



(51) International Patent Classification:

C12N 15/10 (2006.01) C12N 5/10 (2006.01)
C12N 15/90 (2006.01)

(21) International Application Number:

PCT/US2017/014234

(22) International Filing Date:

20 January 2017 (20.01.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/281,435 21 January 2016 (21.01.2016) US

(71) Applicants: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). INRIA (FRENCH INSTITUTE FOR RESEARCH IN COMPUTER SCIENCE AND AUTOMATION) [FR/FR]; 23 avenue d'Italie, CS 81321, 75214 Paris Cedex 13 (FR).

(72) Inventors: WEISS, Ron; 90 Olde Field Rd, Newton, MA 02459 (US). DUPORTET, Xavier, Cesar; 17 rue de la Brèche aux Loups, 75012 Paris (FR). BATT, Gregory; 4 rue Amroise Croizat, 78210 Saint Cyr l'Ecole (FR). LI, Yinqing; 540 Memorial Drive, Apt. 605, Cambridge, MA 02139 (US).

(74) Agent: VAN AMSTERDAM, John, R.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

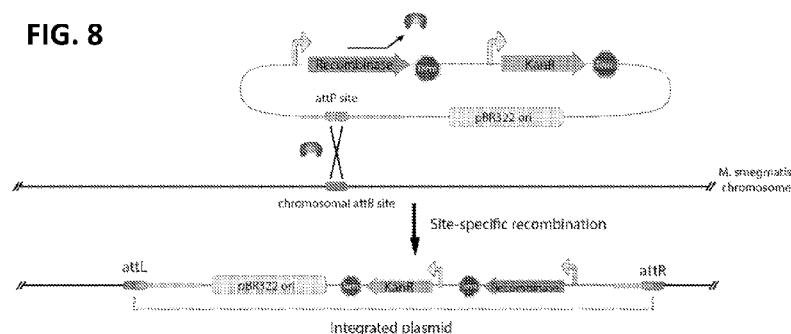
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: NOVEL RECOMBINASES AND TARGET SEQUENCES

FIG. 8



(57) Abstract: The disclosure provides novel large serine recombinases and their respective recognition sites, as well as libraries of orthogonal recombinase recognition sites. Uses of the large serine recombinases, recognition sites, and libraries of orthogonal recombinase recognition sites also are provided.

NOVEL RECOMBINASES AND TARGET SEQUENCES

RELATED APPLICATION

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional
5 Application Serial No. 62/281,435, filed on January 21, 2016, the entire content of which is
herein incorporated by reference.

FIELD OF THE DISCLOSURE

Some aspects of this present disclosure relate to discovering novel large serine
10 recombinases and their respective recognition sites.

BACKGROUND

Programming mammalian cells with large synthetic gene networks is expected to play
a central role in helping elucidate complex regulatory cellular mechanisms (41, 42, 13, 43),
15 implementing new useful biological functions (44, 14, 45) and accelerating the design of
novel tailor-made therapeutic treatments (46, 47, 17). Complex, stable and heritable
programming of mammalian cells by genomic engineering is limited by the requirement to
pre-integrate a natural recombination site at single or multiple genomic (chromosomal) loci,
thus necessitating the identification of programmable orthogonal (independently acting)
20 recombinases that can be directly targeted sequentially or simultaneously to the endogenous
sequences of choice in the mammalian genome.

SUMMARY

Provided herein are newly identified large serine recombinases from recently
25 sequenced Mycobacteriophage genome and the characterization of their respective
recombination sites (attP and attB). A number of these new large serine recombinases had a
significantly different amino acid sequence than the well-characterized BxB1. In addition, the
nucleotide sequences preceding the recombinase coding sequences were also significantly
different compared to the sequence at the same location from BxB1, indicating that these
30 recombinases would use different attP and attB sites than BxB1. A dedicated plasmid rescue
system was designed and used to discover the specific attB/attP recombination sites
recognized by these recombinases. New attP/attB site pairs for Mycobacteriophages were

identified. These sites are used by four of the newly identified large serine recombinases, Theia, Benedict, Veracruz and Rebeuca, to stably integrate their own genome into *M. smegmatis* chromosome. The functioning of the four above-mentioned recombinases in heterologous hosts *E. coli* and mammalian cells were validated. Libraries of orthogonal sites for each of these recombinases were created, which could be used in parallel at the same time, allowing multiplex integration of different genetic circuits into the host genome. These recombinases also provides basis for engineering cell lines step by step, or developing new DNA assembly methods.

Thus, in one aspect, methods for identifying a genomic insertion site of an integrated plasmid are provided. The methods include obtaining genomic DNA from bacteria that include an integrated plasmid, wherein the plasmid includes a promoter operably linked to a nucleotide sequence encoding a large serine recombinase, an attP site recognized by the large serine recombinase, a bacterial origin of replication, one recognition site for a DNA restriction enzyme, and a promoter operably linked to a nucleotide sequence encoding a selectable marker. The methods also include digesting the genomic DNA with the DNA restriction enzyme to obtain fragments of the genomic DNA, ligating the fragments of the genomic DNA to recircularize plasmid sequences, wherein the recircularized plasmid includes the selectable marker, the bacterial origin of replication, and genomic DNA flanking the genomic insertion site, transforming transformation-competent bacteria with the ligated fragments, culturing the transformed bacteria on media that selects for growth of bacteria transformed with the recircularized plasmid that includes the selectable marker, isolating the recircularized plasmid, and determining the sequence of the recircularized plasmid to identify the genomic insertion site.

In some embodiments, the DNA restriction enzyme is *EcoRI*. In some embodiments, the step of ligating the fragments of the genomic DNA is performed under conditions that favor intramolecular ligation. In some embodiments, the transformation-competent bacteria are transformation-competent *E. coli*. In some embodiments, the genomic insertion site is a chromosomal attB site. In some embodiments, the recognition site for the DNA restriction enzyme is located between the nucleotide sequence encoding a large serine recombinase and the attP site recognized by the large serine recombinase. In some embodiments, the methods also include identifying the attB and attP sites at the genomic insertion site.

According to another aspect, reporter plasmids for monitoring recombination events between large serine recombinase recognition sites are provided. The reporter plasmids include, in order, a predicted attB site sequence for a large serine recombinase, a gene cassette expressing a detectable marker, a predicted attP site sequence for the large serine recombinase, a gene cassette expressing the large serine recombinase, and a gene cassette expressing a selectable marker, wherein each gene cassette includes a promoter operably linked to a sequence encoding a polypeptide, followed by a terminator. In some embodiments, the detectable marker is a lacZ nucleotide sequence.

Another aspect provides methods for testing for function of predicted large serine recombinase recognition sites, including obtaining bacterial cells transformed with the reporter plasmid, culturing the bacterial cells on media that permits detection of expression of the detectable marker and to express the large serine recombinase, wherein if the predicted recombinase recognition sites are correct, then recombination between attP and attB sites results in excision of the gene cassette expressing the detectable marker, and wherein if the predicted recombinase recognition sites are not correct, then the gene cassette expressing the detectable marker is not excised. In some embodiments, detectable marker is a lacZ nucleotide sequence and the media that permits detection includes 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal). In some embodiments, the bacteria are *E. coli*.

In another aspect, reporter systems are provided for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells. A reporter system includes a first plasmid and a second plasmid. The first plasmid includes a gene cassette that includes a constitutive promoter operably linked to a nucleotide sequence encoding a first detectable marker, and an attB site of the large serine recombinase placed in between the constitutive promoter and the first detectable marker coding sequence. The second plasmid includes a promoter-less gene cassette that includes an attP site of the large serine recombinase followed by a nucleotide sequence encoding a second detectable marker, wherein the second detectable marker is not the same as the first detectable marker. In some embodiments, the reporter system also includes a plasmid expressing the recombinase. In some embodiments, the first detectable marker is a first detectable protein. In some embodiments, the first detectable protein is a first fluorescent protein. In some embodiments, the fluorescent protein is eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, or mHoneydew. In some

embodiments, the constitutive promoter is a Hef1a promoter, a CMV promoter, or a SV40 promoter. In some embodiments, the second detectable marker is a second detectable protein. In some embodiments, the second detectable protein is a second fluorescent protein. In some embodiments, the second fluorescent protein is eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, or mHoneydew.

Another aspect provides methods for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells. The methods include co-transfecting the first plasmid and the second plasmid together with a plasmid expressing the corresponding large serine recombinase, wherein intermolecular site - specific recombination between the attB and the attP sites results in fusion of the first plasmid and the second plasmid which results in the insertion of the coding sequence of the second detectable marker in frame with the constitutive promoter.

In another aspect, libraries of orthogonal large serine recombinase recognition site sequences, including pairs of attB site sequences and attP site sequences recognized by large serine recombinases, are provided. The attB site sequences in the libraries include a central dinucleotide sequence and at least 5 nucleotides upstream of the central dinucleotide sequence and at least 5 nucleotides downstream of the central dinucleotide sequence. The attP site sequences in the library include a central dinucleotide sequence, at least 5 nucleotides upstream of the central dinucleotide sequence and at least 5 nucleotides downstream of the central dinucleotide sequence. Each pair of attB site sequences and attP site sequences includes one attB site sequence for use in a host genome and one attP site sequence for use in an integrative vector. In the libraries disclosed herein, each pair of attP site sequences and attB site sequences share the same central dinucleotide, and crosstalk between the attB site sequences and the attP site sequences in the libraries is less than about 20%.

In some embodiments, the central dinucleotide is nonpalindromic. In some embodiments, the central dinucleotide is AA, TT, GG, CC, AG, GA, AC, CA, TG, GT, TC, or CT. In some embodiments, the sequences upstream and downstream of the central dinucleotide of the attB site include additional recognition sequences for the large serine recombinase. In some embodiments, the sequences upstream of the central dinucleotide of the attB sites are 5-50 nucleotides in length. In some embodiments, the sequences downstream of the central dinucleotide of the attB sites are 5-50 nucleotides in length. In some embodiments,

the sequences upstream and downstream of the central dinucleotide of the attP site include additional recognition sequences for the recombinase. In some embodiments the sequences upstream of the central dinucleotide of the attP sites are 5-50 nucleotides in length. In some embodiments, the sequences downstream of the central dinucleotide of the attP sites are 5-50 nucleotides in length.

In some embodiments, the large serine recombinase, whose attP and attB sites are included in the libraries, is BxB1, Peaches, Veracruz, Rebeuca, Theia, Benedict, or PhiC31. In some embodiments, a of attB site sequences and attP site sequences can be: SEQ ID NO: 79 and SEQ ID NO: 79, SEQ ID NO: 80 and SEQ ID NO: 80, SEQ ID NO: 81 and SEQ ID NO: 81, SEQ ID NO: 82 and SEQ ID NO: 82, SEQ ID NO: 83 and SEQ ID NO: 83, SEQ ID NO: 84 and SEQ ID NO: 84, SEQ ID NO: 85 and SEQ ID NO: 91, SEQ ID NO: 86 and SEQ ID NO: 92, SEQ ID NO: 87 and SEQ ID NO: 93, SEQ ID NO: 88 and SEQ ID NO: 94, SEQ ID NO: 89 and SEQ ID NO: 95, SEQ ID NO: 90 and SEQ ID NO: 96, SEQ ID NO: 97 and SEQ ID NO: 103, SEQ ID NO: 98 and SEQ ID NO: 104, SEQ ID NO: 99 and SEQ ID NO: 105, SEQ ID NO: 100 and SEQ ID NO: 106, SEQ ID NO: 101 and SEQ ID NO: 107, SEQ ID NO: 102 and SEQ ID NO: 108, SEQ ID NO: 109 and SEQ ID NO: 115, SEQ ID NO: 110 and SEQ ID NO: 116, SEQ ID NO: 111 and SEQ ID NO: 117, SEQ ID NO: 112 and SEQ ID NO: 118, SEQ ID NO: 113 and SEQ ID NO: 119, SEQ ID NO: 114 and SEQ ID NO: 120, SEQ ID NO: 121 and SEQ ID NO: 127, SEQ ID NO: 122 and SEQ ID NO: 128, SEQ ID NO: 123 and SEQ ID NO: 129, SEQ ID NO: 124 and SEQ ID NO: 130, SEQ ID NO: 125 and SEQ ID NO: 131, SEQ ID NO: 126 and SEQ ID NO: 132, SEQ ID NO: 133 and SEQ ID NO: 139, SEQ ID NO: 134 and SEQ ID NO: 140, SEQ ID NO: 135 and SEQ ID NO: 141, SEQ ID NO: 136 and SEQ ID NO: 142, SEQ ID NO: 137 and SEQ ID NO: 143, SEQ ID NO: 144 and SEQ ID NO: 150, SEQ ID NO: 145 and SEQ ID NO: 151, SEQ ID NO: 146 and SEQ ID NO: 152, SEQ ID NO: 147 and SEQ ID NO: 153, SEQ ID NO: 148 and SEQ ID NO: 154, or SEQ ID NO: 149 and SEQ ID NO: 155.

Another aspect provides methods for site-specific recombination between specific attB and attP sites of a large serine recombinase. The methods include providing a nucleic acid molecule including an attB site from a pair of the attB and attP sites from the libraries, providing a nucleic acid molecule including an attP site from the same pair of the attB and attP sites from the library, and combining the nucleic acid molecules in the presence of the large serine recombinase. In some embodiments, the nucleic acid molecule includes an attP

site is a bacteriophage genome or a plasmid. In some embodiments, the nucleic acid molecule includes an attB site is a bacterial genome or a mammalian genome.

In yet another aspect, cell lines comprising a plurality of landing pads integrated into the genomic DNA of a parental cell line are provided. Each landing pad includes a
 5 constitutive promoter operably linked to a nucleotide sequence encoding a detectable marker, followed by a nucleotide sequence encoding a first selectable marker, and an attB site of a large serine recombinase, wherein the attB site is between the promoter and the nucleotide sequence encoding the detectable protein. The attB sites in the plurality of landing pads are orthogonal to each other, and crosstalk between the attB site sequences is less than about
 10 20%.

In some embodiments, the number of landing pads is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12. In some embodiments, each landing pad is integrated into the genomic DNA of the cell line using a zinc-finger nuclease, TALEN or CRISPR-Cas system. In some embodiments, the attB site is SEQ ID NOs: 79-84, 85-90, 97-102, 109-114, 121-126, 133-138, and 144-148, or
 15 149. In some embodiments, the detectable marker is a fluorescent protein. In some embodiments, the fluorescent protein is EYFP, eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, or mHoneydew. In some embodiments, the first selectable marker hydrolyzes a first drug. In some embodiments, first drug is puromycin, hygromycin, G418, neomycin, or bleomycin. In
 20 some embodiments, the large serine recombinase that recognizes the attB sites in the landing pads integrated into the cell lines is BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71),
 25 Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), or Severus (SEQ ID NO: 57).

According to another aspect, methods of integrating a plurality of genetic sequences or circuits into a cell are provided. The methods include delivering into the cell line any one
 30 of claims E1-E8 a plurality of integrative vectors, each including an attP site, one or more genetic sequences or gene cassettes that express components of the genetic circuit, and a promoterless second selectable marker, wherein the cell expresses the large serine

recombinase, and wherein each attP site is selected to allow site - specific recombination between the attB site in one of the landing pads and the attP site in the integrative vector, which results in the insertion of the coding sequence of the second selectable marker in frame with the constitutive promoter in the landing pad, and screening for cells with the integration of the genetic sequence or circuit in media including a drug to which the second selectable marker confers resistance.

In some embodiments, the one or more gene cassettes that express components of the genetic circuit include a promoter operably linked to a sequence encoding a polypeptide that is part of the genetic circuit. In some embodiments the integrative vector is introduced into the cell by CaPO₄ transfection, lipid transfection, electroporation, or lentiviral vector infection. In some embodiments wherein the drug in the media is puromycin, hygromycin, G418, neomycin, or bleomycin.

In some embodiments, the large serine recombinase used to integrate the genetic sequences or circuits into the landing pads is BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), or Severus (SEQ ID NO: 57).

Another aspect provides methods for site-specific recombination between specific attB and attP sites of a large serine recombinase. The methods include providing a first nucleic acid molecule comprising an attB site, providing a second nucleic acid molecule comprising an attP site that selectively recombines with the attB site, and combining the nucleic acid molecules in the presence of a large serine recombinase, which can be: Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), Severus (SEQ ID NO: 57), Sarfire (SEQ ID NO: 53), Rey (SEQ ID NO: 51), Bongo (SEQ ID

NO: 31), Airmid (SEQ ID NO: 23), Hinder (SEQ ID NO: 35), ICleared (SEQ ID NO: 37), Sheen (SEQ ID NO: 59), or Mundrea (SEQ ID NO: 43).

BRIEF DESCRIPTION OF THE DRAWINGS

5 FIG. 1 shows Lambda phage integration and excision. Integration of the circular phage genome into the bacterial genome is mediated by the Lambda phage integrase Int, which interacts with endogenous Integration Host Factor (IHT). Site-specific recombination between the phage designated site (attP) and the bacterial designated site (attB) results in the creation of attL and attR sites flanking the linearized phage genome
10 within the bacterial genome. The reverse reaction happens when Int and Xis are expressed at the same time.

 FIGs. 2A-2B show recombination events mediated by Cre recombinase. FIG. 2A: When both loxP sites are in opposite orientation on the same DNA strand, the sequence located between the sites is constantly inverted by the Cre recombinase. FIG. 2B: When
15 both loxP sites are in the same orientation, the Cre recombinase mediates the excision of the insert. Because the reaction is non-directional, the circular excised insert can be re-linearized, though the intermolecular reaction is less favored compared to intramolecular recombination.

 FIG. 3 shows an overview of the structural and functional variation in the serine
20 recombinases. This figure shows the catalytic domain and E-helix, with S* showing the position of the serine nucleophile; the DNA-binding domain containing a recognizable helix-turn-helix (H-T-H) motif, and conserved domains of unknown function found in subsets of recombinases.

 FIG. 4 shows sequences of PhiC31 and BxB1 attachment sites. For each
25 recombinase, the attB and attP attachment sites share a common central dinucleotide sequence (in gray) that is used for the strand exchange reaction during site-specific recombination. This central dinucleotide sequence ensures the directionality of the reaction. Homologies between the sites are shown with dots between the sequences. This figure depicts SEQ ID NOs: 1 through 8 from top to bottom, respectively.

30 FIG. 5 shows conserved motifs within the catalytic domain and dimerization helix (α E) of large serine recombinases. Motifs A and C contain the critical active site residues of the recombinase. Motif D, contained within the C-terminal portion of the E-helix plus a few

residues beyond, is mostly involved with binding the DNA in the region abutting the cleavage site. Motif B forms a rather mobile loop whose function remains unknown despite the remarkable conservation of the Ser-39, Gly-40, and Arg-45 residues. This figure depicts SEQ ID NOs: 156-159 from left to right, respectively.

5 FIG. 6 shows a phylogenetic tree of identified large serine recombinase sequences, including the 26 recombinases together with Peaches and BxZ2. The scale bar indicates a difference of 2 amino acids.

10 FIG. 7 shows details of the integrative vector architecture. The vector is composed of three main components: the recombinase, the gene cassette encoding the selective marker, and a sequence carrying the attP sequence. The pBR322 ori has been chosen so that the vector can replicate in *E. coli* but not in *M. smegmatis*.

15 FIG. 8 shows the integration of the integrative vector into *M. smegmatis* genome. After transformation of the integrative vector in *M. smegmatis*, the putative serine recombinase is expressed and mediates site-specific recombination between the attP site located on the integrative vector and its counterpart, the attB site on the *M. smegmatis* genome, resulting in the linearization of the integrative vector and its stable integration into the genome of *M. smegmatis*, which then becomes resistant to Kanamycin.

20 FIG. 9 shows the representation of the primers used to check chromosomal integration of the vector. Fw1 and rev1 primer pair was used to confirm the presence of the Kanamycin resistance gene. Fw2 and rev2 primer pair was used to confirm that integration of the vector mediated by site-specific recombination occurred at an attP site within the predicted attP region sequence carried by the vector.

 FIG. 10 shows a summary of functional recombinases in *M. smegmatis*.

25 FIGs. 11A-11E show the plasmid-rescue workflow. FIG. 11A: The genome sequence of *M. smegmatis* carries multiple cleavage sites for EcoRI (arrows), while the linearized integrated vector carries a single EcoRI restriction site. FIG. 11B: Genomic DNA is extracted from a resistant clone and digested with EcoRI, leading to fragments of linearized DNA pieces, among which one contains both part of the *M. smegmatis* genome and part of the integration vector with the origin of replication and the gene cassette encoding the selective marker. FIG. 11C: Ligation of the fragments results in the re-circularization of the linearized DNA sequences. FIG. 11D: Transformation of the re-circularized DNA into competent *E. coli* cells and selection of the bacteria containing the re-circularized DNA

containing the pBR322 origin of replication and the gene cassette encoding the selective marker (from the integrative vector) and the attB site sequence (from *E. coli* genome DNA).

FIG. 11E: Sequencing of the ligated vector reveals the sequence of the recombination site.

FIG. 12 shows the localization of the attB sites for different large site-specific recombinases from Mycobacteriophages.

FIG. 13 shows sequences of the attB and attP sites for Veracruz, Rebeuca, Benedict, and Theia. Central dinucleotides are bolded in gray. Homologies between the attP sites are shown with dots between the sequences. Inverted repeats are represented by gray arrows for attB sites. This figure depicts SEQ ID NOs: 9 through 20 from top to bottom, respectively.

FIGs. 14A-14B show an intramolecular site-specific recombination reporter system.

FIGs. 15A-15B show an intermolecular site-specific recombination mammalian reporter system. FIG. 15A: The reporter system is composed of 2 plasmids: 1 plasmid harboring the attP site followed by the mKate2 coding sequence and 1 plasmid carrying the Hef1a promoter followed by the corresponding attB site and the EYFP coding sequence. A third vector to constitutively express the recombinase was also created. FIG. 15B: After co-transfection of these 3 plasmids in mammalian cells, intermolecular site-specific recombination between the plasmids carrying attB and attP site should result in the constitutive expression of mKate2. Depending on the recombination efficiency, some plasmids may still express EYFP a few hours after transfection.

FIGs. 16A-16B show fluorescent microscopy images of mammalian cells 36h after transfection of the intermolecular site-specific recombination reporter system. FIG. 16A: Co-transfection of both the reporter system and the vector expressing the recombinase leads to efficient intermolecular recombination between the 2 plasmids of the reporter system as depicted by expression of mKate2. FIG. 16B: Negative control of the same experiment: only the 2 plasmids of the reporter system were transfected without the vector expressing the recombinase.

FIG. 17 shows crosstalk of different large serine recombinases. Efficient site-specific recombination between attB/attP pair is represented by light gray rectangles. On the opposite side, rectangles in dark gray illustrate a non-functional recombinase for the corresponding attB/attP site pair.

FIG. 18 shows a library of orthogonal recombination sites for 7 large serine-recombinases.

DETAILED DESCRIPTION

Provided herein are new large serine recombinases identified from recently sequenced Mycobacteriophage genomes and recombinase recognition sites therefor, and a dedicated plasmid rescue system to identify the specific attB/attP recombination sites recognized by these recombinases. New attP/attB site pairs for Mycobacteriophages Theia, Benedict, Veracruz, and Rebeuca were identified and used to stably integrate their own genome into the *M. smegmatis* chromosome. The present disclosure further provides that, Benedict and Theia recombinases use the same attB site within the 2-nitropropane dioxygenase CDS; and Rebeuca and Veracruz use the same attB site within the TmrH RNA methyltransferase CDS. While the attB site specificity for these two pairs of recombinases is extremely high, the attP specificity is less stringent as Theia can use Benedict's attP site and vice-versa; and Rebeuca can use Veracruz's attP site and vice-versa.

Also provided herein are procedures to identify new serine recombinases and discover their attB/attP sites, which can also be used for newly sequenced isolated mycobacteriophages in order to extend the toolbox of available recombinases. The present disclosure further provides methods for using the newly discovered large serine recombinases and their recognition sites in heterologous hosts. Also provided herein is a library of orthogonal sites that could be used in parallel at the same time and therefore broaden the spectrum of applications. Programming mammalian cells with large synthetic gene networks can significantly facilitate elucidating complex regulatory cellular mechanisms (41, 42, 13, 43), implementing new useful biological functions (44, 14, 45) and accelerating the design of novel tailor-made therapeutic treatments (46, 47, 17). For example, transcriptional and post-transcriptional regulatory networks may be integrated into a mammalian genome using the methods disclosed herein, to explore how these networks fine tune mammalian gene expression. In yet another example, trigger-controlled transcription factors, which independently control gene expression, and RNA-binding proteins that inhibit the translation of transcripts harboring specific RNA target motifs maybe included in a set of synthetic transcription-translation control devices that could be rewired in a plug-and-play manner, which can then be integrated into mammalian cells using the methods disclosed herein, and allow the cells to execute input-triggered genetic instructions with precision, robustness and computational logic reminiscent of electronic circuits. In some embodiments, an inducible

positive or negative feedback loop maybe integrated into a mammalian cell using the methods disclosed herein, allowing the monitoring and modeling of cellular behavior. In yet another embodiment, therapeutic peptides, RNAs, proteins maybe integrated into cells using the methods disclosed herein, allowing continuing and controlled expression of these therapeutic agents.

Also provided herein, are methods for integrating complex genetic circuits into mammalian chromosomes using landing pads comprising the recognition sites for large serine recombinases, wherein the genetic circuits to be integrated comprise attP sites that are orthogonal to the large serine recombinase.

While intramolecular site-specific recombination has proven useful for excision of constructs integrated in mammalian genomes, intermolecular site-specific recombination can be used to integrate complex circuits into mammalian chromosomes. The ability of multiplex integration by using orthogonal sites could help to integrate a variety of different sequences, including circuits containing more than one sequence each, at different locations within a genome, which is helpful in preventing interference between circuits or attaining higher levels of expression (with a lower coefficient of variation) by integrating the same construct in multiple copies. The method of sequential integration allows engineering cell lines step by step, integrating one circuit after the other so as to incrementally increase the complexity. Such genomic modifications can also be performed in a simultaneous multiplex ("one shot") procedure, providing savings in time and facilitating inter-related or compensatory modifications which must be synchronized in time.

Further provided herein is the use of multiple serine recombinases with orthogonal sites to develop new DNA assembly methods. For instance, a system can be created based on the Gateway assembly methodology, by flanking each part to be assembled with either an attB on its 5' end and an attP on its 3' end. Having six orthogonal sites for one recombinase would allow assembling five parts per reaction, *i.e.* enough for many mammalian transcription units. The assembled vectors could carry sites from another recombinase, and following the same principle, could be used to assemble five transcription units together. Such strategy could be used to assemble even larger constructs with more orthogonal recombinases. As site-specific recombination with serine recombinases is extremely efficient, fast and reliable, such an assembly method would be extremely convenient.

The present disclosure also provides reliable methods for gene transfer via recombinase-mediated cassette exchange (RMCE), wherein predictable expression patterns are achieved by the non-disruptive insertion of a gene cassette at a pre-characterized genomic insertion site. In some embodiments, the genomic insertion site is marked by a recognition sequence, *e.g.*, the attB site of a large serine recombinase disclosed herein, at the flanks of a selection marker. Providing an integrative vector comprising a gene cassette encoding the gene of interest and the attP site of the corresponding large serine recombinase, can cleanly replace the resident cassette under the influence of a site-specific recombinase.

Recombinases

A “recombinase,” as used herein, is a site-specific enzyme that recognizes short DNA sequence(s), which sequence(s) are typically between about 30 base pairs (bp) and 40 bp, and that mediates the recombination between these recombinase recognition sequences, which results in the excision, inversion, or exchange of DNA fragments between the recombinase recognition sequences. Recombinases can be classified into two distinct families: serine recombinases (*e.g.*, resolvases and invertases) and tyrosine recombinases (*e.g.*, integrases), based on the conserved nucleophilic amino acid residue that they use to attack the DNA and which becomes covalently linked to it during strand exchange.

Large serine recombinases are the most efficient, directional, and specific recombinases for DNA integration in mammalian cells. Examples of large serine recombinases provided herein or useful in the products and methods disclosed herein include, but are not limited to, KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Trouble, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum, Severus, Rey, Bongo, Airmid, Benedict, Theia, Hinder, ICleared, Sheen, Mundrea, Veracruz, and Rebeuca, from the recently sequenced *Mycobacteriophage*, and the previously characterized Bxb1, Peaches, PhiC31, and BxZ2. These large serine recombinases can be used individually, or in combination, such as to integrate complex genetic circuits into a host genome.

In some aspects, a large serine recombinase is required for site-specific recombination, wherein DNA strand exchange takes place between segments possessing attB and attP sites, and wherein the recombinase rearranges DNA segments by recognizing and binding to the attB and attP sites, at which they cleave the DNA backbone, exchange the two DNA helices involved and rejoin the DNA strands.

The recombinases that may be used include, without limitation, BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), and Severus (SEQ ID NO: 57).

Recombinase Recognition Sites

Recombination sites are typically between 30 and 200 nucleotides in length and consist of two motifs with a partial inverted-repeat symmetry, which flank a central crossover sequence at which the recombination takes place (see FIG. 13, arrows above the nucleic acid sequence of attB sites). Recombinases bind to these inverted-repeated sequences, which are specific to each recombinase, and are herein referred to as “recombinase recognition sequences,” “recombinase recognition sites,” “attP sites,” “attB sites” or “genomic insertion sites”. Pairs of attB and attP sites share a certain degree of identity, such as at least 80%. In some embodiments, pairs of attB and attP sites share an identity of at least 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or even 100%. In some embodiments, an attB site is present in the genomic DNA sequence and an attP site is present in the DNA sequence integrated into the genomic DNA sequence. In some embodiments, an attB site is in the genomic DNA of bacteria and an attP site is in the DNA sequence of the respective bacteriophage.

During recombination, crossover occurs at the central dinucleotide of the attB/attP sites. The sequence of the central dinucleotide is the sole determinant of the directionality of the recombination. For the recombination to be directional, the central dinucleotide needs to be nonpalindromic. For example, the central dinucleotide sequence found in the attB/attP sites for large serine recombinases, which are strictly directional, can be AA, TT, GG, CC, AG, GA, AC, CA, TG, GT, TC, or CT.

The outcome of recombination depends, in part, on the location and orientation of two short repeated DNA sequences that are to be recombined, typically less than 30 bp long. For example, inversion recombination happens between two short, inverted, repeated DNA

sequences. A DNA loop formation, assisted by DNA bending 5 proteins, brings the two repeat sequences together, at which point DNA cleavage and ligation occur. This reaction is ATP independent and requires supercoiled DNA. The end result of such an inversion recombination event is that the stretch of DNA between the repeated site inverts (*i.e.*, the stretch of DNA reverses orientation) such that what was the coding strand is now the non-coding strand and vice versa. In such reactions, the DNA is conserved with no net gain or no loss of DNA. Conversely, integration (excision) recombination occurs between two short, repeated DNA sequences that are oriented in the same direction. In this case, the intervening DNA is excised/removed.

Methods for identifying recognition sites

The present disclosure provides, in some aspects, methods for characterizing the recombination sites of large serine recombinases. In some embodiments, the method comprises using a nucleic acid construct designated an “integrative plasmid.” The “integrative plasmid”, as defined herein, comprises a promoter operably linked to a nucleotide sequence encoding a large serine recombinase, an attP site recognized by the large serine recombinase, a bacterial origin of replication, one single recognition site for a DNA restriction enzyme, and a promoter operably linked to a nucleotide sequence encoding a selectable marker. The integrative plasmid is transformed into a cell in which the large serine recombinase is expressed, such as a *M. smegmatis* cell, and the integrative plasmid is inserted (*i.e.*, integrated) into the cell’s genomic DNA at the genomic insertion site.

The methods provided herein further include digesting the genomic DNA containing the integrated plasmid with a DNA restriction enzyme. In some embodiments of the present disclosure, the recognition and cleavage site of the DNA restriction enzyme is only present at one single location in the integrated plasmid, *e.g.*, between the nucleotide sequence encoding a large serine recombinase and the attP site recognized by the large serine recombinase. Recognition and cleavage sites are usually present throughout cell genomes. In some embodiments, digesting the genomic DNA containing the integrated plasmid results in DNA fragments, among which one species of fragments contains the sequences of the integrated plasmid and a segment of the genomic sequence where the recombinase recognition site can be found. Once obtained from restriction digestion of the genomic DNA, the DNA fragments are recircularized, delivered into bacteria, and bacteria are selected that can grow on selective

media based on the presence of the selectable marker. In some embodiments, the way of delivering the recircularized DNA into bacteria is by transforming transformation-competent bacteria. Only bacteria that are transformed with the recircularized plasmid containing sequences from the integrated plasmid, which contains the bacteria origin of replication and the selectable marker, can grow on the selective media. Then the recircularized plasmid is isolated from the bacteria and the sequence of the recircularized plasmid is determined to identify the genomic insertion site of the integrated plasmid.

Using this method, recognition sites in the *M. smegmatis* genome for the exemplary large serine recombinases were identified, which are provided herein. Fifteen of the exemplary large serine recombinases integrated into the same genomic attB site as BxB1 (in the groEL coding sequence), two integrated in the TmrH RNA methyltransferase coding sequence (Rebeuca and Veracruz) and two others integrated into the 2-nitropropane dioxygenase coding sequence (FIG. 12). The recognition sequences at the attP and attB sites are disclosed herein. For the recombinases that integrated into the groEL coding sequence, the attP and attB sites are totally homologous to the one recognized by BxB1 recombinase. However, for Rebeuca and Veracruz, even though integration happens in the same attB site, the attP site recognized by Rebeuca is slightly different from the attP site recognized by Veracruz. Similarly, for Theia and Benedict, even though integration happens in the same attB site, the attP site recognized by Theia is slightly different from the attP site recognized by Benedict (FIG. 13).

The recombinases provided herein are not meant to be exclusive examples of recombinases that can be used in certain embodiments.

Testing the site-specific recombinase activity

The ability of a large serine recombinase to function with its predicted recombination sites in a heterologous host allows the possibility to use these recombinases to engineer the genetic composition of an organism. For example, such recombination events may be used in human cells to allow the precise integration and excision of DNA at a predetermined chromosomal locus, thus facilitating the analysis of gene regulation in a particular chromosomal, physiological, or disease related context. Examples of testing the functionality of the new large serine recombinases and its respective recognition sites include, without

limitation, reporter plasmids or reporter systems that monitor the recombinase activities in heterologous hosts.

Reporter plasmid

5 Provided herein are reporter plasmids for monitoring recombination events between large serine recombinase recognition sites comprising, in order, a predicted attB sequence for a large serine recombinase, a gene cassette expressing a detectable marker, a predicted attP site sequence for the large serine recombinase, a gene cassette expressing the large serine recombinase, and a gene cassette expressing a selectable marker, wherein each gene cassette
10 comprises a promoter operably linked to a sequence encoding a polypeptide, followed by a terminator (FIG. 14). The reporter plasmid, when delivered into an appropriate, e.g., *E. coli*, results in the expression of the large serine recombinase. When the attB and attP sites provided on the reporter plasmid are correct sites for the particular large serine recombinase, intramolecular recombination of the reporter plasmid between the attP and attB sites is
15 triggered, leading to the excision of the gene cassette expressing the detectable marker. In one example, LacZ is the detectable marker, and the loss of LacZ gene can be monitored by plating the *E. coli* on media including 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal) and checking for white colonies. Other detectable markets are described elsewhere herein.

20 In some embodiments, a plasmid having similar features as the reporter plasmid disclosed herein may also be used for exerting precise control over the expression of a gene of interest. For example, a gene of interest can replace the LacZ gene cassette in the reporter plasmid. Without limitation, such gene of interest may be a gene encoding a protein or variants whereof, a gene of a non-coding RNA such as a siRNA, or a ribosomal RNA or
25 variants whereof. Wherein such gene of interest encodes a protein, the protein may be, for example, a detectable marker, a selectable marker, a therapeutic protein, a regulatory protein, or an enzyme.

 In yet another embodiment, the excision of the gene of interest can be timed by manipulating the promoter under which the expression of the large serine recombinase is
30 controlled. For example, the gene of the large serine recombinase can be put under the control of an inducible promoter. At a desired time, the expression of the large serine

recombinase is induced and only then will intramolecular recombination occur, excising the gene of interest.

In some aspects, the reporter plasmid provided herein can be modified to have attP and attB site sequences of different lengths. Recombination will only occur when the lengths of the attP and attB sites are sufficiently long for the large serine recombination, thus identifying the minimal sequence required for attP and attB sites for the large serine recombinase to function. For example, the lengths of the attP and attB sites can be: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, or more base pairs.

Reporter system

The present disclosure also provides a reporter system for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells. The reporter system, defined herein, comprises a first plasmid, a second plasmid, and a plasmid express the recombinase.

In some embodiments, the first plasmid comprises a gene cassette that comprises a constitutive promoter operably linked to a nucleotide sequence encoding a first detectable marker, and an attB site of the large serine recombinase placed in between the constitutive promoter and the detectable maker coding sequence. The second plasmid comprises a promoter-less gene cassette that comprises an attP site of the large serine recombinase followed by a nucleotide sequence encoding a second detectable marker, wherein the second detectable marker is not the same as the first detectable marker. In some embodiments, the first detectable marker is a first detectable protein. The first detectable protein can be, for example, a fluorescent protein chosen from the group consisting of: eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew. The constitutive promoter in the first plasmid may be, for example, a Hef1a promoter, a CMV promoter, or a SV40 promoter. The second detectable protein can be a second detectable protein. For example, the second detectable marker can be a second fluorescent protein chosen from the group consisting of: eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine,

mBanana, and mHoneydew. In some embodiments, the first detectable marker and/or the second detectable marker is an RNA molecule, or an enzyme such as beta-galactosidase (encoded by LacZ), horseradish peroxidase, or luciferase.

Also provided herein, is the use of the of the reporter system to test intermolecular site-specific recombination capacity of a large serine recombinase, wherein the first plasmid, the second plasmid, and the plasmid expressing the corresponding large serine recombinase are co-transfected into mammalian cells. The constitutive or induced expression of the recombination leads to the intermolecular recombination of the first plasmid and the second plasmid, further leading to the insertion of the second detectable marker after the constitute promoter in the first plasmid (and the first detectable marker is promoter-less) (FIG. 15). In some embodiments, the recombinase is under the control of a constitutive promoter. In some embodiments, the recombinase is under the control of a inducible promoter to provide control of when the recombination occurs. Similar to the reporter plasmid described above, the reporter system disclosed herein can also be used to control the expression of one or more exogenous proteins in the host cell, such as when the recombinase is placed under an inducible promoter. In some embodiments, the large serine recombinases whose activity is tested in mammalian cells using the reporter system is Veracruz, Rebeuca, Benedict or Theia. In other embodiments, the large serine recombinase used is any large serine recombinase disclosed herein. In other embodiments, the large serine recombinase is Bxb1, Peaches, PhiC31, or BxZ2.

Host organisms in which the recombinase activity can be tested using the reporter systems provided herein or in which gene expression can be manipulated using the recombinases and recombinase recognition sites encompass prokaryotes and eukaryotes, unicellular organisms and multi-cellular organisms. A prokaryote, as defined herein, a single-celled organism that lacks a membrane-bound nucleus, mitochondria, or any other membrane-bound organelle. A eukaryote is any organism whose cells contain a nucleus and other organelles enclosed within membranes, which may be a unicellular organism or a multicellular organism. Examples of a prokaryote that may be used include but are not limited to all bacteria species and their variants, *e.g.*, *Yersinia spp.*, *Escherichia spp.*, *Klebsiella spp.*, *Bordetella spp.*, *Neisseria spp.*, *Aeromonas spp.*, *Francisella spp.*, *Corynebacterium spp.*, *Citrobacter spp.*, *Chlamydia spp.*, *Hemophilus spp.*, *Brucella spp.*, *Mycobacterium spp.*, *Legionella spp.*, *Rhodococcus spp.*, *Pseudomonas spp.*, *Helicobacter*

spp., *Salmonella spp.*, *Vibrio spp.*, *Bacillus spp.*, *Erysipelothrix spp.*, *Salmonella spp.*, *Streptomyces spp.* In some embodiments, the bacterial cells are *Staphylococcus aureus*, *Bacillus subtilis*, *Clostridium butyricum*, *Brevibacterium 15 lactofermentum*, *Streptococcus agalactiae*, *Lactococcus lactis*, *Leuconostoc lactis*, *Streptomyces*, *Actinobacillus*

5 *actinobycetemcomitans*, *Bacteroides*, *cyanobacteria*, *Escherichia coli*, *Helobacter pylori*, *Selenomonas ruminatium*, *Shigella sonnei*, *Zymomonas mobilis*, *Mycoplasma mycoides*, *Treponema denticola*, *Bacillus thuringiensis*, *Staphylococcus lugdunensis*, *Leuconostoc oenos*, *Corynebacterium xerosis*, *Lactobacillus plantarum*, 20 *Streptococcus faecalis*, *Bacillus coagulans*, *Bacillus ceretus*, *Bacillus popilliae*, *Synechocystis strain PCC6803*, *Bacillus*

10 *liquefaciens*, *Pyrococcus abyssi*, *Selenomonas nominantium*, *Lactobacillus hilgardii*, *Streptococcus ferus*, *Lactobacillus pentosus*, *Bacteroides fragilis*, *Staphylococcus epidermidis*, *Zymomonas mobilis*, *Streptomyces phaeochromogenes*, *Streptomyces ghanaensis*, or *Halobacterium strain GRB*. Examples of eukaryotic cells that maybe used include but are not limited to mammalian cells, insect cells, yeast cells (e.g., *Saccharomyces cerevisiae*) and

15 plant cells. In some embodiments, a eukaryotic cell is a vertebrate cell. Examples of vertebrate cells include, without limitation, reproductive cells including sperm, ova and embryonic cells, and non-reproductive cells, including kidney, lung, spleen, lymphoid, cardiac, gastric, intestinal, pancreatic, muscle, bone, neural, brain and epithelial cells. Stem cells, including embryonic stem cells, can also be used. In some embodiments, a non-cellular

20 system or an in vitro system may be used.

Libraries of orthogonal large serine recombinase recognition sites

Also provided herein are libraries of orthogonal attB and attP sequences (SEQ ID NOs: 79-155) A recombinase recognition site is “orthogonal” when it does not significantly

25 recognize the recognition site or nucleotide sequence of a recombinase. Thus, one attB site of a recombinase can be orthogonal to an attB site of a different recombinase. In addition, one pair of attB and attP sites of a recombinase can be orthogonal to another pair of attB and attP sites recognized by the same recombinase. A pair of recombinases are considered orthogonal to each other, as defined herein, when the recognition of each other’s attB or attP

30 site sequences. The lack of recognition of recombinase recognitions sites or pairs of sites by the same recombinase or a different recombinase (also referred to herein as “crosstalk”) is

less than about 20%. The crosstalk can be less than 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1%, or even 0%.

In some embodiments, the attB and/or attP site sequences in the library comprise a central dinucleotide sequence and at least 5 nucleotides upstream and downstream of the central dinucleotide. For example, the sequence upstream of the central dinucleotide in the attP and/attB site sequences can be 5-50 nucleotides. In some embodiments, the sequence upstream of the central dinucleotide in the attP and/attB site sequences are at least 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, or 45 nucleotides. In some embodiments, the sequence upstream of the central dinucleotide in the attP and/attB site sequences are 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides. Likewise, the sequence downstream of the central dinucleotide in the attP and/or attB site sequences can be 5-50 nucleotides. In some embodiments, the sequence downstream of the central dinucleotide in the attP and/attB site sequences are at least 10 nucleotides, 15 nucleotides, 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, or 45 nucleotides. In some embodiments, the sequence downstream of the central dinucleotide in the attP and/attB site sequences are 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides. In some embodiments, the sequence upstream and downstream of the attB sites provide additional recognition sequence for the corresponding large serine recombinase. . In some embodiments, the sequence upstream and downstream of the attP sites provide additional recognition sequence for the corresponding large serine recombinase. In some embodiments, the attB and attP sites in the library are selected from SEQ ID NOs: 79-155.

A “pair of an attB site sequence and an attP site sequence,” and like terms, refers to attB and attP site sequences that share the same central dinucleotide. In some embodiments, the central dinucleotide is nonpalindromic. In some embodiments, the central dinucleotide is selected from the group consisting of: AA, TT, GG, CC, AG, GA, AC, CA, TG, GT, TC, and CT. In some embodiments, a pair of an attB site sequence and an attP site sequence are used in a recombination event, wherein the attB site sequence is used in a host genome and the attP site sequence for use in an integrative vector. In some embodiments, in a pair of an attB site sequence and an attP site sequence, the sequences upstream and downstream of the central

dinucleotide in the attB site share 100% homology with that of the attP site. In some embodiments, in a pair of an attB site sequence and an attP site sequence, the sequences upstream and downstream of the central dinucleotide in the attB site vary from that of the attP site, *e.g.*, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 91%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, or 80% identical in sequence.

An integrative vector, as defined herein, is a nucleotide sequence that is exogenous to a host DNA sequence and is to be inserted into the host DNA sequence via site-directed recombination, wherein the integrative vector comprises an attP site recognized by the respective large serine recombinase and any additional sequence features it comprises. The integrative vector maybe, without limitation, a bacteriophage DNA, a plasmid, a linear DNA fragment, or a synthetic nucleic acid sequence. A host genome that contains the attB site recognized by the respective large serine recombinase, without limitation, maybe a bacteria genome, an insect cell genome, a mammalian cell genome, or a human cell genome.

In some embodiments, the attP and attB site sequences are recognized by a large serine recombinase, such as BxB1, Peaches, Veracruz, Rebeuca, Theia, Benedict, or PhiC31.

Further provided herein, is a method for site-specific recombination between specific attB and attP sites of a large serine recombinase, wherein a pair of an attB site sequence and an attP site sequence are used in two different nucleic acid molecules, and wherein the two nucleic acid molecules are combined in the presence of a large serine recombinase. In some embodiments, the large serine recombinase is BxB1, Peaches, Veracruz, Rebeuca, Theia, Benedict, or PhiC31. In some embodiments, the attB and attP sites in the library are selected from SEQ ID NOs: 79-155. In some embodiments, the nucleic acid including the attP site is, without limitation, a bacteriophage genome or a plasmid. In some embodiments, the nucleic acid including the attB site is, without limitation, a bacterial genome or a mammalian genome.

Use of the libraries of orthogonal attB and attP sites

The libraries of orthogonal attB and attP sites described herein have a variety of uses. For example, a cell line comprising a plurality of landing pads integrated into the genomic DNA of a parental cell line can be constructed. The parental cell line may be a wild type cell line, or a cell line with existing genomic modification. In the latter, the cell line would be “parental” to the cell line generated from further modification of its genomic DNA. A

“landing pad,” is an exogenous DNA sequence integrated into a location of the host genome that includes an attB site of a large serine recombinase. In some instances, the exogenous DNA sequence includes an attB site of a large serine recombinase, a constitutive promoter operably linked to a nucleotide sequence encoding a detectable marker, followed by a nucleotide sequence encoding a first selectable marker. In certain types of landing pads, the attB site is between the promoter and the nucleotide sequence encoding the detectable protein. When there are more than one landing pads used in a given cell, it is preferred that an attB site of one landing pad is orthogonal to an attB site of the same large serine recombinase in any other landing pad. The landing pad is used for further genetic engineering and integration of a nucleic acid molecule of interest via site-specific recombination. The landing pad can be integrated into the parental genome using any method known in the art, such as by using a zinc finger nuclease, TALEN, or the CRISPR-Cas9 system. In some embodiments, the number of landing pads integrated into a cell line is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12. In some embodiments, the attB site in the landing pad is selected from SEQ ID NOs: 79-84, 85-90, 97-102, 109-114, 121-126, 133-138, and 144-149. In some embodiments, the detectable marker in the landing pad is a fluorescent protein, such as, eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew, luciferase, or LacZ. In some embodiments, a selectable marker hydrolyzes a drug, such as, puromycin, hygromycin, G418, neomycin, or bleomycin.

Further provided herein, is a method of integrating a genetic circuit, or multiple genetic circuits, into a cell comprising a plurality of landing pads. In some embodiments, one genetic circuit is integrated into the cell line. In some embodiments, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more circuits may be integrated into the cell line, provided that the number of the landing pads in the cell line is sufficient to accommodate the number of genetic circuits that are to be integrated into the cell line. In some instances, it may be preferred that the number of landing pads is at least the number of genetic sequences or circuits to be integrated. In other instances, a single landing pad can include multiple circuits under the control of different promoters. A genetic circuit, as used herein, is a rationally designed artificial gene regulatory networks with robust function, comprising primary genetic elements, *i.e.*, building blocks. The primary genetic elements may be, without limitation, insulators, promoters, ribosome binding sites, transcriptional activators and repressors, gene coding sequences, 5'UTRs, 3'UTRs, polyA signals, and terminators. Independent modules of a genetic circuit

can be built using the primary genetic elements. Methods of building these genetic circuits are known to those of skill in the art. A library of these independent modules are available for bacterial systems, *e.g.*, switches, time-delayed circuits, cascades, pulse generators, logic gates, spatial patterning and memory devices (33-40). In some aspects, a plurality of landing pads may be integrated into different locations of the genome, allowing modification at multiple loci of the genome via site-specific recombination.

To integrate a genetic sequence or circuit into a cell, one can obtain a cell line with pre-integrated landing pads, deliver into the cells of the cell line an integrative vector comprising an attP site, the genetic sequence or circuit of interest, and an promoter-less second selectable marker. In some embodiments, the attP site used in the integrative vector is selected from SEQ ID NOs: 79-84, 91-96, 103-108, 115-120, 127-132 and 150-155. In some embodiments, the cell also expresses a large serine recombinase that recognizes the attB site sequences in the landing pads and the attP site sequences in the integrative vector. The site-specific recombination mediated by the large serine recombinase between the landing pad and the integrative vector results in the insertion of the coding sequence of the second selectable marker in frame with the constitutive promoter in the landing pad, thereby allowing expression of the second selectable marker. Cells with the integrative vector incorporated into the genomic DNA may be selected using media including a drug to which the second selectable marker confers resistance. Examples of the drug to which the second selectable marker confers resistance include puromycin, hygromycin, G418, neomycin, or bleomycin. In some embodiments, the large serine recombinase may be BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), or Severus (SEQ ID NO: 57).

Gene cassette

A gene cassette, as defined herein, is a nucleic acid sequence that comprises genetic elements necessary for the expression of a gene. In some embodiments, a gene cassette

comprises a promoter, wherein the promoter is operably linked to the nucleic acid encoding a protein, and a terminator. In some embodiments, a gene cassette comprises a nucleic acid encoding a protein and a terminator.

5 Promoter

A “promoter” refers to a control region of a nucleic acid at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter drives transcription of the nucleic acid sequence that it regulates, thus, it is typically located at or near the transcriptional start site of a gene. A promoter may have, for example, a length of
10 100 to 1000 nucleotides.

In some embodiments, a promoter is operably linked to a nucleic acid, or a sequence of a nucleic acid (nucleotide sequence). A promoter is considered to be “operably linked” to a sequence of nucleic acid that it regulates when the promoter is in a correct functional location and orientation relative to the sequence such that the promoter regulates (*e.g.*, to control
15 (“drive”) transcriptional initiation and/or expression of) that sequence.

The promoters that can be used include constitutive promoters or inducible promoters. In some embodiments, a promoter is a “constitutive” promoter. A constitutive promoter refers to an unregulated promoter that allows for continual transcription of its associated gene. In some embodiments of the present disclosure, a gene is “promoter-less”. A promoter-less gene
20 is not transcribed and does not express.

In some embodiments of the present disclosure, a gene is placed under an “inducible promoter”. An “inducible promoter” may be conditioned to endogenous factors or to environmental conditions and external stimuli that can be artificially controlled. Within this group, there are promoters modulated by abiotic factors such as light, oxygen levels, heat,
25 cold and wounding. Since some of these factors are difficult to control outside an experimental setting, promoters that respond to chemical compounds, not found naturally in the organism of interest, are of particular interest. Along those lines, promoters that respond to antibiotics, copper, alcohol, steroids, and herbicides, among other compounds, have been adapted and refined to allow the induction of gene activity at will and independently of other
30 biotic or abiotic factors. The promoters that can be used include, without limitation, the promoters disclosed in (30), the entire contents of which hereby incorporated by reference,

and the promoters disclosed in (31); the entire contents of which also incorporated by reference.

In some embodiments, the promoters used herein may be, without limitation, a Hef1a promoter, a CMV promoter, an SV40 promoter, a luxI promoter, an Hsp70 promoter, a UREX promoter, a LacZ promoter, or a T7 promoter.

Terminator

Provided herein are terminator sequences for use in some embodiments. A “terminator” or “terminator sequence,” as used herein, is a nucleic acid sequence that causes transcription to stop. A terminator may be unidirectional or bidirectional. It is comprised of a DNA sequence involved in specific termination of an RNA transcript by an RNA polymerase. A terminator sequence prevents transcriptional activation of downstream nucleic acid sequences by upstream promoters. The most commonly used type of terminator is a forward terminator. When placed downstream of a nucleic acid sequence that is usually transcribed, a forward transcriptional terminator will cause transcription to abort. In prokaryotic systems, terminators usually fall into two categories (1) rho-independent terminators and (2) rho-dependent terminators. Rho-independent terminators are generally composed of palindromic sequence that forms a stem loop rich in G-C base pairs followed by several T bases. Without wishing to be bound by theory, the conventional model of transcriptional termination is that the stem loop causes RNA polymerase to pause, and transcription of the poly-A tail causes the RNA:DNA duplex to unwind and dissociate from RNA polymerase. In eukaryotic systems, the terminator region may comprise specific DNA sequences that permit site-specific cleavage of the new transcript so as to expose a polyadenylation site. This signals a specialized endogenous polymerase to add a stretch of about 200 A residues (polyA) to the 3' end of the transcript. RNA molecules modified with this polyA tail appear to more stable and are translated more efficiently. Thus, in some embodiments involving eukaryotes, a terminator may comprise a signal for the cleavage of the RNA. In some embodiments, the terminator signal promotes polyadenylation of the message. The terminator and/or polyadenylation site elements may serve to enhance output nucleic acid levels and/or to minimize read through between nucleic acids. Terminators include any terminator of transcription described herein or known to one of ordinary skill in the art.

Examples of terminators include, without limitation, the termination sequences of genes such as, for example, the bovine growth hormone terminator, and viral termination sequences such as, for example, the SV40 terminator, spy, yejM, secG-leuU, thrLABC, rrnB T1, hisLGDCBHAFI, metZWV, rrnC, xapR, aspA and arcA terminator. In some
 5 embodiments, the termination signal may be a sequence that cannot be transcribed or translated, such as those resulting from a sequence truncation.

Selectable marker

As used herein, a selectable marker is a gene introduced into a cell, especially a
 10 bacterium or to cells in culture, that confers a trait suitable for artificial selection. In some embodiments, a selectable marker is a gene that confers resistance to bacteria against an antibiotic, including but not limited to, ampicillin, amphotericin B, carbenicillin, chloramphenicol, erythromycin, kanamycin, gentamycin, neomycin, nystatin, rifampicin, streptomycin, or tetracycline. In some embodiments, a selectable marker is a gene that
 15 confers resistance to a drug to eukaryotic cells, including but not limited to puromycin, hygromycin, G418, neomycin, or bleomycin.

Detectable marker

As used herein, a reporter refers to a protein that can be used to measure gene
 20 expression and generally produce a measurable signal such as fluorescence, luminescence or color. The presence of a reporter in a cell or organism is readily observed. For example, fluorescent proteins (*e.g.*, GFP) cause a cell to fluoresce when excited with light of a particular wavelength, luciferases or horseradish peroxidase cause a cell to catalyze a reaction that produces light, and enzymes such as galactosidase convert a substrate to a colored
 25 product. Reporters include any reporter described herein or known to one of ordinary skill in the art. In some embodiments, the detectable marker is a fluorescent protein, including, without limitation, eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, or mHoneydew.

Insulator

An insulator, as defined herein, is a control element which insulates the transcription of genes placed within its range of action. In its natural position, the chromatin insulator

element presumably buffers the genes and the regulatory machinery of one domain from the cis-acting influence(s) of the chromatin structure and the regulatory machinery of an adjacent domain. An insulator can be a “pure” insulator from higher eukaryotes, which on its own, does not perturb gene expression, either positively or negatively, and which serves to insulate the expression of a given gene. An insulator can be incorporated into a genetic circuit of the present disclosure to prevent or reduce epigenetic chromatin silencing, and/or to significantly reduce promoter interference between the integrated transgene’s promoter and the surrounding endogenous promoters from the integration locus. The insulator of the present disclosure can be cis-acting or trans-acting. In some embodiments, the insulator is a nucleic acid sequence. In some embodiments, the insulator can be the cHS4 sequence (48). In some embodiments of the present disclosure, the presence of insulators in a procedure for integrating multiple genetic circuits into the cell line containing a plurality of landing pads can providing savings in time and facilitate inter-related or compensatory modifications which must be synchronized in time.

Plasmid

A plasmid is a small DNA molecule within a cell that is physically separated from a chromosomal DNA and can replicate independently. They are most commonly found in bacteria as small, circular, double-stranded DNA molecules; however, plasmids are sometimes present in archaea and eukaryotic organisms.

Bacterial origin of replication

A bacterial origin of replication herein refers to a bacterial plasmid origin of replication. It is a particular sequence in a genome at which replication is initiated. The specific structure of the origin of replication varies somewhat from species to species, but all share some common characteristics such as high AT content (repeats of adenine and thymine are easier to separate because their base stacking interactions are not as strong as those of guanine and cytosine. The origin of replication binds the pre-replication complex, a protein complex that recognizes, unwinds, and begins to copy DNA. The bacteria origin of replication that may be used in the present disclosure include, without limitation, pBR322 and pUC.

Ligation

Ligation refers to the joining of two nucleic acid fragments through the action of an enzyme, most commonly done using DNA ligase, such as T4 DNA ligase. The ends of DNA fragments are joined together by the formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another. A co-factor is generally involved in the reaction, and this is usually ATP or NAD⁺. Factors that affect an enzyme-mediated chemical reaction would naturally affect a ligation reaction, such as the concentration of enzyme and the reactants, as well as the temperature of reaction and the length of time of incubation. Ligation is complicated by the fact that the desired ligation products for most ligation reactions should be between two different DNA molecules and the reaction involves both inter- and intra-molecular reactions, and that an additional annealing step is necessary for efficient ligation.

In some embodiments of the present disclosure, ligation is carried out under conditions that favor intra-molecular ligation, wherein the two ends of one linear DNA molecule joins to form a circularized molecule. The conditions that favor intramolecular ligation comprises, without limitation, using low concentration of linear DNAs and/or carrying out the ligation reaction in the absence of cations or polyamines, and in 6.0% to 10% PEG 6,000 solutions (32).

Restriction enzyme and restriction sites

A restriction enzyme or restriction endonuclease is an enzyme that cuts DNA at or near specific recognition nucleotide sequences known as restriction sites. Restriction enzymes are commonly classified into three types, which differ in their structure and whether they cut their DNA substrate at their recognition site, or if the recognition and cleavage sites are separate from one another. To cut DNA, all restriction enzymes make two incisions, once through each sugar-phosphate backbone (*i.e.* each strand) of the DNA double helix.

These enzymes are found in bacteria and archaea and provide a defense mechanism against invading viruses. Inside a prokaryote, the restriction enzymes selectively cut up foreign DNA in a process called restriction; meanwhile, host DNA is protected by a modification enzyme (a methyltransferase) that modifies the prokaryotic DNA and blocks cleavage. Together, these two processes form the restriction modification system.

The restriction enzymes that may be used include, without limitation, AatII, MfeI, Acc65I, AccI, MscI, AclI, MspA1I, AatII, MfeI, Acc65I, MluI, AccI, MscI, AclI, MspA1, AfeI, NaeI, AflIII, NarI, AgeI, NcoI, ApaI, NdeI, ApaLI, NgoMIV, ApoI, NheI, AscI, NotI, AseI, NruI, AsiSI, NsiI, AvrII, NspI, BamHI, PacI, BclI, PciI, BglII, PmeI, Bme1580I, PmlI, BmtI, PsiI, BsaHI, PspOMI, BsiEI PstI, BsiWI, PvuI, BspEI, PvuII, BspHI, SacI, BsrGI, SacII, BssHII SalI, BstBI, SbfI, BstZ17I, ScaI, BtgI, SfcI, ClaI, SfoI, DraI, SgrAI, EaeI, SmaI, EagI, SmlI, EcoRI, SnaBI, EcoRV, SpeI, FseI, SphI, FspI, SspI, HaeII, StuI, HincII, SwaI, HindIII, XbaI, HpaI, XhoI, KasI, XmaI, and KpnI.

EXAMPLES

Example 1: Expanding the site-specific recombinase toolbox for mammalian synthetic biology

Site-specific recombination systems mediate recombination reactions between two specific DNA sequences termed recombination sites. While most site-specific recombination systems occur in bacteria and their viruses, many of the characterized site-specific recombinases function in heterologous systems, which can be extremely useful to manipulate or engineer their genome via integration, excision or inversion events as demonstrated in many other studies (1-7). To streamline the engineering of mammalian cells, one would benefit from having a programmable recombinase that could be directly targeted to the endogenous sequence of choice in the mammalian genome. This would eliminate the requirement to pre-integrate a natural recombination site. To be useful, the integration reaction would still have to be highly specific and efficient.

Based on amino acid sequence homology, site-specific recombinases fall into one of two mechanistically distinct families: the tyrosine and the serine recombinases(8, 9). The names come from the conserved nucleophilic amino acid residue that they use to attack the DNA, which becomes covalently linked to it during strand exchange.

The better characterized members of the tyrosine recombinase family are the integrases from coliphage I and from prophage lambda (10, 11) which are used either to integrate or excise the phage genome from their bacterial host. The integrase recognize a specific attachment on the phage genome, attP, and its counterpart on the bacterial genome, attB, between which it catalyzes a single DNA crossover (Table 1; SEQ ID NOs: 79-155). The circular phage genome is therefore linearized upon integration and the hybrid attachment

sites formed that flank it are called attL and attR (FIG. 1). The integration reaction also requires an accessory protein, the Integration Host Factor (IHF), which is specific to bacteria and helps folding the DNA molecule in the appropriate way for recombination to happen (12). When and only when an additional protein (Xis) is expressed together with Int and IHF, the excision reaction is triggered.

This system ensures that the integration events are directional and therefore not reversible in the absence of the Xis protein.

A major drawback of these systems that limits their use for mammalian genome engineering is the required intervention of the IHF for recombination to occur. IHF is indeed not naturally expressed in mammalian cells and even though scientists have engineered a mammalian version of IHF, the efficiency of both intra- and intermolecular reactions remained extremely low in mammalian cells (13, 1).

The tyrosine recombinase family also includes many members such as Cre (8), Flp (14,10) and Dre (15, 13) that catalyze non-directional and therefore reversible site-specific recombination reactions between two identical sites of approximately 35bp in length in the absence of accessory proteins. These recombination systems, in particular the widely used Cre-loxP and FLP-frt, demonstrated highly efficient deletion of DNA located between two directly repeated substrates inserted into the genomes of many higher eukaryotes (16, 17). As the recombination sites are regenerated after Cre-loxP or FLP-frt recombination, efficient and stable integration of DNA is greatly compromised by the tendency of the integrated DNA to immediately excise, therefore favoring the deletion reaction (FIGs. 2A and 2B). To overcome the reversibility of these systems, some recombination with a few base pair mutation sites have been created that prevent results in hybrid recombination sites after the integration events, not recognized anymore by the recombinase (15, 14). Although these tricks improved the integration efficiency of DNA in the genomes of heterologous hosts, it stayed significantly lower than what has been achieved with serine recombinases.

The serine recombinase family is composed of multiple structural groups represented by the resolvase/invertases, the large serine recombinases (Table 1; SEQ ID NOs: 21-78) and the transposases (FIG. 3) (18). They all mediate site-specific recombination reactions with strictly controlled directionality, in the absence of an accessory protein (Xis or recombination directionality factor, RDF), and are therefore of great interest for mammalian genome engineering (19, 15, 18).

In their native contexts, serine resolvases and invertases selectively recombine target sites within the same DNA molecule. Although directional, they are poorly suited for accurate genomic recombination because the recognition motifs of their DNA binding domain (DBD) are short (4-6 bp) and degenerate. Recent studies have demonstrated that the DBD of a serine resolvase can be replaced with a zinc finger protein of higher affinity and specificity (20, 15). This substitution retargets recombination to sequences flanked by zinc finger binding sites (ZFBS). However, these zinc finger-recombinases (RecZFs) retain a second, complementary specificity. The serine catalytic domain indeed imposes its own sequence requirements on the interior of the RecZF target site (20-bp core). Functional RecZF recombination sites must then contain sequences compatible with both the zinc finger DNA-binding protein and recombinase catalytic domain. Even though such hybrid recombinases can offer broader targeting capabilities, their low efficacy for DNA integration is still a bottleneck.

To date, the most efficient, directional and specific recombinases are the large serine recombinases, widely used for DNA integration in mammalian cells (19). The integrases from the *Streptomyces* temperate phage Φ C31 (21, 20) or from the *Mycobacteriophage* BxB1 (22, 21) were shown to mediate unidirectional and highly specific recombination between relatively small recombination substrates, attP and attB (< 50 bp). The domains that compose large serine recombinases are, however, not as modular as with the resolvases. It remains unclear how to retarget these recombinases towards new recombination site sequences. To date, a very limited number of highly efficient large serine recombinases have been characterized. As they have been discovered in phage genomes, it is very probable that they are yet many others to discover given the large diversity of phages that exist on the planet.

The identification and characterization of new large serine recombinases would not only help better understand the relation between their structure and their function and therefore give more insight on how to reprogram them, but it would also expand the current toolbox scientists have at their disposal for the engineering of heterologous hosts genomes and for the manipulation of synthetic genetic circuits.

Described here is the discovery of new large serine recombinases from recently sequenced *Mycobacteriophages* (FIG. 4) and the characterization of their respective attachment sites. After having identified putative large serine recombinase from genome

sequences, their function in their natural host, *Mycobacterium smegmatis*, was validated and a plasmid rescue strategy was used to infer their respective attachment sites, which was then validated in *E. coli*. It was then demonstrated that these systems could be used in mammalian cells and libraries of orthogonal site-pairs for each of these recombinases were then created.

5

Example 2: Identification of large serine recombinases

Among the few large serine recombinases already identified, 3 of them have been discovered from Mycobacteriophage genomes: BxB1 (23, 21), Peaches and BxZ2 (24, 22, 25). BxB1 recombinase has been demonstrated to be the most efficient to integrate DNA into the genome of mammalian cells (19).

10

Peaches and BxZ2 have yet never been tested in eukaryotic cells. To extend the limited toolbox of large serine recombinases available for the manipulation of mammalian genomes, putative recombinase sequences among the very recently sequenced 400 new Mycobacteriophages genomes, were analyzed (26, 23).

15

An algorithm was created to scan through all putative protein sequences from these genomes and extract the ones that contain conserved motifs typical of large serine recombinases. Compared to other serine recombinases, many features and motifs are specific to large serine recombinases (FIG. 3). Not only all known large serine recombinases are constituted of more than 400 amino acids, but they also carry very specific conserved residues in their catalytic and dimerization domains (FIG. 5).

20

Taking into account these specificities, the algorithm identified that 26 genomes carried a large serine recombinase (out of the 400 genomes). The 374 other genomes carried a tyrosine recombinase. By comparing the amino-acid sequences of the 26 identified serine recombinases to the ones of BxB1, Peaches and BxZ2, two new clusters emerged, significantly distant from the Peaches and BxZ2 cluster and the BxB1 cluster (FIG. 6). While none of the 26 recombinases was grouped with Peaches and BxZ2, one third of them shared a high degree of homology with the well-characterized BxB1 recombinase: KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Trouble, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum and Severus. The first new cluster contains both Rey and Bongo recombinases; and the second one contains Airmid, Benedict, Theia, Hinder, ICleared, Sheen, Mundrea, Veracruz and Rebeuca recombinases.

25

30

Different amino acid sequences likely result in different tertiary protein structure and therefore different DNA recognition and binding activities.

Example 3: Creation of integrative vectors for *Mycobacterium* species

When a mycobacteriophage infects *M. smegmatis*, its recombinase is expressed to mediate the integration of the phage genome into the *M. smegmatis* via site-specific recombination between the phage attachment site (attP) and the bacterial chromosomal attachment site (attB). Therefore, in order to validate the integration capacity of the 26 putative recombinases identified with the algorithm, a synthetic vector was created mimicking the mycobacteriophage genome and designed in a way such that integration events could be easily monitored (FIG. 7).

This vector carried three essential components: an attP site, a gene cassette to constitutively express the recombinase and a selection marker to enable selection of integrants. While it is almost impossible to predict the exact sequence of the attP site, it is however possible to predict its approximate location on the phage genome. When looking into phage biology, it appears that the attP site is most often located close to the start codon of the recombinase coding sequence. This evolutionary strategy allows the recombinase expression to be under the control of bacterial promoter once the phage has integrated into its host genome. Because the physical DNA of the 26 mycobacteriophages identified for the presence of a large serine recombinase was not available, the DNA region that was potentially carrying the attP site was synthesized (from 750bp before to 50bp after the start codon of the putative recombinase).

For the recombinase cassette, the Golden Gate assembly method was used to assemble the strong Hsp70 constitutive promoter together with a codon-optimized coding sequence of the recombinase followed by an efficient transcription terminator. Again, the coding sequence of the recombinase was synthesized given that the physical DNA of the corresponding phage was not available. The coding sequence was codon-optimized for an efficient expression in mammalian cells since the end goal was to test these recombinases in mammalian cells.

To be able to screen for integration events of the synthetic vector into the chromosome of *M. smegmatis*, a selection marker was added constitutively expressed both in *E. coli* and *M. smegmatis*.

Finally, the origin of replication pBR322 was used to allow for efficient replication of the plasmid in *E. coli* for cloning purpose. However, pBR322 cannot be replicated in *M. smegmatis*.

Due to this specificity, transformation of *M. smegmatis* with the integrative vector would result in resistant cells only if the expressed recombinase would mediate site-specific recombination between the plasmidic attP and the chromosomal endogenous attB (FIG. 8). This would indeed lead to the stable integration of the vector and therefore to the stable expression of the resistance marker. If the vector cannot self-integrate into *M. smegmatis* genome (either because the recombinase is not functional or because the attP sequence is not present on the vector), then the non-replicative vector would be diluted among the population of cell.

Example 4: Transformation of *M. smegmatis* with the integrative vectors

Each of the 27 integrative vectors (26 new recombinases + BxB1 as a positive control) was transformed in wild type *M. smegmatis* and plated the transformed bacteria on a selective media. Out of the 27 integrative vectors tested, 23 led to the growth of resistant clones, which were then picked 4 days after transformation and expanded for further testing.

To confirm the stable integration of the vector into the chromosome of the resistant clones, their chromosomal DNA was extracted and a PCR was performed with primers flanking the resistance cassette (FIG. 9). The PCRs were positive for each of the 3 clones picked for the 23 functional recombinases.

To verify that stable integration into the chromosome was the result of the plasmid linearization via site-specific recombination from a sequence carried by the putative attP region on the plasmid, this attP region with primers flanking the sequence carried by the original integration vector was amplified. While the amplification worked on all 23 integration vectors, it did not work when using the extracted chromosomal DNA as a template.

These first results were encouraging as they meant that both the identified recombinases were functional and that the attP sequences were indeed located in the phage genome region predicted, somewhere close to the start codon of these recombinases.

Example 5: Identification of the integration site in *M. smegmatis* genome

The plasmid rescue method was used to discover the exact location of the insertion site of the integrative vector in the *M. smegmatis* genome (FIG. 11). This method allowed for the identification of the genomic sequences flanking the integrative vectors insertion site.

5 First, the extracted genomic DNA was digested from the resistant clones with EcoRI restriction enzyme. This restriction digest reaction would normally result in 1199 independent fragments. However, only one EcoRI site was intentionally inserted in the integrative vector sequence, in between the putative attP region and the recombinase cassette. Restriction digest of the genomic DNA from the resistant clones would therefore result in 1200 independent
10 fragments.

All the fragments were ligated together and *E. coli* were transformed with the reaction mix and plated on selective media. While the majority of the ligation products would not be able to replicate in *E. coli*, this method allows the ligated fragment that contains both the resistance gene cassette and the origin of replication from the integrative vector to replicate in
15 *E. coli*, conferring the resistance to Kanamycin in the bacteria. Because the integrative vector contained a single EcoRI restriction site, the recircularized product also contains some chromosomal sequences from *M. smegmatis* flanking the insertion site.

Therefore, the resistant clones that grew on selective media were expanded, minipreped and the extracted vectors sequenced. The sequencing results unveiled the
20 sequence flanking the insertion site and allowed for the inference of the sequence of the attB site.

Out of the 23 putative recombinases tested, 15 of them integrated into the same genomic attB site as BxB1 (groEL CDS), two integrated in the TmrH RNA methyltransferase CDS (Rebeuca and Veracruz) and two others integrated in the 2-nitropropane dioxygenase
25 CDS (FIG. 12). No information concerning the importance of these genes in the growth regulation or metabolism of *M. smegmatis* was found. However, the resistant clones that were expanded to prepare the plasmid-rescue workflow grew at a normal rate and with a normal phenotype.

Based on the sequencing results, the point of strand exchange of the site-specific
30 recombination reaction was deduced and then the approximate attB and attP sites sequences were predicted.

For the recombinases that integrated into the groEL CDS, the attP and attB sites were total homologs to the one recognized by BxB1 recombinase.

Although integration happens in the same attB site for Rebeuca and Veracruz, the attP site recognized by Rebeuca is slightly different from the attP site recognized by Veracruz (FIG. 13). This case is actually similar for Theia and Benedict recombinases and such differences had already been observed with the BxZ2/Peaches pair (22).

Example 6: Validation of functional attB/attP recombination sites

To verify that the predicted attP and attB sites were correct and could be recombined by their putative large serine recombinase in a heterologous host, a reporter system to monitor recombination events between the predicted att sites in *E. coli* was created (FIG. 14). The reporter system consists of a plasmid carrying 1) the predicted attB sequence followed by 2) a lacZ gene cassette followed by 3) the predicted attP sequence, 4) a gene cassette expressing the recombinase and 5) a gene cassette expressing a resistance marker. After transformation in *E. coli* cells, the expressed recombinase would mediate the excision of the LacZ cassette after the recognition and site-specific recombination between the predicted attP and attB sites. Efficient recombination between attP and attB sites would result in the development of white colonies (LacZ negative) in an Xgal screen. If the sites were not correct, recombination could not allow for the excision of the lacZ cassette and the colonies would be blue (LacZ positive).

The vectors in *E. coli* were transformed, plated on selective media and incubated the plates overnight. Given that pBR322 is a relatively low copy plasmid, excision of the lacZ gene from all the copies within a cell should be achieved relatively rapidly. Blue or white colonies were screened for the next morning and for all 4 recombinase, more than 99% of the obtained resistant clones were LacZ negative, validating the exactitude of the attP and attB sites.

To further validate that the non-expression of LacZ was due to the excision of the LacZ cassette and not a consequence of instability of the construct, the plasmids miniprepmed from white clones were sequenced. The results were positive as the sequencing showed the presence of the expected deletion and the attL site created by site-specific recombination between attB and attP site.

In some embodiments the exact minimal sequence required for the att sites to enable site-specific recombination is performed. In some embodiments, a library of att sites is synthesized with different lengths and the experiment described above is then repeated. In some embodiments, when the att site tested is too short, resistant clones should all express the LacZ gene even after an extended period of time.

Example 7: Assessment of intermolecular recombination

Because both the activity of these recombinases for intramolecular site-specific recombination in *E. coli* and the correct sequences of their attB and attP sites were confirmed, further testing was performed to confirm whether they could also be used for the manipulation of DNA sequences in mammalian cells. Therefore, a reporter system was created based on two separate plasmids to test their intermolecular site-specific recombination capacity in mammalian cells (FIG. 15). The first plasmid carried a gene cassette allowing the constitutive expression of EYFP. The attB site was placed in between the Hef1a constitutive promoter and the EYFP coding sequence. The second plasmid carried a promoter-less gene cassette composed of the attP site followed by the mKate2 coding sequence. When co-transfecting these two plasmids together with a plasmid expressing the corresponding recombinase, intermolecular site-specific recombination between the attB and the attP sites would result in the fusion of the two plasmids and therefore in the insertion of the promoterless mKate2 coding sequence in frame with the Hef1a promoter. This recombination event would therefore trigger the constitutive expression of mKate2 and interrupt the expression of EYFP as it becomes promoterless.

These two plasmids were co-transfected together with a third plasmid to express the recombinase and assessed the intermolecular recombination events based on the fluorescence output. If the recombinase mediates site-specific recombination between the separated attB and attP sites, the two plasmids fuse together and the mKate2 coding sequence is placed just after the Hef1a promoter while the EYFP coding sequence becomes promoterless.

As depicted in FIG. 16, the 4 new recombinases could mediate intermolecular recombination in mammalian cells.

Example 8: Assessment of recombination crosstalk

To test whether these recombinases were orthogonal to each other, *i.e.* could only mediate site specific recombination between their own attB and attP sites identified in previous experiments, each of the recombinase expression vectors were co-transfected with the reporter systems corresponding to all recombinases. The recombinases that had different attB sites were completely orthogonal to each other (no mKate2 expression). However, the recombinases, which shared the same attB but a different attP site, could mediate site-specific recombination using both attP sites (significant mKate2 expression). The results are summarized in FIG. 17.

Example 9: Creation of a library of orthogonal sites

The nonpalindromic central dinucleotide at the center of attP and attB sites is the sole determinant of the prophage orientation (27), and a single base pair substitution in the two sites inhibits recombination. However, when identical central dinucleotide mutations are in both attP and attB partners, recombination is restored, as demonstrated for BxB1 and PhiC31 in bacterial cells (27, 28). Moreover, for the directionality of the reaction to be retained, these two central dinucleotides have to be non-palindromic.

Based on these requirements, 6 different variants of the central dinucleotide sequence were synthesized for each attP and attB partners that were non-palindromic and tested against each other in transient transfections to see whether they were orthogonal in mammalian cells.

As predicted, only attP and attB that shared the exact same central dinucleotide sequence could recombine for all 7 recombinases tested (*i.e.* 42 orthogonal sites) (FIG. 18).

For this purpose, a stable cell line with two landing pads was created, each one carrying one BxB1 attB site orthogonal to the other. The integration of a circuit in the landing pad of choice depending on the attP site carried by the integrative circuits was then performed (results not shown).

Example 10: Streamlined identification of attP/attB sites

A deeper analysis of these sites could help shorten the pace of att site identification by skipping the required experiments in the natural host *M. smegmatis*. Indeed, all these sites share common features such as 6 to 8 base pair homologies between attB and attP sites core sequence and numerous inverted repeats in the flanking sequences. It could therefore become

useful to develop an algorithm to 1) itemize all possible large serine recombinases within sequenced mycobacteriophage genomes, 2) itemize all putative attP sites within 200bp flanking the start codon of these recombinases, 3) scan mycobacterium genomes to itemize all putative attB sites sharing an homologous core sequence flanked by numerous inverted repeats.

Example 11: Multiplexed integration of genetic circuits

While intramolecular site-specific recombination has proven useful for excision of constructs integrated in mammalian genomes, intermolecular site-specific recombination can be used to integrate complex circuits into mammalian chromosomes. The ability to multiplex integrations by using orthogonal sites could help to integrate a variety of different circuits at different locations within a genome. This could be very helpful to prevent interference between circuits or attain higher levels of expression (with a lower coefficient of variation) by integrating the same construct in multiple copies. This could also allow engineering cell lines step by step, integrating one circuit after the other so as to incrementally increase the complexity.

To integrate multiple genetic circuits into mammalian genome, multiple landing pads containing attB sites orthogonal to the corresponding large serine recombinase are integrated first. Each genetic circuit to be integrated is provided on a integrative vector, together with a unique attP site that only recombines with one landing pad in the mammalian genome, allowing specific integration of each genetic circuit into its desired target location. The multiple genetic circuits may be inserted into a central location in a mammalian genome or into different locations in a mammalian genome. The multiple genetic circuits can also be integrated sequentially or simultaneously to the endogenous sequences of choice in the mammalian genome.

Example 12: New DNA assembly methods

The use of multiple serine recombinases with orthogonal sites could also lead to the development of new DNA assembly methods. Based on the Gateway principle, a system may be created flanking each DNA fragment to be assembled with either an attB on its 5' end and an attP on its 3'. Having 6 orthogonal sites for one recombinase would allow assembling 5 parts per reaction, *i.e.* enough for a mammalian transcription unit. The assembled vectors

could carry sites from another recombinase, and following the same principle, could be used to assemble 5 transcription units together. Such strategy could be used to assemble even larger constructs with 6 orthogonal recombinases. As site-specific recombination with serine recombinases is extremely efficient, fast and reliable, such an assembly method would be extremely convenient.

Materials and Methods

In-silico identification of large serine recombinases

Scripts were written and run with Matlab2013a.

Expression units and plasmids assembly

All expression units and plasmids were assembled with the Golden Gate framework and are listed in the Annexes. For all Golden Gate assembly reactions, used were: 0.4 µl of Type IIS enzyme (either BsaI from NEB, or BpiI from Fermentas), 0.2µl of T4 Ligase HC + 1 µl of T4 Ligase HC buffer (Promega), 1 µl of 10xBSA (NEB), 40 fmol for all vectors used in the reaction, ddH2O up to a final total volume of 10 µl. The thermocycler program used for all assemblies included: 1 step of 15 min at 37°C; then 50 cycles of [2 min at 37°C followed by 5 min at 16°C]; 1 step of 15 min at 37°C, 1 step of 5 min at 50°C and 1 final step of 5 min at 80°C.

Bacterial cell cultures

Liquid cultures of *E. coli* MG1655 were grown in LB Medium (Difco) at 37°C. When appropriate, antibiotics were added as follows: spectinomycin (100 µg/mL), ampicillin (100 µg/mL) and kanamycin (25 µg/mL). For blue/white screening, we used X-gal at a final concentration of 40 µg/mL.

For liquid cultures, *M. smegmatis* was grown in Difco 7H9 liquid medium supplemented with 0.5% glycerol, 0.5% Tween 80 and 10% albumin–dextrose complex (ADC). Carbenicillin and cycloheximide were added to all cultures at concentrations of 50 mg ml⁻¹ and 10 mg ml⁻¹ respectively. When necessary, the following antibiotics were also added; kanamycin (8 mg ml⁻¹), Hygromycin (50 mg ml⁻¹) and tetracycline (0.5 mg ml⁻¹). For solid cultures, *M. smegmatis* was grown on Difco 7H10 agar supplemented with 0.5% glycerol and 10% ADC.

Mammalian cell culture and transfections

HEK293FT cell line was purchased from Invitrogen. HEK293FT cells were maintained in Dulbecco's modified Eagle medium (DMEM, Cellgro) supplemented with 10% FBS (PAA), 0.045 g/mL penicillin/streptomycin and non-essential amino acids (HyClone) at 37°C, 100% humidity and 5% CO₂. HEK293FT transfections were carried out in 24-well plates using Attractene reagent (Qiagen), 200000 cells and 200-300 ng total DNA per well (plasmid ratio 1 : 1 : 1). Media was changed 24 hours after transfection.

Microscope measurements and image processing

Fluorescence microscopy images of live cells were taken in glass-bottom dishes or 12-well plates using Zeiss Axiovert 200 microscope and Plan-Neofluar 10x/0.30 Ph1 objective. The imaging settings for the fluorophores were S430/25x (excitation) and S470/30m (emission) filters for EYFP, and S565/25x (excitation) and S650/70m (emission) for mKate2. Data collection and processing were performed using AxioVision software (Zeiss).

Table 1: Amino-acid and Nucleotide Sequences

SEQ ID NO:	Species	Sequence
21	Abrogate	MRALVVIRLSRVTDATTSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRRRRPNLARWLAFFEEQPFQDVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTPFAAVVIALMGTVQAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQVRERILEVYHRVVDNHEPLHLVAHDLNRR GVLSPKDYFAKLQGREPRGREWSATALKRSLISEAMLGYATLNGKTVRDDD APLVRAEPILTREQLEALRAELVKTDRAKPSVSTPSMLLRVLFCAVCGEPAKFT GGGRKNARYRCRSWGWAQRCGNGTVAMAEWDSFCEEQVLDLLGDSERLE KVWVAGSDSAVELAEVNAELVLTSLIGSPAYRAGSPQREALDARIAALAAARQ EELEGLEARPSGWREWRETGQRFQDWWREQDTAGKNTWLRSMNVRLTFDV RGGLTRTIDFGDLQEYEQHLRLGVSVERLHAGMS
22	Abrogate	ATGAGAGCACTGGTAGTGATCCGACTGTCCCGCGTCACCGATGCTACGACT TCACCGGAGCGCCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG GTGGGACGTCGTCGGGGTAGCAGAGGATCTGGACGTCTCCGGAGCAGTC GATCCGTTGACCGGAGGCGCAGACCGAACCTGGCCCGGTGGCTAGCGTT CGAGGAGCAACCGTTCGACGTGATCGTGGCGTACCGGGTAGACCGGCTGA CCCGATCAATCCGGCATCTGCAGCAGCTGGTCCACTGGGCCGAGGACCAC AAGAAGCTGGTCGTCTCCGCGACCGAAGCGCACTTCGACACGACAACGCC GTTGCGGCGGTGTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAAT TAGAAGCGATCAAAGAGCGGAACCGATCGGCGGCGCATTCAATATCCGC GCCGGTAAATACCGAGGTTCCCTGCCCGCGTGGGGTTACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCGACCCGGTGACGCGAGAGCGC ATCCTCGAGGTGTATACCGCGTCGTCGACAACCACGAGCCGCTGCACCTG

		GTGGCACACGACCTGAACCGGCGTGGTGTCTGTGCGCTAAGGACTACTTC GCAAAGCTGCAAGGTCCGGGAGCCGCGGGGCCGGGAGTGGTCCGGTACCG CGCTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTG AACGGTAAGACCGTCCGAGACGACGACGAGCTCCGCTGGTGC GGCTG AGCCGATCCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTC AAGACCGACCGGGCAAAGCCATCGGTGTCGACTCCGTCGATGTTGCTGCG GGTGTGTTCTGCGCAGTGTGCGGGGAGCCCGCATACAAGTTCACCGGGG GCGGTAGGAAGAACGCACGATACCGCTGCCGGTCTGGGGGCTGGGCACA GCGGTGCGGCAACGGCACGGTTCGCAATGGCAGAGTGGGACTCGTTCTGC GAGGAGCAGGTGCTGGATCTGCTCGGGGACTCGGAGCGCCTGGAGAAAG TCTGGGTAGCAGGCTCGGACTCCGAGTCGAACTCGCGGAGGTGAACCGG GAGCTGGTGGACCTGACGTCGCTGATCGGCTCCCCGGCCTACCGGGCCGG TTCTCCGACGCGGAAGCGCTGGATGCTCGTATTGCGGCGCTGGCAGCAC GGCAGGAGGAGCTGGAGGGTCTAGAGGCACGCCGCTCGGGTGGGAGTG GCGCGAGACTGGGCAGCGGTTCCGGGACTGGTGGCGGGAGCAGGACACC GCGGGTAAGAACACCTGGCTTCGGTCGATGAACGTTCCGGCTGACGTTCA CGTCCGCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGT ACGAGCAGCACCTCAGGCTCGGCAGCGTGGTGAACGACTACACGCCGGG ATGTCGTAG
23	Airmid	MRVLGRIRLSRVMEESTSVRQREIETWARQNDHEIIGWAEDLDVSGSVDPF ETPALGPWLTDHRKHEWDILVAWKLDRLSRAIPMNKLFGWVMENDKTLVC VSENLDLSTWIGRMIANVIAGVAEGELEAIRERTKGSQKKLRELGRWGKPY YGYRAQEREDAAGWELVPDEHASKVLLSIEKVLEGQSTESIARELNERGELSPS DYLHRHAGKPTRGGKWSNAHIRQQLRSKTLGYSTHNGETIRDERGIAVRKGP ALVSQDVFDRQLQAALDSRSFKVTNRSASPLLGVLVCRVCERPMHLRQHHN KKRGKTYRYYQCVGGVEKTHPANLTNADQMEQLVEESFLAELGDRKIQERVYI PAESHRAELDEAVRAVEITPLLGTVSDTMRKRLDQLSALDARISELEKLPE EARWEYREGDETYAEAWNRRGDAEARRQLLLKSGITAAAEKMGREARVNPVG LHFDLRIPEDILERMSA
24	Airmid	ATGCGAGTCTTGGAAGAATACGACTCTCGGGGTCATGGAGGAAT CGACATCGGTGAGAGGCAGCGAGAGATCATCGAGACGTGGGCGC GTCAGAACGACCACGAGATCATCGGCTGGGCTGAGGACCTCGACGT GTCTGGATCGGTGATCCGTTGAGACGCCAGCCTTGGGTCCGTGG CTTACCGACACCGGAAGCACGAGTGGGACATCCTCGTGGCATGGA AGCTCGACCGGTGTCCAGGCGAGCTATCCCGATGAACAACTCTTC GGCTGGGTGATGGAGAACGACAAGACCCTCGTCTGCGTGTGCGAGA ATCTGGACCTGTGACGTGGATCGGTGCGATGATCGCCAACGTATC GCTGGCGTGGCAGAAGGTGAGTTGGAGGCGATACGAGAGAGGACC AAGGGCTCTCAGAAGAAGCTACGTGAGCTTGGCCGCTGGGGAGGA GGCAAGCCCTATTACGGCTACCGCGCGCAAGAGCGTGAGGACGCTG CTGGGTGGGAGCTGGTGCCTGACGAGCACGCCTCGAAGGTCTGCT CTCGATCATCGAGAAGGTCTCGAAGGGCAGTCGACGGAGTCGATA GCTCGTGAGCTGAACGAGAGAGGAGAGCTGTCCCCGTCTGACTACC TTCGGCACAGGGCTGGTAAGCCGACCAGAGGCGGTAAGTGGAGCA ACGCGCACATCCGTGAGCAGCTCCGCTCCAAGACTCTCTGGGTTAC TCCACGCATAACGGCGAAACCATCCGAGACGAGCGGGGGATCGCG GTACGCAAAGGGCCGGCGCTGGTTTCCAGGACGTGTTGACCGCC TCCAGGCGGCGCTTGATTCTCGATCCTTCAAGGTGACGAACAGGTCA GCGAAAGCGTCGCCGTTGCTCGGCGTCTCGTCTGCCGGGTGTGCG AACGACCGATGCACCTGCGTCAGCACCAACAAGAAGCGCGGCAA GACCTACCGCTACTACCAAGTGCCTGGGCGGTGTTGAAAAGACCCAC CTGCCAATCTACCAACGCCGATCAGATGGAGCAGTTGGTGAAG AGTCCTTCTGCTGAACTCGGTGACCGGAAGATCCAAGAGAGGGT TTACATCCCTGCGGAGTCACATCGAGCCGAGTTGGACGAGGCTGTA

		CGGGCCGTTGAGGAGATAACCCCTCTGCTGGGCACCGTCACGTCGG ACACCATGCGAAAGCGTCTCCTGGATCAGCTGAGCGCGTTAGATGC TCGTATCTCCGAGCTGGAGAAGCTGCCCGAGTCCGAAGCTCGGTGG GAGTACCGAGAAGGCGACGAAACCTACGCCGAGGCGTGGAACCGG GGTGACGCGGAAGCCCGTCGACAGCTCCTGCTCAAGTCGGGGATCA CGGCGGCTGCTGAGATGAAGGGCAGAGAGGCCCGAGTCAACCCGG GGGTCTTACACTTCGACCTACGAATACCGGAGGACATCTTAGAAAG GATGAGCGCGTGA
25	Anglerfish	MRALVIRLSRVTDATTSRPERQLESCRQLCAQRGWEVVGVAEDLDVSGAVDP FDRKRRPNLARWLAFFEEQFDFVIVAYRVDRLTRSIRHLQQLVHWAEDHKKLV VSATEAHFDTTTTFAAVVIALMGTVAQMELESIKERNRSAAHFNIRAGKYRGS LPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLHLVAHDLNRR GVLSPKDYFAKLQGREPQGRAWSATALKRLISEAMLGYTTLNGKTVRDDDG APLVRAEPILTREQLAELRAELVKTDRTKPAVSTPSLLRLVFCVCGEPAYKFD AGRKIPRYRCSRFGFAQRCNGTIPIAEWDAFCEEQVLDLLGDSEERLEKVVWA GSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALAARQEELEGL EARPSGWREWRETGQRFQDWWRDQDTAGKNTWLRS MNVRLTFDVRGGTLT RTIDFGDLQEYEQHLRLGSLDLVNAEKPTGR
26	Anglerfish	ATGAGAGCACTGGTAGTGATCCGACTGTCCCGCTCACCGATGCTACGACC TCCCGGAGCGTCAGCTGGAGTCTTGCCGGCAGCTCTGCGCCAGCGCGG CTGGGAGGTGGTCGGGGTAGCAGAGGATCTGGACGTCTCCGGAGCAGTC GATCCGTTGACCGGAAGCGCAGACCGAACCTGGCCCGGTGGCTAGCGTT CGAGGAGCAACCGTTCGACGTGATCGTGGCGTACCGGGTAGACCGACTGA CCCGATCGATCCGTCATCTGCAGCAGCTGGTCCACTGGGCCGAGGACCAC AAGAAGCTGGTCGTTTCCGCGACCGAAGCCCACTTCGATACGACGACGCC GTTCCGCGCGGTGCTCATCGCGCTTATGGGAAGTGTGGCGCAGATGGAAT TAGAATCGATCAAAGAGCGGAACCGATCGGCGGCGCATTTCAATATCCGC GCCGGTAAATACCGCGTTCCCTGCCGCCGTGGGGCTACCTGCCACGCG CGTGGACGGGGAGTGGAGGCTGGTGCCTGACCCGGTGCAGCGCGAGCGC ATCCTCGAGGTGTATACCGCGTCTGTCGACAACCACGAGCCGCTGCACCTG GTGGCACACGACCTGAACCGCGTGGTGTCTGTGCGCGAAGGACTACTT CGCAAAGCTGCAGGGTCGGGAGCCGACGGGGCGGGCGTGGTCAGCCACC GCGCTGAAGCGCTCGCTGATCTCTGAGGCGATGCTCGGGTATACGACGCT GAACGGCAAGACCGTCCGAGACGACGACGGGGCTCCGCTGGTGCGGGCT GAGCCGATCCTGACCCGCGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGT CAAGACCGACCGGACGAAGCCCGCGGTGTCGACGCCGTGCTGCTGCTGC GGGTGTTGTTCTGCGCGGTGTGCGGGGAGCCTGCCTACAAGTTCGACGCA GGCCGGAAGATCCCCCGCTACCGCTGCAGGTGCTTCGGGTTCGACAGCG CTGCGGGAACGGCACCATACCGATCGCAGAGTGGGACGCATTCTGCGAGG AGCAGGTGCTGGATCTGCTCGGGGACTCGGAGCGTCTGGAGAAAGTCTG GGTAGCAGGCTCGGACTCGGCGGTGAACTCGCGGAGGTGAACGCGGAG CTGGTGGACCTGACGTCGTTGATCGGCTCTCCGGCCTACCGGGTCGGTTCT CCGACGCGCAAGCACTGGATGCTCGTATTGCGGCGCTGGCCGCGCGGCA GGAGGAGCTGGAAGGGCTAGAGGCTCGCCGTCGGGTGGGAGTGGCGC GAAACCGGTGAGAGGTTCGGTGACTGGTGGCGGGATCAGGACACCGCGG GTAAGAACACCTGGCTCCGGTGCATGAACGTTCCGGCTGACGTTGACGTCC GCGGCGGGCTGACTCGACGATCGACTTCGGGGATCTGCAGGAGTATGAG CAGCATCTGAGGCTGGGCTCGGCTCTGGACCTCGTAAACGCAGAAAAGCC CCCTACGGGCGCGTAG
27	Benedict	MRVLGRIRLSRVMEESTSVERQREIETWARQNDHEIIGWAEDLDVSGSVDPF ETPALGPWLTDHRKHEWDILVAWKLDRLSRAIPMNKLFQVWMENDKTLVC VSENLDLSTWIGRMIANVIAGVAEGELEAIRERTKGSQKRLRELGRWGGGKPY YGYRAQEREDAAGWELVPDEHASAVLLSIEKVLEGQSTESIARELNERGELSPS DYLRRHAGKPTRGGKWSNAHIRQQLRSKTLGYSTHNGETIRDERGIAVRKGP

		ALVSQDVFDRLQAALDSRSFKVTNRSASPLLGVLICRVCPMHRLRQHNNK KRGKTYRYYQCVGGVEKTHPANLTNADQMEQLVEESFLAELGDRKIQERYVIP AESHRAELDEAVRAVEEITPLLGTVSDTMRKRLLDQLSALDARISELEKLPESE ARWEYREGDETYAEAWNRRGDAEARRQLLLKSGITAAEMKGREARVNPGVL HFDLRIPEDILERMSEA
28	Benedict	ATGCGAGTTCTTGAAGAATACGACTCTCGCGGGTCATGGAAGAATCGAC ATCGGTCGAGAGGCGAGAGATCATCGAAACCTGGGCGCGTCAGAAC GACCACGAGATCATCGGCTGGGCTGAAGACCTAGACGTGTCTGGATCGGT CGATCCGTTTCGAGACGCCAGCCTTGGGTCCGTGGCTTACCGACACCGGA AGCACGAGTGGGACATCCTCGTGGCGTGGAAGCTCGACCGGCTGTCCAGG CGAGCTATCCCGATGAACAACTCTTCGGCTGGGTCATGGAGAACGACAA GACCCTCGTCTGCGTGTGCGAGAATCTGGACCTGTGACATGGATCGGTC GGATGATCGCCAACGTCATCGCTGGCGTGGCAGAGGGGGAGTTGGAGGC GATACGAGAGAGGACCAAGGGCTCTCAGAAGAAGCTACGTGAGCTTGGC CGCTGGGGAGGAGGCAAGCCCTACTACGGCTACCGCGCGCAAGAGCGTG AGGACGCTGCTGGGTGGGAGCTGGTGCCCGACGAGCACGCCTCGGCGGT CCTGCTCTCCATCATCGAGAAGGTCCTCGAAGGGCAGTCGACGGAGTCGA TAGCTCGTGAGCTGAACGAGAGAGGAGAGCTGTCCCCTTCTGACTACCTTC GGCACAGGGCCGGTAAGCCGACCAGGGGCGGTAAGTGAGCAACGCGCA CATCCGTCAGCAGCTCCGCTCAAGACTCTCCTGGGTTACTCCACGCATAAC GGCGAAACCATCCGAGACGAGCGGGGGATCGCGGTACGCAAGGGCCGG CGCTGGTTTCCAGGACGTGTTGACCGCCTCCAGGCGGCGCTTGATTCTC GATCCTTCAAGGTGACGAACAGGTCAGCGAAAGCGTCGCCGTTGCTCGGC GTCCTCATCTGCCGGGTGTGCGAACGACCGATGCACCTGCGTCAGCACAC AACAAGAAGCGCGGCAAGACCTACCGCTACTACAGTGCGTGGGCGGTGT TGAAAAGACCCACCCTGCCAATCTACCAACGCCGATCAGATGGAGCAGTT GGTCGAAGAGTCCTTCTTGCTGAACTCGGTGACCGGAAGATCCAAGAGA GGGTGTACATCCCTGCGGAGTCACATCGGGCCGAGTTGGACGAGGCTGTA CGGGCCGTTGAGGAGATAACCCCACTGCTGGGCACCGTCACATCGGACAC CATGCGAAAGCGTCTCCTGGATCAGCTGAGCGCGTTAGATGCTCGTATCTC CGAGCTGGAGAAGCTGCCCGAGTCCGAAGCTCGGTGGGAGTATCGAGAA GGCGACGAAACCTACGCCGAGGCGTGGAACCGGGGTGACGCGGAAGCGC GTCGACAGCTCCTGCTCAAGTCGGGGATCACGGCGGCTGCTGAGATGAAG GGCAGAGAGGCCGAGTCAACCCGGGGGTCTTACACTTCGACCTACGAAT ACCGGAGGACATCTTAGAAAGGATGAGCGCGTGA
29	Bob3	MRALVIRLSRVTDATTSERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFDVIVAYRVDRLTRSIRHLQQLVHWAEDHKLI VSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRGS LPPWGYMPARVDGEWRLLVDPVQREIRILEVYHRVVDNHEPLHLVAHDNLQR GILSPKDYFAKLQGREPKGREWSATALKRSLEAMLGYYTTLNGKTVRDDDGA PLVRAEPILTREQLSRLAEVLKTDRTKPAVSTPSLLRLVLCFVCGEPAYKFTGG GRKNARYRCRSWGWAQRCGNGTVAMAEWDAFCEEQVLDLLGDAERLEKV WVAGSDAAVELTELNAELVDLTSLIGSPAYRAGSPQREALDARIAALAAARQEEL EGLEARPSGWEWRETGQRFQDWWRDQDTAGKNTWLRSMNVRLTFDVRG GLTRTIDFGDLQEYEQHLRLGSVVEQLHTGMS
30	Bob3	ATGAGAGCACTGGTAGTCATCCGACTGTCCCGCGTCACCGATGCTACGACC TCACCGGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG CTGGGACGTCTGCGGGTAGCAGAGGATCTGGACGTCTCCGGAGCAGTC GATCCGTTTCGACCGGAAGCGCCGCCGAACCTGGCCCCGATGGCTAGCGTT CGAGGAGCAACCGTTCGACGTATCGTGGCGTACCGGGTGGACCGGTTGA CCGCTCGATCCGGCATCTGCAGCAGCTGGTGCCTGGGCGGAGGACCAC AAGAAGCTGATCGTCTCCGCGACCGAAGCGCACTTCGATACGACGACGCC GTTGCGGCGGTCGTGATCGCGCTTATGGGAACGGTGGCGCAGATGGAAT TAGAAGCGATCAAGAGCGGAACCGTTCCGGCTGCGCATTTCAACATCCGC

		GCGGGTAAATACCGCGGTTCCCTGCCGCCGTGGGGTTACATGCCCGCACG CGTGGACGGGAGTGGAGGCTGCTCGTCGACCCCGTGACGCGGAACGC ATCCTCGAGGTCTATCACCGCGTCGTCGACAACCACGAGCCTCTGCATCTG GTCGCACACGACCTGAACCAGCGTGGTATCCTGTCGCCGAAGGACTACTTC GCAAAGCTGCAGGGTCGAGAGCCCAAGGGCCGGGAGTGGTTCGGCTACCG CGCTGAAGCGCTCGCTGATCTCGGAGGCGATGCTCGGGTATACGACGCTG AACGGCAAGACCGTCCGAGACGACGACGGGGCTCCGCTGGTGCGGGCCG AGCCGATCCTGACGCGCGAGCAGCTGGAATCGTGCGGGCGGAACTGGTC AAGACCGACCGGACCAAGCCCGCAGTGTCCACCCCGTCGCTGCTGCTGCG GGTGCTGTTCTGCGCAGTGTGCGGGGAGCCCGCATACAAGTTCACCGGGG GCGGCAGGAAGAACGCTCGTACCGCTGCCGGTCGTGGGGCTGGGCGCA GCGGTGCGGCAACGGCACGGTGGCAATGGCCGAGTGGGACGCATTCTGC GAGGAGCAGGTGTTGGATCTGCTCGGGGACGCAGAGCGTCTGGAGAAAG TCTGGGTAGCAGGCTCGGACGCTGCTGTGGAGTGACGGAGCTCAACGCA GAGCTGGTGGATCTGACGTCGCTGATCGGCTCCCCGGCATACCGGGCAGG TTCTCCGCAGCGCAGGCACTGGATGCTCGTATCGCAGCACTGCCAGCGC GGCAGGAGGAGTTGGAAGGGCTAGAGGCACGCCCCTCGGGCTGGGAGT GGCGCGAAACCGGGCAGAGGTTGCGGGACTGGTGGCGGGATCAGGACAC CGCAGGTAAGAACACCTGGCTCCGGTCGATGAACGTTGCGCTGACGTTTCG ACGTCGCGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAG TACGAGCAGCATCTCAGGCTCGGCAGCGTGGTCGAACAGCTACACACCGG GATGTCGTAG
31	Bongo	MTATLETPPQVVAPPRLRAAVYLRMSTDKELGIDRQREDCVLALCERLWDPV LYVDNDRSAVKENVRREAYEQMCEDIRDGRIDAIITWRSRLYRKMKSLPLID LIQGVNKNKGRIPIETCQTGLIDLTTDAGRMTAKILAAVAENEVEVTRARQMR AYEQIAESGRSLGAPAFGYTNDPKARVREIVPEEAAAIREGYDDVLACTLYSIA KKWNRDLKTPRGNAFVATVVGRILRNPRYAGLYRFRGEIIEGDEWPIVDVE TWAMATAVLDGKNTGPKGPRVRTLLSGIVRCGHCGNRMSASKNSNGEPIY KCKRYEVCNHGVTRVRKKVDKYVEARIVAKLEERKWVVGTKSDADQAKALHT EAETLRARKASFTDALVDGTLTPAQVKEASDKVDKLEERQLARLTKSRVYD GLLGHDDLEAVWVGLPLDRKRAIEQLCDKIVIRHVEITGRAAAKLPLGHNIDIY WHKPSDD
32	Bongo	ATGACAGCAACCCCTGAAACCCACACAGGTCGTCGCGCCGCCCGGCTG AGGGCTGCGGTCTACCTCCGCATGTCCACCGACAAAGAGCTGGGCATCGA CCGCCAACCGCAGGACTGCGTCGCCCTGTGCGAGCGCCTCGGCTGGGATC CCGTGCTCTACGTCGACAACGACCGCAGCGCCGTCAAAGAGAACGTGCGC CGCGAAGCGTACGAGCAGATGTGCGAGGACATCCGCGACGGCCGCATCG ACGCCATCATCACGTGGCGCTCCGACCGGCTCTACCGCAAGATGAAGTCCC TGCTGCCGCTCATCGACCTGATCCAGGGCGTCAACAAGAACGGCAAGCGG ATCCCCATCGAGACGTGCCAGACAGGGCTCATCGATCTACCACCGACGCA GGCCGCATGACGGCTAAGATCCTGGCAGCCGTAGCGGAGAACGAGGGCG AGGTCAGGACAGCCAGGAGATGCGCGCATACGAGCAGATCGCAGAGAG CGGCCGGTCACTGGGGGCCCGAGCGTTGCGGTACACCAACGACCCCAAGG CCCGCGTGCCTGAGATCGTGCCGAAGAAGCCGCCGCAATCCGCGAGGGC TACGACGATGTGCTCGCCGGATGCACGTTGTACTCGATTGCGAAGAAGTG GAACGACCGTGGACTCAAGACACCCCGCGGCAACGCGTTCTGTCGCCACCG TCGTCGCGCCGATCCTCAGGAACCCCGCTATGCGGGCCTGTACCGCTTCA GGGGCGAGATCATCGGTGAGGGCGACTGGGAGCCCATCGTGGACGTTGA GACGTGGGCGATGGCCACAGCGGTCCTCGACGGCAAGAACACCGGCCCA AAGGGCCCCAGGGTGCGCACAACGCTGCTCTCGGGCATCGTGCGGTGCGG ACACTGCGGCAACCGGATGTGCGCCAGCAAGAACAGCAACGGCGAGCCG ATCTACAAGTGCAAGCGCTACGAGGTCTGCAACCACGGTGTTACGCGAGT ACGCAAGAAGGTCGACAAGTACGTCGAGGCGCGGATCGTGGCCAAGCTC GAAGAGCGCAAGTGGGTTGTCGGCACCAAGTCCGACGCGGACCAAGCCA

		AGGCCCTGCACACAGAGGCCGAGACGCTGCGGGCCCCGAAGGCTTCGTTC ACCGACGCCCTGGTCGATGGCACCTGACACCCGCACAGGTGAAGGAGGC CAGCGACAAGGTCGACGCGAAGCTGGAGGAGATTGAACGCCAGCTAGCC CGCCTCACCAAGTCTCGGGTGTATGACGGGCTGCTGGGTACGACGACCT GGAGGCCGTCTGGGTGGGGCTGCCGCTGGACCGCAAGCGGGCCATCATC GAGCAGCTATGCGACAAGATCGTGATACGGCATGTGAGATCACCGGCCG TGCTGCCGCCAAGCTGCCGCTTGGCCACAACATCGACATCTACTGGCATAA GCCCAGCGATGACTGA
33	ConceptII	MRALVIRLSRVTDATTSRPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFQVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQREIRILEVYHRVVDNHEPLHLVAHDNLNR GVLSPKDYFAKLQGREPQGREWSATALKRSLISEAMLGYATLNGKTVRDDDG APLVRAEPILTREQLRALRAELVKADRTKPAVSTPSLLRLVFCVCGEPAYKFT GGGRKNARYRCRSWGWQAQRCNGTVAMAEWDAFCEEQVLDLLGDSERLE KVWVAGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALAARQ EELEGLEARPSGWEWRETGQRFQDWWRDQDTAGKNWLRSMLNVRITFD VRGGLTRTIDFGDLQEYEQHLRLGSALDLVNAEKPTGR
34	ConceptII	ATGAGAGCACTGGTAGTCATCCGCCTGTCCCGCTCACCGATGCTACGACT TCGCCGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG TTGGGACGTCGTCGGTGTAGCAGAGGATCTGGACGCTCTCCGGAGCGGTCTG ATCCGTTCCGACCGGAAGCGCAGACCGAACCTGGCCCGGTGGCTAGCATT GAGGAGCAACCGTTCGATGTCATCGTGGCATAACGGGTAGACCGGCTGAC CCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCTGAGGACCATA AGAAGCTGGTCGTCTCCGCAACCGAAGCCCACTTCGACACGACGACGCCG TTCGCAGCGGTCTGTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCGGCGCATTTCAATATCCGCG CCGGTAAATACCGTGGAAGCCTGCCGCCGTGGGGTTACCTGCCTACGCGC GTGGACGGTGAGTGGCGGTGGTGCCGGACCCGGTGACGAGAGAGCGCA TCCTCGAGGTGTATACCGCGTCGTCGACAACACGAGCCGCTGCACCTGG TGGCCACGACCTGAACCAGCGCGGCTCCTGTCGCCGAAGGACTACTTC GCGAAGCTGCAAGGCCGCGAGCCGAGGGCCGGGAGTGCTCGGCTACCG CGCTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTG AACGGTAAGACCGTCCGAGACGACGAGCGAGCCCCGCTGGTGCGGGCTG AGCCGATCCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTG AAGGCCGACCGGACCAAGCCCGCAGTGTCTACCCGCTCGCTGCTGCTGCG GGTGCTGTTCTGCGCAGTGTGCGGTGAGCCCGCATACAAGTTACCCGGTG GCGGTAGGAAGAACGCTCGCTACCGCTGCCGGTCTGCGGGTTGGGCACA GCGGTGCGGTAACGGCACGGTGGCGATGGCAGAGTGGGACGCATTCTGC GAGGAGCAGGTGTTGGATCTGCTCGGGGACTCGGAGCGTCTGGAGAAAG TCTGGGTAGCAGTTCCGACTCCGAGTAGAACTCGCGGAGGTGAACCGG GAGCTGGTGGACCTGACGTCGCTGATCGGCTCCCCGGCGTACCGGGCCGG TTCTCCGACGCGGAGGCGCTGGATGCTCGTATCGCGGCGCTGGCCGCGC GGCAGGAGGAGTTGGAAGGGCTAGAGGCTCGCCGCTCGGGTTGGGAGTG GCGCGAAACCGGGCAGAGGTTCCGGGACTGGTGCGGGGATCAGGACACC GCGGGTAAGAACGCGTGGCTCCGGTCGATGAACGTTCCGCTGACGTTCA CGTCCGCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGT ATGAGCAGCATCTGAGGCTGGGCTCGGCTCTGGACCTCGTAAACGCAGAA AAGCCCCCTACGGGCCGCTAG
35	Hinder	MRVLGRIRLSRLSDESTSPERQREIIEGWAKSNDHTIVGWAEDLDVSGSVDPF DTPALGPWLSEPKLHEWDILCAWKLDRLSRRAIPMNKLFQWVMDHDKTLVC VNDNIDLSTWIGRMVANVIAGVAEGELEAIRERTTASHRKLRELGRWPGR SYGYRAVEREDAAGWVLEPDVSSVLRSIIDWVLQGSVESIAKDLTAMGE VSPSDYVRQRAGEAPRGHPWHGRTIVKLLRSKTLGLYVTHNGTTVRDENGVP

		VQKGPPPLVDQDTFNRLQAALDDGSRPKTVNRTSKASPLLGVLCWDCEKPLY SRRQTTAGKVYRYHCRDGHQTQSIPADDLQQLVEERFLNALGDQEVHEMVYL PAESHQAEEEAQIAVQELTSALGRMKSNYAQQRIHTQLEALDKRIQELEGPLT SEARSEMRTGGGLYKDAWEEADEQGRRELLIKSGITAKAKLEGRVPNQSGGAL SFDLVVPEDLLARMSV
36	Hinder	ATGCGTGTCTTGGGAGAATCAGGCTGTCCAGGCTCAGCGACGAATCTACC AGTCCCGAGCGGCAGCGAGAGATCATCGAGGGTGGGCAAAGTCCAACG ACCACACCATCGTCGGTGGGCGAGAGGATCTGGACGTGTCCGGTCCGTC GACCCGTTTCGACACCCCGGCCTTGGGTCCGTGGCTGTCCGAGCCGAACT GCACGAGTGGGACATCCTCTGCGCATGGAAGCTCGACCGCCTCTCGCGGC GAGCGATCCCGATGAACAACTCTTCGGTTGGGTGATGGATCACGACAAG ACACTCGTATGCGTCAACGACAACATCGACCTCTCGACGTGGATCGGTTCG ATGGTCGCAAACGTATCGCTGGTGTGGCCGAGGGGGAGTTGGAGGCGA TCCGGGAGAGGACGACGGCCTCTACCGGAAGCTCCGGGAGTTGGGCCG GTGGCCGGGAGGACGACCGTCTACGGCTACCGCGCCGTGAGCGCGAG GACGACGAGGGTGGGTGTTGAGCCTGACCCGGTGTCTCGTGGTGGTCT CCGGTCGATCATCGACTGGGTCTGCGAGGGGAGTCCGTGAGTCGATAG CCAAGGATCTGACCGCGATGGGAGAGGTGTCCCGTCCGACTACGTCCGT CAGCGGGCCGGTGGGCCCCACGGGGACACCCGTGGCATGGCCGCACCA TCGTCAAGCTGTTGCGGTCCAAGACCCTCTGGGGTACGTACGCACAACG GGACGACCGTCCGGGACGAGAACGGTGTGCCGGTCCAGAAAGGCCACC GCTGGTCGACAGGACACCTTCAACCGGTGACGGCCGTCTCGATGACG GCTCCAGACCAAGACCGTCAACCGGACCTCGAAGGCGTCCCGCTACTCG GGGTGCCCTGTGTTGGGACTGCGAGAAGCCGTGTACTCCCGCCGCCAG ACGACGGCGGGGAAGGTGTACCGTACTACCTGCCGCGACGGCCACAC CCAGTCCATCCCGCCGACGACCTACAACAACCTGTGGAGGAGCGGTTCT GAACGCACTCGGGGACCAGGAGGTCCACGAGATGTTTACCTCCAGCGG AATCGCACCAGGCCGAGCTAGAGGAGGCTCAGATCGCGGTACAGGAGTTG ACGAGCGCCCTCGGGAGGATGAAGTCCAACCTACGCGCAGCAGCGCATCCA CACGCAACTGGAGGCTCTGGACAAACGCATACAGGAGCTTGAGGGACTAC CGACCTCCGAGGCCCGGTGCGAGATGCGCCGACAGGTGGGCTGTACAA GGACGCCTGGGAGGAGGCCGACGAGCAAGGTCGCCGGGAGCTTCTGATC AAGTCAGGGATCACGGCCAAGGCCAAGCTGGAGGGCCGGGTGCCCAACC AGTCCGGAGGGGCATTGTCGTTGATCTCGTTGTGCCAGAGGATCTTCTGG CACGAATGTCCGTGTAA
37	ICleared	MRVLGRRLISRATEESTSIERQRELVEQWAAAHEHEIVGWAVDQDVSGSVD FDAPALGPWLSDRKHEDILCAWKLDRLSRRRAIPMNKLGWMIDNDKTLV CVSENLDLGTWVGRMIANVIAGVAEGELEAIRERTTASHKKLRELGRWAGGP TYYGYVPKPRDGAGWELDIDLHAAGVLEIIEKTIAGQSTESIVVELNERGELSP SDYHRKRSKPIRGTKWNTSWLRTQLKSKTLLGHMTHNGETVYDDAGLPVQ KGPALIDRDYKQLQDALQSRGINRTKRRTGASPLLGVAVCDVCDGPLYYRQT KNQKGTAMLRQYICKHGRYGNTKANGGEPYNIQADLLEATVEELFLSKMGD LPRVERVFIPGEGHQHELETAERAVEDLTSLLGTITQEGARKRLLAQLSAAHERL AHLEELPSSEPRWETRETGETYREAWESATTEERRQILLKAGVTLKVQMKGRV PRVHPGVIVANWIEPHDIEKRLAS
38	ICleared	ATGCGAGTTCTTGAAGATTGCGAATCTCACGAGCCACTGAAGAATCTACC AGCATCGAGCGGCAGCGCGAGTTGGTCGAGCAGTGGGCGGCGGCTCACG AGCACGAGATCGTCGGTGGGCCGTGGATCAGGACGTGTCCGGATCGGTC GATCCGTTTCGACGCGCCAGCACTTGGGCCCTGGCTGTCTGACCATCGGAAA CACGAATGGGACATCCTGTGCGCCTGGAAGCTGGACCGCCTGTCCGGCGG AGCGATCCCGATGAACAAGCTGTTCCGTTGGATGATCGACAACGACAAGA CGCTCGTCTGCGTGTGCGAGAATCTGGACCTCGGGACGTGGGTCCGTGCG ATGATCGCCAACGTATCGCCGGGGTTCGAGAGGGTGAATTGGAGGCGAT CCGGGAGCGGACGACTGCCTCGACAAGAAGCTCAGGGAGCTGGGAAGG

		TGGGCCGGTGGGCCACGTAACGGCTACGTGCCGAAGCCCCGCGACGG CGCGGGGTGGGAGCTGGACATCGACCTGCACGCTGCGGGCGTGCTGAGG GAGATCATCGAAAAGACCATCGCCGGTCAGTCGACCGAGTCCATCGTGGT CGAGCTCAACGAGCGAGGGGAGCTGTCCCGTCCGACTACCACCGGAAGC GGAGCGGGAAGCCGATCCGGGGCACGAAGTGGAACACGTCGTGGCTGCG GACGCAGCTCAAGTCCAAGACCCTGCTTGGTCACATGACCACAACGGCG AAACCGTCTACGACGACGCCGGCCTCCCGTCCAGAAGGGCCCTGCGCTG ATCGACCGGGACACCTACAAGCAACTGCAGGACGCACTGCAGAGTCGCGG GATCAACCGGACCAAGCGCAGGACCGGGGCTCGCCGCTGCTCGGTGTCG CCGTCTGCGATGTCTGCGACGGGCGCTGTACTACCGCCAGACCAAGAAC CAGAAGGGCACGGCCATGCTGCGGCAGTACATCTGCAAGCACGGCCGCTA CGGCAACACCAAGGCCAACGGCGGGGAGCCGTACAACATCATTAGGCCG ATCTATTGGAGGCCACAGTCGAAGAGCTGTTCTCTCCAAGATGGGCGATC TGCCCCGCTCGAGCGGGTGTTTCATCCAGGCGAAGGACACCAGCATGAG CTGGAAACGGCTGAGCGAGCCGTGGAGGACTTGACATCGTTCTGGGCAC GATACCCAGGAGGGTGCAGAAAAGCGTCTCTGGCGCAACTCTCAGCCG CCCACGAACGGCTGGCCACCTCGAGGAGCTCCCGAGCTCCGAGCCCCGC TGGGAAACCCGAGAGACAGGGGAGACGTACCGCGAAGCGTGGGAGAGC GCGACGACCGAGGAGCGTCGCGAGATTCTGCTCAAAGCCGGTGTGACACT CAAGGTCCAGATGAAGGGCCGGGTGCCGAGAGTGCATCCCGGCGTCATC GTCGCGAACTGGATCGAGCCGCACGACATCGAGAAGCGCCTGGCTTCCTG A
39	KSSJEB	MRALVIRLSRVTDAATSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEQPFVIVAYRVDRLRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLLVDPVQRRERILEVYHRVVDNHEPLHLVAHDNLQR GILSPKDYFAQLQGREPQGRAWSATALKRSLISEAMLGYATLNGKTVRDDDG APLVRAEPILTREQLEALRAELVKDRTPAVSTPSLLRLVFCVCGEPAYKFA GGGRKHPRYRCRSMGFPHKCGNGTVAMAEWDAFCEEQVLDLLGDAERLEK VWVAGSDAAVELAELNAELVDLTSLIGSPAYRAGSPQREALDARIEALAARQE ELEGLEARPSGWEWRETGQRFGDWWREQDTAAKNTWLRSNMNVRITFDVR GGLTRTIDFGDLQEYEQHLRLGSVVERLHTGMS
40	KSSJEB	ATGAGAGCACTGGTAGTCATCCGCCTGTCCCGCTCACCGATGCTACGACT TCACCGGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCACAGCGCGG TTGGGACGTCGTCGGTGTGGCGGAGGATCTGGACGTCCTCGGAGCAGTCG ATCCGTTCGACCGGAAGCGCAGACCGAACCTGGCACGGTGGCTGGCATT GAGGAGCAACCGTTCGATGTGATCGTGGCGTACCGGGTAGACCGGTTGAC CCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCGGAGGACCACA AGAAGCTGGTCTCTCCGCGACCGAAGCGCACTTCGACACGACGACGCCG TTCGCGGCGGTCTGATCGCGCTTATGGGTACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCACATTTCAATATCCGCG CCGGTAAATACCGAGGTTCCCTGCCCGCTGGGGTTACCTGCCTACGCGCG TGGACGGGGAGTGGAGGCTGCTCGTCGACCCCGTGAACGAGAGCGCAT CCTCGAGGTCTATACCGCGTCGTCGACAACCACGAGCCGCTGCATCTGGT CGCCACGACCTGAACCAGCGCGGCATCCTGTGCGCCAAGGACTACTTCGC GCAGCTGCAGGGCCGGGAGCCGAGGGGCGGGCGTGGTCGGCTACCGCG TTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTGAA CGGTAAGACCGTCCGAGACGACGACGAGCTCCGCTGGTGCGGGCCGAG CCGATCCTGACGCGAGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTCAA GACCGACCGACCAAGCCCGCAGTGTCTACCCCGTCTGCTGCTGCGGG TGCTGTTCTGCGCAGTGTGCGGGGAGCCCGCATACAAGTTCGAGGGGGA GGACGTAAGCACCCGCGCTACCGCTGCCGCTCGATGGGGTCCCGAAGCA CTGCGGGAACGGCACGGTCGCAATGGCAGAGTGGGACGCATTCTGCGAG GAGCAGGTGCTGGATCTGCTCGGGGACGAGAGCGGCTGGAGAAGGTAT

		GGGTCGCAGGCTCGGACGCTGCTGTGGAGCTGGCAGAGCTCAACGCAGA GCTGGTGGACCTGACGTCGCTGATCGGCTCCCCGGCATAACCGCGCAGGTT CCCCGCAGCGGGAGGCATTGGACGCACGTATCGAGGCTCTGGGTGCACGG CAAGAGGAGTTGGAAGGGCTGGAGGCTCGCCCGTCGGGTTGGGAGTGGC GCGAAACCGGGCAGCGGTTCCGGTGACTGGTGGCGAGAGCAGGACACCGC AGCAAAGAACACCTGGCTCCGGTCGATGAACGTCGGGCTGACGTTGCACG TCCGCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGTAC GAGCAGCATCTCAGGCTCGGCAGCGTGGTGAACGGCTACACACCGGGAT GTCGTAG
41	Doom	MRALVIRLSRVTDATTSRPERQLESCQQLCAQRGWEVVGVAEDLDVSGAVD PFDRKRRPNLARWLAEEQPFQVIVAYRVDRRLRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQREIRILEVYHRVVDNHEPLHLVAHDNLNR GVLSPKDYFAKLQGREPQGREWSATALKRSLISEAMLGYATLNGKTVRDDDG APLVRAEPILTREQLAELRAELVKTDRAKPAVSTPSLLRLVFCVCGEPAYKFD AGRKIPRYRCSRFGFAQRCNGTVPIAEWDAFCEEQVLDLLGDSERLEKVVW AGSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALAARQEELEG LEARPSGWEWRETGQRFQDWREQDTSKNTWLRSMNVRLTFDVRGGTLT RTIDFGDLQEYEQHLRLGSVVEQLHAGMS
42	Doom	ATGAGAGCACTGGTAGTCATCCGCCTGTCCCGTGTCACCGATGCTACGACT TCACCTGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCACAGCGCGG TTGGGACGTCGTCGGTGTAGCGGAGGATCTGGACGCTCTCCGGAGCAGTCG ATCCGTTTCGACCGGAAGCGCCGCCGAACCTGGCACGGTGGCTGTCATTCT GAGGAGCAACCGTTCGATGTGATCGTGGCGTACCGGGTAGACCGGTTGAC CCGCTCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCCGAGGACCACA AGAAGCTGATCGTCTCCGCGACTGAATCGCACTTCGACACGACGTCGCCGT TTGCGGCGGTCTGATCGCGCTTATGGGAACGGTGGCGCAGATGGAGTTG GAGGCGATCAAGGAGCGGAACCGCTCGGCAGCACACTTCAACATCCGCGC CGGTAAGTACCGTGGTTCCTGCCGCCGTGGGGTTACATGCCCGCACGCG TGGACGGGGAGTGGCGGCTGGTGCCGGACCCGGTGCAGCGCGAGCGCAT CCTCGAGGTGTATCACCGCGTCGTCGACAACACGAGCCGCTGCACCTGGT GGCACACGACCTGAACCGGCGTGGTGTCTGTGCGCGAAGGACTACTTCG CAAAGCTGCAAGGTCGCGAGCCGCGGGGTCGGGAGTGGTGGGCTACCGC GCTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTGA ACGGAAGACCGTCCGAGACGACGAGCGAGCCCGCTGGTGCAGGCTGA GCCGATCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTCA AGACCGACCGGATGAAGCCCGCAGTGTCACCCCGTCGCTGCTGCTGCGG GTGCTGTTCTGTGAGTGTGTGGGGAGCCCGCATACAAGTTCGACGCCGG CCGGAAGATCCCCCGCTACCGCTGTAGGTCGTTCCGGGTCGCACAGCGCTG TGGGAACGGCACCATACCGATCGCAGAGTGGGACGCATTCTGTGAGGAGC AGGTGCTGGATCTGCTCGGGGACGAGAGCGGCTGGAGAAGGTGTGGGT CGCAGGCTCGGACGACGAGTGGAGCTGGCAGAGCTCAACGCGGAGCTG GTGGATCTGACGTCGCTGATCGGCTCCCCGGCATAACGGGCAGGCTCTCC GCAGCGCGAGGCACTGGACGCTCGTATCGCGGCGCTGGCCGCGCGCAG GAGGAGCTGGAAGGGCTAGAGGCTCGCCCGTCGGGCTGGGAGTGGCGCG AAACCGGGCAGAGGTTCCGGGACTGGTGGCGAGAGCAGGACACCGCGGC AAAGAACACCTGGCTTCGGTCGATGAACGTTCCGGCTGACGTTTCGACGTCC GCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGTACGA GCAGCATCTCAGGCTCGGCAGCGTGGTGAACAGCTACACACCGGGATGT CGTAG
43	Mundrea	MRVLGRLRISRATEESTSIRQREIVEQWASVNDHEIVGWAVDQDVSGSDP FDAPALGPWLSEHRKHEWDILVAWKLDRLSRAIPMNKLFQWMIDNEKTLV CVSENLDLGTWVGRMIANVIAGVAEGELEAIRERTTASHKKLRELGRWAGGP TYYGYVPKPRDGAGWELDIDPHAAGVLTEIIEKTIAGQSTESIVVELNERGELAP

		SDYHRKRNGKPIRGTKWSTSWLRSQKSKTLLGHMTHNGETVYDDAGLPVQ KGPALVDRDITYKQLQDALESRGINRTRRTGASPLLGAVCDVCEGPLYYRQT KNAKGTAMLRQYICKHGRYGNTQANGGEPYNIQADLLEATVEELFLSKMGD LPRVERVFIPEGHQHELETAERAVEDLTSLLGTITQESARKRLLAQLAAAHRL AHLEELPSSEPRWETRETGETYREAWEGATVEERRQILLKAGVTLKVQMKDR VPRVHPGVIVANWIEPHDIEKRLAS
44	Mundrea	ATGCGAGTTCTTGGGAAGATTGCGAATTTCCCGAGCCACTGAGGAATCTACC AGCATCGAGCGGCAGCGCGAGATCGTGGAGCAGTGGGCGTCCGTGAACG ACCACGAGATCGTCGGCTGGGCCGTGGACCAGGACGTGTCCGGATCGGTC GATCCGTTTCGACGCGCCAGCACTGGGCCCTGGCTGTCTGAGCACCGGAA ACACGAGTGGGACATCCTCGTGGCGTGGAAGCTCGACCGCTGTGCGGGC GAGCGATCCCGATGAACAAGCTGTTCTGGTTGGATGATCGACAACGAAAAG ACTCTGGTCTGCGTATCGGAGAATCTGGACCTCGGGACGTGGGTGGTCTG AATGATCGCCAACGTCATCGCCGGGGTCGAGAGGGTGAGCTGGAGGCC ATCCGGGAGCGGACGACCGCCTCGCATAAGAAGCTCAGGGAGCTGGGAA GGTGGGCCGGGGGCCGACGTACTACGGGTACGTGCCGAAGCCCCGCGA CGGCGCGGGGTGGGAGCTGGACATCGACCCGACGCTGCTGGCGTGCTG ACGGAGATCATCGAAAAGACCATCGCCGGTCAGTCGACGGAGTCCATCGT GGTCGAGCTGAACGAGCGCGGGGAGCTGGCTCCGTCCGACTACCACCGG AAGCGGAACGGCAAGCCGATCCGGGGGACGAAGTGAGCACGTCGTGGC TGCGGTGCGAGCTCAAGTCCAAGACCCTGCTCGGCCACATGACCCACAACG GTGAAACCGTCTACGACGACGCCGGCCTCCCTGTCCAGAAGGGCCCTGCCT TGGTCGACCGGGACACATAACAAGCAACTGCAGGACGCGCTGGAGAGTCGT GGGATCAACCGGACCCGGCGCCGAACCGGAGCCTCACCGCTGCTCGGTGT CGCCGTGTGCGATGTCTGCGAGGGGGCCGCTGTACTACCGGCAGACCAAGA ACGCCAAGGGCACGGCCATGCTGCGGCAGTACATCTGCAAGCACGGCCGC TACGGCAACACTCAGGCCAACGGCGGGGAGCCGTACAACATCATCCAGGC CGACCTCTTGAGGGCCACCGTCAAGAGCTGTTCTGTGCAAGATGGGTG ACCTGCCCCGCTCGAGCGGGTGTTTCATCCCCGGCGAAGGACATCAGCAC GAGCTGGAACGGCTGAGCGAGCCGTGGAGGACTTGACATCGTTCTGG GCACGATACCCAGGAGAGTGCGAGAAAGCGTCTCCTGGCTCAACTCGCC GCCGCCACGAGCGGCTGGCTCATCTCGAAGAGCTCCCGAGCTCCGAGCC CCGCTGGGAAACCCGAGAGACAGGGGAGACGTACCGCGAAGCGTGGGA GGGCGCGACGGTCGAGGAGCGTCGGCAGATTCTGCTCAAAGCCGGTGTG ACATTGAAGGTCCAGATGAAGGACCGAGTGCCGAGAGTGATCCCGGTGT CATCGTCGCGAACTGGATCGAGCCACACGACATCGAGAAGCGCCTGGCCT CTTGA
45	Museum	MRALVIRLSRVTDATTSPEQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLSFEEQPFVIVAYRVDRLTRSIRHLQQLVHWAEDHKLI VSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRGS LPPWGYMPARVDGEWRLLVDPVQREIRILEVYHRVVDNHEPLHLVAHDNQR GILSPKDYFAKLQGREPKGREWSATALKRSISEAMLGYATLNGKTVRDDDGA PLVRAEPILTREQLEALRAELVKTDRTKPAVSTPSLLRLVFCVAVCGEPAYKFTG GGRKNARYRCRSWGWQAQRGNGTVAMAEWDAFCEEQVLDLLGDAERLEK VWVAGSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALARQE ELEGLEARPSGWEWRETGQRFGDWWREQDTAAKNTWLRSMNVRLTFDVR GGLTRTIDFGDLQEYEQHLRLGSALDLVNAEKPTGR
46	Museum	ATGCGCGCTTTGGTAGTGATCCGCTTGTCCTCGTGACCGATGCTACGACT TCACCCGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG CTGGGACGTGTCGGGGTGGCGGAGGATCTGGACGTGTCCGGAGCGGTC GATCCGTTTCGACCGGAAGCGCCGCCAACCTGGCCCCGTGGCTGTCTGTT CGAGGAGCAACCGTTCGATGTGATCGTGGCGTACCGGGTGGATCGGTTGA CCCGCTCGATCCGGCATCTTCAGCAGCTGGTCCACTGGGCCGAGGACCACA AGAAGCTGATCGTCTCCGCGACCGAAGCGCACTTCGATACGACGACGCCG

		TTCGCGGCGGTCTGATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCGCATTTCAACATCCGCG CGGGGAAATACCGCGGCTCCCTGCCGCCGTGGGGTTACATGCCCCCGCGC GTGGACGGGGAGTGGAGGCTGCTCGTCGACCCCGTGCAGCGCGAACGCA TCCTCGAGGTCTATCACC CGTCTGTCGACAACCACGAGCCTCTGCATCTGG TCGCCCACGACCTGAACCAGCGTGGCATCCTGTCGCCGAAGGACTACTTCG CGAAGCTGCAGGGCCGAGAGCCCAAGGGCCGGGAGTGGTGGCTACCGC GCTGAAGCGCTCGTGTATCTCCGAGGCGATGCTCGGGTACGCGACTCTGA ACGGTAAGACCGTCCGAGACGACGACGGAGCCCCGCTGGTGGGGCTGA GCCGATCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTCA AGACCGACCGGACCAAGCCCGCAGTGTCTACCCCGTCTGCTGCTGCGG GTGTTGTTCTGCGCAGTGTGCGGTGAGCCCGCATACAAGTTCACCGGTGG CGGTAGGAAGAACGCTCGCTACCGCTGCCGGTCTGTTGGGGTTGGGCACAGC GGTGGCGTAACGGCACGGTGGCAATGGCGGAGTGGGACGCGTTCTGCGA GGAGCAGGTGCTGGATCTGCTCGGGGACGACAGCGTCTGGAGAAAGTC TGGGTAGCAGGTTCGGACTCGGCAGTCAACTCGCGAGGTGAACGCGG AGCTGGTGGACCTGACGTCGCTGATCGGCTCTCCGGCGTACCGGGTCTGGT TCTCCGCAGCGCGAAGCACTGGATGCTCGTATTGCGGCGCTGGCCGCGCG GCAAGAGGAGTTGGAAGGGCTAGAGGCTCGTCCGTCTGGCTGGGAGTGG CGCGAAACCGGGCAGCGGTTCCGGGACTGGTGGCGGGAGCAGGACACCG CGGCAAAGAACACCTGGCTTCGGTTCGATGAACGTTCCGGCTGACGTTTCGAC GTCCGCGGCGGGCTGACTCGGACGATCGACTTCGGGGATCTGCAGGAGTA TGAGCAGCATCTGAGGCTGGGCTCGGCTCTAGACCTCGTAAACGCAGAAA AGCCCCCTACGGGCCGCTAG
47	PattyP	MRALVVIRLSRVTDAATSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEQPFVIVAYRVDRLRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQRRERILEVYHRVVDNHEPLHLVAHDLNRR GVLSPKDYFAKLQGREPRGREWSATALKRS LISEAMLGYTTLNGKTVRDDDG APLVRAEPILTREQLSLRAELVKTDRTKPAVSTPSLLLRLVFCAVCGEPAYKFAG GQRKNPRYRCRSMGFPHCGNGTVAMAEWDAFCEEQVLDLLGDAERLEKV WVAGSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALAAARQEE LEGLEARPSGWEWRETGQRFQDWREQDASGKNTWLRS MNVRLTFDVRG GLTRTIDFGDLQEYEQHLRLGVSVERLHTGMS
48	PattyP	ATGAGAGCACTGGTGGTCATCCGACTGTCCCGCTCACCGATGCTACGACT TCACCCGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCACAGCGCGG TTGGGACGTCGTCGGTGTAGCGGAGGATCTGGACGCTCTCCGGAGCAGTCG ATCCGTTCGACCGGAAGCGCCGCCGAACCTGGCACGATGGCTAGCATTC GAGGAGCAACCGTTCGACGTCATCGTGGCGTACCGGGTGGACCGGTTGAC CCGCTCGATCCGGCATCTGCAGCAGCTGGTGCCTGGGCCGAGGACCATA AGAAGCTGGTCTCTCCGCGACCGAAGCCCACTTCGACACGACGACGCCG TTCGCGGCGGTCTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCAGCACATTTCAATATCCGCG CCGGTAAATACCGCGGTTCCCTGCCGCCGTGGGGTTACCTGCCTACGCGCG TGGACGGTGAGTGGCGGCTGGTGCCGGACCCGGTGCAGCGAGAGCGCAT CCTCGAGGTGTATCACC CGTCTGTCGACAACCACGAGCCGCTGCACCTGGT GGCCACGACCTGAACCGCGTGGTGTCTGTCGCCGAAGGACTACTTCG CGAAGCTGCAAGGCCGCGAGCCGCGGGGCGGGGAGTGGTGGCTACCGC GCTGAAGCGCTCGTGTATCTCGGAGGCGATGCTCGGGTATACGACGCTGA ACGGCAAGACCGTCCGAGACGACGACGGGGCTCCGCTGGTGGGGCCGA GCCGATCCTGACGCGCGAGCAGCTGGAATCGCTGCGGGCGGAACTGGTCA AGACCGACCGGACCAAGCCCGCAGTGTCCACCCCGTCTGCTGCTGCGG GTGTTGTTCTGCGCAGTGTGCGGGGAGCCCGGTACAAGTTCGCCGGTGG TCAGCGCAAGAACCCGCGCTACCGCTGCCGCTCGATGGGGTTCCCGAAGC

		ACTGCGGTAACGGTACGGTGGCGATGGCCGAGTGGGACGCGTTCTGCGA GGAGCAGGTGCTGGATCTGCTCGGGGACGCGGAGCGTCTGGAGAAAGTC TGGGTAGCCGGTTCGGACTCGGCAGTCGAACTCGCAGAGGTGAACGCGG AGCTGGTGGACCTGACGTCGCTGATCGGCTCTCCGGCATAACCGGGTTGGT TCTCCGAGCGCGAGGCGCTCGACGCTCGTATCGCAGCACTGGCCGCACG GCAAGAGGAGTTGGAAGGGCTAGAGGCTCGTCCGTCGGGCTGGGAGTGG CGAGAAACCGGGCAGAGGTTCCGGGACTGGTGGCGGGAGCAGGACGCCT CGGGTAAGAACACCTGGCTTCGGTTCGATGAACGTTCCGGTACGTTTCGAC GTCCGCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGTA CGAGCAGCATCTCAGGCTCGGCAGCGTGGTCAACGGCTACACACCGGGA TGTCGTAG
49	Rebeuca	MRVLGRLRLSRSTEESTSIERQREIVTAWAESNGHTLVGWAEDVDVSGAIDPF DTPSLGPWLDERRGEWDILCAWKLDRLGRDAIRLNKLFQWCEHKGTVASC SEGIDLSTPVGRILANVIAFLAEGEIRERVTSSKQKLREVGRWGKKPPFG YMGIPNPDGQHILVDPVAKPVVRRIVDDILDGKPLTRLCTELTEERYLTPAE YYATLKAGAPRQKAEPDETPAKWRPTALRNLLRSKALRGYAHKQGTVRDLK GQPVRLAEPLVDADEWELLQETLDRVQANWSGRRVEGVSPLSGVVVCITCDR PLHHDRLVLRPYGDYPYRYRCRDRHGKNLPAEMVETLMEESFLARVGDP VRERVVWQGTNWADLKEAVAAYDELVQAAGRAKSATAKERLQRQLDALD ERIAELESAPATEAHWEYRPTGGTYRDAWETADTDERREILRRSGIVLAVGVD GVDGRRSKHNPGALHFDPRVPEELTQRLGVS
50	Rebeuca	ATGCGAGTGCTCGGAAGATTGCGTTTGTCAGGTCAACGGAGGAATCTAC CTCCATCGAGCGGCAGCGGGAGATTGTCACCGCTTGCGGAGAGTCTAACG GTCACACCCTCGTGGGTGGGCCGAGGACGTTGACGTGTCTGGTGCAATC GACCCGTTGACACCCCGTCGCTCGGTCCGTGGCTGGACGAGCGGCGAGG CGAGTGGGACATCCTCTGCGCGTGAAGCTCGACCGGCTGGGCCGTGATG CCATCCGGCTCAACAAGCTCTCGGTTGGTGCCAGGAGCACGGCAAGACC GTGGCCTCGTGACGAGGGCATCGACCTCAGCACGCCGGTTCGGTTCGGCT CATCGCCAACGTCATCGCCTTCCTGGCCGAGGGTGAGCGGGAAGCGATCC GCGAGCGCGTCACGTCCTCGAAGCAGAAGCTGCGCGAGGTAGGCCGGTG GGGTGGTGGTAAGCCGCCGTTTCGGTTACATGGGCATCCCCAACCCGGACG GTCAGGGACACATCCTCGTGGTTCGATCCGGTGGCCAAGCCGGTGGTGCGC CGGATCGTGGACGACATCCTCGACGGCAAGCCGCTCACGCGGCTCTGCAC GGAGCTGACCGAGGAGCGATACCTCACGCCGGCAGAGTACTACGCAACCC TCAAGGCAGGGGCACCGAGGCGAAGGCCGAGCCGACGAAACCCCTGC AAAGTGGCGGCCGACCGCTCTCCGCAACCTTCTCCGGAGCAAGGCCCTGC GGGGTTACGCCCATCACAAGGGGCAGACCGTCAGGGACCTCAAGGGGCA GCCTGTGCGCCTCGCTGAGCCACTGGTGGACGCCGACGAATGGGAATAC TGCAGGAGACTCTGACCGCGTACAGGCAAACCTGGTGGGTCGACGGGTC GAGGGTGTGAGCCCGCTGTCCGGTGTCTGGTGTGATCACCTGTGACCGT CCGCTGCACCACGACCGGTATCTGGTGAAGCGGCCATACGGTACTACCC CTACCGGTACTACCGGTGCCGTGACCGCCACGGCAAGAACCTCCGGCAG AGATGGTCAAAACCTCATGGAAGAGTCCTTCTGGCACGCGTGGGTGAC TATCCGGTGCGGGAGAGGGTCTGGGTCCAGGGTGACACGAATTGGGCAG ATCTGAAGGAGGCTGTGGCAGCCTACGACGAACTGGTGCAGGACGCCG CCGCGCCAAGAGCGCAACGGCGAAGGAGAGACTGCAGAGGCAGCTCGAT GCACTAGACGAGCGGATCGCGGAGCTTGAGTCCGACCCGCCACCGAGGC CCACTGGGAGTACCGGCCGACCGGAGGTACCTACCGGGACGCTGGGAG ACAGCAGACACCGACGAGCGCCGCGAGATCCTGCGGCGGTGCGGGATCG TCCTGGCGGTGCGAGTCGACGGCGTGGATGGCCGTGCTCCAAGCACAAC CCAGGAGCCCTGCACTTCGACTTCGGGTCCCGAGGAACTGACCCAGCG GCTCGGAGTCTCTGA
51	Rey	MTATLERHLDTPQQEALRVGVYLRMSTDKELGIDRQREDCLALAERLWVPV EYIDNDRSATKENVKREGFDALSEDIRDGRIDSIIITWRSRDLRYKMKDLLPLIDI

		QGVNKTGKRIPIETCQTGLIDLTTDAGRMTAKILAAVSENEGEVVRTARQMRAY EQIADSGRRLLGAPAFGWTHDPRDPQIVPEEAAAIRQAYADVLGCTLSIAKK WNEDGLRTRTGNQFVGSVVGKILRSPRNAGLLTFRDEIVGEGTWEPIVDRET WEAACAVLDQKNTGKKGPRVRSTLLSGIVRCGACGNKMAAGKNSNGEPIYK CKRYEVCKHGVTRVRKKVDKYVEMSMVAKLEQRKWIVGTQVDAEQAKGLH AEAEALRARKASFGEALADGTLTPAQVKDATDRVNAKLEEIDAKLARLTRSRVF DGLLGHDDEKVLWGLDLERKRAIIESLCDKIVIQHVGGTGRSAAKLPLGHAIKI HWHDPNSND
52	Rey	ATGACAGCGACCCTCGAGCGACACCTCGACACCCCGCAGCAGGAGGCCCT GCGGGTGGGTGTCTACCTGCGCATGTCCACCGACAAGGAGCTGGGCATCG ATCGCCAGCGCGAGGACTGCCTGGCGCTGGCTGAGCGTCTGGGCTGGGTT CCCGTCGAGTACATCGACAACGACCGCAGCGCCACCAAGGAGAACGTCAA GCGCGAAGGCTTCGATGCGCTGAGCGAGGACATCCGCGATGGCCGGATC GACAGCATCATCACGTGGCGCAGCGACCGGCTGTACCGAAAGATGAAGG ATCTACTCCCCCTGATCGATCTGATCCAGGGTGTAAATAAGACAGGCAAGC GGATCCCCATCGAAACCTGCCAGACTGGCCTGATCGATCTCACCACCGACG CCGGTCGGATGACAGCGAAGATCCTTGCCGCTGTCTCGGAGAACGAGGGC GAGGTAAGGACGGCGCGACAGATGCGAGCCTACGAGCAGATCGCCGACA GTGGCCGTCGCTGGGCGCTCCCGCTTCGGCTGGACCCATGACCCAGG GACCCGAGATCGTGCCCGAGGAGGCCGCTGCGATCCGGCAGGCGTACG CCGACGTGCTCGCTGGCTGCACCCTCTACTCCATCGCGAAGAAGTGAACG AGGACGGTCTGCGCACCAACGCGTGGCAACCAGTTCGTCGGGTGGTCTGTC GGGAAGATCCTGCGAAGCCCCGCAATGCGGGCCTGCTGACCTCAGGGA CGAGATCGTGGGCGAAGGCACCTGGGAGCCCATCGTTGATCGGGAAACCT GGGAGGCAGCCTGTGCGGTGCTCGACCAGAAGAACACCGGCAAGAAGGG TCCACGTGTACGGTCGACCCTCCTGTCGGGGATCGTGCGTGCGGTGCCTG CGGCAACAAGATGGCTGCCGGGAAGAACTCCAACGGTGAGCCCATCTACA AGTGCAAGCGCTACGAGGTCTGCAAGCACGGTGTACCCGTGTGCGTAAG AAGGTCGACAAGTATGTGAGATGTCCATGGTGGCGAAGCTGGAGCAGC GCAAGTGGATTGTGCGCACCCAGGTGACGCGGAGCAGGCCAAGGGGCT TCACGCTGAGGCTGAAGCTCTACGCGCTCGCAAGGCATCGTTCCGGTGAGG CCCTGGCCGACGGCACCTCACACCGGCTCAGGTGAAGGACGCCACCGAC CGGGTGAACGCCAAGCTGGAAGAGATCGATGCCAAGCTAGCCCGCCTTAC ACGATCTCGCGTGTTCGACGGGCTGCTCGGTCATGACGATCTGAAAAGGT TTGGCTGGGACTGGATCTCGAGCGTAAGCGTGCGATCATCGAGTCGCTAT GCGACAAGATCGTGATCCAGCATGTGCGGGCAGACGGGTCGCTCTGCGGCC AACTGCCCTCGGTACGCCATCAAGATCCACTGGCATGATCCAGCAAC GACTGA
53	Sarfire	MRALVIRLSRVTDATTSRPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFEEQPFQVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQREIRILEVYHRVVDNHEPLHLVAHDNRR GVLSPKDYFAKLQGREPQGREWSATALKRSLISEAMLGYATLNGKTVRDDD APLVRAEPILTREQLRALRAELVKTSRAKPAVATPSLLRLVFCVCGEPAYKFD AGRKIPRYRCRSFGFAVRCGNGTVPIAEWDAFCEEQVLDLLGDSERLEKVVV AGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALAARQEELEG LEARPSGWEWRETGQRFQDWWRDQDTAGKNTWLRSMNVRLTFDVRGGL TRTIDFGDLQEYEQHLRLGSLDLVNAEKPPTGR
54	Sarfire	ATGAGAGCACTGGTAGTCATCCGCTGTCCCGCTCACCGATGCTACGACT TCACCGGAGCGCCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG TTGGGACGTGCTCGGTGTAGCAGAGGATCTGGACGTCTCCGGAGCAGTCG ATCCGTTGACCGGAAGCGCCGCCAACCTGGCCCGGTGGCTAGCGTTC GAGGAGCAGCCGTTGATGTGATCGTGGCGTACCGGGTGGACCGGTTGAC CCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCCGAGGACCACA

		AGAAGCTGGTCGTCTCCGCAACCGAAGCACACTTCGACACGACGACGCCG TTCGCAGCAGTCGTATCGCGTTATGGGAACGGTGGCACAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCGCATTTCAATATCCGCG CCGGGAAATACCGCGGCTCCCTGCCGCCGTGGGGTTACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCGGACCCTGTGCAGCGAGAGCGCA TCCTCGAGGTGTATCACCGCGTCGTCGACAACCACGAGCCGCTGCACCTGG TGGCCACGACCTGAACCGGCGAGGCGTCCTGTCGCCTAAGGACTACTTC GCGAAGCTGCAGGGCCGCGAGCCGAGGGCCGGGAGTGCTCGGCTACCG CGCTGAAGCGCTCGCTGATCTCTGAGGCGATGCTCGGGTACGCGACGCTG AACGGTAAGACCGTCCGAGACGACGACGGAGCCCCGCTGGTGCGGGCTG AGCCGATCCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTC AAGACCTCCCGGGCAAAGCCCGCAGTGGCTACTCCGTCGCTGCTGCTGCG GGTGTTGTTCTGCGCGGTGTGCGGGGAGCCCGCTACAAGTTCGACGCCG GTCGGAAGATCCCCGCTACCGCTGCCGGTCTTCGGGTTGCGGTTTCGCT GCGGGAACGGCACGGTGCCGATCGCGGAGTGGGACGCGTTCTGCGAGGA GCAGGTGCTCGATCTGCTCGGGGATTCGAGCGGCTGGAGAAAAGTCTGG GTAGCCGGCTCGGACTCCGCGGTAGAACTCGCGGAGGTGAACGCGGAGC TGGTGGATCTGACGTCGCTGATCGGCTCCCGGCATACCGGGCCGTTCTC CGCAGCGGGAGGCACTGGACGCTCGTATCGCAGCGCTGGCCGACGGCA GGAGGAGCTGGAAGGGCTAGAGGCTCGCCGTCGGGTTGGGAGTGGCGC GAAACCGTCAAGAGTTCGGGGACTGGTGCGGGATCAGGACACCGCAG GTAAGAACACCTGGCTACGGTCGATGAACGTTGCGCTGACGTTGACGTC CGCGGTGGTCTGACTCGCACGATCGACTTCGGTGATCTGCAGGAGTACGA GCAGCATCTGAGGCTGGGCTCGGCTCTAGACCTCGTAAACGCAGAAAAGC CCCCTACGGGTCGCTAG
55	Scowl	MRALVIRLSRVTDATTSERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLSFEDQPFVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLHLVAHDNRR GVLSPKDYFAQLQGREPQGRAWSATALKRSISEAMLGYTTLNGKTVRDDD APLVRAEPILTRELEALRAELVKDRTPAVSAPSLRLVLCVAVGEPAYKFD AGRKIPRYRCSRFGFAQRGNGTIPIAEWDAFCEEQVLDLLGDSEKLVWVA GSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDSRIAALAAARQEEGLE ARPSGWEWRETGQRFQDWWRREQDIAAKNTWLRSMNVRLTFDVRGGLTRT IDFGDLQEYEQHLRLGSVVERLHTGMS
56	Scowl	ATGAGAGCACTGGTAGTCATCCGACTGTCCCGCTACCGATGCTACGACC TCACCGGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG TTGGGACGTCGTCGGTGTAGCAGAGGATCTGGACGCTCTCGGGAGCAGTCG ATCCGTTGACCGGAAGCGTCGCCGAACCTGGCCCGGTGGCTGTCGTTT GAGGATCAACCGTTCGATGTGATCGTGGCGTACCGGGTAGACCGGTTGAC CCGCTCGATCCGGCATCTTCAGCAGCTGGTCCACTGGGCCGAGGACCACA AGAAGCTGGTCGTCTCCGCGACCGAAGCGCACTTCGATACGACGACGCCG TTCGCAGCAGTCGTATCGCGTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCGGCGCATTTCAATATCCGCG CCGGGAAATACCGAGGCTCCCTGCCGCCGTGGGGATACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCAGACCCGGTGCAGCGAGAGCGCA TCCTCGAGGTGTATCACCGCGTCGTCGACAACCACGAGCCGCTGCACCTGG TGGCCACGACCTGAACCGGCGTGGTGTCTGTCGCCGAAGGACTACTTC GCGCAGCTGCAGGGCAGGGAGCCGAGGGCCGGGCGTGCTCGGCTACCG CGCTGAAGCGCTCGCTGATCTCTGAGGCGATGCTCGGGTATACGACGCTG AACGGCAAGACCGTCCGAGACGACGACGGGGCTCCGCTGGTGCGGGCTG AGCCGATCCTGACCCGCGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTC AAGACCGACCGACCAAGCCCGCAGTGTCCGACCGCTCGCTGCTGCTGCG GGTGTTGTTCTGCGCGGTGTGCGGGGAGCCTGCCTACAAGTTCGACGCCG

		GCCGGAAGATCCCCGCTACCGCTGCAGGTCGTTCCGGGTTCCGACAGCGC TGCGGGAACGGCACCATAACCGATCGCAGAGTGGGACGCATTCTGCGAGG AGCAGGTGCTGGATCTGCTCGGGGACTCGGAGCGTCTGGAGAAAGTCTG GGTAGCAGGCTCGGACTCGGCAGTCGAACTCGCAGAGGTGAACGCAGAG CTGGTGGACCTGACGTCGCTGATCGGCTCTCCGGCATAACCGGTCGGTTCT CCGCAGCGCGAGGCACTCGACTCCCGGATCGCAGCACTGGCCGCACGGCA GGAGGAGTTGGAAGGGCTGGAGGCACGTCCGTCGGGTTGGGAGTGGCGC GAAACCGGGCAGCGGTTCCGGGACTGGTGGCGGGAGCAGGACATCGCAG CAAAGAACACCTGGCTTCGGTCGATGAACGTTCCGGTTGACGTTTCGACGTCC GCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGTACGA GCAGCATCTCAGGCTCGGCAGCGTGGTGAACGGCTACACACCGGGATGT CGTAG
57	Severus	MPSKRALLVIRLSRVTDATTSPERQLADCQALCAQRGYEVAGVAEDLDVSGSI DPFDRKKRPNLAHWLHDRHNEFDVVVAYRVDRLTRSVRYLQKLVNWAEDH DKLVVSATEPHFDTTSPFAAVLIALLGTVAQMELEIAERNRSAARNHNRAGKY RGSKPPWGYMPQRDDEGVWRLVQDPDQVKIIHEVVQRVLEGEPTQRIANDL TLRGIPTPKDAFAISQGRKPEGLAWNMTTLKRSLKSEAMLGRVTNAAGKSIRK EDGAPVIRSTPILTREVFDVQVELETRARSGGPTRSTALLRVIYCAICGMPA YQYSGGTSGBKASRYRCSSSPRSNTDPNIKKCGNRTFVAEADAVVVKTLGLL GDSEKKEKIWESGCDHSAELADLDATLADLTDQLGTGVFARGTPQRRARDARI AELAARQAALSAEAVRPAGWTWHGTGERFSDWVERQDVTAKNVWLRS NIQLTFDRERFYLDLGDIVQLTEQFDPQGPVAQWQGLLAAMQAEGIAQVEIR GGEAQATPRDDMRGALGMTGAPLVD
58	Severus	ATGCCATCGAAGAGAGCATTGCTGGTCATCCGGCTCAGCCGGGTGACGGA TGCAACGACCTCCCCGAGCGTCAGCTCGCGGACTGCCAGGCGCTGTGTG CTCAGCGCGGATACGAGGTGGCAGGCGTGGCCGAGGACCTCGACGTGTC CGGGTCCATCGACCCGTTCCGACCGGAAGAAGCGGCCAACCTGGCCCACT GGCTCCACGACCGTCACAACGAGTTCGACGTGGTGGTTCCTACAGGGTA GATCGCCTCACCCGATCCGTGCGTTACCTGCAGAAGCTGGTCAACTGGGCA GAGGACCACGACAAGCTCGTCGTCTCCGCAACCGAGCCGCACTTCGACAC CACGAGCCCGTTCGCCGCCGTGCTGATCGCACTGCTGGGCACCGTGGCCC AGATGGAGCTGGAGGCCATCGCAGAGCGCAACCGCTCTGCTGCCCACAC AACATCAGGGCAGGGAAGTACCGAGGGTCCAAGCCCCCATGGGGTACAT GCCCCAGAGAGACGATGAGGGGGTCTGGCGGCTGGTGCAGGACCCTGAT CAGGTCAAGATCATTACGAGGTCTGTCAGCGGGTGTGGAGGGTGAGCC GACGCAGCGGATCGCCAATGACCTGACCTGCGCGGGATTCCGACGCCGA AGGATGCGTTCGCGATCTCCAGGGCCGCAAGCCCCGAGGGGCTGGCGTG GAACATGACGACGCTCAAGCGGTGCTTGAAGTCGGAGGCAATGCTCGGCC GCGTGACGAACGCGGCGGGGAAGTCCATCCGGAAGGAGGACGGGGCCCC GGTGATCCGGTCCACCCCGATCTGACGAGAGAGGTGTTCCGACCGGTGTC AGGTGAGCTTGAGACTCGGGCTCGGAGCGGAGGTCCGACGACGCGATC TACCGCACTGCTGCTGCGGGTATCTACTGCGCAATCTGCGGTATGCCCGC ATACCAGTACAGCGGTGGTACCAGCGGCAAGGCATCTCGGTACCGGTGCA GCTCGTCTCCCCGAGCAACACAGACCCGAACATCAAGAAGTGCGGTAAC CGAACATTGCGGGTGGCCGAGGCCGACGAGTGGTGGTCAAGACCCTGCT GGGTCTGCTCGGTGACTCCGAGCGCAAGGAGAAGATCTGGGAATCCGGCT GTGACCACTCGGCAGAGCTGGCCGACCTCGACGCCACGCTGGCCGATCTG ACCGACCAGCTTGGTACCGGAGTGTTCGCCAGGGGGACGCCACAGCGGG CCCGTCTGGACGCCAGGATCGCAGAGCTGGCCGCGAGGCAGGCTGCCCTA TCGGCAGAGGCCGTTCCGCCCCGAGGATGGACGTGGCACGGCACTGGTG AGCGTTTCAGTGATTGGTGGGAGCGGCAGGACGTTACCGCGAAGAATGTC TGGCTGCGGTGATGAACATCCAACACGTTTCGACCGTGAGCGGTTCTAT CTCGACCTCGGAGACATCGTCCAACGACCGAGCAGTTTGACCCGACGGG CCCGGTGGCCAGTGGCAGGGTCTGCTCGCAGCGATGCAAGCCGAAGGC

		ATTGCCGGCGTGGAGATTCGAGGGGGTGAAGCACAGGCGACACCGAGGG ATGACGACATGAGAGGGGGCCCTCGGAATGACAGGAGCCCCCTGGTCGAT TAG
59	Sheen	MRVLGRLRLSRSTEESTSIERQREIVTAWADSNIGHTVVGWAEDVDVSGAVDP FDTPSLGPWLDERRGEWDILCAWKLDRLGRDAIRLNKFLWCQENGKTVA STSEGIDLGTVPVGRLIANVIAFLAEGEREAIRERVTSKQKLRVGRWGGGKPPFG YMGIPNPDGQGYILVVDPIAKPVVRRIVDDVVDGKPLTRLCAELTEERYLTPAE YYATLKAGAPRQQADPGEVIAKWRPTAVRNLLRSKALRGHAHHRGQTVRDD QGRPVLRAEPLVDSDEWELLQETLDGIQADFSGRRVEGASPLSGVAVCMTCG RPLHFSRHMVCRPYGDYPYQYRCQDRHGKNVPADVLDELVEENFLLKVG DYPVRERVWVQGD TNWADMKEAVAAYDELVQAAGRAKSATAKERLQRQLDA LDTRIAELESAPATEAHWEYRETGSTYRDAWENS DTDQRRELLKSGITVAVGI DGVEGRRSKHNPGALRFDIRVPAELTQRLGAS
60	Sheen	ATGCGTGTACTAGGTAGACTGCGTCTGTCCAGGTCAACGGAGGAATCTAC CTCCATCGAGAGGCAACGAGAGATCGTCACCGCATGGGCCGATTCTAACG GCCACACGGTCGTCGGTTGGGCAGAGGACGTTGACGTCTCGGGTGCAGTG GACCCGTTGACACACCGTCGCTGGGGCCGTGGCTGGATGAGCGCCGGG GCGAGTGGGACATCCTCTGCGCCTGGAACTGGACCGCTGGGTCTGTGAT GCCATCCGGCTCAACAAGCTGTTCTGTGGTGCCAGGAGAACGGCAAGAC GGTGGCGTCCACGAGTGAGGGTATCGACCTCGGTACGCCGTTGGTCGGC TCATCGAAACGTGATCGCGTTCCTTGCCGAGGGTGAGCGTGAGGCCATC CGGGAGAGGGTCACGTCTCGAAGCAGAAGCTGCGGGAGGTAGGTGGT GGGGTGGCGGTAAGCCGCCCTTCGGTTACATGGGCATCCCCAACCCCGAC GGACAGGGCTACATCCTCGTGGTCGATCCCATCGCCAAGCCGGTGGTGCG CCGGATCGTGGACGACGTTGTCGACGGCAAGCCGTGACGCGGCTGTGCG CCGAGCTGACCGAGGAGCGGTACCTGACGCCTGCGGAGTACTACGCCACC CTCAAGGCCGGTGCCCCGAGGCAGCAGGCCGACCCCGGTGAGGTGATCG CCAAGTGGCGTCCGACCGCCGTCCGCAACCTGCTCCGACGAAAGCCCTCC GGGGGCACGCCACCACAGGGGACAGACCGTCAGAGACGACAGGGGACG CCCTGTGCGGCTCGCTGAGCCGCTTGTGGACTCCGATGAATGGGAGTTGC TGCAGGAGACTCTGGACGGCATCCAGGCAGACTTCTCCGGTCGTCGGGTC GAGGGGGCAAGCCCGCTCTCCGGTGTCGCCGTCTGCATGACGTGCGGACG CCCCCTGCACTTCTCCCGGCACATGGTGAAGCGCCCTACGGCGACTACCC GTACCACTACTACCGCTGCCAGGACCGGCACGGGAAGAACGTGCCCGCCG ACGTACTCGATGAATTGGTCGAGGAGAACTTCTCCTGAAGGTGGGCGAC TACCCGGTGCGTGAGCGGGTGTGGGTCCAGGGTGACACCACTGGGCCG ACATGAAGGAGGCAGTAGCCGCATACGACGAACCTGGTGAGGCCGCTGG CCGTGCTAAGTCGGCAACCGCCAAGGAACGCCTGCAGCGCCAACCTGGACG CCCTCGATACCCGCATCGCAGAGCTGGAGTCGGCCCCGGCAACCGAGGCC CACTGGGAGTACCGGGAAACCGGTAGCACCTACCGGGACGCCTGGGAGA ACTCCGACACCGATCAGCGCCGGGAGCTGTTGAAGAAGTCTGGCATCACC GTGGCCGTCGGCATCGACGGCGTGGAAGGCCGTGCTCCAAACACAATCC CGGTGCCCTGCGCTTCGACATCCGAGTGCCGGCAGAACTGACCCAGCGCC TGGGAGCGTCCTGA
61	SkiPole	MRALVVIRLSRVTDATTSERQLESCRQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLA FEEQPFDVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQRE RILEVYHRVVDNHEPLHLVAHD LNRR GVLSPKDYFAKLQGREPQGREWSATALKRS LISEAMLGYATLNGKTVRDDD G APLVRAAPILTREQLEALRAELVKTDRTKPAVSTPSLLLRLVFCVCGEPAYKFA GGGRKHPRYRCRSMGFPKHCGNGTVMAAEWD AFCEEQVLDLLGDAERLEK VWVAGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALARQE ELEGLEARPSGWEWRETGQRF GDWWRDQDTAGKNTWLRSMNVRLTFDVR GGLTRTIDFGDLQEYEQHLRLGAALDLVNAEKPTGR

62	SkiPole	ATGAGAGCTCTCGTCGTGATCCGCTTGTCGCCGCTCACCGATGCTACGACC TCACCGGAGCGTCAGCTGGAGTCTTGCCGGCAGCTCTGCGCCAGCGCGG TTGGGACGTCGTCGGTGTAGCGGAGGATCTGGACGTCTCCGGAGCAGTCG ATCCGTTCGACCGGAAGCGCAGACCGAACCTGGCACGGTGGCTAGCATT GAGGAGCAACCGTTCGACGTCATCGTGGCGTACCGGGTGGACCGGTTGAC CCGCTCGATCCGGCATCTTCAGCAGCTGGTCCACTGGGCGGAGGACCACA AGAAGCTGGTCGTCTCCGCGACCGAAGCCCACTTCGACACGACGACGCCG TTCGCGGCGGTCTGTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCACATTTCAATATTCGCG CCGGGAAATACCGCGGTTCCCTGCCGCCGTGGGGTTACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCGACCCGGTGACGCGGAGCGCA TCCTCGAGGTGTATCACCGCGTCGTCGACAACCACGAGCCGTGACCTGG TGGCACACGACCTGAACCGGCGTGGTGCTCTGCGCCGAAGGACTACTTC GCAAAGCTGCAAGGTCGCGAGCCGAGGGGCGGGAGTGGTTCGGCTACCG CACTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTG AACGGTAAGACCGTCCGAGACGACGACGGAGCCCCGCTGGTGCGGGCTG CGCCGATCCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTC AAGACCGACCGGACCAAGCCCGCAGTGTCTACCCCGTCTGCTGCTGCG GGTGCTGTTCTGCGCAGTGTGCGGTGAGCCCGCATACAAGTTCGACAGGT GAGGACGCAAGCACCCGCGCTACCGCTGCCGCTCGATGGGTTTCCGAAG CACTGCGGTAACGGTACGGTCGCAATGGCCGAGTGGGACGCATTCTGCGA GGAGCAGGTGCTGGATCTGCTCGGGGACGACAGCGTCTGGAGAAAGTC TGGGTAGCTGGCTCGGACTCAGCAGTAGAACTCGCAGAGGTGAACGCAG AGCTGGTGGACCTGACGTCGCTGATCGGTTCCCCGGCGTACCGGGCCGT TCTCCGACGCGGAGGCGCTGGATGCTCGTATCGCGGCGCTGGCCGCGCG GCAGGAGGAGTTGGAAGGGCTAGAGGCTCGCCCGTGGGCTGGGAGTGG CGCGAAACCGGGCAGAGGTTGCGGGACTGGTGGCGGGATCAGGACACCG CGGGTAAGAACACCTGGCTCCGGTCGATGAACGTTGCGCTGACGTTTCGAC GTCCGAGGCGGCCTGACTCGTACGATCGACTTCGGGGATCTGCAGGAGTA TGAGCAGCATCTGAGGCTGGGCGCGGCTCTAGACCTCGTAAACGCAGAAA AGCCCCCTACGGGCCGCTAG
63	Switzer	MRALVIRLSRVTDATTSRPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFVIVAYRVDRLRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQREIRILEVYHRVVDNHEPLHLVAHDLNRR GVLSPKDYFAKLQGREPQGREWSATALKRSLISEAMLGYATLNGKTVRDDDG APLVRSEPILTREQLEALRAELVKTDRTKPAVSTPSLLLRLVFCVCGEPAYKFTG GGRKNARYRCRSWGWAQRCNGTVMAAEWDAFCEEQVLDLLGDAERLEK VWVAGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALAARQE ELEGLEARPSGWEWRETGQRFGDWWRDQDTAAKNTWLRSNMNVRITFDVR GGLTRTIDFGDLQEYEQHLRLGSVVERLHTGMS
64	Switzer	ATGAGAGCACTGGTAGTCATCCGCCTGTCCCGCTCACCGATGCTACGACT TCACCGGAGCGTCAGCTGGAGTCTTGCCAGCAGCTCTGCGCCAGCGCGG TTGGGACGTCGTCGGTGTAGCGGAAGATCTGGACGTCTCCGGAGCAGTCG ATCCGTTCGACCGGAAGCGCGCCCGAACCTGGCACGGTGGCTAGCATT GAGGAGCAACCGTTCGATGTCATCGTGGCGTACCGGGTAGACCGGCTGAC CCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCGGAGGACCACA AGAAGCTGGTCGTCTCCGCGACCGAAGCCCACTTCGACACGACGACGCCG TTCGCGGCGGTCTGTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCACATTTCAATATCCGCG CCGGGAAATACCGAGGTTCCCTGCCGCCGTGGGGATACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCGACCCGGTGACGCGAGAGCGC ATCTCGAGGTGTATCACCGCGTCGTCGACAACCACGAGCCGCTGACCTG GTGGCCACGACCTGAACCGGCGTGGTGCTCTGTCGCTAAGGACTACTTC

		GCGAAGCTGCAAGGCCGAGAGCCGAGGGCCGGGAGTGGTCGGCTACCG CGCTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTG AACGGTAAGACCGTACGAGACGACGACGGGGCTCCGCTGGTGCGGTCTG AGCCGATCCTGACCCGCGAGCAACTAGAGGCGCTGCGCGCCGAGCTCGTC AAGACCGACCGGACCAAGCCCGCAGTGTCTACTCCGTCGCTGCTGCTGCG GGTGCTGTTCTGCGCAGTGTGCGGGGAGCCCGCATACAAGTTCACCGGTG GCGGTAGGAAGAACGCACGCTACCGCTGCCGGTCGTGGGGCTGGGCGCA GCGGTGCGGGAACGGAACGGTCGCGATGGCTGAGTGGGACGCGTTCTGC GAGGAGCAGGTGCTGGATCTGCTCGGTGACGCAGAGCGTCTGGAGAAAG TCTGGGTAGCCGGTTCGGACTCGGCGGTGAACTCGCAGAGGTGAACCGG GAGCTGGTGGACCTGACGTCGCTGATCGGTTCTCCGGCCTACCGGGCAGG TTCTCCGACGCGGAGGCGCTGGATGCTCGTATCGCGGCGCTGGCCGCGC GGCAGGAGGAGTTGGAAGGGCTAGAGGCTCGCCCGTCGGGCTGGGAGTG GCGCGAAACCGGTACAGAGTTTCGGTGACTGGTGGCGGGATCAGGACACC GCGGCAAAGAACACCTGGCTCCGGTCGATGAACGTTTCGGCTGACGTTCA CGTCCGCGGCGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGT ACGAGCAGCATCTCAGGCTCGGTAGCGTGGTGAACGGCTACACACCGGT ATGTCGTAG
65	Theia	MRVLGRIRLSRMMEESTSVRQREFIETWARQNDHEIVGWAEDLDVSGSVD PFDTQGLGPWLKEPKLREWDILCAWKDLRLARRAVPLHKLFGMCQDEQKVL VCVSDNIDLSTWVGRLVASVIAGVAEGELEAIRERTLSSQRKLRELGRWAGGK PAYGFKAQEREDSAGYELVHDEHAANVMLGVIEKVLQSTESVARELNEAG ELAPSDYIRARAGRKTRGTKWSNAQIRQLLSKTLGHVTHNGATVRDDDGIP IRKGPALISEEKFDQLQAALDARSFKVTNRSKASPLLGVAICGLGRPMHIRQ HRRNGNLRYRYRCDSGSHSGGGGAPEHPSNIIKADDLEALVEEHFLDEVGRF NVQEKVYVPASDHRAELDEAVRAVEELTQLLGTMTSATMKSRLMGQLTALD ERIALLENLPSEEARWDYRATDQTYAEAWEEADTEGRRQLLIRSGITAEVKVT GGDRGVRGVLEFHLKVPEDVRERLSA
66	Theia	ATGCGGGTCTTGGGAGAATACGACTCTCCAGAATGATGGAGGAGTCTAC CAGTGTGGAACGCCAACGTGAGTTCATCGAGACGTGGGCGCGGCAGAAC GATCACGAAATCGTCGGATGGGCTGAGGACCTGGACGTGTCGGCTCGGT CGATCCGTTTCGACACGCAGGGGCTAGGTCCGTGGCTCAAGGAGCCGAAGC TCAGGGAGTGGGACATCCTCTGCGCCTGGAAGCTGGACCGACTTGCTCGA CGCGCCGTACCGCTGCACAACTGTTGGGATGTGCCAGGACGAGCAAAA GGTCTCTGTGTGCGTCAGCGACAACATCGACCTGTCGACGTGGGTGCGGAC GCCTCGTGGCGTCGGTCATCGCAGGTGTTGCCGAGGGAGAGCTGGAGGC GATCAGGGAGCGGACTCTGTCATCCAGCGGAAGCTTCGGGAGCTTGGCC GGTGGGCGGTGGTAAACCTGCCTACGGCTTCAAGGCACAGGAGCGGGA GGACTCAGCCGGGTACGAGCTGGTGCATGACGAGCACGCGGCCAACGTG ATGCTCGGTGTCATCGAGAAGGTGCTGGCCGGTCACTCGACGGAGTCTGT CGCCAGGGAGCTCAACGAGGCAGGGGAGCTGGCACCGTCCGACTACATC AGGGCCAGGGCCGGTCGCAAGACCAGGGGCACGAAGTGAGCAACGCTC AGATCCGGCAGCTCCTCAAGTCCAAGACTCTGCTCGGGCACGTACCCACA ACGGGGCCACCGTGCGGGACGACGACGCATACCGATCCGGAAGGGTCC TGCGCTCATCTCAGAGGAGAAGTTCGATCAGCTCCAGGCCGCGCTCGATG CGCGGTGTTCAAGGTCACCAACCGATCAGCCAAGGCGTCCCCTCTGCTCG GGGTGGCGATCTGTGGCCTGTGCGGTGTCGCGATGCACATACGTGAGCAT CGTCGCAACGGCAACCTCTATCGGTACTACCGCTGCGATAGCGGCAGCCAC TCCGGAGGCGGCGGTGCCGCCCTGAACACCGTGAACATCATCAAGGC CGACGACCTGGAGGCCCTGGTCGAGGAACACTTCTAGACGAAGTCGGCA GGTTCAACGTGCAGGAGAAGGTGTACGTCCAGCGTCCGATCACCGGGCA GAGCTGGACGAGGCTGTACGGGCCGTAGAGGAAGTACGCGAGCTTCTCG GGACGATGACATCGGCCAGATGAAATCTCGCCTAATGGGCCAACTCACG GCGTAGACGAGCGCATCGCGAGACTGGAGAACCTCCCGAGTGAGGAAG

		CCCGATGGGACTACCGTGCAACCGACCAGACGTACGCCGAAGCGTGGGA GGAAGCCGACACCGAAGGCAGACGGCAGCTCCTCATCAGATCCGGAATCA CCGCTGAGGTCAAGGTGACTGGCGGTGACCGAGGAGTCAGGGGTGTGCT GGAGTTTCACCTGAAGGTCCCCGAGGACGTGAGAGAACGCCTCTCCGCTT AA
67	Trouble	MRALVVIRLSRVTDATTSPERQLESCRQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFDVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTTFAAVVIALMGTVAQMELEAIKERNRSTAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVSDPVQRELERILEVYHRVVDNHEPLHQIAHDNLQR GLPSPKDYFAKLQGREPKGREWSATALKRSILSEAMLGYATLNGKTVRDDDG APLVRAEPILTREQLALRAELVKTDRTKPAVATPSLLLRVLFCAVCGEPAYKFT GGGRKNARYRCRSWGWAQRCNGTVAMAEWDAFCEEQVLDLLGGSERLE KVWVAGSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALAAARQ EELEGLEARPSGWEWRETGQRFQGDWWREQGTAAKNTWLRSMNVRLTFDV RGGLTRTIDFQDLQEQHLRLGVSVERLHAGMS
68	Trouble	ATGAGAGCACTGGTAGTCATCCGACTGTCCCGCGTCACCGATGCTACGACC TCACCGGAGCGCCAGCTGGAGTCTTGCCGACAGCTCTGCGCCCAGCGCGG TTGGGACGTCGTCGGTGTAGCGGAGGATCTGGACGCTCTCCGGAGCAGTCG ATCCGTTTCGACCGGAAGCGCCGCCGAACCTGGCACGGTGGCTAGCATTCT GAGGAGCAACCGTTCGATGTGATCGTGGCGTACCGGGTAGACCGGTTGAC CCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCTGAGGACCACA AGAAGCTGGTCGTCTCCGCGACCGAAGCGCACTTCGATACGACGACGCCG TTCGCGCGGTCGTGATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGCTCGACTGCACATTTCAATATCCGCG CCGGGAAATACCGAGGTTCCCTGCCGCCGTGGGGTTACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGTGCGGACCCGGTGACGCGCAACGCA TCCTCGAGGTCTATCACCGCGTCGTCGACAACCACGAGCCTCTGCACCAGA TCGCCCACGACCTGAACCAGCGCGGCCTACCGTCGCGCAAGGACTACTTCG CGAAGCTGCAGGGCCGAGAGCCCAAGGGTCGGGAGTGGTGGCTACCGC GCTGAAGCGGTGCTGATCTCGGAGGCGATGCTCGGGTACGCGACGCTGA ACGTAAGACCGTCCGAGACGACGACGCGCGCCCGCTGGTGCGGGGCCGA GCCGATCTGACGCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTCA AGACCGACCGGACCAAGCCCGCAGTGGCTACTCCGTCGCTGCTGCTGCGG GTGTTGTTCTGCGCAGTGTGCGGTGAGCCCGCTACAAGTTCACCGGTGG CGGTAGGAAGAACGCTCGCTACCGCTGCCGGTCTGGGGATGGGCGCAG CGGTGCGGCAACGGAACGGTCGCGATGGCGGAGTGGGATGCGTTCTGCG AGGAGCAGGTGCTGGATCTGCTCGGTGGCTCGGAGCGTCTGGAGAAAAGT CTGGGTAGCAGGCTCGGACTCGGCAGTCGAACTCGCAGAGGTGAACGCA GAGCTGGTGGACCTGACGTCGCTGATCGGTTCCCCCGCATACCGGGTCGG TTCTCCGACGCGGAGGCACTGGATGCTCGTATCGCGGCACTGGCCGCAC GGCAGGAGGAGTTGGAAGGGCTGGAGGCTCGCCCGTGGGGTTGGGAGTG GCGCGAGACTGGGCAGAGGTTCCGTGACTGGTGGCGGGAGCAGGGTACC GCGGCAAAGAACACCTGGCTCCGGTCGATGAACGTTCCGGCTGACGTTTGA CGTCCGCGGTGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGT ACGAGCAGCATCTCAGGCTCGGTAGCGTGGTGAACGGCTACACGCAGGG ATGTCGTAG
69	Veracruz	MYPTFVRVLGRLRLSRSTEESTSIRQREIVTAWADSNHGTVVGWAEDVDVS GAIDPFDTPSLGVWLDERRGEWDILCAWKLDRLGRDAIRLNKFLWCQEHGK TVTSCSEIDLGTPVGRLIANVIAFLAEGEREAIRERVASSKQKLREIGRWGGGK PPFGYMGVRNPDGQGHILVVDPAKPVVRRIVEDILEGKPLTRLCTELTEERYL TPAEYYATLKAGAPRQQAEEGEVTAKWRPTAVRNLLRSKALRGHAHHKGQT VRDDQGRAIQLAEPLVDADEWELLQETLDGIAADFSGRRVEGASPLSGVAVC MTCDKPLHHDRLVLRPYGDYPYRYRCRDRHGKNVPAETLEELVEDAFLQR VGDFPVRERVVWQGDNTWADLKEAVAAYDELVQAAGRAKSATARERLQRQ

		LDILDERIAELESAPNTEAHWEYQPTGGTYRDAWENSADDERRELLRRSGIVV AVHIDGVEGRRSKHNPGALHFDIRVPHELTQRLIAP
70	Veracruz	ATGTATCCTACCTTCGTGCGTGTGTTGGGTAGACTGCGTCTGTGAGGTCA ACGGAGGAATCCACCTCGATTGAGAGGCAAAGGGAGATCGTACCCGCTG GGCCGATTCTAACGGTCACACCGTCGTGCGATGGGCAGAGGACGTAGACG TATCGGGTGCCATCGACCCCTTCGACACCCCGTCTCTGGGGGTGTGGTTGG ACGAGCGCCGGGGCGAGTGGGACATCCTGTGCGCCTGGAAGTGGACCG CCTGGGCGGTGATGCCATCCGGCTGAACAAGCTCTTCTGTGGTGCCAGGA GCACGGCAAGACGGTGACATCGTGCAGTGAGGGAATCGACCTCGGCACG CCGGTCGGCCGGCTCATCGCCAACGTGATCGCATTCTGGCCGAGGGGGA GCGGGAGGCCATCCGCGAGCGGGTCGCATCCTCGAAGCAGAAGCTGCGC GAGATCGGTGCGTGGGGTGGCGGTAAGCCGCCCTTCGGATACATGGGTG TCCGCAACCCAGACGGACAGGGACACATCCTTGTGGTTCGATCCCGTCGCA AAGCCAGTCTGCGCCGGATCGTGGAGGACATCCTGGAGGGCAAGCCGC TCACGCGGTCTGCACCGAGCTGACCGAGGAGCGGTACCTGACCCCTGCG GAGTACTACGCCACCCTCAAGGCCGGTGCCCCGAGGCAGCAGGCCGAGG AGGGGGAGGTGACCGCAAGTGGCGTCCGACCGCCGTGAGGAACCTGCT CCGAGCAAAGCCCTCCGGGGCCACGCCCATCACAAGGGCCAGACCGTCA GAGACGACCAGGGGAGGGCTATCCAGCTCGCTGAGCCCTTGTGACGCC GACGAGTGGGAGTTGCTGCAGGAAACCCTGGACGGTATCGCCGCCGACTT CTCCGGTCGGCGCGTCGAGGGTGCCAGCCCGCTCTCCGGTGTGCGAGTCT GCATGACGTGTGACAAGCCCCTGCACCATGACCGGTATCTGGTGAAGAGG CCCTACGGTGACTACCCCTACCGGTACTACCGGTGCCGTGATCGGCACGGC AAGAACGTCCCAGCCGAAACCCTGGAGGAGTTGGTCGAGGACGCATTCT GCAGCGCGTAGGTGACTTCCCGGTGCGCGAGCGGGGTGTGGTCCAGGGT GACACCAACTGGGCAGACCTGAAGGAGGCAGTGGCAGCGTATGACGAAC TGGTGCAGGCCGCTGGCCGTGCCAAGTCGGCCACGGCGCGAGAGCGCCTC CAGAGGCAACTGGACATCCTGGACGAGCGGATCGCGGAGCTGGAGTCCG CACCGAACACCGAGGCCCACTGGGAGTACCAGCCGACCGGAGGGACCTAC CGGGACGCCTGGGAGAACTCCGACGCCGACGAGCGCCGCGAGCTACTGC GGCGGTCCGGGATCGTCGTGGCGGTCCACATCGACGGCGTGGAAGGCCG ACGCTCCAAGCACAACCCCGGAGCCCTCCACTTCGACATCCGGGTCCGCA CGAACTGACACAGAGACTCATCGCCCCATGA
71	Lockley	MRALVIRLSRVTDATTSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFVIVAYRVDRLTRSIRHLQQLVHWAEDHKLI VSATEAHFDTTTFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRGS LPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLHLVAHDLNRR GVLSPKDYFAKLQGRDPQGREWSATALKRLISEAMLGYATLNGKTVRDDD APLVRAEPILTREQLALRAELVKTDRTKPAVSTPSLLRLVFCVCGEPAYKFTG GGRRNARYRCRSWGWAQRCNGTVAMAEDAFCEEQVLDLLGDAERLEK VWVAGSDSAVELAEVNAELVDLTSLIGSPAYRVGSPQREALDARIAALAARQE ELEGLEARPSGWEWRETGQRFGDWWREQDTS GKNTWLRS MNVRLTFDVR GGLTRTIDFGDLQEYEQHLRLGSVVERLHTGMS
72	Lockley	ATGAGAGCTCTCGTCGTGATCCGCTTGTCCCGCGTCACCGATGCTACGACC TCACCGGAGCGTCAGCTGGAGTCTTGCCGGCAGCTCTGCGCCAGCGCGG TTGGGACGTCGTGCGGTGTAGCGGAGGATCTGGACGTCTCCGGAGCAGTCG ATCCGTTCCGACCGGAAGCGCAGACCGAACCTGGCACGGTGGCTAGCATT GAGGAGCAACCGTTCGACGTCATCGTGGCGTACCGGGTGGACCGGTTGAC CCGCTCGATCCGGCATCTTCAGCAGCTGGTCCACTGGGCGGAGGACCACA AGAAGCTGGTCGTCTCCGCGACCGAAGCCCACTTCGACACGACGACGCCG TTCGCGGCGGTGTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATT AGAAGCGATCAAAGAGCGGAACCGTTCGGCTGCACATTTCAATATTCGCG CCGGGAAATACCGCGGTTCCCTGCCGCCGTGGGGTTACCTGCCTACGCGC GTGGACGGGGAGTGGCGGCTGGTGCCGGACCCGGTGACGCGGAGCGCA

		TCCTCGAGGTGTATCACCGCGTCGTCGACAACCACGAGCCGCTGCACCTGG TGGCACACGACCTGAACCGGCGTGGTGTCTGTGCGCCAAGGACTACTTC GCAAAGCTGCAAGGTCGCGAGCCGAGGGGCGGGAGTGGTCGGCTACCG CACTGAAGCGCTCGCTGATCTCCGAGGCGATGCTCGGGTACGCGACTCTG AACGGTAAGACCGTCCGAGACGACGACGAGCCCCGCTGGTGCGGGCTG CGCCGATCCTGACCCGTGAGCAGCTGGAGGCGCTGCGCGCCGAGCTCGTC AAGACCGACCGGACCAAGCCCCGAGTGTCTACCCCGCTCGCTGCTGCG GGTGCTGTCTGCGCAGTGTGCGGTGAGCCCGCATACAAGTTCGCAGGTG GAGGACGCAAGCACCCGCGCTACCGCTGCCGCTCGATGGGTTTCCCGAAG CACTGCGGTAACGGTACGGTCGCAATGGCCGAGTGGGACGCATTCTGCGA GGAGCAGGTGCTGGATCTGCTCGGGGACGCAGAGCGTCTGGAGAAAGTC TGGGTAGCTGGCTCGGACTCAGCAGTAGAACTCGCAGAGGTGAACGCAG AGCTGGTGGACCTGACGTCGCTGATCGGTTCCCCGGCGTACCGGGCCGGT TCTCCGAGCGCGAGGCGCTGGATGCTCGTATCGCGGCGCTGGCCGCGCG GCAGGAGGAGTTGGAAGGGCTAGAGGCTCGCCGTCGGGCTGGGAGTGG CGCGAAACCGGGCAGAGGTTCTGGGACTGGTGGCGGGATCAGGACACCG CGGGTAAGAACACCTGGCTCCGGTCGATGAACGTTCTGGCTGACGTTTCGAC GTCCGAGGCGGCCTGACTCGTACGATCGACTTCGGGGATCTGCAGGAGTA TGAGCAGCATCTGAGGCTGGGCGCGGCTCTAGACCTCGTAAACGCAGAAA AGCCCCCTACGGGCCGCTAG
73	BxB1	MRALVIRLSRVTDAATSPERQLESCQQLCAQRGWDVVGVAEDLDVSGAVD PFDRKRRPNLARWLAFFEEQPFVIVAYRVDRLTRSIRHLQQLVHWAEDHKKL VVSATEAHFDTTTPFAAVVIALMGTVAQMELEAIKERNRSAAHFNIRAGKYRG SLPPWGYLPTRVDGEWRLVPDPVQRERILEVYHRVVDNHEPLHLVAHDNRR GVLSPKDYFAQLQGREPQGREWSATALKRSMISEAMLGYATLNGKTVRDDD GAPLVRAEPILTREQLEALRAELVKTSRAKPAVSTPSLLRLVFCVCGEPAYKF AGGGRKHPRYRCRSMGFPHKCGNGTVAMAEWDAFCEEQVLDLLGDAERLE KVWVAGSDSAVELAEVNAELVDLTSLIGSPAYRAGSPQREALDARIAALAAARQ EELEGLEARPSGWEWRETGQRFQDWWREQDTAAKNTWLRSMNVRLTFDV RGGLTRTIDFQDLQEQYHLRLGSVVERLHTGMS
74	BxB1	ATGAGAGCACTGGTGGTCATCCGACTGAGTAGGGTCACAGACGCAACAAC AAGCCCCGAGAGGCAGCTGGAATCATGTCAGCAGCTGTGCGCACAGCGA GGATGGGACGTGGTCGGAGTGGCAGAGGATCTGGACGTGAGCGGCGCTG TCGATCCATTCGACAGAAAGCGGAGGCCAACCTGGCAAGGTGGCTGGCT TTCGAGGAACAGCCCTTTGATGTGATCGTCGCCTACAGAGTGGACAGGCT GACACGCTCTATTGACATCTGCAGCAGCTGGTGCATTGGGCCGAGGACC ACAAGAACTGGTGGTCAGTGCAACTGAAGCCCACTTCGATACCACAACCTC CTTTGCCGCTGTGGTCATCGCACTGATGGGCACCGTGGCCAGATGGAG CTGGAAGCTATCAAGGAGCGAAACCGGAGTGCAGCCATTTCAATATTCG GGCCGGGAAATACAGAGGATCACTGCCCCCTGGGGCTATCTGCCTACCC GGGTGGATGGGGAGTGGAGACTGGTGCCAGACCCGTCAGAGAGAGAG GATTCTGGAAGTGTACCACAGGGTGGTCGATAACCACGAACCACTGCATCT GGTCGCCACGACCTGAATAGGCGCGGCGTGCTGAGCCCAAAAGATTATT TTGCTCAGCTGCAGGGAAGGGAGCCACAGGGACGAGAATGGTCCGCTAC CGCCCTGAAGCGGAGCATGATCAGTGAGGCTATGCTGGGCTACGCAACTC TGAATGGGAAAACCGTCCGGGACGATGACGGAGCACCCTGGTGAGGGC TGAGCCTATTCTGACACGCGAGCAGCTGGAAGCTCTGCGGGCAGAACTGG TGAAAACCTCCAGAGCCAAACCTGCCGTGAGCACCCCAAGCCTGCTGCTGA GGGTGCTGTTCTGCGCCGTCTGTGGGGAGCCAGCATACAAGTTTGCCGGC GGGGGAAGAAAACATCCCCGCTATCGATGCCGCTCTATGGGATTCCCTAA GCACTGTGGAAACGGCACTGTGGCTATGGCCGAGTGGGACGCCTTTTGTG AGGAACAGGTGCTGGATCTGCTGGGAGACGCCGAGAGGCTGGAAAAAGT GTGGGTGCTGGCAGCGACTCCGCTGTGGAGCTGGCAGAAAGTCAATGCCG AGCTGGTGGATCTGACCTCCCTGATCGGATCTCCTGCATATAGGGCAGGCT

		CACCACAGCGAGAAGCTCTGGACGCACGAATTGCTGCACTGGCAGCTCGA CAGGAGGAAGTGGAGGGGCTGGAAGCACGACCTAGCGGATGGGAGTGG CGAGAAACAGGCCAGCGGTTTGGGGATTGGTGGAGAGAGCAGGACACAG CAGCCAAGAACAAGTGGCTGAGAAGTATGAATGTCAGGCTGACTTTCGAT GTGCGCGGGGGCTGACCCGAACAATCGATTTTGGCGACCTGCAGGAGTA TGAACAGCACCTGAGACTGGGGAGCGTGGTCAAAGACTGCACACTGGG ATGTCATAG
75	Peaches	METMPQPLRALVGARVSVVQGPQKVSQQAQLETARKWAEAQGHEIVGTFE DLGVSASVRPDERPDLGKWLTDGASKWWDVIVWSKMDRAFRSTKHCVDF QWAEERQKVVMFAEDNLRDYLDPGAAGKIDAMMAELFVYLSFFAQLELNR FKSRAQDSHRVLRQTDRWASGLPPLGYKTVPHPSGKGFGLDDEDTKAVLYD MAGKLLDGWSLIGIAKDLNDRGVLGSRSRARLAKGKPIDQAPWNVSTVKDAL TNLKTQGIKMTGKGKHAKPVLDDKGEQIVLAPPTFDWDTWKQIQDAVALRE QAPRSRVHTKNPMLGIGICGKCGATLAQQHSRKS DKS VVYR YRCSRTPVN CDGVFIVADEADTLLEAFLYEWADQPVTRRVFVPGEDHTYELEQINETIARLR RES DAGLIVSDEDERIYLERMRLITRRKLEAMP RRSAGWVEETT GQTYGEA WETEDHQQLLDAKVKFILYSNKPRNIEVVVPQDRVAVDLAI
76	Peaches	ATGGAGACAATGCCTCAGCCACTGAGGGCTCTGGTGGGAGCTAGAGTGA GTGTGGTGCAGGGGCTCAGAAAGTGTACAGCAGGCCAGCTGGAGAC AGCTAGAAAGTGGCCGAGGCTCAGGGCCACGAAATCGTGGGGACTTTT GAAGATCTGGCGTGAGCGCTTCCGTGAGGCCAGATGAGCGCCCCGACCT GGGGAAGTGGCTGACAGATGAAGGAGCCAGCAATGGGACGTGATTGTG TGGTCCAAGATGGATCGCGCTTCCGGTCTACCAAACATTGCGTGACTTT GCCCAGTGGGCTGAGGAAAGACAGAAGGTGGTCATGTTCCGCCGAGGATA ACCTGCGGCTGGACTACAGACCTGGAGCCGCTAAAGGCATCGACGCCATG ATGGCTGAGCTGTTGTGTATCTGGGGAGTTCTTTGCCAGCTGGAAGTGA AATAGGTTCAAGTCCCGCGCTCAGGATTCTCACAGAGTGTGAGGCAGAC TGACAGATGGGCTAGCGGACTGCCCCCTCTGGGCTACAAAACCGTGCCCC ATCCTTCCGGCAAGGGGTTTGGACTGGACACAGATGAGGACACTAAAGCC GTGCTGTATGATATGGCTGGGAAGCTGCTGGACGGATGGTCCCTGATCGG CATTGCCAAAGATCTGAACGACCGCGGGGTGCTGGGATCTCGGAGTAGAG CCAGGCTGGCTAAAGGCAAGCCTATCGATCAGGCCCATGGAACGTGTCT ACCGTGAAGGACGCTCTGACCAATCTGAAAACACAGGGAATTAAGATGAC AGGCAAAGGGAAGCACGCCAAACAGTGCTGGACGATAAGGGAGAGCAG ATCGTGCTGGCTCCACCAACTTTCGATTGGGACACCTGGAACAGATTGAG GATGCTGTGGCTCTGAGGGAACAGGCTCCTCGCAGCCGGGTGCACACTAA AAACCCAATGCTGGGAATCGGCATTGCGGCAAGTGTGGAGCTACCCTGG CTCAGCAGCATTACGCAAAAAGTCAGATAAGAGCGTGGTGTACCGCTAC TATCGGTGCAGCAGAAACCCGTGAATTGTGACGGCGTGTTCATCGTGGC CGATGAGGCTGACACTCTGCTGGAGGAAGCCTTCTGTACGAATGGGCTG ATCAGCCCGTGACCCGAGAGTGTTCTGTCCTGGGGAGGACCACACCTAT GAGCTGGAACAGATCAATGAGACAATTGCCCGCTGAGGCGCGAAAGTG ACGCTGGACTGATCGTGTGATGAGGACGAAAGAATCTACCTGGAGAGA ATGCGGAGCCTGATCACTCGGAGAACCAAGCTGGAAGCTATGCCAAGGAG GAGTGCTGGATGGGTGGAGGAAACACAGGGCAGACATACGAGAGGGCC TGGGAGACTGAAGATCATCAGCAGCTGCTGAAAGACGCTAAAGTGAAGTT TATCCTGTATTCCAACAAGCCAAGGAATATTGAAGTGGTGGTGCCACAGG ATAGGGTGGCTGGACCTGGCTATTTGA
77	Bx22	MAQPLRALVGARVSVVQGPQKVSHIAQQETGAKWVAEQGHTVVGSFKDL VSATVSPFERPDLGPWLSPELEGEWDILVFSKIDRMFRSTRDCVKFAEWAEAH GKILVFAEDNMTLNYRDKDRSGSLESMMSELFYIGSFFAQLELNRFKSRADS HRVLRGMDRWASGVPLGFRIVDHPGKKGKGLDTPDGKAILEDMAAKLLD GWSFIRIAQDLNQRKVLTNMDKAKIAKGKPPHPNPWTVNTVIESLTSPTQGI KMTKHGTRGGSKIGTTVLDAEGNPIRLAPPTFDPATWKQIQEAAARRQGNRR

			SKTYTANPMLGVGHGACGASLAQQFTHRKLADGTEVYRTRYRCGRTPLNCG GISMRGDEADGLLEQLFLEQYGSQPVTEKVFVPGEDHSEEEQVRATIDRLRR ESDAGLIATAEDERIYFERMKSIDRRTRLEAQPRRASGWVTQETDKTNADEW TKASTPDERRRLLMKQGIRFELVRGKPDPEVRLFTPGEIPEGEPLPEPSPR
78	BxZ2		ATGGCACAGCCACTGAGGGCACTGGTCGGAGCACGGGTGACGCTCGTCCA GGGGCCACAGAAAGTCTCCACATCGCACAGCAGGAGACTGGAGCAAAG TGGGTGGCAGAGCAGGGACACACCGTGGTCGGAAGCTTCAAAGACCTGG ACGTGAGCGCCACCGTCTCTCTTTGAACGCCAGATCTGGGACCTGGC TGTCCTGAGCTGGAAGGGGAGTGGGACATCCTGGTGTTTCAGCAAGATT GATAGGATGTTTAGATCCACTCGGGACTGCGTGAAGTTTCGAGAATGGGC AGAGGGCCATGGAAAAATCCTGGTCTTTGCCGAGGACAACATGACCCTGA ATTACCGCGACAAGGATCGAAGCGGATCCCTGGAATCTATGATGAGTGAG CTGTTTCATCTATATTGGGTCTTTCTTTGCTCAGCTGGAGCTGAACAGGTTCA AGAGTCGCGCACGAGATTACACCGCGTGCTGCGAGGCATGGACAGATG GGCATCTGGAGTGCCACCTCTGGGCTTCCGGATCGTCGACCATCCCACTGG GAAGGGCAAAGGACTGGACACAGATCCTGAAGGCAAGGCTATTCTGGAG GACATGGCCGCTAAACTGCTGGATGGATGGAGCTTTATCCGCATTGCACA GGATCTGAACCAGAGGAAGGTGCTGACTAATATGGACAAGGCCAAAATCG CTAAGGGCAAACCAACCCACCCCAACCTTGGACAGTGAATACTGTCATCG AGTCACTGACCAGCCCCCGCACACAGGGAATTAAGATGACTAAACATGGG ACCAGGGGCGGATCCAAGATCGGCACCACAGTGCTGGATGCCGAGGGAA ATCCTATTCGGCTGGCACCTCAACCTTCGACCCAGCCACATGGAAGCAGA TCCAGGAAGCAGCAGCTAGGAGACAGGGAAACCGGCGCAGTAAACTTA CACCGCCAATCCTATGCTGGGAGTGGGACACTGCGGAGCATGTGGAGCCT CACTGGCTCAGCAGTTTACCCATAGAAAGCTGGCTGATGGCACCGAGGTC ACATACAGGACTTATAGATGCGGACGGACACCACTGAAGTGAATGGAAT TTCTATGCGGGGGGACGAAGCTGATGGCCTGCTGGAACAGCTGTTCTGG AGCAGTACGGGTACAGCCAGTGACCGAAAAGGTGTTGTCCCGGCGAG GACCACAGCGAGGAACTGGAACAGGTGAGGGCCACAATCGACAGACTGC GAAGGGAGAGTGATGCTGGGCTGATCGCTACTGCAGAAGATGAGCGGAT CTACTTCGAGCGCATGAAATCCCTGATTGACAGACGGACAAGGTGGAAG CTCAGCCACGCCGAGCATCTGGTGGGTGACCCAGGAGACAGACAAGACT AACGCCGATGAATGGACAAAAGCTAGCACTCCCGATGAGAGGAGACGGC TGCTGATGAAGCAGGGGATCCGATTGAGCTGGTGCGGGGCAAACCAGA CCCCGAAGTCAGACTGTTACCCCCGCGAAATCCCCGAAGGCGAGCCCC GCCCCAGCCCTCCCCAGATGATAA
79	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCACATCGAGCAGCATGCCGACCAG
80	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCAGATCGAGCAGCATGCCGACCAG
81	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCCATCGAGCAGCATGCCGACCAG
82	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCTCATCGAGCAGCATGCCGACCAG
83	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCAG
84	Peaches	attB/attP	GCGGTCTCCATCGGGATCTGCTTATCGAGCAGCATGCCGACCAG
85	BxB1	attB	GGCTTGTCGACGACGGCacCCTCCGTCGTCAGGATCAT
86	BxB1	attB	GGCTTGTCGACGACGGCagCCTCCGTCGTCAGGATCAT
87	BxB1	attB	GGCTTGTCGACGACGGCccCCTCCGTCGTCAGGATCAT
88	BxB1	attB	GGCTTGTCGACGACGGCtcCCTCCGTCGTCAGGATCAT
89	BxB1	attB	GGCTTGTCGACGACGGCtgCCTCCGTCGTCAGGATCAT
90	BxB1	attB	GGCTTGTCGACGACGGCttCCTCCGTCGTCAGGATCAT
91	BxB1	attP	GTGGTTTGTCTGGTCAACCACCGCacGCTCAGTGGTGTACGGTACAAACCC A
92	BxB1	attP	GTGGTTTGTCTGGTCAACCACCGCagGCTCAGTGGTGTACGGTACAAACCC A
93	BxB1	attP	GTGGTTTGTCTGGTCAACCACCGCccGCTCAGTGGTGTACGGTACAAACCC A

94	BxB1	attP	GTGGTTTGTCTGGTCAACCACCGCtGCTCAGTGGTGTACGGTACAAACCC A
95	BxB1	attP	GTGGTTTGTCTGGTCAACCACCGCtgGCTCAGTGGTGTACGGTACAAACCC A
96	BxB1	attB	GTGGTTTGTCTGGTCAACCACCGCttGCTCAGTGGTGTACGGTACAAACCCA
97	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGCacGAAGAACATCACCTGAACATCG
98	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGCagGAAGAACATCACCTGAACATCG
99	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGCCcGAAGAACATCACCTGAACATCG
100	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGCtGAAGAACATCACCTGAACATCG
101	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGctgGAAGAACATCACCTGAACATCG
102	Rebeuca	attB	CGGTATTCGGCGCGATCCGCGGcttGAAGAACATCACCTGAACATCG
103	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGacGAAGAACAT
104	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGagGAAGAACAT
105	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGccGAAGAACAT
106	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGtGAAGAACAT
107	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGtgGAAGAACAT
108	Rebeuca	attP	AGAGCATCGGAGCCTTTGCGGGGATGTGATGTTGAGttGAAGAACAT
109	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGCacGAAGAACATCACCTGAACATCG
110	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGCagGAAGAACATCACCTGAACATCG
111	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGCCcGAAGAACATCACCTGAACATCG
112	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGCtGAAGAACATCACCTGAACATCG
113	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGctgGAAGAACATCACCTGAACATCG
114	Veracruz	attB	CGGTATTCGGCGCGATCCGCGGcttGAAGAACATCACCTGAACATCG
115	Veracruz	attP	GTATTGGGGGAACGCGATATTGAGacGTAGAACATCACCTTCACCAAATT C
116	Veracruz	attP	GTATTGGGGGAACGCGATATTGAGagGTAGAACATCACCTTCACCAAATT C
117	Veracruz	attP	GTATTGGGGGAACGCGATATTGAGccGTAGAACATCACCTTCACCAAATT C
118	Veracruz	attP	GTATTGGGGGAACGCGATATTGAGtGTAGAACATCACCTTCACCAAATT C
119	Veracruz	attP	GTATTGGGGGAACGCGATATTGAGtgGTAGAACATCACCTTCACCAAATT C
120	Veracruz	attB	GTATTGGGGGAACGCGATATTGAGttGTAGAACATCACCTTCACCAAATT C
121	Theia	attB	TGGGTGAACGCAAAGATGGGGacCTCGATGCCGAGCTCGTCGCA
122	Theia	attB	TGGGTGAACGCAAAGATGGGGagCTCGATGCCGAGCTCGTCGCA
123	Theia	attB	TGGGTGAACGCAAAGATGGGGccCTCGATGCCGAGCTCGTCGCA
124	Theia	attB	TGGGTGAACGCAAAGATGGGGtCTCGATGCCGAGCTCGTCGCA
125	Theia	attB	TGGGTGAACGCAAAGATGGGGtgCTCGATGCCGAGCTCGTCGCA
126	Theia	attB	TGGGTGAACGCAAAGATGGGGttCTCGATGCCGAGCTCGTCGCA
127	Theia	attP	TTGTCAAAGTCTAAAGATGGGGacCTCAATATTCATGCTTTGCGAA
128	Theia	attP	TTGTCAAAGTCTAAAGATGGGGagCTCAATATTCATGCTTTGCGAA
129	Theia	attP	TTGTCAAAGTCTAAAGATGGGGccCTCAATATTCATGCTTTGCGAA
130	Theia	attP	TTGTCAAAGTCTAAAGATGGGGtCTCAATATTCATGCTTTGCGAA
131	Theia	attP	TTGTCAAAGTCTAAAGATGGGGtgCTCAATATTCATGCTTTGCGAA
132	Theia	attB	TTGTCAAAGTCTAAAGATGGGGttCTCAATATTCATGCTTTGCGAA
133	Benedict	attB	TGGGTGAACGCAAAGATGGGGacCTCGATGCCGAGCTCGTCGCA
134	Benedict	attB	TGGGTGAACGCAAAGATGGGGagCTCGATGCCGAGCTCGTCGCA
135	Benedict	attB	TGGGTGAACGCAAAGATGGGGccCTCGATGCCGAGCTCGTCGCA
136	Benedict	attB	TGGGTGAACGCAAAGATGGGGtCTCGATGCCGAGCTCGTCGCA
137	Benedict	attB	TGGGTGAACGCAAAGATGGGGtgCTCGATGCCGAGCTCGTCGCA
138	Benedict	attB	TGGGTGAACGCAAAGATGGGGttCTCGATGCCGAGCTCGTCGCA

139	Benedict	attP	TTCGCAAAGCCTCAAAATCGGGacCTCGATATTCATGCTTTGTGAA
140	Benedict	attP	TTCGCAAAGCCTCAAAATCGGGagCTCGATATTCATGCTTTGTGAA
141	Benedict	attP	TTCGCAAAGCCTCAAAATCGGGccCTCGATATTCATGCTTTGTGAA
142	Benedict	attP	TTCGCAAAGCCTCAAAATCGGGtcCTCGATATTCATGCTTTGTGAA
143	Benedict	attP	TTCGCAAAGCCTCAAAATCGGGtgCTCGATATTCATGCTTTGTGAA
144	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCacGGGCTCCCCGGGCGCGTACTCC
145	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCagGGGCTCCCCGGGCGCGTACTCC
146	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCccGGGCTCCCCGGGCGCGTACTCC
147	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCctGGGCTCCCCGGGCGCGTACTCC
148	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCtgGGGCTCCCCGGGCGCGTACTCC
149	PhiC31	attB	GTGCGGGTGCCAGGGCGTGCCttGGGCTCCCCGGGCGCGTACTCC
150	PhiC31	attP	GTGCCCCAACTGGGGTAACCTacGAGTTCTCTCAGTTGGGGG
151	PhiC31	attP	GTGCCCCAACTGGGGTAACCTagGAGTTCTCTCAGTTGGGGG
152	PhiC31	attP	GTGCCCCAACTGGGGTAACCTccGAGTTCTCTCAGTTGGGGG
153	PhiC31	attP	GTGCCCCAACTGGGGTAACCTtcGAGTTCTCTCAGTTGGGGG
154	PhiC31	attP	GTGCCCCAACTGGGGTAACCTtgGAGTTCTCTCAGTTGGGGG
155	PhiC31	attB	GTGCCCCAACTGGGGTAACCTttGAGTTCTCTCAGTTGGGGG
160	PhiC31		MDTYAGAYDRQSRERENSSAASPATQRSANEDKAADLQREVERDGGFRFES AFGTAERPEFERILNECRAGRLNMIIVYDVSRLKVMDAIPIVSELLALGVTV STQEGVFRQGNVMDLIHLIMRLDASHKESLSAKILDKNLQRELGGYVGGK APYGFELVSETKEITRNGRMVNVINKLAHSTPLTGPFEFEPDVIRWWWREI KTHKHLPFKPGSQAAIHPGSITGLCKRMDADAVPTRGETIGKKTASSAWDPAT VMRILRDPRIAGFAAEVIYKKKPDGTPTTKIEGYRIQRDPITLRPVELDCGPIEP AEWYELQAWLDGRGRGKGLSRGQAILSAMDKLYCECGAVMTSKRGEESIKD SYRCRRRKVVDPSPAPGQHEGTCNVSMALDKFVAERIFNKIRHAEGDEETLAL LWEAARRFGKLTEAPEKSGERANLVAERADALNALEELYEDRAAGAYDGPVG RKHFRKQQAALT LRQQGAERLAELEAAEAPKLPLDQWFPEDADADPTGPKS WWGRASVDDKR VFVGLFVDKIVVTKSTTGRGQGTPIEKRASITWAKPPTDDD EDDAQDGTEDVAA
161	PhiC31		GTGGACACGTACGCGGGTGCTTACGACCGTCAGTCGCGCGAGCGCGAGA ATTCGAGCGCAGCAAGCCCAGCGACACAGCGTAGCGCCAACGAAGACAA GGCGGCCGACCTTCAGCGCGAAGTCGAGCGCGACGGGGGCCGGTTCAGG TTCGTCGGGCATTTTCAGCGAAGCGCCGGGCACGTGCGCGTTTCGGGACGGC GGAGCGCCCGGAGTTCGAACGCATCCTGAACGAATGCCGCGCCGGGGCGG CTCAACATGATCATTGTCTATGACGTGTGCGCTTCTCGCGCCTGAAGGTC ATGGACGCGATTCCGATTGTCTCGGAATTGCTCGCCCTGGGCGTGACGATT GTTTCCACTCAGGAAGGCGTCTTCCGGCAGGGAAACGT CATGGACCTGATT CACCTGATTATGCGGCTCGACGCGTCGCACAAAGAATCTTCGCTGAAGTCG GCGAAGATTCTCGACACGAAGAACCCTTCAGCGCGAATTGGGCGGGTACGT CGGCGGGAAGGCGCCTTACGGCTTCGAGCTTGTTTCGGAGACGAAGGAG ATCACGCGCAACGGCCGAATGGTCAATGTCGTCATCAACAAGCTTGCGCAC TCGACCACTCCCCTTACCGGACCCCTTCGAGTTCGAGCCCGACGTAATCCGG TGGTGGTGGCGTGAGATCAAGACGCACAAACACCTTCCCTTCAAGCCGGG CAGTCAAGCCGCCATTACCCGGGCAGCATCACGGGGCTTTGTAAGCGCA TGGACGCTGACGCCGTGCCGACCCGGGGCGAGACGATTGGGAAGAAGAC CGCTTCAAGCGCCTGGGACCCGGCAACCGTTATGCGAATCCTTCGGGACCC GCGTATTGCGGGCTTCGCCGCTGAGGTGATCTACAAGAAGAAGCCGGACG GCACGCCGACCAAGATTGAGGGTTACCGCATTACGCGCGACCCGATC ACGCTCCGGCCGGTCGAGCTTGATTGCGGACCGATCATCGAGCCCGCTGA GTGGTATGAGCTTCAGGCGTGGTTGGACGGCAGGGGGCGCGGCAAGGGG CTTTCCCGGGGGCAAGCCATTCTGTCCGCCATGGACAAGCTGTAAGTGCAG TGTGGCGCCGT CATGACTTCGAAGCGCGGGGAAGAATCGATCAAGGACTC TTACCGCTGCCGTCGCCGGAAGGTGGTCGACCCGTCCGCACCTGGGCAGC ACGAAGGCACGTGCAACGTCAGCATGGCGGCACTCGACAAGTTCGTTGCG

		GAACGCATCTTCAACAAGATCAGGCACGCCGAAGGCGACGAAGAGACGTT GGCGCTTCTGTGGGAAGCCGCCGACGCTTCGGCAAGCTCACTGAGGCGC CTGAGAAGAGCGGCGAACGGGCGAACCTTGTTGCGGAGCGCGCCGACGC CCTGAACGCCCTTGAAGAGCTGTACGAAGACCGCGCGGCAGGCGCGTACG ACGGACCCGTTGGCAGGAAGCACTTCCGGAAGCAACAGGCAGCGCTGAC GCTCCGGCAGCAAGGGGCGGAAGAGCGGCTTGCCGAACCTGAAGCCGCC GAAGCCCCGAAGCTTCCCCTTGACCAATGGTTCCCCGAAGACGCCGACGCT GACCCGACCGGCCCTAAGTCGTGGTGGGGGCGCGCGTCAGTAGACGACA AGCGCGTGTTCGTGCGGGCTCTTCGTAGACAAGATCGTTGTACGAAGTCGA CTACGGGCGAGGGGCGAGGGAACGCCCATCGAGAAGCGCGCTTCGATCAC GTGGGCGAAGCCGCCGACCGACGACGAAGACGACGCCAGGACGGC ACGGAAGACGTAGCGGCGTAG
--	--	---

REFERENCES

1. Malchin, N., Goltsman, J., Dabool, L., Gorovits, R., Bao, Q., Dröge, P., Yagil, E. and Kolot, M., Optimization of coliphage HK022 Integrase activity in human cells. (2009) Gene, 437, 9–13.
2. Krishnakumar, R., Grose, C., Haft, D. H., Zaveri, J., Alperovich, N., Gibson, D. G., Merryman, C., and Glass, J. I., Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. (2014) Nucl. Acids Res., 42, 1-8.
3. Park JT, Leach SD. TAILOR: transgene activation and inactivation using lox and rox in zebrafish. (2013) PLoS One. 8(12):e85218.
4. Long D, Zhao A, Xu L, Lu W, Guo Q, Zhang Y, Xiang Z. In vivo site-specific integration of transgene in silkworm via PhiC31 integrase-mediated cassette exchange. (2013) Insect Biochem Mol Biol. 43(11):997-1008.
5. Turan S, Qiao J, Madden S, Benham C, Kotz M, Schambach A, Bode J. Expanding Flp-RMCE options: the potential of Recombinase Mediated Twin-Site Targeting (RMTT). Gene. (2014) 546(2):135-44.
6. Osterwalder, M., Galli, A., Rosen, B., Skarnes, W.C., Zeller, R. and Lopez-Rios, J. Dual RMCE for efficient re-engineering of mouse mutant alleles. (2010) Nat Meth, 7, 893 – 895.

7. Jaina, P., Hsua, T., Araia, M., Biermanna, K., Thaler, D.S., Nguyena, A., González, P.A., Tufariello, J.M., Kriakova, J., Chena, B., Larsena, M.H., and Jacobs Jr., W.R., Specialized transduction designed for precise high-throughput unmarked deletions in *Mycobacterium tuberculosis*. (2014) *mBio*. 5, e01245–14.
8. Wang, Y., Krushel, L.A. and Edelman, G.M. Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. (1996) *Proc. Natl. Acad. Sci. U.S.A.*, 93, 3932– 3936.
9. Grindley ND, Whiteson KL, Rice PA. Mechanisms of site-specific recombination. *Annu Rev Biochem.* (2006)75:567-605.
10. G Pan, K.L.P.D.S. Mechanism of cleavage and ligation by FLP recombinase: classification of mutations in FLP protein by in vitro complementation analysis. *Molecular and Cellular Biology*, 13, 3167.
11. Esposito, D. and Scoocca, J.J. (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.*, 25, 3605–3614.
12. Mumm, J.P., Landy, A., and Gelles, J. Viewing single lambda site-specific recombination events from start to finish. (2006). 25, 4586 – 4595.
13. Anastassiadis K, Fu J, Patsch C, Hu S, Weidlich S, Duerschke K, Buchholz F, Edenhofer F, Stewart AF. Dre recombinase, like Cre, is a highly efficient site-specific recombinase in *E. coli*, mammalian cells and mice. *Dis Model Mech.* (2009) 2 (9-10):508-15.
14. Zuwen Zhang, B.L. Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. (2002) *Nucleic Acids Res.*, 30, e90.

15. Gaj,T., Gaj,T., Mercer,A.C., Mercer,A.C., Gersbach,C.A., Gersbach,C.A.,
Gordley,R.M., Gordley,R.M., Barbas,C.F. and Barbas,C.F. Structure-guided
reprogramming of serine recombinase DNA sequence specificity. (2011) Proc. Natl.
Acad. Sci. U.S.A., 108, 498 – 503.
- 5
16. Hubbard EJ. FLP/FRT and Cre/lox recombination technology in *C. elegans*. (2014)
Methods. 68(3):417-24.
17. Turan S, Zehe C, Kuehle J, Qiao J, Bode J. Recombinase-mediated cassette exchange
10 (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. Gene.
2013 Feb 15;515(1):1-27.
18. Smith,M.C.M. and Thorpe,H.M. Diversity in the serine recombinases. (2002) Mol
Microbiol, 44, 299–307.
- 15
19. Xu,Z., Thomas,L., Davies,B., Chalmers,R., Smith,M. and Brown,W. Accuracy and
efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases
for the integration of DNA into the human genome. (2013) BMC Biotechnol, 13, 87.
- 20
20. Thyagarajan,B., Ausländer,S., Thyagarajan,B., Olivares,E.C., Ausländer,D.,
Olivares,E.C., Hollis,R.P., Müller,M., Hollis,R.P., Ginsburg,D.S., et al. Site-specific
genomic integration in mammalian cells mediated by phage phiC31 integrase. (2001)
Molecular and Cellular Biology, 21, 3926–3934.
- 25
21. Kim,A.I., Ghosh,P., Aaron,M.A., Bibb,L.A., Jain,S. and Hatfull,G.F.
Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene.
(2003) Mol Microbiol, 50, 463– 473.
22. Singh,S., Rockenbach,K., Dedrick,R.M., VanDemark,A.P. and Hatfull,G.F. Cross -
30 talk between Diverse Serine Integrases. (2014) Journal of Molecular Biology, 426,
318 – 331.

23. Hatfull, G.F., the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science Program, the KwaZulu-Natal Research Institute for Tuberculosis and HIV Mycobacterial Genetics Course Students the Phage Hunters Integrating Research and Education Program Complete Genome Sequences of 138 Mycobacteriophages. (2012) *Journal of Virology*, 86, 2382–2384.
24. Park, A., Jeong, H., Lee, J., Lee, C. The inhibitory effect of phloretin on the formation of *Escherichia coli* O157:H7 biofilm in a microfluidic system. (2012) *BioChip J.* 6, 299–305.
25. Barreca D, Bellocco E, Laganà G, Ginestra G, Bisignano C. Biochemical and antimicrobial activity of phloretin and its glycosylated derivatives present in apple and kumquat. (2014) *Food Chem.* 160:292-7.
26. Terán, W., Felipe, A., Segura, A., Rojas, A., Ramos, J., and Gallegos, M. Antibiotic-Dependent Induction of *Pseudomonas putida* DOT-11E TtgABC Efflux Pump Is Mediated by the Drug Binding Repressor TtgR (2003) *Antimicrob Agents Chemother.* 47, 3067 – 3072.
27. Rutherford K, Yuan P, Perry K, Sharp R, Van Duyne GD. Attachment site recognition and regulation of directionality by the serine integrases. (2013) *Nucleic Acids Res.* 41(17):8341-56.
28. Singh, S., Ghosh, P., Hatfull, G.F. Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. (2013) *PLOS Genetics.* 9, e1003490.
29. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. (2002) *Nature* 420, 224-230.

30. Shimada T, Yamazaki Y, Tanaka K, Ishihama A. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of *Escherichia coli*. (2014) PLoS One. 9(3):e90447.
- 5 31. Johansen, T.E., Scholler, M.S., Tolstoy, S., Schwartz, T. Biosynthesis of peptide precursors and protease inhibitors using new constitutive and inducible eukaryotic expression vectors. (1990) FEBS. 267(2), 289-94.
- 10 32. Hayashi, K., Nakazawa, M., Ishizaki, Y., Hiraoka, N., Regulation of inter- and intramolecular ligation with T4 DNA ligase in the presence of polyethylene glycol. (1986) Nucl. Acids Res. 14(19):7617-31.
- 15 33. Gardner,T.S., Cantor,C.R. and Collins,J.J. Construction of a genetic toggle switch in *Escherichia coli*. (2000) Nature, 403, 339-342.
34. Danino,T., Mondragón-Palomino,O., Tsimring,L. and Hasty,J. A synchronized quorum of genetic clocks. (2010) Nature, 463, 326-330.
- 20 35. Moon,T.S., Lou,C., Tamsir,A., Stanton,B.C. and Voigt,C.A. Genetic programs constructed from layered logic gates in single cells. (2012) Nature, 491, 249-253.
36. Purnick,P.E.M. and Weiss,R. The second wave of synthetic biology: from modules to systems. (2009) Nature Reviews Molecular Cell Biology, 10, 410-422.
- 25 37. Wang,B., Kitney,R.I., Joly,N. and Buck,M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. (2011) Nat Commun, 2, 508.
- 30 38. Friedland, A.E., Lu, T.K., Wang, X., Shi, D., Church, G., Collins, J.J., Synthetic Gene Networks That Count (2009) Synthetic Gene Networks That Count. 324, 1199- 1202.

39. Elowitz,M.B. and Leibler,S. A synthetic oscillatory network of transcriptional regulators. (2000) *Nature*, 403, 335-338.
40. Siuti,P., Yazbek,J. and Lu,T.K. Synthetic circuits integrating logic and memory in living cells. (2013) *Nature Biotechnology*, 31, 448-452.
41. Bleris,L., Xie,Z., Glass,D., Adadey,A., Sontag,E. and Benenson,Y. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. (2011) *Mol Syst Biol*, 7, 519.
42. Mukherji,S., Ebert,M.S., Zheng,G.X.Y., Tsang,J.S., Sharp,P.A. and van Oudenaarden,A. MicroRNAs can generate thresholds in target gene expression. (2011) *Nat Genet*, 43, 854–859.
43. Tigges,M., Marquez-Lago,T.T., Stelling,J. and Fussenegger,M. A tunable synthetic mammalian oscillator. (2009) *Nature*, 457, 309 – 312.
44. Siciliano,V., Menolascina,F., Marucci,L., Fracassi,C., Garzilli,I., Moretti,M.N. and di Bernardo,D. Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell- - line. (2011) *PLoS Comp Biol*, 7, e1002074.
45. Lanitis,E., Poussin,M., Klattenhoff,A.W., Song,D., Sandaltzopoulos,R., June,C.H. and Powell,D.J. Chimeric antigen receptor T Cells with dissociated signaling domains exhibit focused antitumor activity with reduced potential for toxicity in vivo. (2013) *Cancer Immunol Res*, 1, 43–53.
46. Grushkin,D. The new drug circuit. (2012) *Nat Med*, 18, 1452–1454.
47. Ye,H., Baba,M.D.E., Peng,R.W. and Fussenegger,M. A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. (2011) *Science*, 332, 1565 – 1568.

48. Ishihara,K., Oshimura,M. and Nakao,M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. (2006) Mol. Cell, 23, 733–742.

All publications, patents and sequence database entries mentioned herein, including those items listed below, are hereby incorporated by reference for the teachings referenced herein as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

EQUIVALENTS AND SCOPE

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above description, but rather is as set forth in the appended claims.

Articles such as “a,” “an,” and “the,” as used herein, may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, *etc.*, from one or more of the claims or from relevant portions of the description is introduced into another claim or another portion of the description. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim.

Furthermore, where the claims recite a composition, it is to be understood that methods of using the composition for any of the purposes disclosed herein are included, and methods of making the composition according to any of the methods of making disclosed herein or other

methods known in the art are included, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise.

Where elements are presented as lists, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It is also noted
5 that the term “comprising” is intended to be open and permits the inclusion of additional elements or steps. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, steps, *etc.*, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, steps, *etc.* For purposes of simplicity those embodiments have not
10 been specifically set forth *in haec verba* herein. Thus for each embodiment of the invention that comprises one or more elements, features, steps, *etc.*, the invention also provides embodiments that consist or consist essentially of those elements, features, steps, *etc.*

Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the
15 understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed
20 as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

In addition, it is to be understood that any particular embodiment of the present invention may be explicitly excluded from any one or more of the claims. Where ranges are
25 given, any value within the range may explicitly be excluded from any one or more of the claims. Any embodiment, element, feature, application, or aspect of the compositions and/or methods of the invention, can be excluded from any one or more claims. For purposes of brevity, all of the embodiments in which one or more elements, features, purposes, or aspects is excluded are not set forth explicitly herein.

CLAIMS

What is claimed is:

1. A library of orthogonal large serine recombinase recognition site sequences,
5 comprising pairs of attB site sequences and attP site sequences recognized by large serine recombinases,

wherein the attB site sequences comprise a central dinucleotide sequence, at least 5 nucleotides upstream of the central dinucleotide sequence and at least 5 nucleotides downstream of the central dinucleotide sequence,

10 wherein the attP site sequences comprise a central dinucleotide sequence, at least 5 nucleotides upstream of the central dinucleotide sequence, and at least 5 nucleotides downstream of the central dinucleotide sequence.

wherein each pair of attB site sequences and attP site sequences comprises one attB site sequence for use in a host genome and one attP site sequence for use in an integrative
15 vector,

wherein in each pair of attP site sequences and attB site sequences comprise the same central dinucleotide, and

wherein crosstalk between the attB site sequences and the attP site sequences in the library is less than about 20%.

20 2. The library claim 1, wherein the central dinucleotide is nonpalindromic.

3. The library of claim 1 or claim 2, wherein the central dinucleotide is selected from the group consisting of: AA, TT, GG, CC, AG, GA, AC, CA, TG, GT, TC, and CT.

25 4. The library of any one of claims 1-3, wherein the sequences upstream and downstream of the central dinucleotide of the attB site comprise additional recognition sequences for the large serine recombinase.

30 5. The library of any one of claims 1-4, the sequences upstream of the central dinucleotide of the attB sites are 5-50 nucleotides in length.

6. The library of any one of claims 1-5, wherein the sequences downstream of the central dinucleotide of the attB sites are 5-50 nucleotides in length.

7. The library of any one of claims 1-6, wherein the sequences upstream and downstream of the central dinucleotide of the attP site comprise additional recognition sequences for the recombinase.

8. The library of any one of claims 1-7, wherein the sequences upstream of the central dinucleotide of the attP sites are 5-50 nucleotides in length.

9. The library of any one of claims 1-8, wherein the sequences downstream of the central dinucleotide of the attP sites are 5-50 nucleotides in length.

10. The library of any one of claims 1-9, wherein the large serine recombinase is selected from the group consisting of: BxB1, Peaches, Veracruz, Rebeuca, Theia, Benedict, and PhiC31.

11. The library of any one of claims 1-10, wherein the pairs of attB site sequences and attP site sequences comprise a pair of sequences selected from the group consisting of:

SEQ ID NO: 79 and SEQ ID NO: 79;

SEQ ID NO: 80 and SEQ ID NO: 80;

SEQ ID NO: 81 and SEQ ID NO: 81;

SEQ ID NO: 82 and SEQ ID NO: 82;

SEQ ID NO: 83 and SEQ ID NO: 83;

SEQ ID NO: 84 and SEQ ID NO: 84;

SEQ ID NO: 85 and SEQ ID NO: 91;

SEQ ID NO: 86 and SEQ ID NO: 92;

SEQ ID NO: 87 and SEQ ID NO: 93;

SEQ ID NO: 88 and SEQ ID NO: 94;

SEQ ID NO: 89 and SEQ ID NO: 95;

SEQ ID NO: 90 and SEQ ID NO: 96;

SEQ ID NO: 97 and SEQ ID NO: 103;

SEQ ID NO: 98 and SEQ ID NO: 104;
SEQ ID NO: 99 and SEQ ID NO: 105;
SEQ ID NO: 100 and SEQ ID NO: 106;
SEQ ID NO: 101 and SEQ ID NO: 107;
5 SEQ ID NO: 102 and SEQ ID NO: 108;
SEQ ID NO: 109 and SEQ ID NO: 115;
SEQ ID NO: 110 and SEQ ID NO: 116;
SEQ ID NO: 111 and SEQ ID NO: 117;
SEQ ID NO: 112 and SEQ ID NO: 118;
10 SEQ ID NO: 113 and SEQ ID NO: 119;
SEQ ID NO: 114 and SEQ ID NO: 120;
SEQ ID NO: 121 and SEQ ID NO: 127;
SEQ ID NO: 122 and SEQ ID NO: 128;
SEQ ID NO: 123 and SEQ ID NO: 129;
15 SEQ ID NO: 124 and SEQ ID NO: 130;
SEQ ID NO: 125 and SEQ ID NO: 131;
SEQ ID NO: 126 and SEQ ID NO: 132;
SEQ ID NO: 133 and SEQ ID NO: 139;
SEQ ID NO: 134 and SEQ ID NO: 140;
20 SEQ ID NO: 135 and SEQ ID NO: 141;
SEQ ID NO: 136 and SEQ ID NO: 142;
SEQ ID NO: 137 and SEQ ID NO: 143;
SEQ ID NO: 144 and SEQ ID NO: 150;
SEQ ID NO: 145 and SEQ ID NO: 151;
25 SEQ ID NO: 146 and SEQ ID NO: 152;
SEQ ID NO: 147 and SEQ ID NO: 153;
SEQ ID NO: 148 and SEQ ID NO: 154;
SEQ ID NO: 149 and SEQ ID NO: 155.

30 12. A method for site-specific recombination between specific attB and attP sites of a large serine recombinase, comprising

providing a nucleic acid molecule comprising an attB site from a pair of the attB and attP sites of claim 11,

providing a nucleic acid molecule comprising an attP site from a pair of the attB and attP sites of claim 11,

5 combining the nucleic acid molecules in the presence of the large serine recombinase.

13. The method of claim 12, wherein the nucleic acid molecule comprising an attP site is a bacteriophage genome or a plasmid.

10 14. The method of claim 12 or claim 13, wherein the nucleic acid molecule comprising an attB site is a bacterial genome or a mammalian genome.

15. A cell line comprising a plurality of landing pads integrated into the genomic DNA of a parental cell line, wherein each landing pad comprises

15 a constitutive promoter operably linked to a nucleotide sequence encoding a detectable marker, followed by a nucleotide sequence encoding a first selectable marker, and an attB site of a large serine recombinase, wherein the attB site is between the promoter and the nucleotide sequence encoding the detectable protein, and wherein the attB sites in the plurality of landing pads are orthogonal to each other,
20 wherein crosstalk between the attB sites is less than about 20%.

16. The cell line of claim 15, wherein the number of landing pads is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12.

25 17. The cell line of claim 15 or claim 16, wherein each landing pad is integrated into the genomic DNA of the cell line using a zinc-finger nuclease, TALEN or CRISPR-Cas system.

18. The cell line of any one of claims 15-17, wherein the attB site is selected from the group consisting of: SEQ ID NOs: 79-84, 85-90, 97-102, 109-114, 121-126, 133-138, and
30 144-149.

19. The cell line of any one of claims 15-18, wherein the detectable marker is a fluorescent protein.

20. The cell line of any one of claims 15-19, wherein the fluorescent protein is selected from the group consisting of EYFP, eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew.

21. The cell line of any one of claims 15-20, wherein the first selectable marker hydrolyzes a first drug.

22. The cell line of claim 21, wherein the first drug is selected from a group consisting of: puromycin, hygromycin, G418, neomycin, and bleomycin.

23. The cell line of any one of claims 15-22, wherein the large serine recombinase is selected from the group consisting of: BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), and Severus (SEQ ID NO: 57).

24. A method of integrating a plurality of genetic sequences or circuits into a cell, comprising:

delivering into the cell line any one of claims 15-23 a plurality of integrative vectors, each comprising an attP site, one or more genetic sequences or gene cassettes that express components of the genetic circuit, and a promoterless second selectable marker,

wherein the cell expresses the large serine recombinase, and

wherein each attP site is selected to allow site-specific recombination between the

attB site in one of the landing pads and the attP site in the integrative vector,

which results in the insertion of the coding sequence of the second selectable marker in frame with the constitutive promoter in the landing pad;

screening for cells with the integration of the genetic sequence or circuit in media comprising a drug to which the second selectable marker confers resistance.

25. The method of claim 24, wherein the one or more gene cassettes that express components of the genetic circuit comprise a promoter operably linked to a sequence encoding a polypeptide that is part of the genetic circuit.

26. The method of claim 24 or claim 25, wherein the integrative vector is introduced into the cell by a method selected from the group consisting of: CaPO₄ transfection, lipid transfection, electroporation, and lentiviral vector infection.

27. The method of any one of claims 24-26, wherein the drug in the media is selected from the group consisting of puromycin, hygromycin, G418, neomycin, and bleomycin.

28. The method of any one of claims 24-27, wherein the large serine recombinase is selected from the group consisting of: BxB1 (SEQ ID NO: 73), BxZ2 (SEQ ID NO: 77), PhiC31 (SEQ ID NO: 160), Peaches (SEQ ID NO: 75), Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), and Severus (SEQ ID NO: 57).

29. A method for identifying a genomic insertion site of an integrated plasmid, comprising

obtaining genomic DNA from bacteria that include an integrated plasmid, wherein the plasmid comprises a promoter operably linked to a nucleotide sequence encoding a large serine recombinase, an attP site recognized by the large serine recombinase, a bacterial origin of replication, one recognition site for a DNA restriction enzyme, and a promoter operably linked to a nucleotide sequence encoding a selectable marker,

digesting the genomic DNA with the DNA restriction enzyme to obtain fragments of the genomic DNA,

ligating the fragments of the genomic DNA to recircularize plasmid sequences,
wherein the recircularized plasmid comprises the selectable marker, the bacterial origin of
replication, and genomic DNA flanking the genomic insertion site,

transforming transformation-competent bacteria with the ligated fragments,
5 culturing the transformed bacteria on media that selects for growth of bacteria
transformed with the recircularized plasmid that comprises the selectable marker,
isolating the recircularized plasmid, and
determining the sequence of the recircularized plasmid to identify the genomic
insertion site.

10

30. The method of claim 29, wherein the DNA restriction enzyme is *EcoRI*.

31. The method of claim 29 or claim 30, wherein the step of ligating the fragments of the
genomic DNA is performed under conditions that favor intramolecular ligation.

15

32. The method of any one of claims 29-31, wherein the transformation-competent
bacteria are transformation-competent *E. coli*.

20

33. The method of any one of claims 29-32, wherein the genomic insertion site is a
chromosomal attB site.

34. The method of any one of claims 29-33, wherein the recognition site for the DNA
restriction enzyme is located between the nucleotide sequence encoding a large serine
recombinase and the attP site recognized by the large serine recombinase.

25

35. The method of any one of claims 29-34, further comprising identifying the attB and
attP sites at the genomic insertion site.

30

36. A reporter plasmid for monitoring recombination events between large serine
recombinase recognition sites comprising, in order, a predicted attB site sequence for a large
serine recombinase, a gene cassette expressing a detectable marker, a predicted attP site

sequence for the large serine recombinase, a gene cassette expressing the large serine recombinase, and a gene cassette expressing a selectable marker,

wherein each gene cassette comprises a promoter operably linked to a sequence encoding a polypeptide, followed by a terminator.

5

37. The reporter plasmid of claim 36, wherein the detectable marker is a lacZ nucleotide sequence.

10

38. A method for testing for function of predicted large serine recombinase recognition sites, comprising

obtaining bacterial cells transformed with the reporter plasmid of claim B1, culturing the bacterial cells on media that permits detection of expression of the detectable marker and to express the large serine recombinase,

wherein if the predicted recombinase recognition sites are correct, then recombination between attP and attB sites results in excision of the gene cassette expressing the detectable marker, and

wherein if the predicted recombinase recognition sites are not correct, then the gene cassette expressing the detectable marker is not excised.

20

39. The method of claim 38, wherein detectable marker is a lacZ nucleotide sequence and the media that permits detection comprises 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (Xgal).

25

40. The method of claim 38 or claim 39, wherein the bacteria are *E. coli*.

41. A reporter system for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells comprising

a first plasmid comprising a gene cassette that comprises a constitutive promoter operably linked to a nucleotide sequence encoding a first detectable marker, and an attB site of the large serine recombinase placed in between the constitutive promoter and the first detectable marker coding sequence,

a second plasmid comprising a promoter-less gene cassette that comprises an attP site of the large serine recombinase followed by a nucleotide sequence encoding a second detectable marker, wherein the second detectable marker is not the same as the first detectable marker.

5

42. The reporter system of claim 41, further comprising a plasmid expressing the recombinase.

43. The reporter system of claim 41 or claim 42, wherein the first detectable marker is a first detectable protein.

10

44. The reporter system of any one of claims 41-43, wherein the first detectable protein is a first fluorescent protein.

45. The reporter system of any one of claims 41-44, wherein the fluorescent protein is selected from the group consisting of eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew.

15

46. The reporter system of any one of claims 41-45, wherein the constitutive promoter is selected from the group consisting of a Hef1a promoter, a CMV promoter, and a SV40 promoter.

20

47. The reporter system of any one of claims 41-46, wherein the second detectable marker is a second detectable protein.

25

48. The reporter system of any one of claims 41-47, wherein the second detectable protein is a second fluorescent protein.

49. The reporter system of any one of claims 41-48, wherein the second fluorescent protein is selected from the group consisting of , eGFP, eYFP, eCFP, mKate2, mCherry, mPlum, mGrape2, mRaspberry, mGrape1, mStrawberry, mTangerine, mBanana, and mHoneydew.

30

50. A method for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells comprising

co-transfecting the first plasmid and the second plasmid of any one of claims 41-49 together with a plasmid expressing the corresponding large serine recombinase,

wherein intermolecular site-specific recombination between the attB and the attP sites results in fusion of the first plasmid and the second plasmid which results in the insertion of the coding sequence of the second detectable marker in frame with the constitutive promoter.

51. A method for site-specific recombination between specific attB and attP sites of a large serine recombinase, comprising

providing a first nucleic acid molecule comprising an attB site,

providing a second nucleic acid molecule comprising an attP site that selectively recombines with the attB site,

combining the nucleic acid molecules in the presence of a large serine recombinase selected from the group consisting of: Veracruz (SEQ ID NO: 69), Rebeuca (SEQ ID NO: 49), Theia (SEQ ID NO: 65), Benedict (SEQ ID NO: 27), PattyP (SEQ ID NO: 47), Trouble (SEQ ID NO: 67), KSSJEB (SEQ ID NO: 39), Lockley (SEQ ID NO: 71), Scowl (SEQ ID NO: 55), Switzer (SEQ ID NO: 63), Bob3 (SEQ ID NO: 29), Abrogate (SEQ ID NO: 21), Doom (SEQ ID NO: 41), ConceptII (SEQ ID NO: 33), Anglerfish (SEQ ID NO: 25), SkiPole (SEQ ID NO: 61), Museum (SEQ ID NO: 45), Severus (SEQ ID NO: 57), Sarfire (SEQ ID NO: 53), Rey (SEQ ID NO: 51), Bongo (SEQ ID NO: 31), Airmid (SEQ ID NO: 23), Hinder (SEQ ID NO: 35), ICleared (SEQ ID NO: 37), Sheen (SEQ ID NO: 59), or Mundrea (SEQ ID NO: 43).

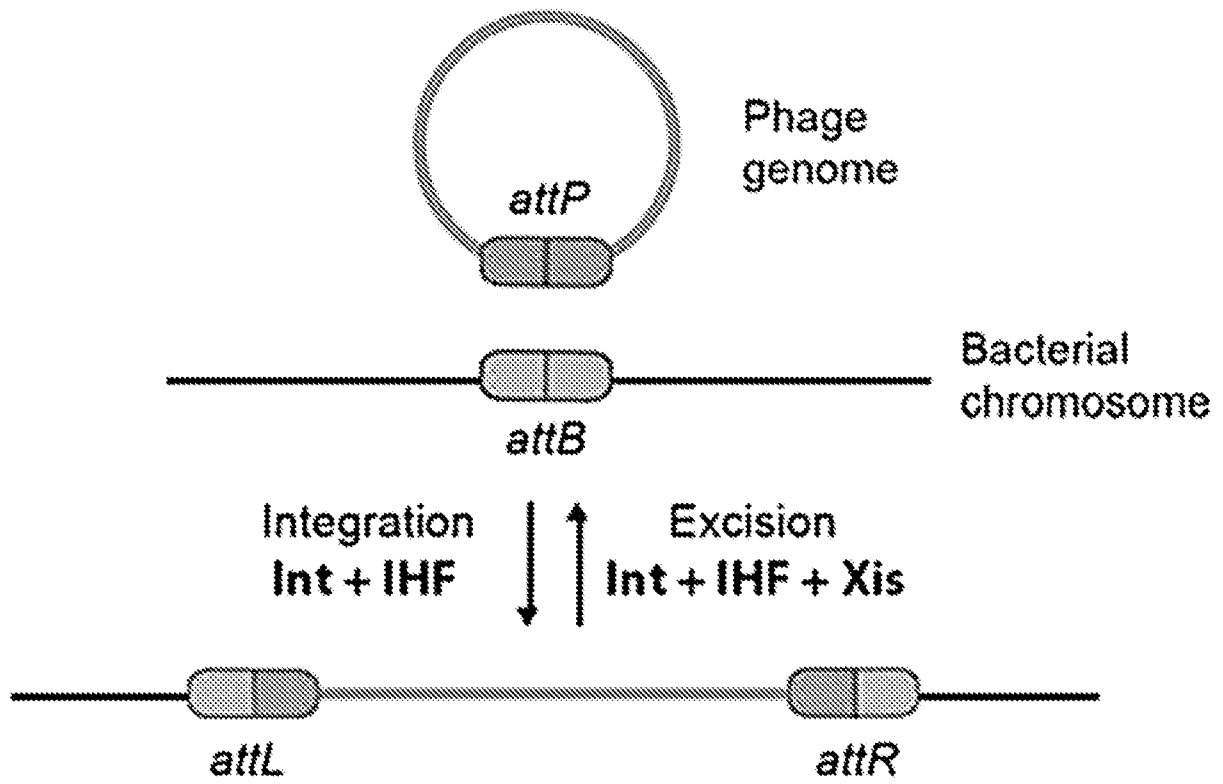


FIG. 1

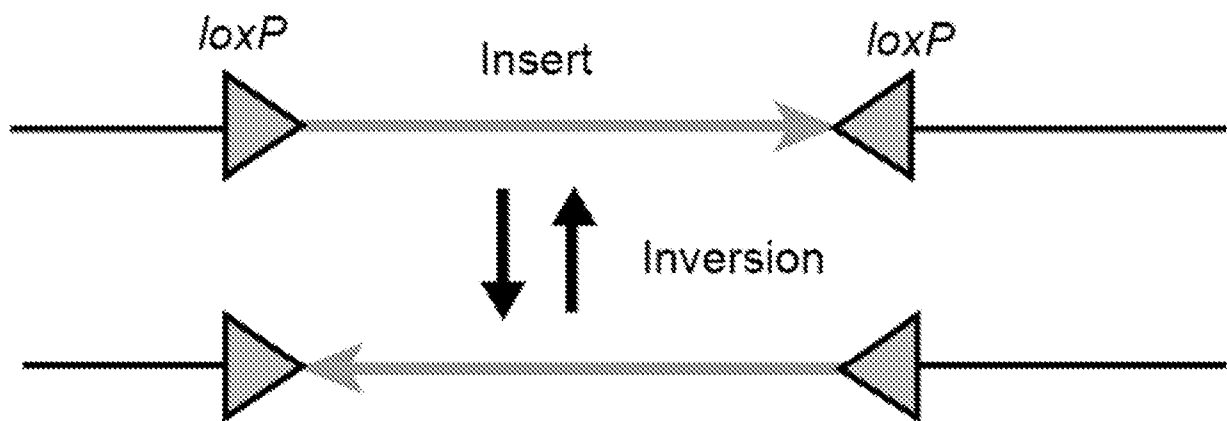
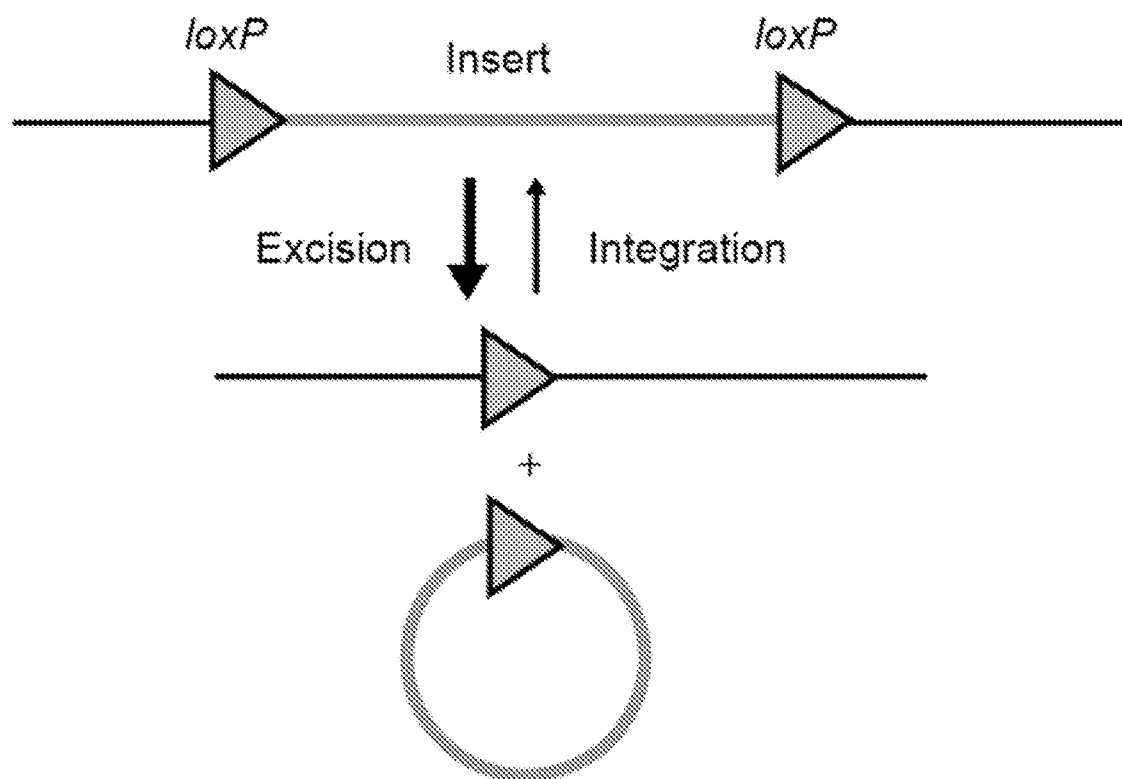


FIG. 2A

**FIG. 2B**

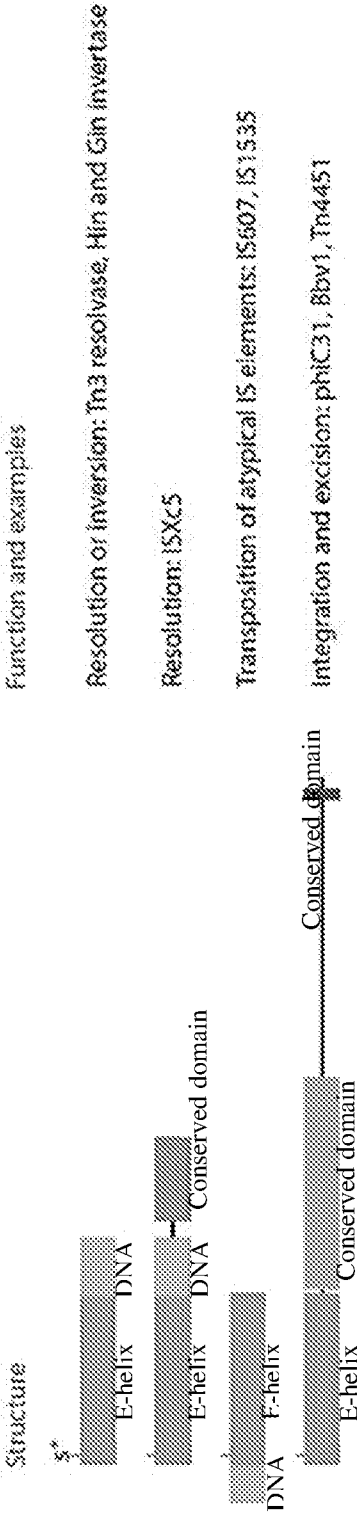


FIG. 3



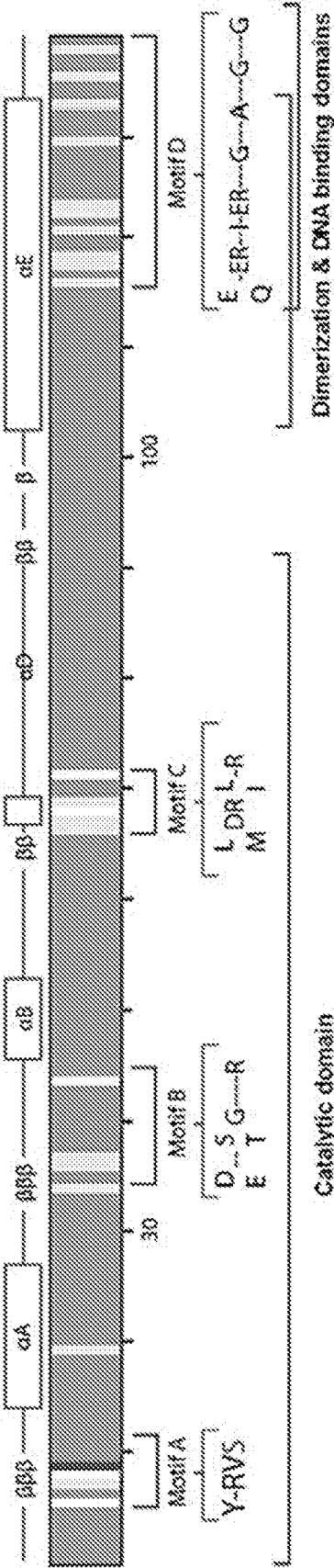


FIG. 5

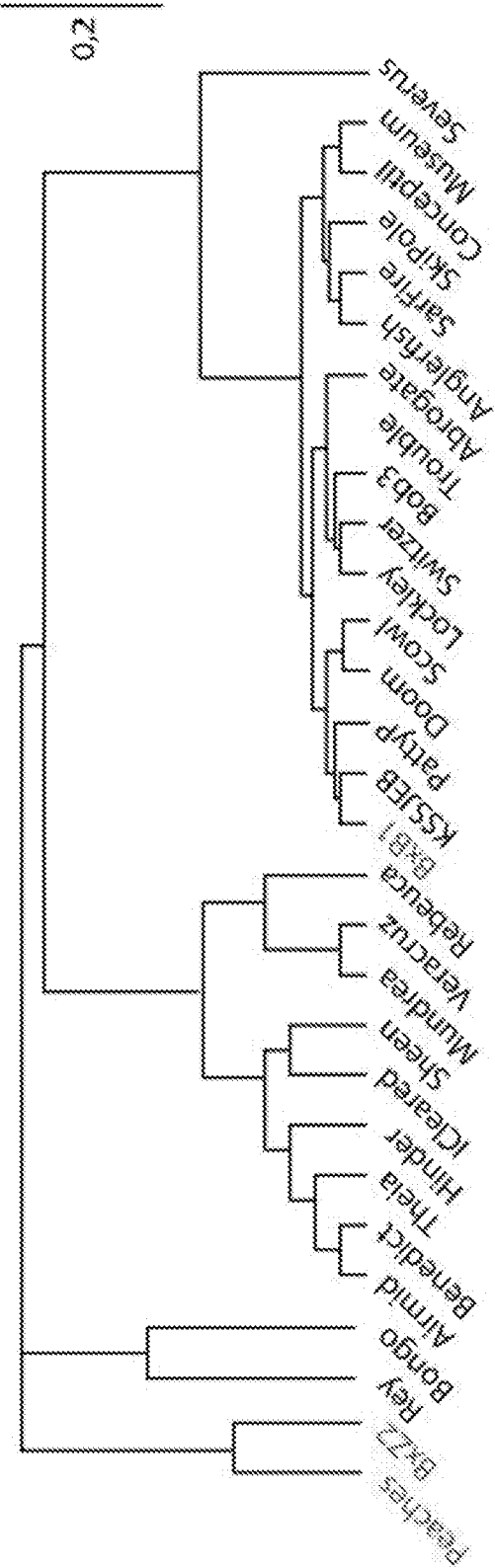
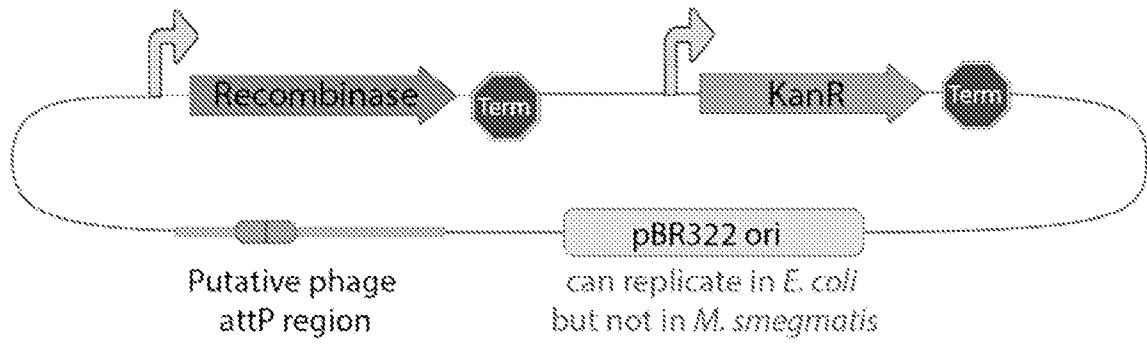
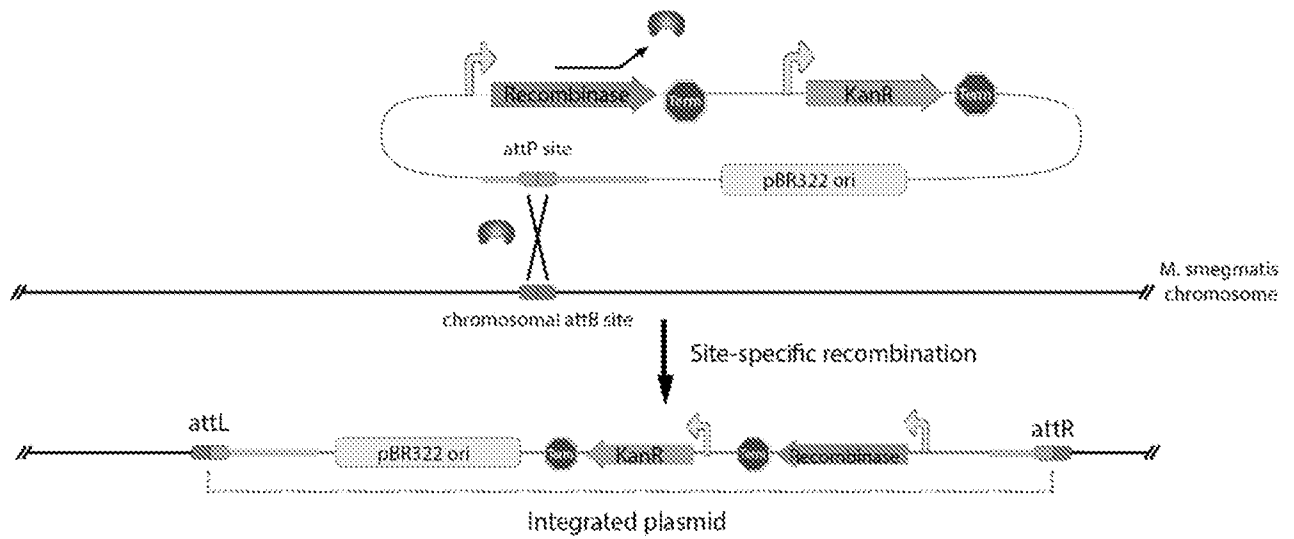
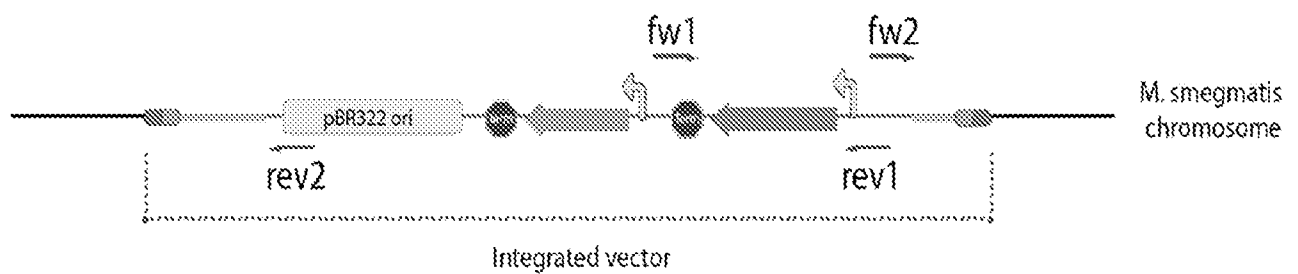


FIG. 6

**FIG. 7****FIG. 8****FIG. 9**

	Recombinases	Resistant clones	PCR verified
Rey	✓	✗	-
Bongo	✓	✓	✓
Almid	✓	✓	✓
Benedict	✓	✓	✓
Thela	✓	✓	✓
Hinder	✗	-	✓
Cleared	✓	✓	✓
Shreen	✓	✓	✓
Mundrea	✓	✓	✓
Veracruz	✓	✓	✓
Rebeuca	✓	✓	✓
B-81	✓	✓	✓
K5J1B	✓	✓	✓
Patyp	✓	✓	✓
Doom	✓	✓	✓
Scowl	✓	✓	✓
Lodley	✓	✓	✓
Switzer	✗	-	✓
Bab3	✗	-	✓
Trouble	✓	✓	✓
Abrogate	✓	✓	✓
Anglerfish	✓	✓	✓
Saffire	✓	✓	✓
Skipole	✓	✓	✓
Concepti	✓	✓	✓
Museum	✓	✓	✓
Severus	✓	✓	✓

FIG. 10

FIG. 11A

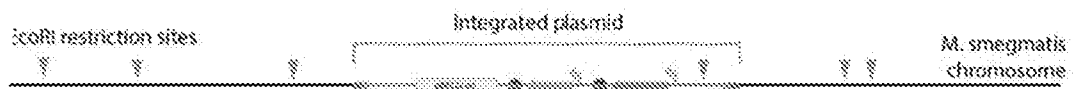


FIG. 11B

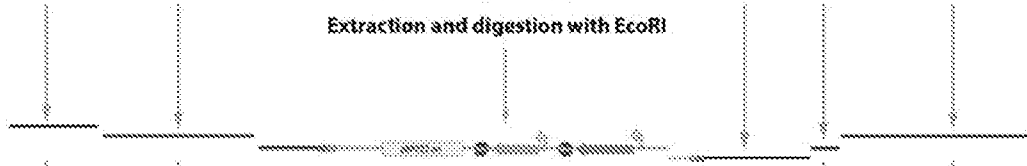


FIG. 11C

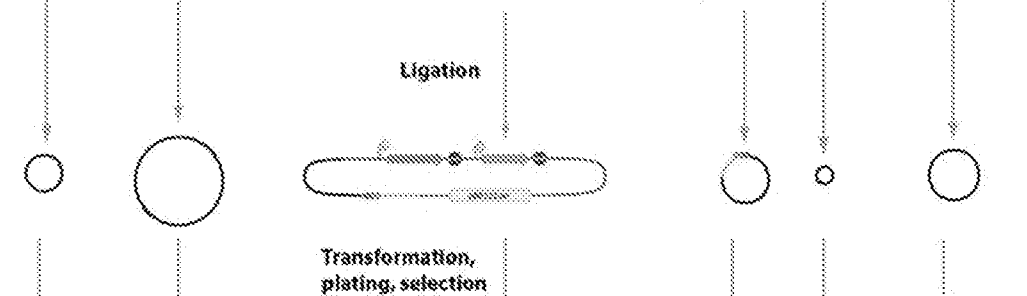


FIG. 11D



FIG. 11E

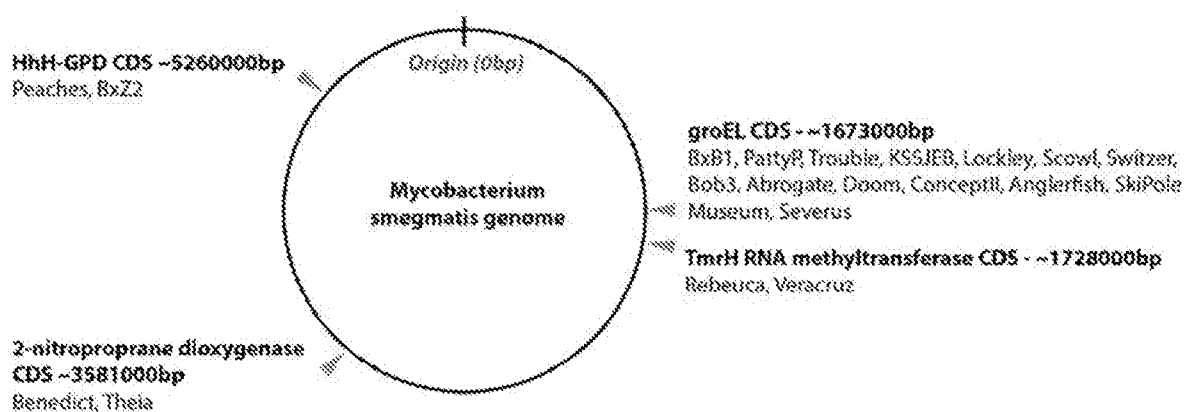
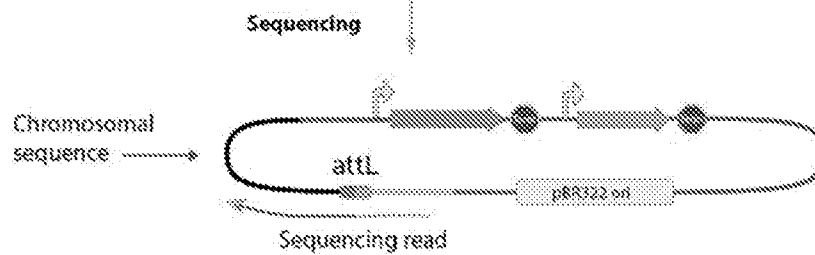


FIG. 12

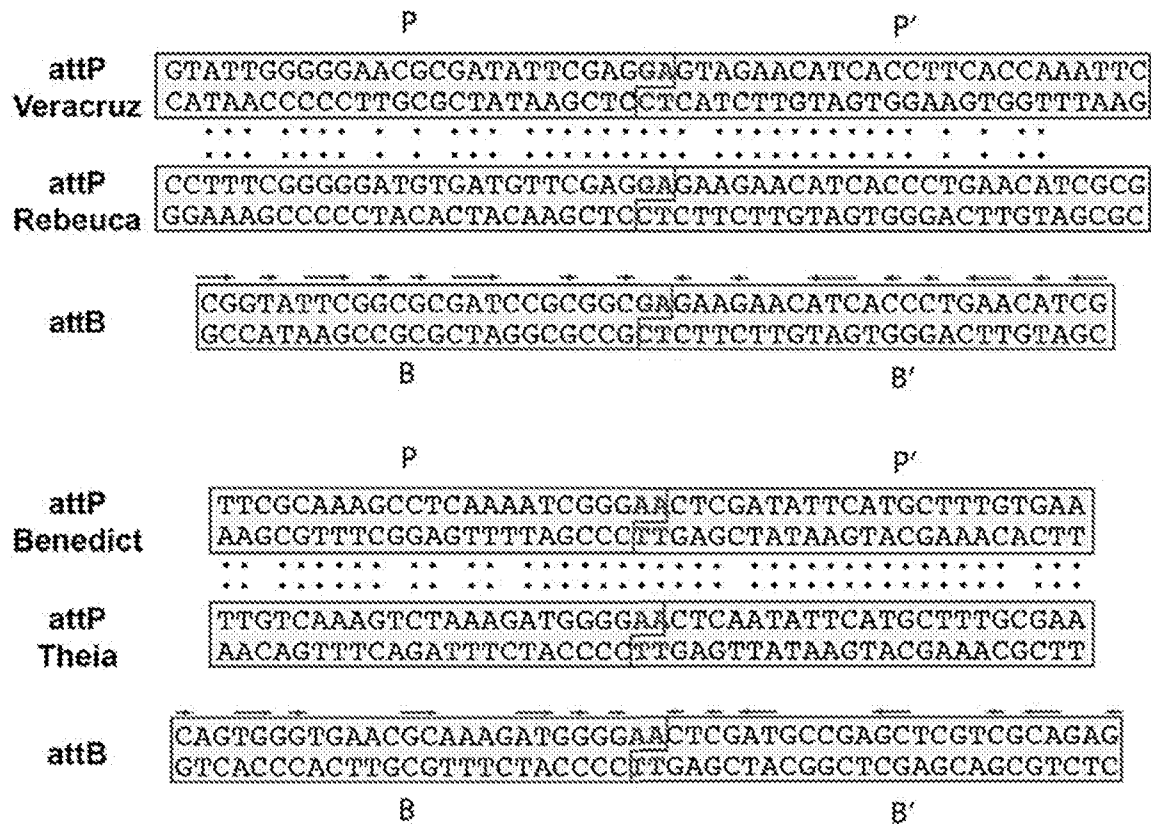


FIG. 13

FIG. 14A

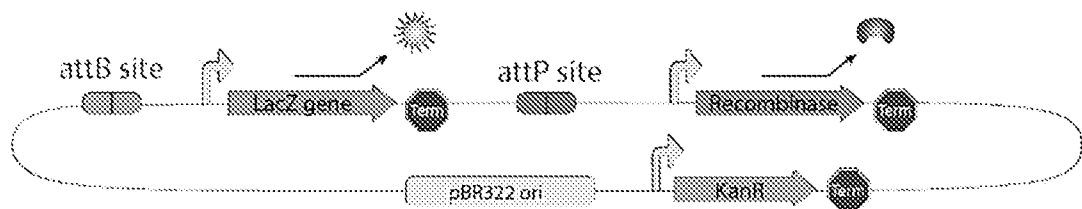


FIG. 14B

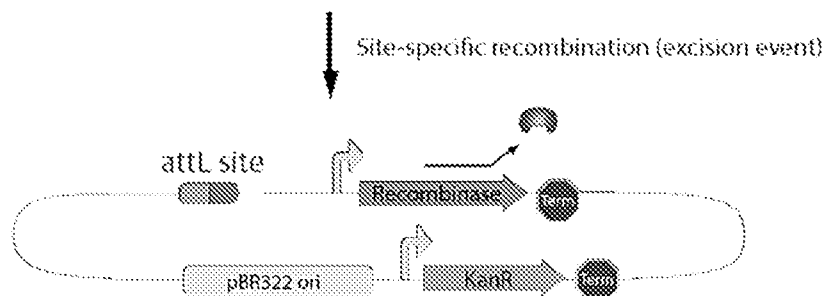


FIG. 15A

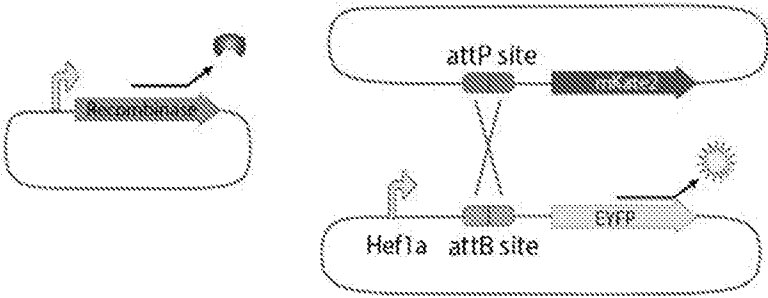


FIG. 15B

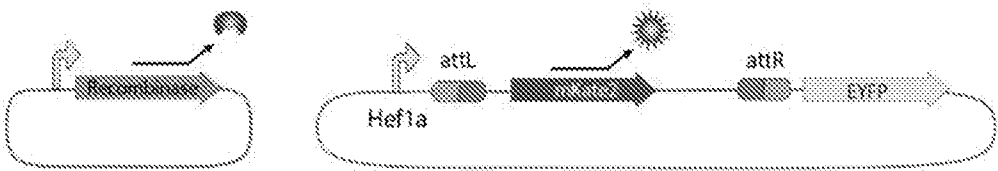


FIG. 16A

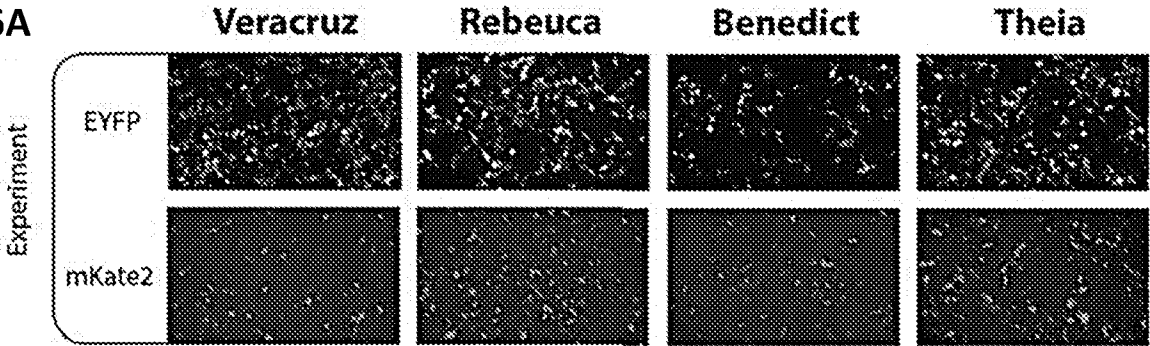
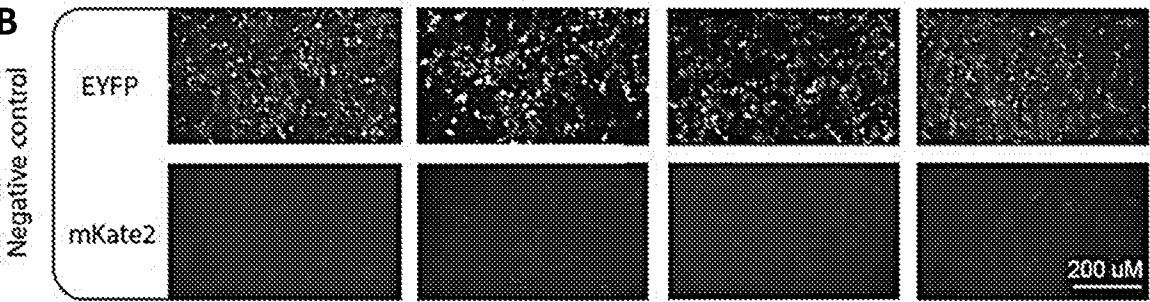


FIG. 16B



		Recombinase					
		Peaches	BxB1	Benedict	Theia	Veracruz	Rebeuca
attB/attP	Peaches						
	BxB1						
	Benedict						
	Theia						
	Veracruz						
	Rebeuca						

FIG. 17

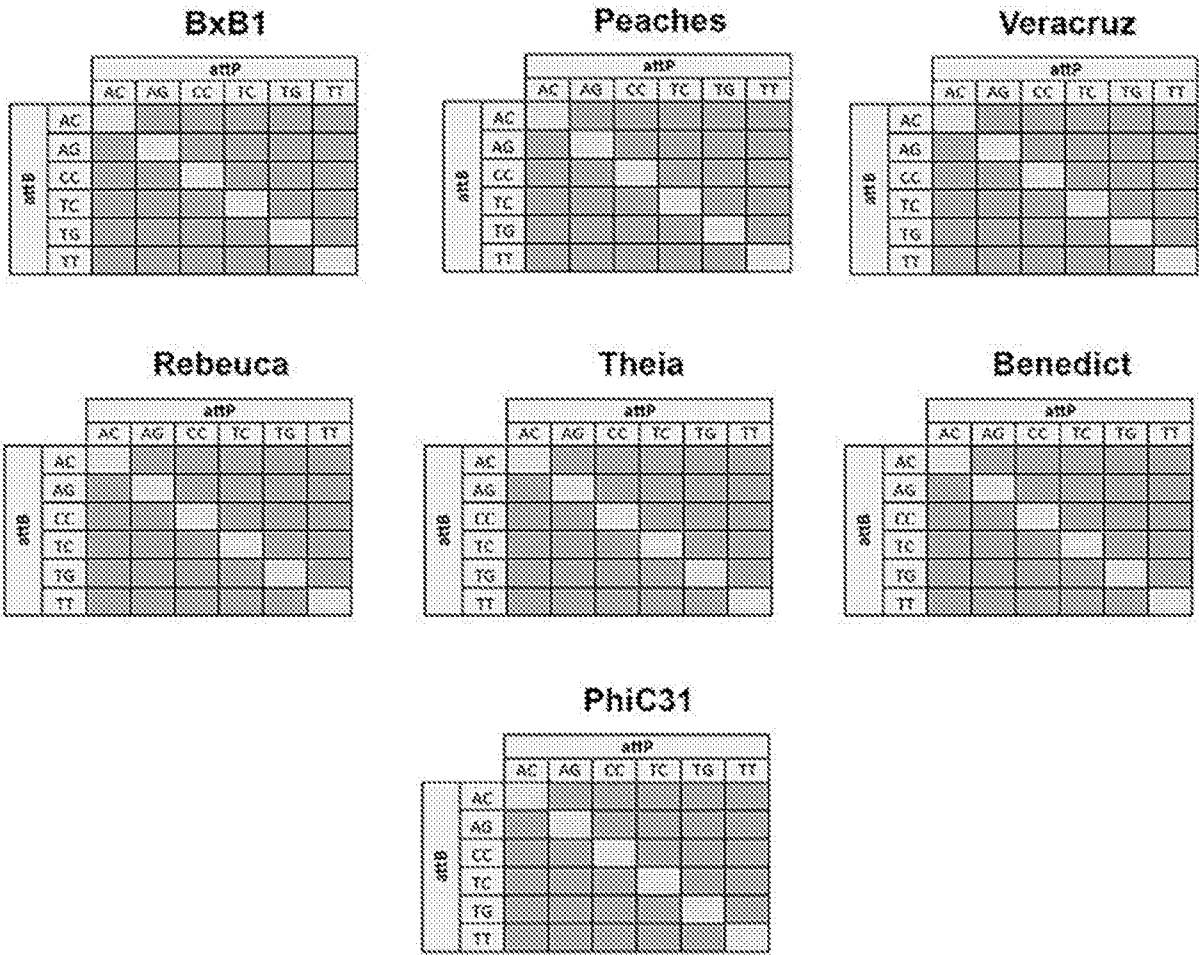


FIG. 18

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2017/014234

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/10 C12N15/90 C12N5/10
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

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

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

EPO-Internal, BIOSIS, CAB Data, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Xavier Duportet: "Developing new tools and platforms for mammalian synthetic biology: From the assembly and chromosomal integration of complex DNA circuits to the engineering of artificial intercellular communication systems", Universite Paris Diderot (Paris7) Ecole Doctorale Frontieres du vivant; DOCTORAT Discipline: Biologie de Synthèse, 14 November 2014 (2014-11-14), pages 1-213, XP002768716, Paris, France Retrieved from the Internet: URL:https://hal.archives-ouvertes.fr/tel-01108520 [retrieved on 2017-03-28] the whole document ----- -/-	1-50

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

☐ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

24 April 2017

Date of mailing of the international search report

19/06/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Hornig, Horst

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2017/014234

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SINGH SHWETA ET AL: "Cross-talk between diverse serine integrases.", JOURNAL OF MOLECULAR BIOLOGY 23 JAN 2014, vol. 426, no. 2, 23 January 2014 (2014-01-23), pages 318-331, XP002768717, ISSN: 1089-8638 the whole document	1-10
X	----- SINGH SHWETA ET AL: "SUPPLEMENTARY INFORMATION", JOURNAL OF MOLECULAR BIOLOGY 23 JAN 2014, vol. 426, no. 2, January 2014 (2014-01), pages 1-7, XP002768718, ISSN: 1089-8638 the whole document	1-10
X	----- SINGH SHWETA ET AL: "Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination.", PLOS GENETICS MAY 2013, vol. 9, no. 5, May 2013 (2013-05), page e1003490, XP002768719, ISSN: 1553-7404 the whole document	1-10
X	----- X. DUPORTET ET AL: "A platform for rapid prototyping of synthetic gene networks in mammalian cells", NUCLEIC ACIDS RESEARCH, vol. 42, no. 21, 1 December 2014 (2014-12-01), pages 13440-13451, XP055346806, ISSN: 0305-1048, DOI: 10.1093/nar/gku1082 the whole document	15,24
X	----- X. DUPORTET ET AL: "SUPPLEMENTARY INFORMATION", A PLATFORM FOR RAPID PROTOTYPING OF SYNTHETIC GENE NETWORKS IN MAMMALIAN CELLS, vol. 42, no. 21, 1 December 2014 (2014-12-01), pages 1-45, XP055346797, ISSN: 0305-1048, DOI: 10.1093/nar/gku1082 the whole document ----- -/--	15,24

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2017/014234

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	AMY C. GROTH ET AL: "Phage Integrases: Biology and Applications", JOURNAL OF MOLECULAR BIOLOGY, vol. 335, no. 3, 1 January 2004 (2004-01-01), pages 667-678, XP055359406, United Kingdom ISSN: 0022-2836, DOI: 10.1016/j.jmb.2003.09.082 the whole document -----	1-50
A	P. GUYE ET AL: "Rapid, modular and reliable construction of complex mammalian gene circuits", NUCLEIC ACIDS RESEARCH, vol. 41, no. 16, 1 September 2013 (2013-09-01), pages e156-e156, XP055359409, ISSN: 0305-1048, DOI: 10.1093/nar/gkt605 the whole document -----	1-50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/014234

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-50

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-50

A library of orthogonal large serine recombinase recognition site sequences, comprising pairs of attB site sequences and attP site sequences recognized by large serine recombinases; A cell line comprising a plurality of landing pads integrated into the genomic DNA of a parental cell line; A method of integrating a plurality of genetic sequences or circuits into a cell; A method for testing for function of predicted large serine recombinase recognition sites; A method for identifying a genomic insertion site of an integrated plasmid; A reporter plasmid for monitoring recombination events between large serine recombinase recognition sites; A method for testing for function of predicted large serine recombinase recognition sites; A reporter system for testing intermolecular site-specific recombination capacity of a large serine recombinase in mammalian cells ;

2-27. claim: 51(partially)

Invention 2: A method for site-specific recombination between specific attB and attP sites of a large serine recombinase, comprising providing a first nucleic acid molecule comprising an attB site, providing a second nucleic acid molecule comprising an attP site that selectively recombines with the attB site, combining the nucleic acid molecules in the presence of a large serine recombinase selected from the group consisting of: Veracruz (SEQ ID NO: 69); Invention 3: Idem as invention 2 but limited to Rebeuca (SEQ ID NO. 49); Invention 4: Idem as invention 2 but limited to Theia (SEQ ID NO. 65; Invention 27: Idem as invention 2 but limited to Mundrea (SEQ ID NO. 43).
