. This metagenomic workflow resulted in the delineation of the MG92 family described herein.

Example 2 – Discovery of MG92 Family of Transposases

[00156] Analysis of the data from the metagenomic analysis of Example 1 revealed a new cluster of previously undescribed putative transposase systems comprising 1 family (MG92). The corresponding protein sequences for these new enzymes and their example subdomains are presented as SEQ ID NOS: 1-349.

Example 3 –Integrase *in vitro* activity (prophetic)

[00157] Integrase activity can be conducted via expression in an *E. coli* lysate based expression system (for example, myTXTL, Arbor Biosciences). The required components for *in vitro* testing are three plasmids: an expression plasmid with the transposon gene(s) under a T7 promoter, a target plasmid, and a donor plasmid which contains the required left end (LE) and right end (RE) DNA sequences for transposition around a cargo gene (e.g. Tet resistance gene). The lysate-based expression products, target DNA, and donor DNA are incubated to allow for transposition to occur. Transposition is detected via PCR. In addition, the transposition product will be tagged with T5 and sequenced via NGS to determine the insertion sites on a population of transposition events. Alternatively, the *in vitro* transposition products can be transformed into *E. coli* under antibiotic (e.g. Tet) selection, where growth requires the transposition cargo to be stably inserted into a plasmid. Either single colonies or a population of *E. coli* can be sequenced to determine the insertion sites.

[00158] Integration efficiency can be measured via ddPCR or qPCR of the experimental output of target DNA with integrated cargo, normalized to the amount of unmodified target DNA also measured via ddPCR.

[00159] This assay may also be conducted with purified protein components rather than from lysate-based expression. In this case, the proteins are expressed in *E. coli* protease-deficient B strain under T7 inducible promoter, the cells are lysed using sonication, and the His-tagged protein of interest is purified using HisTrap FF (GE Lifescience) Ni-NTA affinity chromatography on the AKTA Avant FPLC (GE Lifescience). Purity is determined using densitometry in ImageLab software (Bio-Rad) of the protein bands resolved on SDS-PAGE and

InstantBlue Ultrafast (Sigma-Aldrich) coomassie stained acrylamide gels (Bio-Rad). The protein is desalted in storage buffer composed of 50 mM Tris-HCl, 300 mM NaCl, 1 mM TCEP, 5% glycerol; pH 7.5 (or other buffers as determined for maximum stability) and stored at -80°C. After purification the transposon gene(s) are added to the target DNA and donor DNA as described above in a reaction buffer, for example 26 mM HEPES pH 7.5, 4.2 mM TRIS pH 8, 50 µg/mL BSA, 2 mM ATP, 2.1 mM DTT, 0.05 mM EDTA, 0.2 mM MgCl₂, 28 mM NaCl, 21 mM KCl, 1.35% glycerol, (final pH 7.5) supplemented with 15 mM MgOAc₂.

Example 4 – Transposon end verification via gel shift (prophetic)

[00160] The transposon ends are tested for transposase binding via an electrophoretic mobility shift assay (EMSA). In this case, the potential LE or RE is synthesized as a DNA fragment (100-500 bp) and end-labeled with FAM via PCR with FAM-labeled primers. The transposase protein is synthesized in an *in vitro* transcription/translation system (e.g. PURExpress). After synthesis, 1 µL of protein is added to 50 nM of the labeled RE or LE in a 10 µL reaction in binding buffer (e.g. 20 mM HEPES pH 7.5, 2.5 mM Tris pH 7.5, 10 mM NaCl, 0.0625 mM EDTA, 5 mM TCEP, 0.005% BSA, 1 µg/mL poly(dI-dC), and 5% glycerol). The binding is incubated at 30° for 40 minutes, then 2 µL of 6X loading buffer (60 mM KCl, 10 mM Tris pH 7.6, 50% glycerol) is added. The binding reaction is separated on a 5% TBE gel and visualized. Shifts of the LE or RE in the presence of transposase protein can be attributed to successful binding and are indicative of transposase activity. This assay can also be performed with transposase truncations or mutations, as well as using *E. coli* extract or purified protein.

Example 5 – Cleavage of donor DNA verification (prophetic)

[00161] To confirm that the transposase is involved in cleavage of donor DNA, short (~ 140 bp) fragments containing RE-LE junctions separated by up to 10 bp are labelled at both ends with FAM via PCR with FAM-labeled primers. Labeled DNA fragments are incubated with *in vitro* transcription/translation transposase products and the DNA is analyzed on a denaturing gel. Cleavage at each end of the junction can result in two labelled single-strand fragments which migrate at different rates on the gel.

Example 6 – Integrase activity in *E. coli* (prophetic)

[00162] Engineered *E. coli* strains are transformed with a plasmid expressing the transposon genes and a plasmid containing a temperature-sensitive origin of replication with a selectable marker flanked by left end (LE) and right end (RE) transposon motifs for integration. To confirm donor ssDNA preference by the transposase components, ssDNA plasmid supercoiling can be used as donor. Transformants induced for expression of these genes are then screened for transfer

of the marker to a genomic target by selection at restrictive temperature for plasmid replication and the marker integration in the genome is confirmed by PCR.

[00163] Integrations are screened using an unbiased approach. In brief, purified gDNA is tagmented with Tn5, and DNA of interest is then PCR amplified using primers specific to the Tn5 tagmentation and the selectable marker. The amplicons are then prepared for NGS sequencing. Analysis of the resulting sequences is trimmed of the transposon sequences and flanking sequences are mapped to the genome to determine insertion position, and insertion rates are determined.

[00164] Alternatively, a *polA* mutant *E. coli* strain, MM383, which produces a DNA polymerase I (PolI) that is defective at 42°C, is used to detect integration as described previously (Brandtsma *et al.*, 1981). Resistance to a selectable marker after growth at 42°C indicates incorporation of donor DNA into the chromosome. The pUC19 plasmid without donor is used as a control **following growth for 24 hours at 42°C without antibiotic selection.**

[00165] *E. coli* strains that successfully grow in selection media are presumed to have integrated the donor DNA encoding the cargo resistance gene. Colonies growing in antibiotic selection plates are genotyped for cargo presence and NGS of whole genome sequence is performed.

Example 7 – Integrase activity in mammalian cells (prophetic)

[00166] To show targeting and cleavage activity in mammalian cells, each of the transposon proteins is purified with 2 NLS peptides on either terminus of the protein sequence. A plasmid containing a selectable neomycin resistance marker (NeoR) or a fluorescent marker flanked by the left end (LE) and right end (RE) motifs is synthesized. Cells are then transfected with the plasmid, recovered for 4-6 hours, and subsequently electroporated with transposon proteins. Antibiotic resistance integration into the genome is quantified by G418-resistant colony counts, and positive transposition by the fluorescent marker is assayed by fluorescence activated cell cytometry. 72 hours after cotransfection, genomic DNA is extracted and used for the preparation of an NGS-library. Integration frequency is assayed by Tn5 tagmentation.

Example 8 – *In silico* analysis

[00167] An extensive assembly-driven metagenomic database of microbial, viral and eukaryotic genomes was mined to retrieve predicted proteins with ssDNA transposase function. Over 400 predicted proteins had a significant e-value ($< 1 \times 10^{-5}$) hit to TnpA transposases of the insertion sequences IS200/IS605. After filtering for complete ORFs and confirming presence of catalytic residues (Y1 and HuH), the TnpA-like protein sequences were aligned with MAFFT with parameters G-INSI (*Mol Biol Evol* **30**, 772–780 (2013)) and the alignment was used to infer a

phylogenetic tree with FastTree2 (*Plos One* **5**, e9490 (2010)). Phylogenetic analysis of TnpA transposases uncovered high diversity of novel TnpA-like protein sequences associated with IS200/IS605 insertion sequences (**FIG. 2**).

[00168] In order to predict the left and right ends (LE and RE) of the insertion sequence, covariance models were built from active LE and RE sequences available in the ISFinder database (<https://www-is.biotoul.fr/>). Specifically, a multiple sequence alignment (MSA) of LE and RE sequences was built with MAFFT with parameters X-INSI (*Mol Biol Evol* **30**, 772–780 (2013)) and the secondary structure of the alignment was inferred from the MSA with RNAalifold 2.5.0 with parameters -p --aln-stk (Vienna Package). Covariance models were built with Infernal packages (<http://eddylab.org/infernal/>) and genomic fragments containing candidate TnpA transposases were searched using the covariance models with the Infernal command ‘cmsearch’. Covariance models predicted LE and RE for over 70 candidate IS200/IS605 insertion sequences (**FIG. 3**).

Example 9 – Generation of ssDNA cargos

[00169] Each TnpA-like candidate had a unique cargo comprising the putative left end (LE) and right end (RE) sequences identified in the metagenomic contig. These putative LE and RE sequences were cloned to flank a kanamycin (Kan) resistance cargo gene via Gibson assembly. The ssDNA cargo was generated via PCR of the Kan cargo plasmid with common primers outside of the LE/RE regions with forward primer GTGCGGTAGTAAAGGTTAATACTGTT and a 5'-phosphate-modified reverse primer CTATAGTGAGTCGTATTA using standard cycling conditions with Phusion HF (NEB). After PCR amplification, the DNA bottom strand was degraded using Lambda exonuclease (NEB) and the remaining top strand was purified using a DCC-5 spin column with manufacturer's recommended changes for purifying ssDNA (Zymo Research). The single stranded DNA was checked on an agarose gel to verify complete conversion of dsDNA and quantified by the ssDNA Qubit kit (Thermofisher), yielding an average concentration of 20 nM.

Example 10 – Design of TnpA *in vitro* expression constructs

[00170] For *in vitro* activity, each TnpA-like protein gene was synthesized in pET21(+) codon-optimized for *E. coli* translation under control of a T7 promoter and flanked by C-terminal HA and His tags, with the exception of 92-1 that lacks the HA tag. The TnpA-like protein plasmids were then amplified using primers that bind ~150 bp upstream of the T7 promoter and downstream of the T7 terminator (primers TGGCGAGAAAGGAAGGGAAG and CCGAAACAAGCGCTCATGAG) and purified via SPRI bead clean-up (MagBio HighPrep) to give final template concentrations >80 ng/μL.

Example 11 – *In vitro* transposition activity

[00171] For *in vitro* activity, TnpA-like protein candidates were first expressed in an *in vitro* transcription-translation (IVTT) kit following manufacturer's recommended conditions at 37 °C for 2 hours with a minimum template concentration of 8 ng/μL (PURExpress, NEB). Expression was verified via Western blot to the HA tag, with the exception of 92-1, which lacks this tag. **(FIG. 4)**. Transposition assays were set up with 1 μL of IVTT product added per 10 μL reaction, an average of 5 nM of ssDNA cargo and 50 nM of a 161 nt "target" ssDNA containing an 8N randomized sequence in reaction buffer (20 mM HEPES (pH 7.5), 160 mM NaCl, 5 mM MgCl₂, 5 mM TCEP, 20 μg/mL BSA, 0.5 μg/mL of poly-dIdC, and 20% glycerol). Control reactions contained a no-template control (NTC) reaction of IVTT where Tris buffer was added instead of PCR template to the IVTT. Reactions were incubated at 37 °C for 1 hour to allow transposition to occur, then the reaction was diluted 10-fold in water and transposition was detected via PCR. The LE junction was detected via a forward primer on the 5' end of the target and reverse primer within the Kan cargo, and the RE junction via a forward primer in the Kan cargo and a reverse primer on the 3' end of the target. PCR products were run on an agarose gel to detect transposition **(FIGS. 5A and 5B)**, and sequenced via Sanger and NGS sequencing. Chimeric reads that contained both target and cargo sequence were analyzed to determine the junction of transposition, the insertion motif, and the cleavage sites on the cargo **(FIGS. 6-9)**.

[00172] For the LE PCR product, the insertion motif can be identified from overlapping sequence identity between the cargo and the target. For example, the junction between target and the LE for MG92-3 is identified as the point where sequences for the target and cargo no longer overlap **(FIG. 6)**. The insertion motif can be identified via analysis of the flanking sequence of the target DNA without transposition. In the case of insertion into the 8N, the target motif can only be identified without ambiguity in the LE read, not the RE read. For MG92-3, the insertion motif was identified as AATGAC or a subset of nucleotides therein, for example TGAC **(FIGS. 6-7)**. For the RE PCR product, the RE junction is identified via the breakpoint where reads switch between mapping to the cargo and the target **(FIG. 7)**. Sequencing for the LE junction and the RE junction shows the same insertion location. The LE junction was further confirmed via NGS, which identified the same cleavage point in the LE as determined via Sanger sequencing **(FIG. 8)**.

[00173] From these data, the LE boundary can be determined as:

TGAAAACAAACATTTTACCAAGGCCCGCAGGCTCCGTCTATAGCGACAAGCGCTAAC
TTTGGCTACGCTTGTCGTTTAGGCGGGGTTAGT. This is a subset of the full MG92-3 LE and will be recognized by MG92-3 only when flanked by the recognition motif AATGAC, or a subset of nucleotides therein. Similarly, the RE boundary can be identified as:

GTTTGCCTGTATCTGTGGTCAGGTATCCACTCCTACCTAAAGTAGCAGGCATGAAC
GAAAGTTTATGCGGAGTTTGGAAAGCCCCGTCTATATTCGCGAAAGCGGATTAGGCGG
GGAGGGTTCAC, some or all of which is required for recognition, excision, and insertion by
TnpA-like proteins. Both of the sequences contain predicted hairpins for TnpA-like protein
recognition flanked by non-canonical base pairing interactions which TnpA and TnpA-like
proteins recognize (FIGs. 6-7), as described in *Cell* **132**, 208–220 (2008) and *Nucleic Acids Res*
39, 8503–8512 (2011).

[00174] Similarly, activity of MG92-4 was confirmed via NGS detection, with a weaker signal
not detectable in Sanger sequencing, showing RE cleavage and insertion (FIG. 9). As this signal
was only detectable by NGS, these results suggest that this insertion motif is possible but may
not be the optimal insertion sequence.

Example 12 – *In vitro* excision assay (prophetic)

[00175] To determine *in vitro* excision activity, TnpA-like protein candidates are expressed in an
in vitro transcription-translation (IVTT) kit following manufacturer's recommended conditions at
37 °C for 2 hours with a minimum template concentration of 8 ng/μL (PURExpress, NEB).
Excision assays are set with 1 μL of IVTT product added per 10 μL reaction and 100 ng of LE-
Kan-RE ssDNA (about 2.2 kb) for 60 minutes at 37 °C in TnpA reaction buffer (20 mM HEPES
(pH 7.5), 160 mM NaCl, 5 mM MgCl₂, 10 mM TCEP, 20 mg/mL BSA, 0.5 mg of poly-dIdC,
and 20% glycerol). Reactions are terminated with the addition of 0.1% SDS and incubation of an
additional 15 minutes at 37 °C. Reactions are subsequently RNase treated and run on a DNA
agarose gel to determine if excision of the LE-Kan-RE ssDNA has occurred. The excised Kan
sequence is then gel extracted and submitted for sequencing for determination of the LE and RE
cleavage motifs.

Example 13 – *In vivo* excision assay (prophetic)

[00176] *In vivo* excision assays are also performed by co-transforming *E. coli* with 2 plasmids,
one containing the LE-Kan-RE cargo and the other TnpA. Following transformation and
overnight growth, excision is determined by mini-prep of overnight culture and detection of
reclosed donor backbone molecules from which the Kan sequence has been removed on a DNA
gel. Controls for this experiment include the transformation of a single plasmid or the
transformation of both the TnpA-containing plasmid and the cargo plasmid with an inverted
origin of replication. The excised DNA backbone is gel extracted and subjected to sequencing to
yield the RE and LE boundaries of the TnpA transposon. The insertion motif remains in the
excised backbone and can also be identified at the sealed junction.

Example 14 – Changing insertion site specificity (prophetic)

[00177] Engineering of the insertion recognition site has been demonstrated by *Cell* **132**, 208–220 (2008) without requiring engineering of the TnpA protein. The insertion site recognized by a metagenomics-derived TnpA-like protein described herein is modified via sequence mutations to the insertion site motif and compensatory mutations to the base pairing partners in the LE ssDNA flanking the LE hairpin sequence. A series of single, double, and triple sequence mutations are introduced at rationally designed positions in the insertion site and LE sequence. Recognition and cleavage of the mutated insertion site by wild-type TnpA-like protein is tested concurrently with the wild-type LE insertion sequence using the excision/insertion assays and subsequent sequencing steps described above to compare activity levels.

Example 15 – TnpA can be used with sequence-specific endonucleases for programmable integrations (prophetic)

[00178] IS200/IS605 transposons are a type of mobile genetic element that integrate at specific target sites. These transposons are mobilized by their encoded TnpA-like transposase, an enzyme that belongs to the family of tyrosine (Y) transposases (reviewed in *Microbiol Spectr* **3**, (2015)). The mechanism of IS200/IS605 transposon mobilization involves its excision by TnpA or a TnpA-like protein, followed by its integration at a recognized target site during host replication, when target sites are accessible as ssDNA at the replication fork (*Cell* **142**, 398–408 (2010)).

[00179] The RNA-guided binding ability of certain sequence-specific (e.g., Cas) endonuclease effectors to a target site that is shared with TnpA-like proteins may aid TnpA-like effector-mediated integration of a desired cargo by making ssDNA and target site available through formation of the R-loop. Specifically, a desired cargo (for example, a fluorescence marker gene) flanked by TnpA-like-recognizable LE and RE is excised from a donor template by TnpA or a TnpA-like effector and integrated into a desired target site (which contains the TnpA or TnpA-like protein recognizable motif) that is made available by the binding of a (fused) sequence-specific endonuclease. The sequence-specific endonuclease may be engineered to be catalytically dead or have reduced or altered endonuclease (e.g., nickase) activity. Therefore, TnpA-like proteins can be “programmed” to insert a desired cargo into a TAM-dependent target site made available by fused, engineered (e.g., dead or nickase) sequence-specific endonuclease effectors.

Example 16 – *In vitro* testing of TnpA-like insertion into R-loops in dsDNA (prophetic)

[00180] The ability of TnpA-like proteins to insert into ssDNA generated as an R-loop in dsDNA can be tested using active TnpA-like proteins identified *in vitro* and their corresponding LE and RE sequences. The R-loop can be generated via a sequence-specific endonuclease, such as an RNA-directed nuclease-dead enzyme or nickase that is expressed in an IVTT reaction or added

as purified RNP. The TnpA-like protein is tested as described in the *in vitro* insertion assay, except the target ssDNA is replaced by the dsDNA and RNP. Insertion activity is assayed via PCR with a primer in the dsDNA target and the ssDNA cargo, flanking either the LE junction or the RE junction. The optimal location of the insertion site is tested by placing the insertion motif at various positions along the R-loop to determine the site with best accessibility by the TnpA-like protein. Insertion into ssDNA bubbles in dsDNA where mismatched DNA strands are annealed can also be tested.

Table 2 – Protein and nucleic acid sequences referred to herein

| Cat. | SEQ ID NO: | Description | Type |
|-----------------------------|-------------------|--------------------------------|-------------|
| MG92 transposition proteins | 1 | MG92-1-A transposition protein | protein |
| MG92 transposition proteins | 2 | MG92-1-B transposition protein | protein |
| MG92 transposition proteins | 3 | MG92-2-A transposition protein | protein |
| MG92 transposition proteins | 4 | MG92-2-B transposition protein | protein |
| MG92 transposition proteins | 5 | MG92-3-A transposition protein | protein |
| MG92 transposition proteins | 6 | MG92-3-B transposition protein | protein |
| MG92 transposition proteins | 7 | MG92-4-A transposition protein | protein |
| MG92 transposition proteins | 8 | MG92-4-B transposition protein | protein |
| MG92 transposition proteins | 9 | MG92-5-A transposition protein | protein |
| MG92 transposition proteins | 10 | MG92-5-B transposition protein | protein |
| MG92 transposition proteins | 11 | MG92-6-A transposition protein | protein |
| MG92 transposition proteins | 12 | MG92-6-B transposition protein | protein |
| MG92 transposition proteins | 13 | MG92-7-A transposition protein | protein |
| MG92 transposition proteins | 14 | MG92-7-B transposition protein | protein |
| MG92 transposition proteins | 15 | MG92-8-A transposition protein | protein |
| MG92 transposition proteins | 16 | MG92-9-A transposition protein | protein |
| MG92 transposition proteins | 17 | MG92-9-B transposition protein | protein |
| MG92 transposition proteins | 18 | MG92-10 transposition protein | protein |
| MG92 transposition proteins | 19 | MG92-11 transposition protein | protein |
| MG92 transposition proteins | 20 | MG92-12 transposition protein | protein |
| MG92 transposition proteins | 21 | MG92-13 transposition protein | protein |
| MG92 transposition proteins | 22 | MG92-14 transposition protein | protein |
| MG92 transposition proteins | 23 | MG92-15 transposition protein | protein |
| MG92 transposition proteins | 24 | MG92-17 transposition protein | protein |
| MG92 transposition proteins | 25 | MG92-19 transposition protein | protein |
| MG92 transposition proteins | 26 | MG92-20 transposition protein | protein |
| MG92 transposition proteins | 27 | MG92-21 transposition protein | protein |
| MG92 transposition proteins | 28 | MG92-22 transposition protein | protein |
| MG92 transposition proteins | 29 | MG92-23 transposition protein | protein |
| MG92 transposition proteins | 30 | MG92-24 transposition protein | protein |
| MG92 transposition proteins | 31 | MG92-25 transposition protein | protein |
| MG92 transposition proteins | 32 | MG92-26 transposition protein | protein |
| MG92 transposition proteins | 33 | MG92-27 transposition protein | protein |
| MG92 transposition proteins | 34 | MG92-28 transposition protein | protein |
| MG92 transposition proteins | 35 | MG92-29 transposition protein | protein |
| MG92 transposition proteins | 36 | MG92-30 transposition protein | protein |
| MG92 transposition proteins | 37 | MG92-31 transposition protein | protein |
| MG92 transposition proteins | 38 | MG92-32 transposition protein | protein |
| MG92 transposition proteins | 39 | MG92-33 transposition protein | protein |
| MG92 transposition proteins | 40 | MG92-34 transposition protein | protein |
| MG92 transposition proteins | 41 | MG92-35 transposition protein | protein |
| MG92 transposition proteins | 42 | MG92-36 transposition protein | protein |
| MG92 transposition proteins | 43 | MG92-37 transposition protein | protein |
| MG92 transposition proteins | 44 | MG92-38 transposition protein | protein |
| MG92 transposition proteins | 45 | MG92-39 transposition protein | protein |
| MG92 transposition proteins | 46 | MG92-40 transposition protein | protein |
| MG92 transposition proteins | 47 | MG92-41 transposition protein | protein |
| MG92 transposition proteins | 48 | MG92-42 transposition protein | protein |
| MG92 transposition proteins | 49 | MG92-43 transposition protein | protein |
| MG92 transposition proteins | 50 | MG92-44 transposition protein | protein |
| MG92 transposition proteins | 51 | MG92-45 transposition protein | protein |
| MG92 transposition proteins | 52 | MG92-46 transposition protein | protein |
| MG92 transposition proteins | 53 | MG92-47 transposition protein | protein |
| MG92 transposition proteins | 54 | MG92-48 transposition protein | protein |
| MG92 transposition proteins | 55 | MG92-49 transposition protein | protein |
| MG92 transposition proteins | 56 | MG92-50 transposition protein | protein |

| Cat. | SEQ ID NO: | Description | Type |
|----------------------|------------|--|------------|
| MG92 transposon ends | 352 | MG92-2-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 353 | MG92-2-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 354 | MG92-3-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 355 | MG92-3-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 356 | MG92-4-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 357 | MG92-4-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 358 | MG92-5-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 359 | MG92-5-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 360 | MG92-6-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 361 | MG92-6-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 362 | MG92-7-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 363 | MG92-7-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 364 | MG92-9-A transposon left end (LE) | nucleotide |
| MG92 transposon ends | 365 | MG92-9-A transposon right end (RE) | nucleotide |
| MG92 transposon ends | 366 | MG92-11 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 367 | MG92-11 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 368 | MG92-17 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 369 | MG92-17 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 370 | MG92-20 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 371 | MG92-20 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 372 | MG92-21 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 373 | MG92-21 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 374 | MG92-27 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 375 | MG92-27 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 376 | MG92-28 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 377 | MG92-28 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 378 | MG92-37 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 379 | MG92-37 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 380 | MG92-86 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 381 | MG92-86 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 382 | MG92-136 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 383 | MG92-136 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 384 | MG92-138 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 385 | MG92-138 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 386 | MG92-155, MG92-160 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 387 | MG92-155, MG92-160 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 388 | MG92-157 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 389 | MG92-157 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 390 | MG92-159 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 391 | MG92-159 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 392 | MG92-162 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 393 | MG92-162 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 394 | MG92-163 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 395 | MG92-163 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 396 | MG92-164 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 397 | MG92-164 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 398 | MG92-165 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 399 | MG92-165 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 400 | MG92-172 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 401 | MG92-172 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 402 | MG92-174 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 403 | MG92-174 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 404 | MG92-177 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 405 | MG92-177 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 406 | MG92-183 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 407 | MG92-183 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 408 | MG92-185 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 409 | MG92-185 transposon right end (RE) | nucleotide |

| Cat. | SEQ ID NO: | Description | Type |
|----------------------|------------|--|------------|
| MG92 transposon ends | 410 | MG92-187 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 411 | MG92-187 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 412 | MG92-188 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 413 | MG92-188 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 414 | MG92-189 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 415 | MG92-189 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 416 | MG92-196 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 417 | MG92-196 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 418 | MG92-222 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 419 | MG92-222, MG92-266 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 420 | MG92-224 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 421 | MG92-221 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 422 | MG92-226 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 423 | MG92-226 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 424 | MG92-264 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 425 | MG92-264 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 426 | MG92-266 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 427 | MG92-267 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 428 | MG92-267 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 429 | MG92-272 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 430 | MG92-272 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 431 | MG92-274 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 432 | MG92-274 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 433 | MG92-284 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 434 | MG92-284 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 435 | MG92-288 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 436 | MG92-288 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 437 | MG92-291 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 438 | MG92-291 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 439 | MG92-295 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 440 | MG92-295 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 441 | MG92-302 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 442 | MG92-302 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 443 | MG92-310 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 444 | MG92-310 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 445 | MG92-311 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 446 | MG92-311 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 447 | MG92-312 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 448 | MG92-312 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 449 | MG92-322 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 450 | MG92-322 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 451 | MG92-323 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 452 | MG92-323 transposon right end (RE) | nucleotide |
| MG92 transposon ends | 453 | MG92-344 transposon left end (LE) | nucleotide |
| MG92 transposon ends | 454 | MG92-344 transposon right end (RE) | nucleotide |

[00181] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS**WHAT IS CLAIMED IS:**

1. An engineered transposase system, comprising:
 - (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein said cargo nucleotide sequence is configured to interact with a transposase; and
 - (b) a transposase, wherein:
 - (i) said transposase is configured to transpose said cargo nucleotide sequence to a target nucleic acid locus; and
 - (ii) said transposase is derived from an uncultivated microorganism.
2. The engineered transposase system of claim 1, wherein said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
3. The engineered transposase system of claim 1 or claim 2, wherein said transposase is not a TnpA transposase or a TnpB transposase.
4. The engineered transposase system of any one of claims 1 to 3, wherein said transposase has less than 80% sequence identity to a TnpA transposase.
5. The engineered transposase system of any one of claims 1 to 4, wherein said transposase has less than 80% sequence identity to a TnpB transposase.
6. The engineered transposase system of any one of claims 1 to 5, wherein said transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19.
7. The engineered transposase system of any one of claims 1 to 6, wherein said transposase comprises a catalytic tyrosine residue.
8. The engineered transposase system of any one of claims 1 to 7, wherein said transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.

9. The engineered transposase system of any one of claims 1 to 8, wherein said transposase is configured to transpose said cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.
10. The engineered transposase system of any one of claims 1 to 9, wherein said transposase comprises one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of said transposase.
11. The engineered transposase system of any one of claims 1 to 10, wherein said NLS comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NO: 455-470.
12. The engineered transposase system of any one of claims 1 to 11, wherein said sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm.
13. The engineered transposase system of claim 12, wherein said sequence identity is determined by said BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.
14. An engineered transposase system, comprising:
 - (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein said cargo nucleotide sequence is configured to interact with a transposase; and
 - (b) a transposase, wherein:
 - (i) said transposase is configured to transpose said cargo nucleotide sequence to a target nucleic acid locus; and
 - (ii) said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
15. The engineered transposase system of claim 14, wherein said transposase is derived from an uncultivated microorganism.

16. The engineered transposase system of claim 14 or claim 15, wherein said transposase is not a TnpA transposase or a TnpB transposase.
17. The engineered transposase system of any one of claims 14 to 16, wherein said transposase has less than 80% sequence identity to a TnpA transposase.
18. The engineered transposase system of any one of claims 14 to 17, wherein said transposase has less than 80% sequence identity to a TnpB transposase.
19. The engineered transposase system of any one of claims 14 to 18, wherein said transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19.
20. The engineered transposase system of any one of claims 14 to 19, wherein said transposase comprises a catalytic tyrosine residue.
21. The engineered transposase system of any one of claims 14 to 20, wherein the transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.
22. The engineered transposase system of any one of claims 14 to 20, wherein said transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence.
23. The engineered transposase system of any one of claims 14 to 22, wherein said transposase is configured to transpose said cargo nucleotide sequence as single-stranded deoxyribonucleic acid polynucleotide.
24. The engineered transposase system of any one of claims 14 to 22, wherein said sequence identity is determined by a BLASTP, CLUSTALW, MUSCLE, MAFFT, or CLUSTALW with the parameters of the Smith-Waterman homology search algorithm.

25. The engineered transposase system of claim 24, wherein said sequence identity is determined by said BLASTP homology search algorithm using parameters of a wordlength (W) of 3, an expectation (E) of 10, and a BLOSUM62 scoring matrix setting gap costs at existence of 11, extension of 1, and using a conditional compositional score matrix adjustment.
26. A deoxyribonucleic acid polynucleotide encoding said engineered transposase system of any one of claims 1 to 25.
27. A nucleic acid comprising an engineered nucleic acid sequence optimized for expression in an organism, wherein said nucleic acid encodes a transposase, and wherein said transposase is derived from an uncultivated microorganism, wherein said organism is not said uncultivated microorganism.
28. The nucleic acid of claim 27, wherein said transposase comprises a variant having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
29. The nucleic acid of claim 27 or claim 28, wherein said transposase comprises a sequence encoding one or more nuclear localization sequences (NLSs) proximal to an N- or C-terminus of said transposase.
30. The nucleic acid of claim 29, wherein said NLS comprises a sequence selected from SEQ ID NOs: 455-470.
31. The nucleic acid of claim 29 or 30, wherein said NLS comprises SEQ ID NO: 456.
32. The nucleic acid of claim 31, wherein said NLS is proximal to said N-terminus of said transposase.
33. The nucleic acid of claim 29 or 30, wherein said NLS comprises SEQ ID NO: 455.
34. The nucleic acid of claim 33, wherein said NLS is proximal to said C-terminus of said transposase.
35. The nucleic acid of any one of claims 27 to 34, wherein said organism is prokaryotic,

bacterial, eukaryotic, fungal, plant, mammalian, rodent, or human.

36. A vector comprising said nucleic acid of any one of claims 27 to 35.
37. The vector of claim 36, further comprising a nucleic acid encoding a cargo nucleotide sequence configured to form a complex with said transposase.
38. The vector of claim 36 or claim 37, wherein said vector is a plasmid, a minicircle, a CELiD, an adeno-associated virus (AAV) derived virion, or a lentivirus.
39. A cell comprising said vector of any one of any one of claims 36 to 38.
40. A method of manufacturing a transposase, comprising cultivating said cell of claim 39.
41. A method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide comprising a cargo sequence, comprising:
 - (a) contacting said double-stranded deoxyribonucleic acid polynucleotide with a transposase configured to transpose said cargo nucleotide sequence to a target nucleic acid locus; and
 - (b) wherein said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
42. The method of claim 41, wherein said transposase is derived from an uncultivated microorganism.
43. The method of claim 41 or claim 42, wherein said transposase is not a TnpA transposase or a TnpB transposase.
44. The method of any one of claims 41 to 43, wherein said transposase has less than 80% sequence identity to a TnpA transposase.
45. The method of any one of claims 41 to 44, wherein said transposase has less than 80% sequence identity to a TnpB transposase.
46. The method of any one of claims 41 to 45, wherein said transposase has at least about

80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19.

47. The method of any one of claims 41 to 46, wherein said transposase comprises a catalytic tyrosine residue.
48. The method of any one of claims 41 to 47, wherein said transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.
49. The method of any one of claims 41 to 47, wherein said transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence.
50. The method of any one of claims 41 to 49, wherein said double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide.
51. The method of any one of claims 41 to 50, wherein said double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.
52. A method of modifying a target nucleic acid locus, said method comprising delivering to said target nucleic acid locus said engineered transposase system of any one of claims 1 to 25, wherein said transposase is configured to transpose said cargo nucleotide sequence to said target nucleic acid locus, and wherein said complex is configured such that upon binding of said complex to said target nucleic acid locus, said complex modifies said target nucleic acid locus.
53. The method of claim 52, wherein modifying said target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing said target nucleic acid locus.

54. The method of claim 52 or claim 53, wherein said target nucleic acid locus comprises deoxyribonucleic acid (DNA).
55. The method of claim 54, wherein said target nucleic acid locus comprises genomic DNA, viral DNA, or bacterial DNA.
56. The method of any one of claims 52 to 55, wherein said target nucleic acid locus is *in vitro*.
57. The method of any one of claims 52 to 55, wherein said target nucleic acid locus is within a cell.
58. The method of claim 57, wherein said cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell.
59. The method of claim 57 or 58, wherein said cell is a primary cell.
60. The method of claim 59, wherein said primary cell is a T cell.
61. The method of claim 59, wherein said primary cell is a hematopoietic stem cell (HSC).
62. The method of any one of claims 52 to 61, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering said nucleic acid of any one of claims 27 to 35 or said vector of any of claims 36 to 38.
63. The method of any one of claims 52 to 62, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a nucleic acid comprising an open reading frame encoding said transposase.
64. The method of claim 63, wherein said nucleic acid comprises a promoter to which said open reading frame encoding said transposase is operably linked.
65. The method of any one of claims 52 to 64, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a capped mRNA

containing said open reading frame encoding said transposase.

66. The method of any one of claims 52 to 65, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a translated polypeptide.
67. The method of any one of claims 52 to 66, wherein said transposase induces a single-stranded break or a double-stranded break at or proximal to said target nucleic acid locus.
68. The method of claim 67, wherein said transposase induces a staggered single stranded break within or 5' to said target locus.
69. A host cell comprising an open reading frame encoding a heterologous transposase having at least 75% sequence identity to any one of SEQ ID NOs: 1-349 or a variant thereof.
70. The host cell of claim 69, wherein said transposase has at least 75% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 18-19.
71. The host cell of claim 69, wherein said transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 18-19.
72. The host cell of claim 69, wherein said transposase has at least 75% sequence identity to any one of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, or 17.
73. The host cell of any one of claims 69 to 71, wherein said host cell is an *E. coli* cell.
74. The host cell of claim 73, wherein said *E. coli* cell is a λ DE3 lysogen or said *E. coli* cell is a BL21(DE3) strain.
75. The host cell of claim 73 to 74, wherein said *E. coli* cell has an *ompT lon* genotype.

76. The host cell of any one of claims 69 to 75, wherein said open reading frame is operably linked to a T7 promoter sequence, a T7-lac promoter sequence, a lac promoter sequence, a tac promoter sequence, a trc promoter sequence, a ParaBAD promoter sequence, a PrhaBAD promoter sequence, a T5 promoter sequence, a *cspA* promoter sequence, an *araP_{BAD}* promoter, a strong leftward promoter from phage lambda (pL promoter), or any combination thereof.
77. The host cell of any one of claims 69 to 76, wherein said open reading frame comprises a sequence encoding an affinity tag linked in-frame to a sequence encoding said transposase.
78. The host cell of claim 77, wherein said affinity tag is an immobilized metal affinity chromatography (IMAC) tag.
79. The host cell of claim 78, wherein said IMAC tag is a polyhistidine tag.
80. The host cell of claim 77, wherein said affinity tag is a myc tag, a human influenza hemagglutinin (HA) tag, a maltose binding protein (MBP) tag, a glutathione S-transferase (GST) tag, a streptavidin tag, a FLAG tag, or any combination thereof.
81. The host cell of any one of claims 77 to 80, wherein said affinity tag is linked in-frame to said sequence encoding said transposase via a linker sequence encoding a protease cleavage site.
82. The host cell of claim 81, wherein said protease cleavage site is a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof.
83. The host cell of any one of claims 69 to 82, wherein said open reading frame is codon-optimized for expression in said host cell.
84. The host cell of any one of claims 69 to 83, wherein said open reading frame is provided on a vector.

85. The host cell of any one of claims 69 to 83, wherein said open reading frame is integrated into a genome of said host cell.
86. A culture comprising said host cell of any one of claims 69 to 85 in compatible liquid medium.
87. A method of producing a transposase, comprising cultivating said host cell of any one of claims 69 to 85 in compatible growth medium.
88. The method of claim 87, further comprising inducing expression of said transposase by addition of an additional chemical agent or an increased amount of a nutrient.
89. The method of claim 88, wherein said additional chemical agent or increased amount of a nutrient comprises Isopropyl β -D-1-thiogalactopyranoside (IPTG) or additional amounts of lactose.
90. The method of any one of claims 87 to 89, further comprising isolating said host cell after said cultivation and lysing said host cell to produce a protein extract.
91. The method of claim 90, further comprising subjecting said protein extract to IMAC, or ion-affinity chromatography.
92. The method of claim 91, wherein said open reading frame comprises a sequence encoding an IMAC affinity tag linked in-frame to a sequence encoding said transposase.
93. The method of claim 92, wherein said IMAC affinity tag is linked in-frame to said sequence encoding said transposase via a linker sequence encoding protease cleavage site.
94. The method of claim 93, wherein said protease cleavage site comprises a tobacco etch virus (TEV) protease cleavage site, a PreScission® protease cleavage site, a Thrombin cleavage site, a Factor Xa cleavage site, an enterokinase cleavage site, or any combination thereof.
95. The method of claim 93 or claim 94, further comprising cleaving said IMAC affinity tag by contacting a protease corresponding to said protease cleavage site to said transposase.

96. The method of claim 95, further comprising performing subtractive IMAC affinity chromatography to remove said affinity tag from a composition comprising said transposase.
97. A method of disrupting a locus in a cell, comprising contacting to said cell a composition comprising:
- (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein said cargo nucleotide sequence is configured to interact with a transposase; and
 - (b) a transposase, wherein:
 - (i) said transposase is configured to transpose said cargo nucleotide sequence to a target nucleic acid locus;
 - (ii) said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349; and
 - (iii) said transposase has at least equivalent transposition activity to TnpA transposase in a cell.
98. The method of claim 97, wherein said transposition activity is measured *in vitro* by introducing said transposase to cells comprising said target nucleic acid locus and detecting transposition of said target nucleic acid locus in said cells.
99. The method of claim 97 or claim 98, wherein said composition comprises 20 picomoles (pmol) or less of said transposase.
100. The method of claim 99, wherein said composition comprises 1 pmol or less of said transposase.
101. An engineered transposase system, comprising:
- (a) a double-stranded nucleic acid comprising a cargo nucleotide sequence, wherein said cargo nucleotide sequence is configured to interact with a transposase; and
 - (b) a transposase, wherein
 - (i) said transposase is configured to transpose said cargo nucleotide sequence to a target nucleic acid locus; and

- (ii) said double-stranded nucleic acid comprises a flanking sequence flanking said cargo sequence, wherein said flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.
102. The engineered transposase system of claim 101, wherein said transposase is derived from an uncultivated organism.
103. The engineered transposase system of claim 101 or claim 102, wherein said transposase is not a TnpA transposase or a TnpB transposase.
104. The engineered transposase system of any one of claims 101 to 103, wherein said transposase has less than 80% sequence identity to a TnpA transposase.
105. The engineered transposase system of any one of claims 101 to 104, wherein said transposase has less than 80% sequence identity to a TnpB transposase.
106. The engineered transposase system of any one of claims 101 to 105, wherein said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
107. The engineered transposase system of claim 106, wherein said transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19.
108. The engineered transposase system of any one of claims 101 to 107, wherein said transposase comprises a catalytic tyrosine residue.
109. The engineered transposase system of any one of claims 101 to 108, wherein said transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.
110. The engineered transposase system of any one of claims 101 to 109, wherein said double-

stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide.

111. The engineered transposase system of any one of claims 101 to 110, wherein said transposase comprises one or more nuclear localization signals (NLSs) proximal to an N- or C-terminus of said transposase.
112. The engineered transposase system of claim 111, wherein a NLS of said one or more NLSs comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NOs: 455-470.
113. The engineered transposase system of any one of claims 101 to 112, wherein said double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.
114. The engineered transposase system of any one of claims 101 to 113, wherein said flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350, 352, 355, 356, 359, 361, 362, and 367.
115. The engineered transposase system of any one of claims 101 to 114, wherein said double-stranded nucleic acid comprises another flanking sequence flanking said cargo sequence, wherein said another flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.
116. The engineered transposase system of claim 115, wherein said another flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 351, 353, 354, 357, 358, 360, 363, and 366.

117. The engineered transposase system of claim 115 or claim 116, wherein said flanking sequence flanks a left end of said cargo nucleic acid sequence and wherein said another flanking sequence flanks a right end of said cargo nucleic acid sequence.
118. The engineered transposase system of any one of claims 101 to 117, wherein said transposase is configured to recognize an insertion motif adjacent to said target nucleic acid locus.
119. The engineered transposase system of claim 118, wherein said insertion motif comprises at least three, four, five, or six consecutive nucleotides of the sequence AATGAC.
120. A deoxyribonucleic acid polynucleotide encoding said engineered transposase system of any one of claims 101 to 119.
121. A method for binding, nicking, cleaving, marking, modifying, or transposing a double-stranded deoxyribonucleic acid polynucleotide comprising a cargo sequence, the method comprising:
 contacting said double-stranded deoxyribonucleic acid polynucleotide with a transposase configured to transpose said cargo nucleotide sequence to a target nucleic acid locus; wherein
 said double-stranded deoxyribonucleic acid polynucleotide comprises a flanking sequence flanking said cargo sequence, wherein said flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.
122. The method of claim 121, wherein said transposase is derived from an uncultivated organism.
123. The method of claim 122, wherein said transposase is not a TnpA transposase or a TnpB transposase.
124. The method of any one of claims 121 to 123, wherein said transposase has less than 80% sequence identity to a TnpA transposase

125. The method of any one of claims 121 to 124, wherein said transposase has less than 80% sequence identity to a TnpB transposase.
126. The method of any one of claims 121 to 125, wherein said transposase comprises a sequence having at least 75% sequence identity to any one of SEQ ID NOs: 1-349.
127. The method of claim 126, wherein said transposase has at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to any one of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 18-19.
128. The method of any one of claims 121 to 127, wherein said transposase comprises a catalytic tyrosine residue.
129. The method of any one of claims 121 to 128, wherein said transposase is configured to bind a left-hand region comprising a subterminal palindromic sequence and a right-hand region comprising a subterminal palindromic sequence.
130. The method of any one of claims 121 to 129, wherein said transposase is compatible with a left-hand recognition sequence or a right-hand recognition sequence.
131. The method of any one of claims 121 to 130, wherein said double-stranded deoxyribonucleic acid polynucleotide is transposed as a single-stranded deoxyribonucleic acid polynucleotide.
132. The method of any one of claims 121 to 131, wherein said transposase comprises one or more nuclear localization signals (NLSs) proximal to an N- or C-terminus of said transposase.
133. The method of any one of claims 121 to 132, wherein a NLS of said one or more NLSs comprises a sequence at least 80% identical to a sequence from the group consisting of SEQ ID NOs: 455-470.

134. The method of any one of claims 121 to 133, wherein said double-stranded deoxyribonucleic acid polynucleotide is a eukaryotic, plant, fungal, mammalian, rodent, or human double-stranded deoxyribonucleic acid polynucleotide.
135. The method of any one of claims 121 to 134, wherein said flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350, 352, 355, 356, 359, 361, 362, and 367.
136. The method of any one of claims 121 to 135, wherein said double-stranded deoxyribonucleic acid polynucleotide comprises another flanking sequence flanking said cargo sequence, wherein said another flanking sequence has at least about 70% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 350-454.
137. The method of claim 135, wherein said another flanking sequence has at least about 75%, at least about 80%, at least about 85%, at least about 86%, at least about 87%, at least about 88%, at least about 89%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100% sequence identity to at least 90 consecutive nucleotides of any one of SEQ ID NOs: 351, 353, 354, 357, 358, 360, 363, and 366.
138. The method of claim 135 or claim 137, wherein said flanking sequence flanks a left end of said cargo nucleic acid sequence and wherein said another flanking sequence flanks a right end of said cargo nucleic acid sequence.
139. The method of any one of claims 121 to 138, wherein said transposase is configured to recognize an insertion motif adjacent to said target nucleic acid locus.
140. The method of claim 139, wherein said insertion motif comprises at least three, four, five, or six consecutive nucleotides of the sequence AATGAC.

141. A method of modifying a target nucleic acid locus, said method comprising delivering to said target nucleic acid locus said engineered transposase system of any one of claims 101 to 119, wherein said transposase is configured to transpose said cargo nucleotide sequence to said target nucleic acid locus, and wherein said complex is configured such that upon binding of said complex to said target nucleic acid locus, said complex modifies said target nucleic acid locus.
142. The method of claim 141, wherein modifying said target nucleic acid locus comprises binding, nicking, cleaving, marking, modifying, or transposing said target nucleic acid locus.
143. The method of claim 141 or claim 142, wherein said target nucleic acid locus comprises deoxyribonucleic acid (DNA).
144. The method of claim 143, wherein said target nucleic acid locus comprises genomic DNA, viral DNA, or bacterial DNA.
145. The method of any one of claims 141 to 144, wherein said target nucleic acid locus is *in vitro*.
146. The method of any one of claims 141 to 145, wherein said target nucleic acid locus is within a cell.
147. The method of claim 146, wherein said cell is a prokaryotic cell, a bacterial cell, a eukaryotic cell, a fungal cell, a plant cell, an animal cell, a mammalian cell, a rodent cell, a primate cell, a human cell, or a primary cell.
148. The method of claim 146 or claim 147, wherein said cell is a primary cell.
149. The method of claim 148, wherein said primary cell is a T cell.
150. The method of claim 148, wherein said primary cell is a hematopoietic stem cell (HSC).
151. The method of any one of claims 141 to 150, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a nucleic acid

comprising an open reading frame encoding said transposase.

152. The method of claim 151, wherein said nucleic acid comprises a promoter to which said open reading frame encoding said transposase is operably linked.
153. The method of claim 151 or 152, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a capped mRNA containing said open reading frame encoding said transposase.
154. The method of any one of claims 141 to 153, wherein delivering said engineered transposase system to said target nucleic acid locus comprises delivering a translated polypeptide.
155. The method of any one of claims 141 to 154, wherein said transposase induces a single-stranded break or a double-stranded break at or proximal to said target nucleic acid locus.
156. The method of claim 155, wherein said transposase induces a staggered single stranded break within or 5' to said target locus.

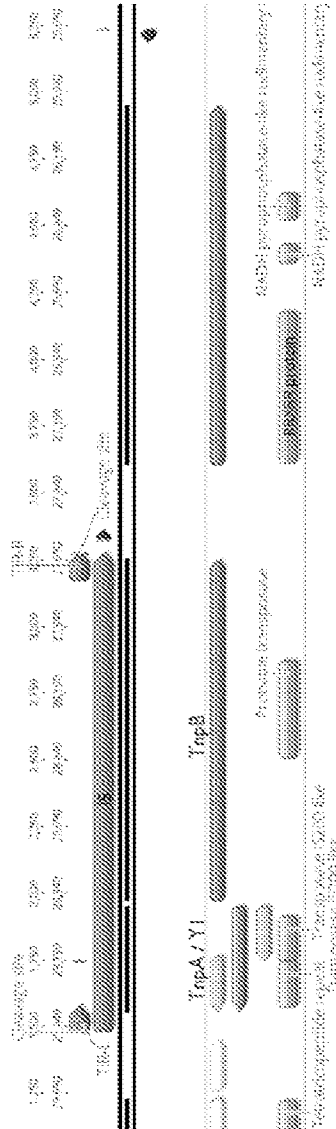
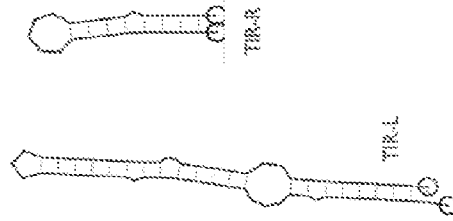


FIG. 1A

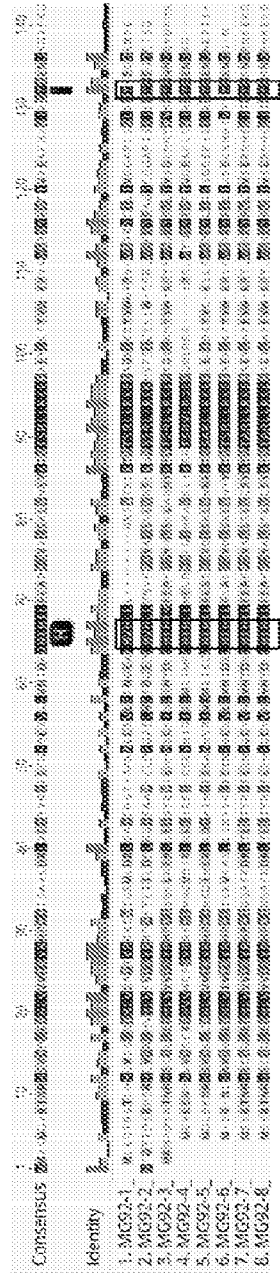


FIG. 1B

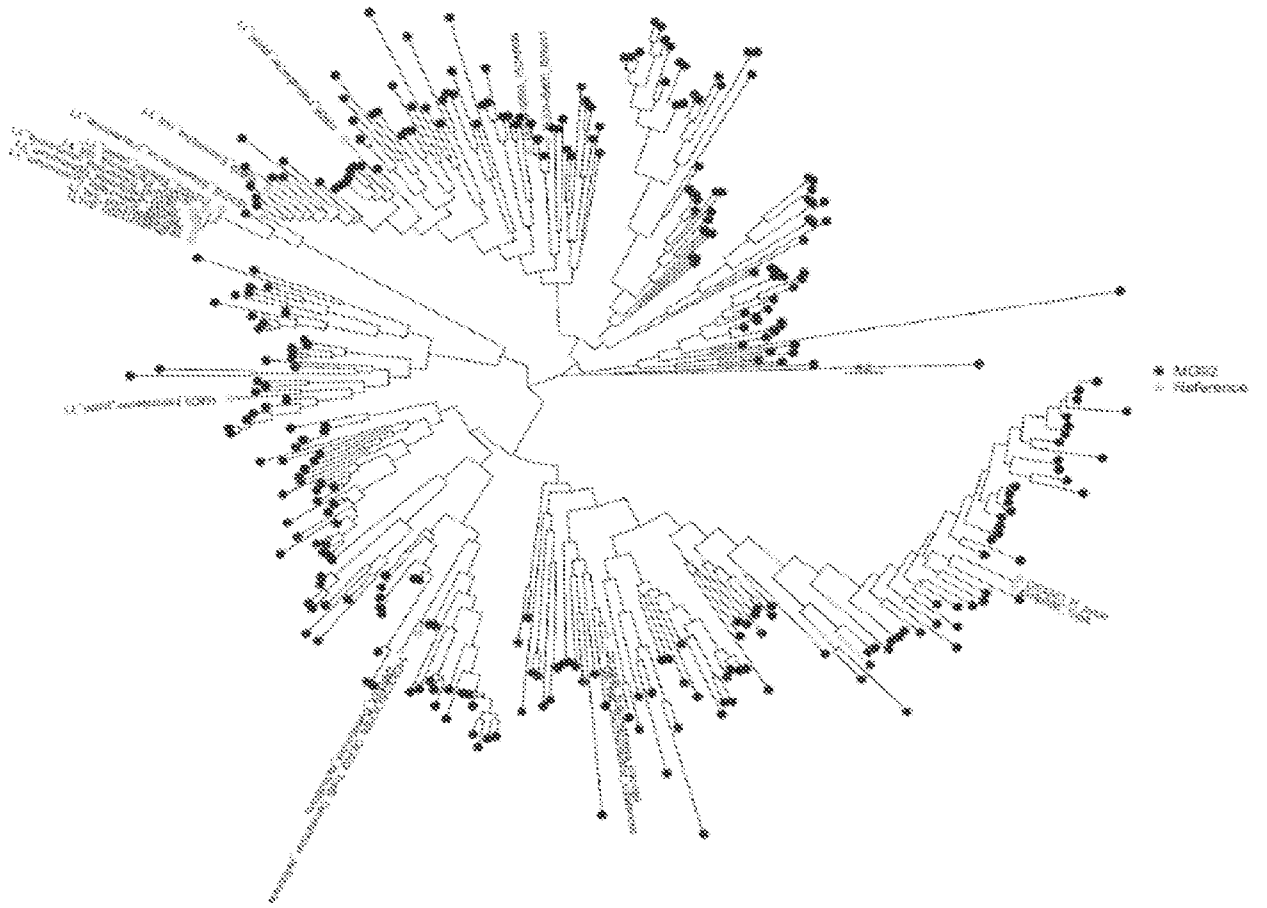


FIG. 2

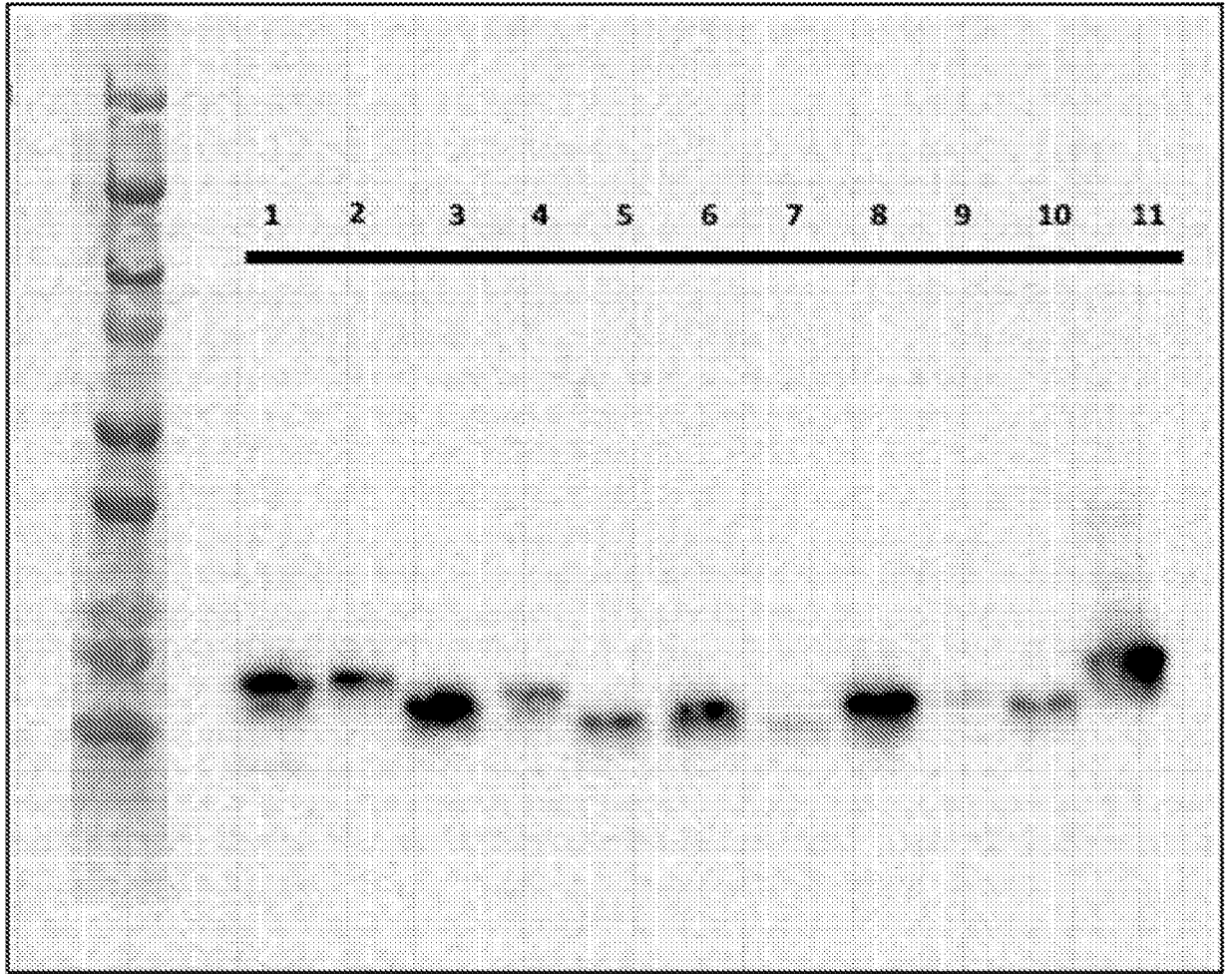


FIG. 4

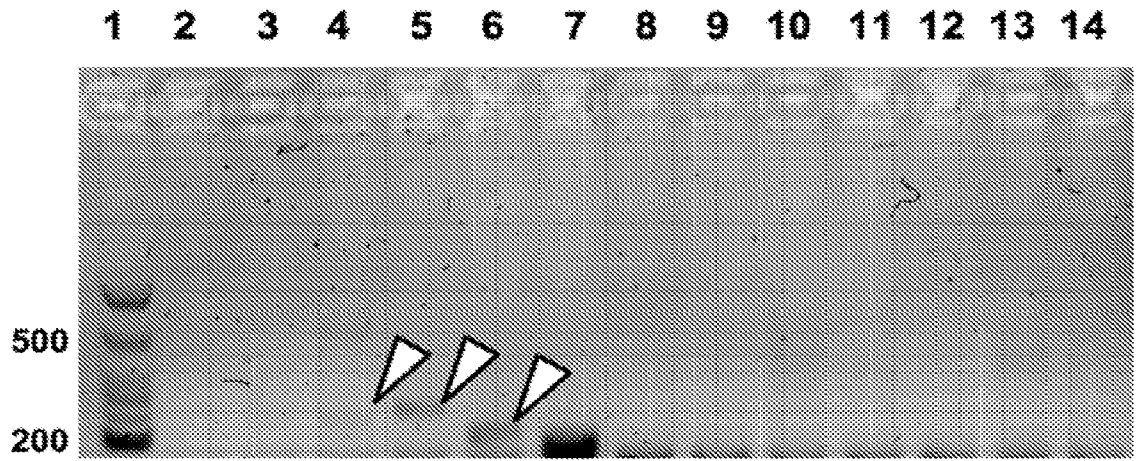


FIG. 5A

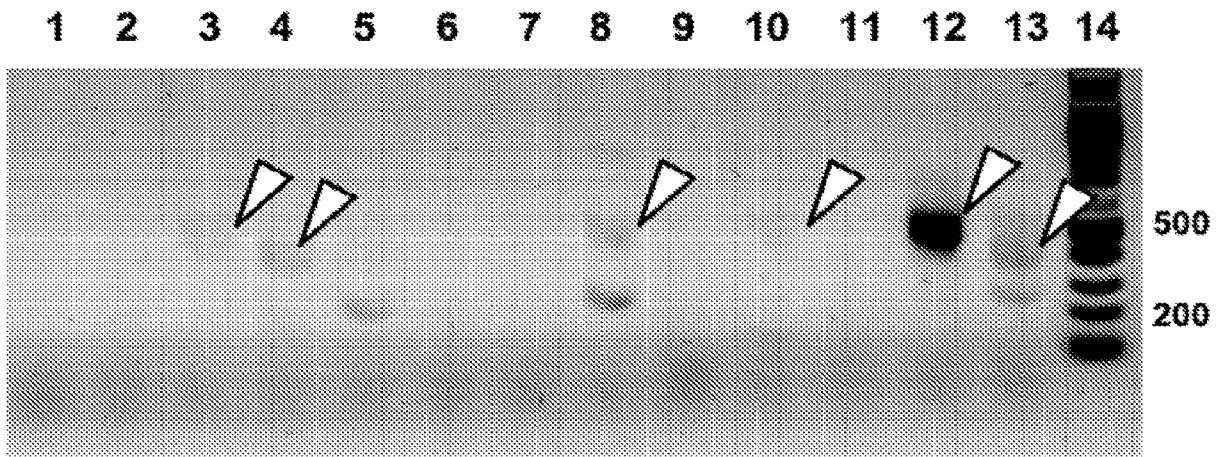


FIG. 5B

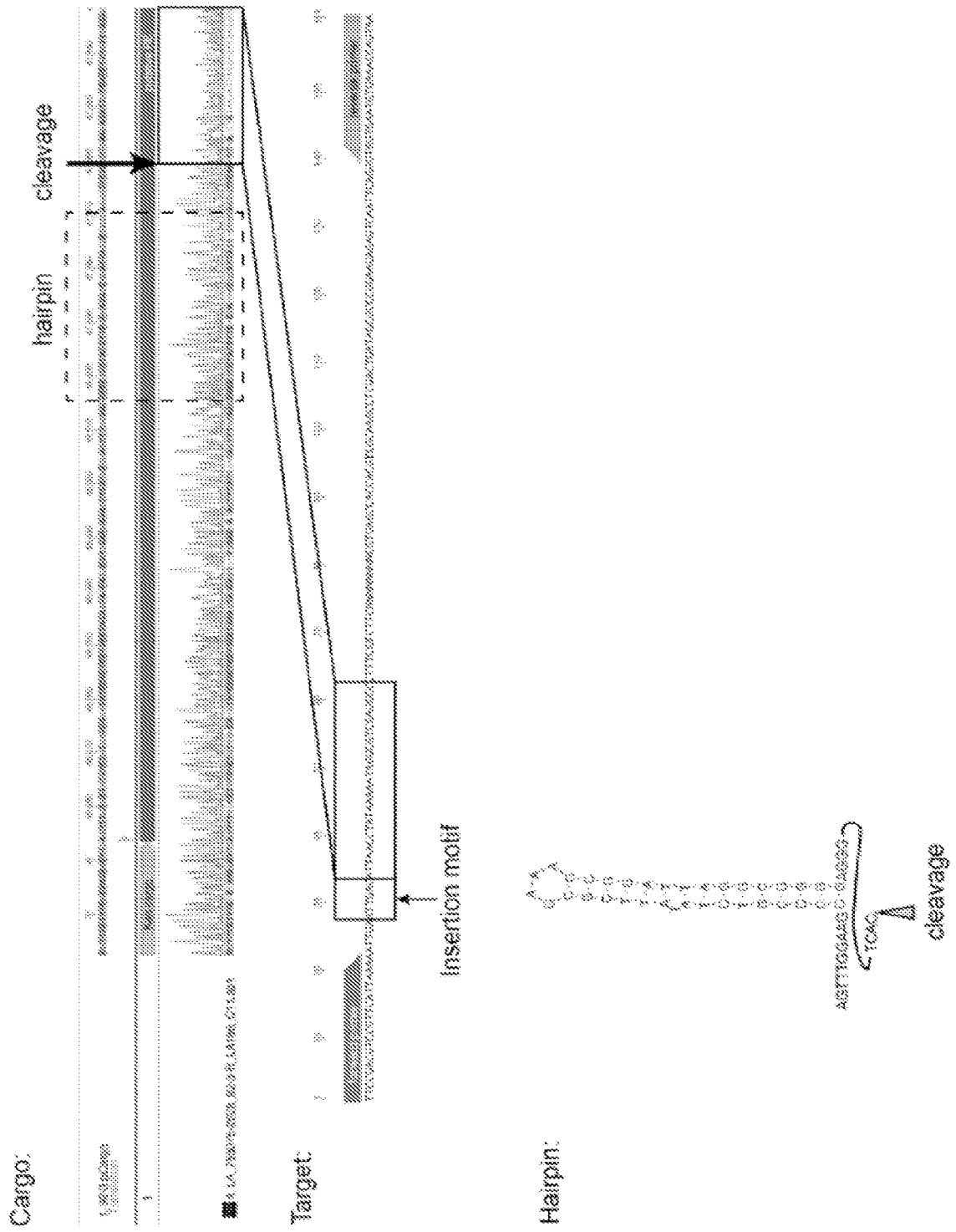


FIG. 7

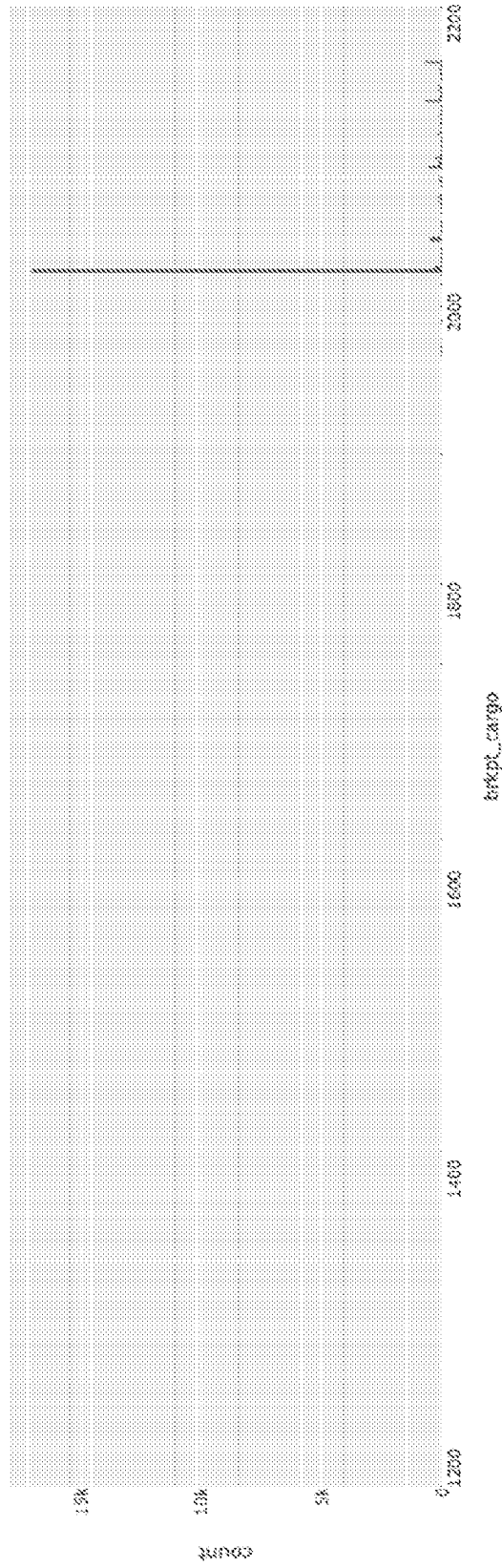


FIG. 8

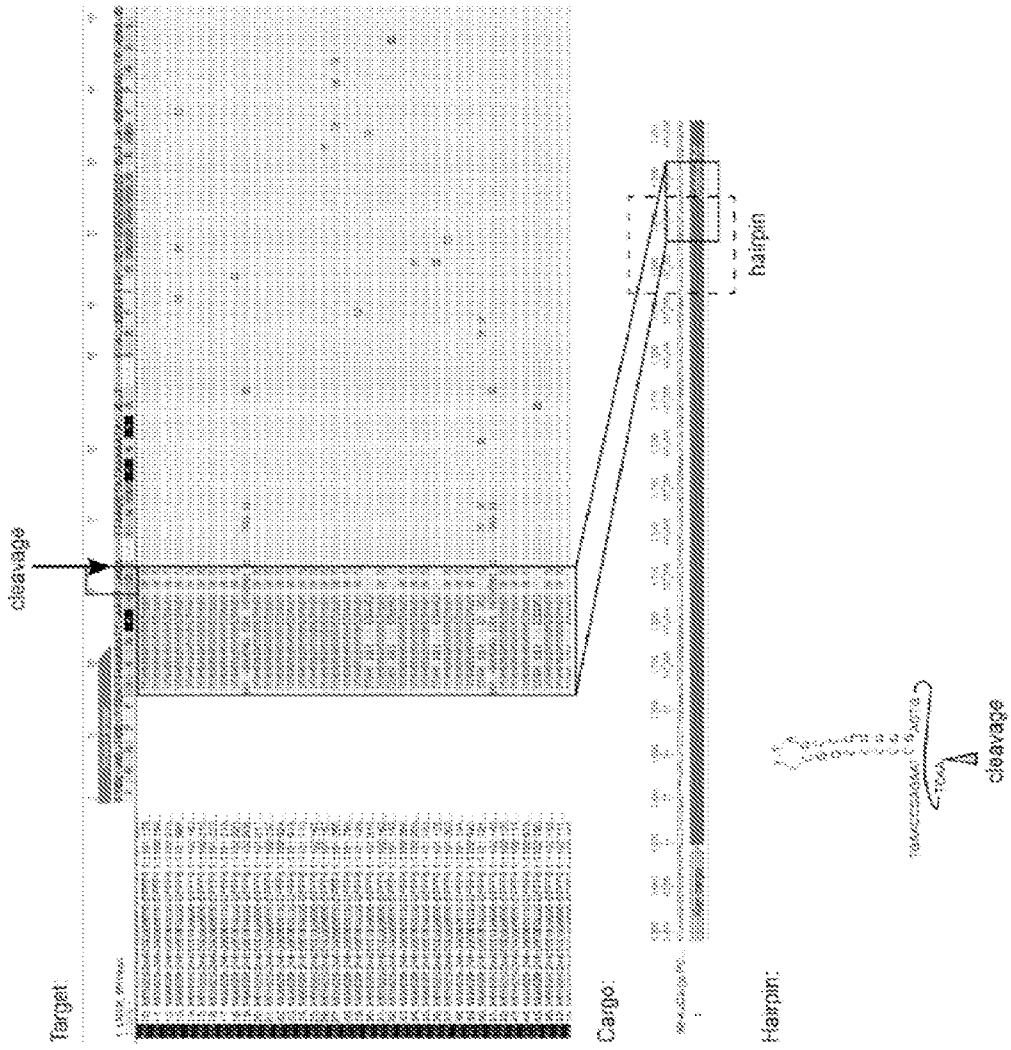


FIG. 9

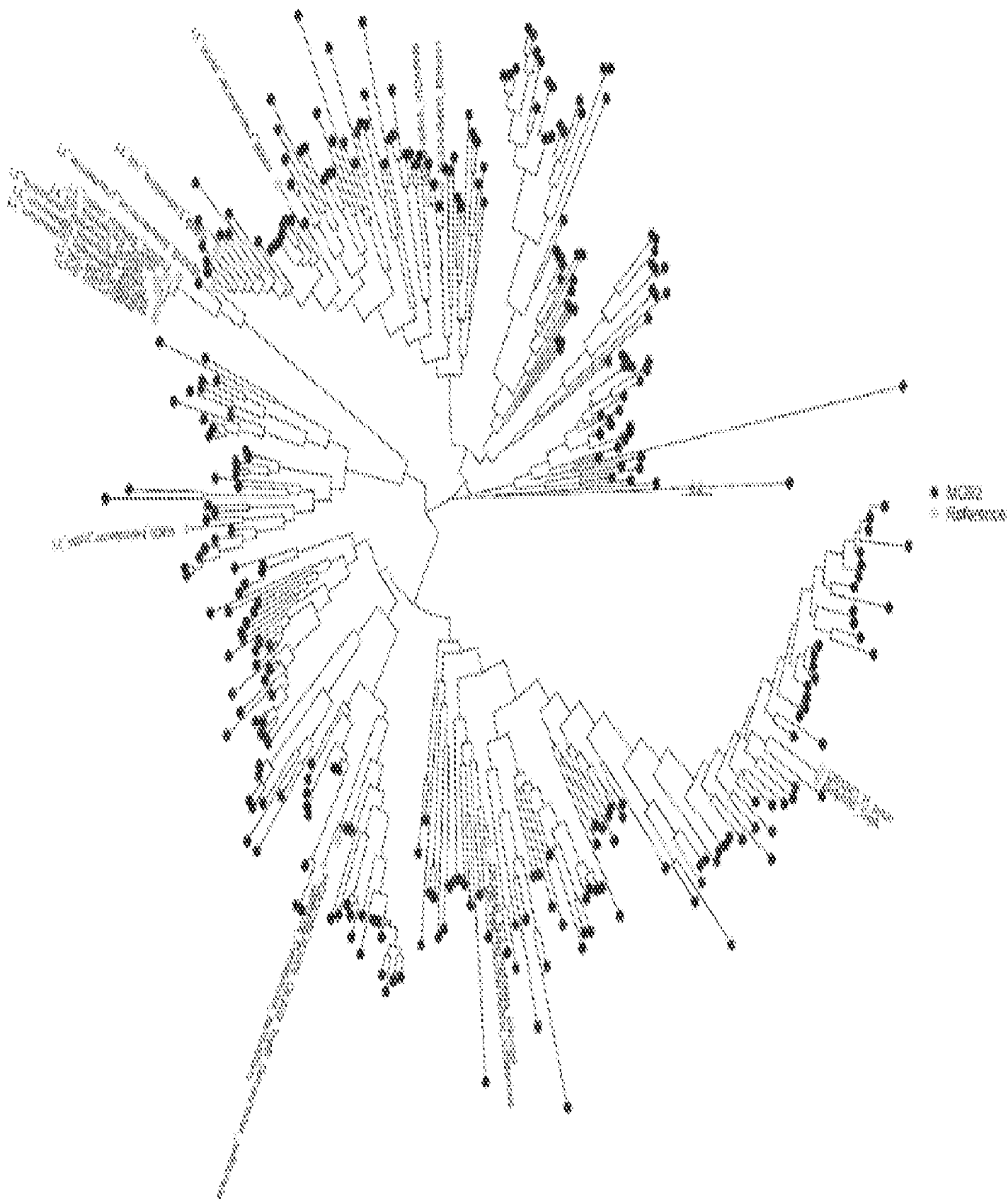


FIG. 2