(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 March 2017 (16.03.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/044476 A1

(51) International Patent Classification:
C12N 15/70 (2006.01) *C12N 1/21* (2006.01)
C12N 15/76 (2006.01) *C12Q 1/68* (2006.01)
C12N 15/31 (2006.01)

(21) International Application Number:
PCT/US2016/050533

(22) International Filing Date:
7 September 2016 (07.09.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/215,281 8 September 2015 (08.09.2015) US

(71) Applicant: **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).

(72) Inventors: **LU, Timothy, Kuan-Ta**; 36 Spring Street, Cambridge, MA 02141 (US). **DA LUZ AREOSA CLETO, Sara**; Apt. 1, Quincy, MA 02169 (US).

(74) Agent: **DIPIETRANTONIO, Heather, J.**; 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, 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, KN, KP, KR, 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))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS AND COMPOSITIONS FOR RECOMBINASE-BASED GENETIC DIVERSIFICATION

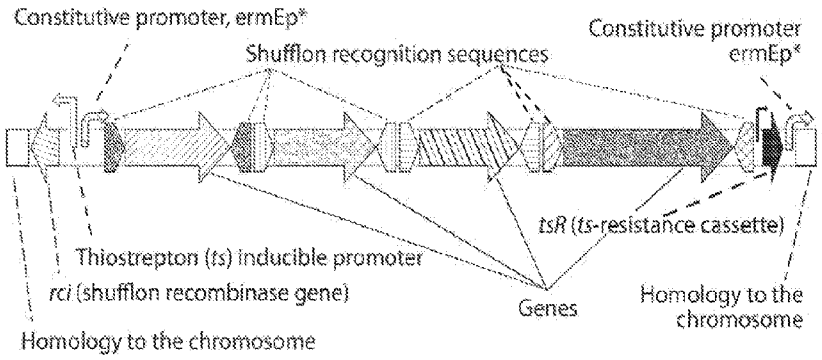


Fig. 4

(57) Abstract: Compositions and methods using shufflon recombinases are presented for use in generating genetic diversity in molecules of interest.

WO 2017/044476 A1

METHODS AND COMPOSITIONS FOR RECOMBINASE-BASED GENETIC DIVERSIFICATION

RELATED APPLICATION

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. provisional application number 62/215,281, filed September 8, 2015, which is incorporated by reference herein in its entirety.

BACKGROUND

There has been a steady increase in bacterial resistance to antibiotics, particularly in the clinic. Nonetheless, generating new antibiotics is challenging, in part because it is difficult to generate new molecules using existing technologies.

SUMMARY

One of the biggest challenges in engineering new molecules, such as antibiotics, is the large amount of manual labor required when using traditional cloning techniques. Provided herein are methods and compositions (*e.g.*, nucleic acids, cells and cell-free compositions) that streamline and expedite biosynthetic pathway engineering and new molecule production. The technology provided herein integrates synthetic biology with natural gene shuffling systems to scale up and engineer genetically diverse pathways containing multiple genes that individually, or in combination, encode a variety of bioactivities of interest, which in turn function together to produce a variety of new molecules. In particular, the technology provided herein uses modified components of a recombinase-based bacterial shufflon system that enables flipping (inverting) the orientation of genes in a controlled and predictable manner to turn expression of genes "on" (expressed) or "off" (not expressed), thereby avoiding leaky gene expression, which often accompanies conventional inducible systems.

Thus, the present disclosure provides engineered nucleic acids that include a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs.

In some embodiments, at least two heterologous genes encode enzymes, regulatory proteins, transport proteins, or combinations thereof.

In some embodiments, an engineered nucleic acid further includes a promoter operably linked to a gene encoding a shufflon recombinase.

The present disclosure also provides cells and cell-free compositions that include the engineered nucleic acids, as described herein.

The present disclosure further provides methods of producing new molecules using the nucleic acids, cells and compositions, as described herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows a schematic of a two-YAC (yeast artificial chromosome) assembly of parts and restriction sites permitting their release from the YAC backbone.

Fig. 2 shows maps of a pSETSC1 plasmid used to generate pSETSC2 and a final pSETSC3 plasmid. Fig. 2 also shows a map of an integrative pSET152 plasmid that includes a *gusA* reporter gene.

Fig. 3 shows the pikromycin pathway (tailoring enzymes are not displayed). Molecules 10-deoxy-methynolide (2) and narbonolide (3) are converted to methymycin and pikromycin, respectively (Kittendorf, J. *et al. Bioorg. Med. Chem.* 17:2137-46).

Fig. 4 shows the organization and description of a biosynthetic pathway assembled.

Fig. 5 shows the organization of genes to be shuffled, respective recognition sequences assigned, and bioactivity score predicted for pathway complementation with each individual gene (not combinations). Higher scores are preferred.

Fig. 6 shows a subdivision of sequences designated B, D, and E for use in obtaining respective PCR products.

Fig. 7 shows a schematic of a complete engineered nucleic acid of the present disclosure. Blocks represent DNA fragments and the letters indicate pieces to be obtained by PCR.

Fig. 8 shows a schematic of two YACs containing the assemblies of interest and the regions amplified by PCR to determine proper assembly. P1-5 represent the primer sets used for the amplifications.

Fig. 9 shows a schematic of an initial PCR approach and the amplicons effectively obtained upon experimentation. Alphanumeric nomenclature corresponds to the primers used (*see* Table 5).

Fig. 10 shows a representation of the double crossover event and integration of the synthetic shufflon system, which originates a new semi-synthetic pikromycin pathway.

Fig. 11 shows examples of gene combinations as a result of selective inversion.

Fig. 12 shows a map of plasmid pSETSC3 and the approach to the cloning of the two shufflon parts.

Fig. 13 shows the two shufflon parts (of Fig. 12) ligated into the dedicated backbone pSETSC3.

Fig. 14 shows a map of pSETSC3-shufflon.

Fig. 15 shows amplicon (PCR-amplified by KOD polymerase) and schematic of the shufflon part without the *rci* gene.

Fig. 16 shows a schematic of the shuffling system without the *rci* gene.

Fig. 17 shows a cloning approach for moving the *rci* gene to a temperature-sensitive plasmid and a final hygromycin resistant construct.

Fig. 18 shows four-FP shuffling system with different distances between recognition sites.

Fig. 19 shows an integrative expression plasmid pSETSCexp, based on pSET152.

Fig. 20 shows a schematic of the cloning scheme of each of the gene encoding the FPs and respective PCR confirmation.

Fig. 21 shows construction of a YAC-E. coli-*Streptomyces* integrative vector.

Fig. 22 shows construction of a temperature sensitive plasmid carrying *tsR*, *rci* and an hygromycin-resistance cassette.

Fig. 23 shows a schematic of the shuffling system (top panel). Primers for verification of shuffling represented by black arrows and roman numerals. PCR results using Herculase II polymerase (middle panel) and the combinations of primers listed (bottom panel) in Arabic numerals. qPCR-estimates % of flipping depicted in the table (bottom panel, right).

Fig. 24 is a schematic of the primer design for determination of flipping events by qPCR.

Fig. 25 shows shuffling systems to test the RCI flipping efficiency with increased inter-recognition sites distance. Systems with 1, 5 and 10 kbp between each pair of recognition sites.

Fig. 26 shows the first step of the 2-step approach to assembling the shuffling systems.

Fig. 27 is a schematic of the first step of the 2-step assembly of the 4-FP shuffling system.

Fig. 28 shows a plasmid for the integration of the large-size FP shuffling systems in *Streptomyces* spp.

Fig. 29 shows a map of the pFX583 plasmid, a version of which (with a gentamycin cassette instead of neo) was used to deliver the *rci* and *tsr* to the *Streptomyces* spp.

Fig. 30 shows a map of the pKC1139 plasmid. New versions of this plasmid, with a hygromycin and then with a gentamycin cassette were built to deliver *rci* and *tsr* to the *Streptomyces* spp.

Fig. 31 shows the equation for the calculation of the DNA inversion frequencies.

Fig. 32 shows the DNA-inversion efficiency of each pair of recognition sites over time.

Fig. 33 shows a map of a plasmid built for integration of the PKS-shufflon system in *Streptomyces* spp.

Fig. 34 is a schematic of PKS-shufflon cloning into a backbone.

Fig. 35 shows a schematic of the process of adding the T7 promoter to the cosmid backbone, for cloning the PKS-shufflon system.

Fig. 36 shows a schematic of the pikromycin operon: gene *pikAII* was not amplified, as the goal was to replace it with a synthetic operon.

Figs. 37A-37C show schematics of the 4-color shuffling constructs for assessing flipping efficiencies: unflipped (Fig. 37A), flipped (Fig. 37B), and schematic of the gBlock (all flipped amplicons) cloned into pSET152 (Fig. 37C).

Figs. 38-39 shows standards correlating Ct values and number of DNA molecules.

Fig. 40 shows a *S. venezuelae* growth curve with and without Rci recombinase.

Fig. 41 shows a schematic of the PrimeTime probes and primers annealing, for probe-based qPCR (IDT DNA Technologies).

Fig. 42 shows a schematic of the construct to be built for using GFP to determine Rci's flipping efficiencies.

Fig. 43 shows MALDI-ToF trace of the samples analyzed. The test samples are labeled 1 through 5, corresponding to the number of the colony displaying a different phenotype that was picked. "No Rci" strain carries only the pathway only (without Rci), "Rci" strains carries only the Rci (no pathway) and "WT" is the wild type strain of *S. coelicolor*. WT is the top trace, Rci is the second from top trace, No Rci is the third from top trace. Test sample 1 is the bottom trace, test sample 2 is the second from bottom trace, test sample 3 is the third from bottom trace, test sample 4 is the fourth from bottom trace, and test sample 5 is the fifth from bottom trace.

DETAILED DESCRIPTION

Provided herein is a system that uses modified components of a bacterial shufflon system to engineer genetically diverse pathways. Several rounds of shuffling (inversion) of a

particular engineered pathway results in expression of different combinations genes (genetic diversity), each combination potentially producing a new molecule (molecular diversity).

Some bacteria, such as *Salmonella typhimurium*, have the capacity to 'shuffle' (vary) the type of conjugative pili they produced, as a function of their conjugative partner in liquid matings (Komano, T. Annu Rev Genet. 1999;33:171-91; Gyohda, A. *et al. J. Bacteriol.* 1997;179:1867–71; Gyohda, A. *et al. J. Mol. Biol.* 2002;318:975–83; and Komano, T. *et al. Nucleic Acids Res.* 1987;15(3):1165-72). The plasmid R64 shufflon contains seven recombination sites, which flank and separate four DNA segments. Site-specific recombinations mediated by the product of the *rci* gene between any two inverted recombination sites result in the inversion of four DNA segments independently or in groups. The shufflon functions as a biological switch to select one of seven C-terminal segments of the PilV proteins, which is a minor component of R64 thin pilus. The shufflon determines the recipient specificity in liquid matings of plasmid R64 (Komano, T. 1999).

The present disclosure describes (1) the assembly of a complete genetically engineered shufflon system encoding heterologous genes (*e.g.*, enzymes), (2) integration of this system into a bacterial (*e.g.*, *Streptomyces* spp.) target locus, (3) induction of *rci* gene/protein expression and subsequent shuffling events, and (4) production of new molecules.

Biosynthetic Pathways of Interest

Common approaches to new molecule discovery and production (*e.g.*, antibiotic production), such as screening of new bacterial isolates, mix-and-match of genes through traditional cloning, chemical synthesis and medicinal chemistry, are slow to yield results. Provided herein is a pathway-based system that generates its own diversity and, in some embodiments, permits "hands-free" molecular engineering, maximizes the number of possible pathways without an increase in labor, can undergo several rounds of engineering with minimal labor, and permits an exponential increase of possible molecules.

The present disclosure provides a system that is able to generate molecular diversity, in some embodiments, by replacing a single gene of a biosynthetic pathway with a new biosynthetic pathway (multiple heterologous genes) and shuffling genes of the new biosynthetic pathway to produce new molecules (*e.g.*, new antibiotics).

The methods as provided herein are not limited by any particular biosynthetic pathway. Polyketide biosynthetic pathways (*e.g.*, pikromycin, erythromycin, lovastatin, discodermolide, aflatoxin B1, avermectins, nystatin and rifamycin) are used as examples

throughout the disclosure, as shown in Fig. 10 and Fig. 11. A (one or more) gene of a biosynthetic pathway of interest may be replaced with at least two heterologous genes that encode particular bioactivities. As shown, for example, in Fig. 10, the pikAII gene of the pikromycin pathway is replaced with *tylG*, *momAVI*, *amphK* and *momAVIII*, each of which encodes an enzyme having a particular bioactivity. Each of *tylG*, *momAVI*, *amphK* and *momAVIII* is flanked by a different pair of shufflon recombinase recognition sequences (RRSs), and each RRS recombines at a different rate in the presence of shufflon recombinase activity. Fig. 11 shows the different combinations of gene expression following selective inversion (shuffling). Each combination can potentially produce a new molecule (*e.g.*, when expressed in combination with other genes of the particular biosynthetic pathway of interest).

The expression of these one or more genes is shuffled to produce new molecules.

For example, as shown in Fig. 10 and Fig. 11, the *pikAII* gene of the pikromycin pathway may be replaced with one or more genes encoding selected enzymatic activities.

Examples of other biosynthetic pathways of interest include other small molecule biosynthetic pathways, such as, but not limited to, alkaloid biosynthetic pathways (*e.g.*, hyoscyamine, atropine, cocaine, scopolamine, codeine, morphine, tetrodotoxin, vincristine and vinblastine), terpenoid biosynthetic pathways (*e.g.*, azadirachtin, artemisinin, tetrahydrocannabinol, steroids and saponins), glycoside biosynthetic pathways (*e.g.*, nojirimycin and glucosinolates), phenazine biosynthetic pathways (*e.g.*, pyocyanin and phenazine-1-carboxylic acid (and derivatives)), biphenyl biosynthetic pathways, dibenxofuran biosynthetic pathways, fatty acid biosynthetic pathways (*e.g.*, FR-900848, U-106305 and phloroglucinols), nonribosomal peptide biosynthetic pathways (*e.g.*, vancomycin, ramoplanin, teicoplanin, gramicidin, bacitracin and ciclosporin), and ribosomal peptide biosynthetic pathways (*e.g.*, microcin-J25).

A "heterologous gene of a biosynthetic pathway of interest" refers to a gene that is not normally a component of a naturally-occurring biosynthetic pathway. For example, the wild-type pikromycin biosynthetic pathway includes the following genes: the *pikRa* and *pikR1* genes, the *pikAI-V* genes, the *pikBI-VIII genes*, the *pikBR* gene, the *pikC* gene and the *pikD* gene. The *pikAIII* gene, as discussed below, may be replaced with other genes, such as *tylG*, *momAVI*, *amphK* and *momAVIII*, that encode bioactivities of interest. The genes that replace the *pikAIII* gene are considered to be heterologous genes because they do not normally/naturally participate in the pikromycin biosynthetic pathway.

In some embodiments, the *pikAII* gene is replaced with at least one, at least two, at least three or at least four genes selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

At least one gene (wild-type gene) of a biosynthetic pathway of interest may be replaced with at least one heterologous gene (shuffling gene). The selection of heterologous genes (to be shuffled) may be based on a number of factors, including phylogenetic distance between the wild type gene and the heterologous gene (increased relatedness is preferred, in some embodiments), the size of the gene (*e.g.*, between 5-15 kilobases) and the predicted bioactivity score (*e.g.*, using an algorithm-based prediction of molecule structure based on genetic information (NP.searcher), used in combination with a bioactivity prediction algorithm (Molinspiration)). In some embodiments, an engineered nucleic acid comprises at least two heterologous genes of a biosynthetic pathway of interest. For example, an engineered nucleic acid may comprise (a promoter operably linked to) at least 2, at least 3, at least 4, or at least 5 heterologous genes of a biosynthetic pathway of interest. In some embodiments, an engineered nucleic acid may comprise (a promoter operably linked to) 1-5, 2-5, 1-10 or 2-10 heterologous genes of a biosynthetic pathway of interest. In some embodiments, an engineered nucleic acid may comprise (a promoter operably linked to) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more heterologous genes of a biosynthetic pathway of interest.

Heterologous genes of a biosynthetic pathway of interest generally encode proteins, such as enzymes, regulatory proteins (proteins that regulate gene/protein expression) and transport proteins (proteins that move materials within an organism, *e.g.*, carrier proteins, membrane proteins). In some embodiments, genes of a biosynthetic pathway of interest encode receptors, ligands, lytic proteins, or antimicrobial proteins/peptides.

Examples of enzymes include, but are not limited to, oxidoreductases (EC 1) (*e.g.*, dehydrogenase, oxidase), transferases (EC 2) (*e.g.*, transaminase, kinase), hydrolases (EC 3) (*e.g.*, lipase, amylase, peptidase), lyases (EC 4) (*e.g.*, decarboxylase), isomerases (EC 5) (*e.g.*, isomerase, mutase) and ligases (EC 6) (*e.g.*, synthetase). The International Union of Biochemistry and Molecular Biology developed a nomenclature for enzymes, the EC numbers; each enzyme is described by a sequence of four numbers preceded by "EC". Specific enzymes encompassed by the present disclosure may be selected from and identified based on their EC number retrieved from the website of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (Moss GP. "Recommendations of the Nomenclature Committee". International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes by the Reactions they Catalyse).

Examples of regulatory proteins include, but are not limited to, transcriptional activators and transcriptional repressors. Transcriptional activators typically bind nearby to transcriptional promoters and recruit RNA polymerase to directly initiate transcription. Transcriptional repressors typically bind to transcriptional promoters and sterically hinder transcriptional initiation by RNA polymerase.

Examples of transport proteins include, but are not limited to, transporters, channels and pumps. Transporters are membrane proteins responsible for transport of substances across the cell membrane. Channels are made up of proteins that form transmembrane pores through which selected ions can diffuse. Pumps are membrane proteins that can move substances against their gradients in an energy-dependent process known as active transport. In some embodiments, nucleic acid sequences encoding proteins and protein domains whose primary purpose is to bind other proteins, ions, small molecules, and other ligands may be used in accordance with the present disclosure.

Receptors, ligands and lytic proteins encoded by at least one heterologous gene include any receptor, ligand and lytic protein, described herein or known to one of ordinary skill in the art. Receptors tend to have three domains: an extracellular domain for binding ligands such as proteins, peptides or small molecules, a transmembrane domain and an intracellular or cytoplasmic domain, which frequently can participate in some sort of signal transduction event such as phosphorylation.

Examples of antimicrobial proteins and/or peptides include, but are not limited to, genes that encode instructions for cell death, or genes that encode bactericidal proteins. Examples of such gene modules include *pemI-pemK* genes of plasmid R100, the *phd-doc* genes of phage P1, the *ccdA-ccdB* genes of plasmid F, *mazE-mazF* (or *chpAI-chpAK*), *sof-gef*, *kicA-kicB*, *relB-relE*, *chpBI-chpBK* and *gef*. Other examples of antimicrobial proteins and/or peptides include, but are not limited to, bacteriocins, hydramacin-1, cecropins, moricins, papiliocins, poneratoxins, mastoparans, melittins, spinigerins, cupiennins, oxyopinins, magainins, dermaseptins, cathelicidins, defensins and protegrins. Other antimicrobial proteins and/or peptides are encompassed by the present disclosure.

Shufflon Recombinase and Recombinase Recognition Sites

A "shufflon recombinase" is the product of a site-specific recombinase gene, *rci*. The rci gene encodes a basic protein belonging to the integrase (Int) family of site-specific recombinases (Kubo A., *et al. Mol. Gen. Genet.* 213:30–35, 1988). A shufflon recombinase (also referred to as a shufflon-specific DNA recombinase) catalyzes site-specific

recombination between a pair of inverted repeat sequences of the multiple DNA inversion system in plasmid R64. There are several pairs of inverted repeat sequences that recombine in the presence of shufflon recombinase activity, any of which may be used as provided herein. Table 3 provides examples of shufflon recombinase recognition sequences (RRSs), including RRS *a* GTGCCAATCCGGTACGTGG (SEQ ID NO: 5), RRS *b* GTGCCAATCCGGTACCTGG (SEQ ID NO: 6), RRS c GTGCCAATCCGGTCGGTGG (SEQ ID NO: 7), and RRS *d* GTGCCAATCCGGTACTTGG (SEQ ID NO: 8). Thus, a "pair" of shufflon RRSs may include (a) GTGCCAATCCGGTACGTGG (SEQ ID NO: 5) and CCACGTACCGGATTGGCAC (SEQ ID NO: 66), (b) GTGCCAATCCGGTACCTGG (SEQ ID NO: 6) and CCAGGTACCGGATTGGCAC (SEQ ID NO: 67), (c) GTGCCAATCCGGTCGGTGG (SEQ ID NO: 7) and CCACCGACCGGATTGGCAC (SEQ ID NO: 68), or (d) GTGCCAATCCGGTACTTGG (SEQ ID NO: 8) and CCAAGTACCGGATTGGCAC (SEQ ID NO: 69). Other shufflon RRSs are encompassed by the present disclosure.

In some embodiments, different shufflon recombinase restriction sites (*e.g.*, RRSs *a-d*) recombine (are inverted) at different rates relative to one another. Thus, in some embodiments, an engineered nucleic acid of the present disclosure comprises at least two (*e.g.*, at least 2, at least 3, or at least 4) pairs of shufflon recombinase recognitions sites that recombine, in the presence of shufflon recombinase activity, at different rates relative to each other. For example, RRS *a* may recombined in 29% of the molecules, while RRS *b* may recombine in only 0.025% of the molecules. In some embodiments, RRS *c* recombines in 0.02% of the molecules and, in some embodiments, RRS *d* recombines in 0.1% of the molecules.

A shufflon recombinase (and/or shufflon recombination recognition sites), in some embodiments, is obtained from a *Salmonella* spp. (*e.g.*, *Salmonella enterica*), however, shufflon recombinase may be obtained from other bacterial species. For example, a shufflon recombinase may be obtained from an *Escherichia* spp. (*e.g.*, *Escherichia coli*), a *Yersinia* spp. (*e.g.*, *Yersinia enterocolitica*), a *Klebsiella spp.* (*e.g.*, *Klebsiella pneumoniae*), a *Acinetobacter spp.* (*e.g.*, *Acinetobacter baumannii*), a *Bordetella spp.* (*e.g.*, *bordetella pertussis*), a *Neisseria spp.* (*e.g.*, *Neisseria meningitidis*), an *Aeromonas spp.*, a *Franciesella spp.*, a *Corynebacterium spp.*, a *Citrobacter spp.*, a *Chlamydia spp.*, a *Hemophilus spp.*, a *Brucella spp.*, a*Mycobacterium spp.*, a *Legionella spp.*, a *Rhodococcus spp.*, a *Pseudomonas spp.*, a *Helicobacter spp.*, a *Salmonella spp.*, a *Vibrio spp.*, a*Bacillus spp.*, a *Erysipelothrix*

spp., a *Salmonella spp.*, a *Streptomyces spp., a Bacteroides* spp., a *Prevotella* spp., a *Clostridium* spp., a *Bifidobacterium spp.*, or *a Lactobacillus spp.*

In some embodiments, a shufflon recombinase is a *Salmonella enterica* shufflon recombinase.

The gene encoding a shufflon recombinase as provided herein may be located on the same nucleic acid comprising heterologous genes of a biosynthetic pathway of interest. Alternatively, the gene encoding a shufflon recombinase may be located on nucleic acid that is separate from a nucleic acid comprising heterologous genes of a biosynthetic pathway of interest. For example, a gene encoding a shufflon recombinase may be located in a chromosome of a cell that includes a plasmid containing an engineered nucleic acid comprising a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest. In other instances, a gene encoding a shufflon recombinase may be located on one plasmid (or other vector), while an engineered nucleic acid comprising a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest may be located on another plasmid (or other vector).

Typically, a single gene is located between a pair of shufflon recombinase recognition sites (*e.g.*, RRS *a* – *gene X* – RRS *a*). In some embodiments, however, more than one gene is located between a pair of shufflon recombinase recognition sites (*e.g.*, RRS *a* – *gene X, gene Y* – RRS *a*). In some embodiments, an engineered nucleic acid comprises a single gene located between a pair of shufflon recombinase recognition sites (*e.g.*, RRS *a* – *gene X* – RRS *a*) and at least two genes located between a different pair of shufflon recombinase recognition sites (*e.g.*, RRS *b* – *gene Y, gene Z* – RRS *b*).

A gene encoding a shufflon recombinase, in some embodiments, is codon-optimized for expression in a host cell of interest (*e.g.*, a bacterial host cell). Thus, a gene encoding a shufflon recombinase may be obtained from one species of bacterial cell (*e.g.*, *Escherichia* spp.) and then expressed in another species of bacterial cell (*e.g.*, *Streptomyces* spp.). There are several online tools available for codon optimization of a gene (*see, e.g.*, Chin, J.X. *et al.* Bioinformatics, *Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene desig*n, 2014; idtdna.com/CodonOpt; jcat.de; dna20.com/services/genegps; and genscript.com/gsfiles/techfiles/Webinar_Codon_ Optimization_Rachel_Speer_GenScript.pdf.

In some embodiments, a gene encoding the shufflon recombinase comprise a nucleotide sequence as set forth in SEQ ID NO: 35 or SEQ ID NO: 36 (codon optimized for expression in *Streptomyces* spp.

A gene encoding a shufflon recombinase, in some embodiments, is located on the same engineered nucleic acid that contains heterologous genes of a biosynthetic pathway of interest. In such embodiments, the gene encoding a shufflon recombinase (operably linked to a promoter) may be oriented in the forward 5' to 3' direction, or in the reverse 3' to 5' direction, as shown, for example, in Fig. 4. In Fig. 4, the *rci* gene and its promoter are oriented in the 3' to 5' direction, while the heterologous genes and their promoter are oriented in the 5' to 3' direction.

Engineered Nucleic Acids

The engineered nucleic acids of the present disclosure contain genetic elements that can regulate gene/protein expression. A "genetic element," as used herein, refers to a nucleotide sequence that has a role in gene expression. For example, nucleic acids (*e.g.*, recombinant nucleic acids) encoding proteins, promoters, enhancers and terminators are genetic elements. The nucleic acids of the present disclosure may be engineered using, for example, standard molecular cloning methods (see, e.g., *Current Protocols in Molecular Biology*, Ausubel, F.M., *et al.,* New York: John Wiley & Sons, 2006; Molecular Cloning: A Laboratory Manual, Green, M.R. and Sambrook J., New York: Cold Spring Harbor Laboratory Press, 2012; Gibson, D.G., *et al., Nature Methods* 6(5):343-345 (2009), the teachings of which relating to molecular cloning are herein incorporated by reference).

The term "nucleic acid" refers to at least two nucleotides covalently linked together, and in some instances, may contain phosphodiester bonds (*e.g.*, a phosphodiester "backbone"). In some embodiments, a nucleic acid of the present disclosure may be considered to be a nucleic acid analog, which may contain other backbones comprising, for example, phosphoramide, phosphorothioate, phosphorodithioate, O-methylphophoroamidite linkages, and/or peptide nucleic acids. Nucleic acids (*e.g.*, components, or portions, of the nucleic acids) of the present disclosure may be naturally occurring or engineered. Engineered nucleic acids include recombinant nucleic acids and synthetic nucleic acids. "Recombinant nucleic acids" may refer to molecules that are constructed by joining nucleic acid molecules and, in some embodiments, can replicate in a living cell. "Synthetic nucleic acids" may refer to molecules that are chemically or by other means synthesized or amplified, including those that are chemically or otherwise modified but can base pair with naturally occurring nucleic acid molecules. Recombinant and synthetic nucleic acids also include those molecules that result from the replication of either of the foregoing.

The nucleic acids may be single-stranded (ss) or double-stranded (ds), as specified, or may contain portions of both single-stranded and double-stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid/chimeric, where the nucleic acid contains any combination of deoxyribonucleotides and ribonucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine, hypoxanthine, isocytosine, and isoguanine.

Some aspects of the present disclosure provide methods and systems that use multiple components, such as a gene encoding shufflon recombinase and gene(s) encoding proteins or other molecules of a biosynthetic pathway. It should be understood that components may be encoded by a single nucleic acid (*e.g.*, on the same plasmid or other vector) or by multiple different (*e.g.*, independently-replicating) nucleic acids.

Engineered nucleic acids of the present disclosure may contain promoter sequences (promoters) operably linked to a heterologous gene of a biosynthetic pathway of interest. As used herein, a "promoter" refers to a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter may also contain sub-regions at which regulatory proteins and molecules may bind, such as RNA polymerase and other transcription factors. Promoters may be constitutive, inducible, activatable, repressible, tissue-specific or any combination thereof.

A promoter drives expression or drives transcription of the nucleic acid sequence that it regulates. Herein, a promoter is considered to be "operably linked" when it is in a correct functional location and orientation in relation to a nucleic acid sequence it regulates to control ("drive") transcriptional initiation and/or expression of that sequence. In some embodiments, a promoter is operably linked to a gene that is in the "on position" (*e.g.*, forward, 5' to 3') or "off position" (*e.g.*, reverse, 3' to 5'). A gene is in the "on position" if the gene and the promoter to which it is operably linked are oriented in the same direction along a nucleic acid (*e.g.*, both oriented 5' to 3', or both oriented 3' to 5'). By contrast, a gene is in the "off position" if the gene and the promoter to which it is operably linked are oriented in opposition directions along a nucleic acid (*e.g.*, the promoter oriented 5' to 3' and the gene oriented 3' to 5'). In such embodiments, the gene (located between (flanked by) a pair of recombinase recognition sites) may be turned "on" following an inversion event. In Figs. 10 and 11, for example, a block arrow is representative of a gene of a pathway of interest. Gene X, flanked by a pair of shufflon recombinase recognition sites (short inverted arrows) is operably linked to the promoter represented by a starred arrow. In Fig. 10, gene X is in the "on position." Gene X is also in the "on position" in Fig. 11, top and bottom panels.

Following recombination of the recombinase recognition sites that flank Gene X, gene X is flipped (inverted) and turned "off." Thus, in the middle panels of Fig. 11, gene X is in the "off position."

In some embodiments, a promoter operably linked to at least two heterologous genes of a synthetic pathway of interest is oriented in the 5' to 3' direction, and each gene operably linked to the promoter is oriented in a 5' to 3' direction. Thus, each gene is in the "on position" when the promoter is active.

In some embodiments, a promoter operably linked to at least two heterologous genes of a synthetic pathway of interest is oriented in the 5' to 3' direction, and each gene operably linked to the promoter is oriented in a 3' to 5' direction. Thus, each gene is in the "off position" when the promoter is active.

In some embodiments, a promoter operably linked to at least two heterologous genes of a synthetic pathway of interest is oriented in the 5' to 3' direction, and at least one gene operably linked to the promoter is oriented in a 5' to 3' direction and at least one gene operably linked to the promoter is oriented in a 3' to 5'.

A promoter may be classified as strong or weak according to its affinity for RNA polymerase (and/or sigma factor); this is related to how closely the promoter sequence resembles the ideal consensus sequence for the polymerase. The strength of a promoter may depend on whether initiation of transcription occurs at that promoter with high or low frequency. Different promoters with different strengths may be used herein.

A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment of a given gene or sequence.

A promoter is "recombinant" or "heterologous" if it is not naturally/normally associated with (does not naturally/normally control transcription of) a gene to which is it operably linked. In some embodiments, a gene (or other nucleotide sequence) may be positioned under the control of a recombinant or heterologous promoter. Such promoters may include promoters of other genes; promoters isolated from any prokaryotic cell; and synthetic promoters that are not "naturally occurring" such as, for example, those that contain different elements of different transcriptional regulatory regions and/or mutations that alter expression through methods of genetic engineering. In addition to producing nucleic acid sequences of promoters synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including polymerase chain reaction (PCR) (see U.S. Pat. No. 4,683,202 and U.S. Pat. No. 5,928,906).

As used herein, an "inducible promoter" is one that is characterized by initiating or enhancing transcriptional activity when in the presence of, influenced by or contacted by an inducer or inducing agent. An "inducer" or "inducing agent" may be endogenous or a normally exogenous condition, compound or protein that contacts a programmable nuclease circuit in such a way as to be active in inducing transcriptional activity from the inducible promoter.

Inducible promoters for use in accordance with the present disclosure, in some embodiments, function in a bacterial cell. Examples of inducible promoters for use herein include, without limitation, bacteriophage promoters (*e.g.* Pls1con, T3, T7, SP6, PL) and bacterial promoters (*e.g.* Pbad, PmgrB, Ptrc2, Plac/ara, Ptac, Pm), or hybrids thereof (*e.g.* PLlacO, PLtetO). Examples of bacterial promoters for use in accordance with the present disclosure include, without limitation, positively regulated *E. coli* promoters such as positively regulated σ^{70} promoters (*e.g.,* inducible pBad/araC promoter, Lux cassette right promoter, modified lamdba Prm promote, plac Or2-62 (positive), pBad/AraC with extra REN sites, pBad, P(Las) TetO, P(Las) CIO, P(Rhl), Pu, FecA, pRE, cadC, hns, pLas, pLux), σ^{S} promoters (*e.g.,* Pdps), σ^{32} promoters (*e.g.,* heat shock) and σ^{54} promoters (*e.g.,* glnAp2); negatively regulated *E. coli* promoters such as negatively regulated σ^{70} promoters (*e.g.,* Promoter (PRM+), modified lamdba Prm promoter, TetR - TetR-4C P(Las) TetO, P(Las) CIO, P(Lac) IQ, RecA_DlexO_DLacO1, dapAp, FecA, Pspac-hy, pcI, plux-cI, plux-lac, CinR, CinL, glucose controlled, modifed Pr, modifed Prm+, FecA, Pcya, rec A (SOS), Rec A (SOS), EmrR_regulated, BetI_regulated, pLac_lux, pTet_Lac, pLac/Mnt, pTet/Mnt, LsrA/cI, pLux/cI, LacI, LacIQ, pLacIQ1, pLas/cI, pLas/Lux, pLux/Las, pRecA with LexA binding site, reverse BBa_R0011, pLacI/ara-1, pLacIq, rrnB P1, cadC, hns, PfhuA, pBad/araC, nhaA, OmpF, RcnR), σ^{S} promoters (*e.g.,* Lutz-Bujard LacO with alternative sigma factor σ^{38}), σ^{32} promoters (*e.g.,* Lutz-Bujard LacO with alternative sigma factor σ^{32}), and σ^{54} promoters (*e.g.,* glnAp2); negatively regulated *B. subtilis* promoters such as repressible *B. subtilis* σ^{A} promoters (*e.g.,* Gram-positive IPTG-inducible, Xyl, hyper-spank) and σ^{B} promoters. Other inducible promoters may be used in accordance with the present disclosure.

The administration or removal of an inducer results in a switch between activation and inactivation of the transcription of the operably linked gene or other nucleic acid sequence (*e.g.*, heterologous gene). Thus, as used herein, the active state of a promoter operably linked to a gene refers to the state when the promoter is actively driving transcription of the gene (*i.e.*, the linked gene is expressed). Conversely, the inactive state of

a promoter operably linked to a gene refers to the state when the promoter is not actively driving transcription of the gene (*i.e.*, the linked gene is not expressed).

An inducible promoter for use in accordance with the present disclosure may be induced by (or repressed by) one or more physiological condition(s), such as changes in pH, temperature, radiation, osmotic pressure, saline gradients, cell surface binding, and the concentration of one or more extrinsic or intrinsic inducing agent(s). The extrinsic inducer or inducing agent may comprise, without limitation, amino acids and amino acid analogs, saccharides and polysaccharides, nucleic acids, protein transcriptional activators and repressors, cytokines, toxins, petroleum-based compounds, metal containing compounds, salts, ions, enzyme substrate analogs, hormones or combinations thereof.

Inducible promoters for use in accordance with the present disclosure include any inducible promoter described herein or known to one of ordinary skill in the art. Examples of inducible promoters include, without limitation, chemically/biochemically-regulated and physically-regulated promoters such as alcohol-regulated promoters, tetracycline-regulated promoters (*e.g.*, anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems, which include a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)), steroid-regulated promoters (*e.g.*, promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily), metal-regulated promoters (*e.g.*, promoters derived from metallothionein (proteins that bind and sequester metal ions) genes from yeast, mouse and human), pathogenesis-regulated promoters (*e.g.*, induced by salicylic acid, ethylene or benzothiadiazole (BTH)), temperature/heat-inducible promoters (*e.g.*, heat shock promoters), and light-regulated promoters (*e.g.*, light responsive promoters from plant cells).

In some embodiments, the inducer used in accordance with the present disclosure is an *N*-acyl homoserine lactone (AHL), which is a class of signaling molecules involved in bacterial quorum sensing. Quorum sensing is a method of communication between bacteria that enables the coordination of group based behavior based on population density. AHL can diffuse across cell membranes and is stable in growth media over a range of pH values. AHL can bind to transcriptional activators such as LuxR and stimulate transcription from cognate promoters. In some embodiments, the inducer used in accordance with the present disclosure is anhydrotetracycline (aTc), which is a derivative of tetracycline that exhibits no antibiotic activity and is designed for use with tetracycline-controlled gene expression systems, for example, in bacteria.

Other inducible promoter systems are known in the art and may be used in accordance with the present disclosure.

Some engineered nucleic acids may include nucleotide sequences homologous to a chromosomal locus of a host cell of interest. Such sequences facilitate integration of an engineered nucleic acid into a chromosomal locus of a host cell. It should be understood, however, than chromosomal integration of an engineered nucleic acid encoding heterologous genes of a biosynthetic pathway is optional.

In some embodiments, an engineered nucleic acid also comprises an antibiotic resistance gene (*see*, *e.g.*, online database: card.mcmaster.ca) to facilitate cloning and selection of the nucleic acid. Thus, in some embodiments, an engineered nucleic acid comprises a kanamycin resistance gene, spectinomycin resistance gene, streptomycin resistance gene, ampicillin resistance gene, carbenicillin resistance gene, bleomycin resistance gene, erythromycin resistance gene, polymyxin B resistance gene, tetracycline resistance gene, chloramphenicol resistance gene, hygromycin resistance gene and/or a ts-resistance gene resistance gene.

Cells

Engineered nucleic acids of the present disclosure may be introduced into a host cell of interest, and in some embodiments, are optimized (*e.g.*, codon-optimized) for expression in a host cell of interest. In some embodiments, the host cell is a bacterial cell. In some embodiments, the host cell is a mammalian cell, a yeast cell or an insect cell.

Bacteria are small (typical linear dimensions of around 1 micron), non-compartmentalized organisms, with at least one circular DNA chromosomes and ribosomes of 70S. As used herein, the term "bacteria" encompasses all variants of bacteria (*e.g.*, endogenous bacteria, which naturally reside in a closed system, environmental bacteria or bacteria released for bioremediation or other efforts). Bacterial cells of the present disclosure include bacterial subdivisions of *Eubacteria* and *Archaebacteria*. *Eubacteria* can be further subdivided into Gram-positive and Gram-negative *Eubacteria*, which depend upon a difference in cell wall structure. Also included herein are those classified based on gross morphology alone (*e.g.*, cocci, bacilli). In some embodiments, bacterial cells are Gram-negative cells, and in some embodiments, the bacterial cells are Gram-positive cells. Examples of bacterial cells of the present disclosure include, without limitation, cells classified as *Yersinia spp.*, *Escherichia spp.*, *Klebsiella spp.*, *Acinetobacter spp.*, *Bordetella spp.*, *Neisseria spp.*, *Aeromonas spp.*, *Franciesella 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.*, *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.*, or *Lactobacillus spp.* In some embodiments, the bacterial cells are classified as *Bacteroides thetaiotaomicron*, *Bacteroides fragilis*, *Bacteroides distasonis*, *Bacteroides vulgatus*, *Clostridium leptum*, *Clostridium coccoides*, *Staphylococcus aureus*, *Bacillus subtilis*, *Clostridium butyricum*, *Brevibacterium lactofermentum*, *Streptococcus agalactiae*, *Lactococcus lactis*, *Leuconostoc lactis*, *Actinobacillus actinobycetemcomitans*, *cyanobacteria*, *Escherichia coli*, *Helicobacter pylori*, *Selnomonas ruminatium*, *Shigella sonnei*, *Zymomonas mobilis*, *Mycoplasma mycoides*, *Treponema denticola*, *Bacillus thuringiensis*, *Staphlococcus lugdunensis*, *Leuconostoc oenos*, *Corynebacterium xerosis*, *Lactobacillus plantarum*, *Lactobacillus rhamnosus, Lactobacillus casei, Lactobacillus acidophilus, Streptococcus spp., Enterococcus faecalis*, *Bacillus coagulans*, *Bacillus ceretus*, *Bacillus popillae*, *Synechocystis strain PCC6803*, *Bacillus liquefaciens*, *Pyrococcus abyssi, Selenomonas nominantium*, *Lactobacillus hilgardii*, *Streptococcus ferus*, *Lactobacillus pentosus*, *Bacteroides fragilis*, *Staphylococcus epidermidis*, *Zymomonas mobilis*, *Streptomyces phaechromogenes*, or *Streptomyces ghanaenis*. Thus, the engineered nucleic acids of the present disclosure may be introduced into a bacterial cell from any one or more of the foregoing genus and/or species of bacteria. Other bacterial cells and microbes may also be used. As used herein, "endogenous" bacterial cells may refer to non-pathogenic bacteria that are part of a normal internal ecosystem such as bacterial flora.

In some embodiments, a host cell is a *Streptomyces* spp. cell, such as a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

In some embodiments, bacterial cells of the present disclosure are anaerobic bacterial cells (*e.g.*, cells that do not require oxygen for growth). Anaerobic bacterial cells include facultative anaerobic cells such as, for example, *Escherichia coli*, *Shewanella oneidensis* and *Listeria monocytogenes*. Anaerobic bacterial cells also include obligate anaerobic cells such as, for example, *Bacteroides* and *Clostridium* species. In humans, for example, anaerobic bacterial cells are most commonly found in the gastrointestinal tract. Thus, engineered nucleic acids of the present disclosure may be introduced into anaerobic bacterial cells.

Cells of the present disclosure are generally considered to be modified. A modified cell is a cell that contains an exogenous nucleic acid or a nucleic acid that does not occur in nature (*e.g.*, an engineered nucleic acid of the present disclosure). In some embodiments, a

cell contains a deletion or a modification (*e.g.*, mutation) in a genomic/chromosomal nucleic acid. For example, a cell may comprise a deletion or modification in a (at least one) gene of a biosynthetic pathway of interest, such as a deletion or a modification in the pikromycin biosynthetic pathway (*e.g.*, in a pikAI-V , pikBI-VIII, pikBR gene, pikC gene or pikD gene). In some embodiments, a host cell comprises a deletion in the *pikAII* or *pikAIII* gene.

In some embodiments, a cell contains an exogenous independently replicating nucleic acid (*e.g.*, an engineered nucleic acid located on an episomal vector). In some embodiments, a cell is produced by introducing a foreign or exogenous nucleic acid (*e.g.*, engineered nucleic acid) into a cell. Thus, provided herein are methods of introducing engineered nucleic acid into a cell. A nucleic acid may be introduced into a cell by conventional methods, such as, for example, electroporation (*see*, *e.g.*, Heiser W.C. *Transcription Factor Protocols: Methods in Molecular Biology*™ 2000; 130: 117-134), chemical (*e.g.*, calcium phosphate or lipid) transfection (*see*, *e.g.*, Lewis W.H., *et al.*, *Somatic Cell Genet.* 1980 May; 6(3): 333-47; Chen C., et al., *Mol Cell Biol.* 1987 August; 7(8): 2745–2752), fusion with bacterial protoplasts containing recombinant plasmids (*see*, *e.g.*, Schaffner W. *Proc Natl Acad Sci USA*. 1980 Apr; 77(4): 2163-7), transduction, conjugation, or microinjection of purified DNA directly into the nucleus of the cell (*see*, *e.g.*, Capecchi M.R. *Cell*. 1980 Nov; 22(2 Pt 2): 479-88).

Engineered nucleic acids of the present disclosure may be transiently expressed or stably expressed. "Transient cell expression" refers to expression by a cell of a nucleic acid that is not integrated into the nuclear genome of the cell. By comparison, "stable cell expression" refers to expression by a cell of a nucleic acid that remains in the nuclear genome of the cell and its daughter cells. Typically, to achieve stable cell expression, a cell is co-transfected with a marker gene and an exogenous nucleic acid that is intended for stable expression in the cell. The marker gene gives the cell some selectable advantage (e.g., resistance to a toxin, antibiotic, or other factor). Few transfected cells will, by chance, have integrated the exogenous nucleic acid into their genome. If a toxin, for example, is then added to the cell culture, only those few cells with a toxin-resistant marker gene integrated into their genomes will be able to proliferate, while other cells will die. After applying this selective pressure for a period of time, only the cells with a stable transfection remain and can be cultured further. Expression of nucleic acids in transiently-transfected and/or stably-transfected cells may be constitutive or inducible. Inducible promoters for use as provided herein are described above.

Cell-free Compositions

Also provided herein are cell-free compositions (e.g., in vitro transcription and/or translation reactions) that may be used, for example, to generate/produce new molecules. In some embodiments, a cell-free composition comprises an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a synthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs. A cell-free composition typically further comprises an engineered nucleic acid comprising a promoter operably linked to a gene encoding a shufflon recombinase, a cell extract and a polymerase.

A "cell extract" (cell lysate) refers to the contents of a cell without the intact cell wall. In some embodiments, a cell extract contains portions of the cell wall, for example, in the form of inverted membrane vesicles. A typical bacterial cell extract may include, for example, at least one component selected from the group consisting of: ribosomes, amino acids, NTPs, phosphoenolpyruvate, pyruvate kinase, polyethylene glycol, ammonium acetate, potassium acetate and folinic acid. In some embodiments, a cell extract comprises a S30 cell fraction obtained from a bacterial cell. In some embodiments, a cell-free composition comprises ribosomes, amino acids, NTPs, phosphoenolpyruvate, pyruvate kinase, polyethylene glycol, ammonium acetate, potassium acetate and folinic acid.

A cell-free composition may further comprise an engineered nucleic acid encoding additional proteins of the biosynthetic pathway of interest, or may further comprise additional proteins of the biosynthetic pathway of interest. For example, if the pikromycin biosynthetic pathway is the pathway of interest, a cell may have a deletion in the pikAII gene (replaced with an engineered nucleic acid encoding heterologous (shuffling) genes) and may express some or all of the remaining genes of the pikromycin biosynthetic pathway (*e.g.*, in a pikAI-V , pikBI-VIII, pikBR gene, pikC gene or pikD gene). Thus, a cell-free composition may also comprises the remaining genes of the pikromycin biosynthetic pathway (*e.g.*, in a pikAI-V , pikBI-VIII, pikBR gene, pikC gene or pikD gene).

Cell-free compositions may also contain polymerase, such as RNA polymerase (*e.g.*, RNA T7, T3 or SP6 polymerase). Other RNA polymerases may be used.

Methods

Provided herein are methods of producing molecules (*e.g.*, new molecules, such as new antibiotic molecules). The methods may comprise incubating a cell-free composition

under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the synthetic pathway of interest.

In some embodiments, the methods comprise culturing cells that comprise a gene encoding a shufflon recombinase and an engineered nucleic acid that includes (a) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and (b) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs, under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the synthetic pathway of interest.

In some embodiments, the methods comprise culturing cells that comprise (a) an engineered nucleic acid that includes a promoter operably linked to a gene encoding a shufflon recombinase, and (b)an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a synthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs, under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the synthetic pathway of interest. In some embodiments, the methods further comprise isolating the molecules. In some embodiments, the molecules comprise (or are) antibiotics.

The present disclosure provide methods that include delivering to cells at least one of the engineered nucleic acid constructs as provided herein. Constructs may be delivered by any suitable means, which may depend on the residence and type of cell. For example, if cells are located *in vivo* within a host organism (e.g., an animal such as a human), engineered nucleic acid constructs may be delivered by injection into the host organism of a composition containing engineered nucleic acid constructs. Constructs may be delivered by a vector, such as a viral vector (*e.g.*, bacteriophage or phagemid). For cells that are not located within a host organism, for example, for cells located *ex vivo/in vitro* or in an environmental (*e.g.*, outside) setting, engineered nucleic acid constructs may be delivered to cells by electroporation, chemical transfection, fusion with bacterial protoplasts containing recombinant, transduction, conjugation, or microinjection of purified DNA directly into the nucleus of the cells.

The present disclosure further provides embodiments encompassed by the following numbered paragraphs:

1. An engineered nucleic acid, comprising: (a) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest; and (b) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs.

2. The engineered nucleic acid of paragraph 1, wherein the at least two heterologous genes encode proteins selected from the group consisting of: enzymes, regulatory proteins and transport proteins.

3. The engineered nucleic acid of paragraph 1 or 2 further comprising (c) a promoter operably linked to a gene encoding a shufflon recombinase.

4. The engineered nucleic acid of paragraph 3, wherein the shufflon recombinase is a *Salmonella enterica* shufflon recombinase.

5. The engineered nucleic acid of paragraph 3 or 4, wherein the gene encoding the shufflon recombinase is codon-optimized for expression in a host cell of interest.

6. The engineered nucleic acid construct of any one of paragraphs 1-5, wherein the gene encoding the shufflon recombinase comprise a nucleotide sequence as set forth in SEQ ID NO: 35 or SEQ ID NO: 36.

7. The engineered nucleic acid of any one of paragraphs 1-7 further comprising (d) nucleotide sequences homologous to a chromosomal locus of a host cell of interest.

8. The engineered nucleic acid of any one of paragraphs 5-7, wherein the host cell of interest is a bacterial cell.

9. The engineered nucleic acid of paragraph 8, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

10. The engineered nucleic acid of paragraph 9, wherein the bacterial cell is a *Streptomyces* spp. cell.

11. The engineered nucleic acid of paragraph 10, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

12. The engineered nucleic acid of any one of paragraphs 3-11, wherein the gene encoding the shufflon recombinase is oriented in a 3' to 5' direction.

13. The engineered nucleic acid of any one of paragraphs 3-12, wherein the promoter of (a) and/or (c) is an inducible promoter.

14. The engineered nucleic acid of any one of paragraphs 1-13, wherein the engineered nucleic acid construct further comprises an antibiotic resistance gene.

15. The engineered nucleic acid of any one of paragraphs 1-14, wherein the RRSs recombine, in the presence of shufflon recombinase activity, at different rates relative to each other.

16. The engineered nucleic acid of paragraph 15, wherein the RRSs comprises a nucleotide sequence selected from sequences set forth as SEQ ID NO: 5-8.

17. The engineered nucleic acid of any one of paragraphs 1-16, wherein at least two of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

18. The engineered nucleic acid of any one of paragraphs 1-17, wherein the promoter of (a) is operably linked to at least three heterologous genes of a biosynthetic pathway of interest.

19. The engineered nucleic acid of paragraph 18, wherein at least three of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

20. The engineered nucleic acid of paragraph 18 or 19, wherein the promoter of (a) is operably linked to at least four heterologous genes of a biosynthetic pathway of interest.

21. The engineered nucleic acid of paragraph 20, wherein at least four of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

22. The engineered nucleic acid of any one of paragraphs 1-21, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 5' to 3' direction.

23. The engineered nucleic acid of any one of paragraphs 1-21, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 3' to 5' direction.

24. A cell comprising the engineered construct of any one of paragraphs 1-23.

25. The cell of paragraph 24, wherein the cell is bacterial cell.

26. The cell of paragraph 25, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

27. The cell of paragraph 26, wherein the bacterial cell is a *Streptomyces* spp. cell.

28. The cell of paragraph 27, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

29. The cell of any one of paragraphs 24-28, wherein the cell comprises a deletion or modification in a gene of a biosynthetic pathway of interest.

30. The cell of paragraph 29, wherein the biosynthetic pathway of interest is a pikromycin biosynthetic pathway.

31. The cell of paragraph 30, wherein the cell comprises a deletion or modification in a *pikAI-V* , *pikBI-VIII*, *pikBR* gene, *pikC* gene or *pikD* gene.

32. The cell of paragraph 31, wherein the cell comprises a deletion in a *pikAII* gene.

33. The cell of any one of paragraphs 24-32, wherein the engineered nucleic acid is present on an episomal vector or integrated into a chromosome of the cell.

34. A cell comprising: (a) an engineered nucleic acid comprising a promoter operably linked to a gene encoding a shufflon recombinase; and (b) an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs.

35. The cell of paragraph 34, wherein the at least two heterologous genes encode proteins selected from the group consisting of: enzymes, regulatory proteins and transport proteins.

36. The cell of paragraph 34 or 35, wherein the shufflon recombinase is a *Salmonella enterica* shufflon recombinase.

37. The cell of any one of paragraph 34-36, wherein the gene encoding the shufflon recombinase is codon-optimized for expression in the cell.

38. The cell of any one of paragraphs 34-37, wherein the gene encoding the shufflon recombinase comprise a nucleotide sequence as set forth in SEQ ID NO: 35 or SEQ ID NO: 36.

39. The cell of any one of paragraphs 34-38, wherein the promoter of (a) and/or (b) is an inducible promoter.

40. The cell of any one of paragraphs 34-39, wherein the engineered nucleic acid construct of (a) and/or (b) further comprises an antibiotic resistance gene.

41. The cell of any one of paragraphs 34-40, wherein the RRSs recombine, in the presence of shufflon recombinase activity, at different rates relative to each other.

42. The cell of paragraph 41, wherein the RRSs comprises a nucleotide sequence selected from sequences set forth as SEQ ID NO: 5-8.

43. The cell of any one of paragraphs 34-42, wherein at least two of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

44. The cell of any one of paragraphs 34-43, wherein the promoter of (b) is operably linked to at least three heterologous genes of a biosynthetic pathway of interest.

45. The cell of paragraph 44, wherein at least three of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

46. The cell of paragraph 44 or 45, wherein the promoter of (b) is operably linked to at least four heterologous genes of a biosynthetic pathway of interest.

47. The cell of paragraph 46, wherein at least four of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

48. The cell of any one of paragraphs 34-47, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 5' to 3' direction.

49. The engineered nucleic acid of any one of paragraphs 34-47, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 3' to 5' direction.

50. The cell of paragraph any one of paragraphs 34-49, wherein the cell is bacterial cell.

51. The cell of paragraph 50, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp.,

Neisseria spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

52. The cell of paragraph 51, wherein the bacterial cell is a *Streptomyces* spp. cell.

53. The cell of paragraph 52, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

54. The cell of any one of paragraphs 34-53, wherein the cell comprises a deletion or modification in a gene of a biosynthetic pathway of interest.

55. The cell of paragraph 54, wherein the biosynthetic pathway of interest is a pikromycin biosynthetic pathway.

56. The cell of paragraph 55, wherein the cell comprises a deletion or modification in a *pikAI-V* , *pikBI-VIII*, *pikBR* gene, *pikC* gene or *pikD* gene.

57. The cell of paragraph 56, wherein the cell comprises a deletion in a *pikAII* gene.

58. The cell of any one of paragraphs 34-57, wherein the engineered nucleic acid of (a) and/or (b) is present on an episomal vector or integrated into a chromosome of the cell.

59. A cell-free composition, comprising

(a) an engineered nucleic acid comprising a promoter operably linked to a gene encoding a shufflon recombinase;

(b) an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs;

(c) a cell extract; and

(d) a polymerase.

60. The cell-free composition of paragraph 59, wherein the cell extract comprises at least one component selected from the group consisting of: ribosomes, amino acids, NTPs, phosphoenolpyruvate, pyruvate kinase, polyethylene glycol, ammonium acetate, potassium acetate and folinic acid.

61. The cell-free composition of paragraph 59 or 60 further comprising an engineered nucleic acid encoding additional proteins of the biosynthetic pathway of interest, or further comprising additional proteins of the biosynthetic pathway of interest.

62. The cell-free composition of any one of paragraphs 59-61, wherein the cell extract comprises a S30 cell fraction obtained from a bacterial cell.

63. The cell-free composition of paragraph 62, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

64. The cell-free composition of paragraph 63, wherein the bacterial cell is a *Escherichia* spp. cell or a *Streptomyces* spp. cell.

65. The cell-free composition of any one of paragraphs 59-64, wherein the polymerase is a RNA polymerase.

66. The cell-free composition of any one of paragraphs 59-64, wherein the RNA polymerase is a RNA T7 polymerase.

67. A method of producing molecules, comprising incubating the cell-free composition under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the biosynthetic pathway of interest.

68. A method of producing molecules, comprising:

culturing the cells of any one of paragraphs 24-58 under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the biosynthetic pathway of interest.

69. The method of paragraph 68 further comprising isolating the molecules.

70. The method of paragraph 69 or 70, wherein the molecules comprise antibiotics.

71. A cell, comprising (a) an engineered nucleic acid, comprising a promoter operably linked to at least one heterologous gene encoding an antimicrobial peptide or an antimicrobial protein located between a pair of shufflon recombinase recognition sites (RRSs), and (b) a nucleic acid comprising a gene encoding a shufflon recombinase.

72. The cell of paragraph 71, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein at least one heterologous gene encoding an antimicrobial peptide or an antimicrobial protein is oriented in a 3' to 5' direction.

73. The cell of paragraph 71 or 72, wherein the cell is a bacterial cell.

74. The cell of paragraph 73, wherein the bacterial cell is an *Escherichia* spp. cell.

75. The cell of paragraph 73, wherein the bacterial cell is an *Salmonella* spp. cell.

EXAMPLES

Example 1

Selection of a base pathway

Upon a thorough bibliographic search, the biosynthetic pathway assembling Pikromycin (Fig. 3) was selected for shuffling. In the basis of such decision were the following facts: it is a very well characterized pathway (39–42), the promiscuity of its enzymes has been studied and enables the generation of pikromycin-derived molecules (39–41), the pathway naturally assembles two different but related molecules (39, 42), and dedicated glycosyltransferases further transfer sugar moieties to the pikromycin aglycon (narbonolide) (39–42).

Selection of genes to shuffle

The final choice of genes to shuffle was made based on several considerations. First, the phylogenetic distance was analyzed. An increased relatedness can potentially result in an increased probability of enzymes from different pathways recognizing each other, thus participating in the assembly of the molecules instead of stopping the assembly line (43–46). Both more distantly and closer related species were used as gene donors. Gene size was also taken into account. The extremely repetitive and high GC polyketide biosynthetic genes can be a major barrier to gene synthesis and even PCR. To obtain such amplicons, to manipulate the resulting very large constructs *in vitro*, and to promote inversion of the sequences, genes of conservative size (6 - 12 kbp) were selected. The predicted bioactivity score was also examined. Looking to increase the likelihood of generating bioactive molecules, an algorithm-based prediction of molecule structure based on genetic information (NP.searcher) was used in combination with a bioactivity prediction algorithm (Molinspiration) for a more educated selection of genes to incorporate into the synthetic shufflon pathway (Table 1). Once a decision had been made regarding the best pathway to manipulate and potential genes

to use, the question was which of the pathway's elongation Pik enzymes should be replaced. Strain availability was also considered. Biosynthetic genes from organisms of unascertainable ID were discarded, having the attention been placed on genes from well-characterized pathways and organisms, available from any of the major organisms collection. The strains used in this project were obtained from the German collection, DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen).

Four genes were selected for shuffling. Using the aforementioned algorithms *piKAII* was selected over *piKAIII*.

Table 1. List of strains providing the genes to be shuffled

Species	Strain	Gene Name (ID)	Pathway	Weblink
S. venezuelae	DSM 41110	-	Pikromycin	biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=PWY-7421
S. fradiae	DSM 41456	*tylG* (AAB66508)	Tylactone	uniprot.org/uniprot/033958
S. cinnamonensis	DSM 1042	*momAVI* (AAO65801)	Monensis	uniprot.org/uniprot/Q846X1
		momAVIII (AAO65807)		uniprot.org/uniprot/Q846W5
S. nodosus	DSM 40109	*amphK* (AAK75303)	Amphotericin	Uniprot.org/uniprot/Q93NX7

In silico design of full construct

The full shufflon system (Fig. 4) used an array of parts, obtained following different strategies as described below.

Genes to be shuffled

Based on the predicted bioactivity it was concluded that the best gene to be replaced, within the pikromycin pathway, would be *piKAII*. It was decided as well to use *tylG*, *momVI*, *momVIII* and *amphK* as replacements for *pikAII* and as genes to be shuffled, as they have all the aforementioned properties. Furthermore, these genes, when considered individually as additions to the *pikAII* defective pathway, should display high values of bioactivity (Fig. 5).

Given the previous report regarding the shuffling frequency per recognition sequence (23), the recognition sequence "a" was assigned to *momAVIII*, "b" to *tylG*, "c" to *momAVI* and "d" to *amphK* (Fig. 5). It was expected to find a higher number of inversion events for *momAVIII* > *amphK* > *tylG* > *momAVIII*.

As ribosomal bidding sites (RBSs) for transcription of these genes it was decided

to use their native sequences (Table 2). Therefore, the DNA sequence between the upstream stop codon and the genes start codon was cloned in together with the gene sequence. As the genes utilized in this assembly are organized in operons, the resulting mRNAs are polycistronic. In addition, not being the first gene of the respective operons, these upstream sequences should not contain promoters, only RBSs. To further attest this, the typical *Streptomyces* spp. -10 (TAGGAT) and -35 (GGCT) promoter sequences were found not to be present.

Table 2. DNA sequences upstream initiation codon, which includes the respective RBSs (putative sequences underlined)

Gene	RBS-inclusive sequence upstream the gene	SEQ ID NO:
tylG	CCGGCCGGCGGAGCACACCCGGCCGTCTCCGGCCCGGCCGCGG CCGGGCCGGAAGCCATCCGCCGCCCACCCGGTACCGACCCCTC AAGCCCTTCAAGCCCTTCGACCCGTCCGATCAGTCAGTCCGGC GGTCCTCCACGACCGGTCCGGAATCGCCCCCACACGAGTCAGG AAGCACACC	1
momAVI	TCAGGAGTGGAGAAGAC	2
momAVIII	GCAGCGATCAGTCCTGAGCGCGATGAGCCCCGGCCACACCGGC CCCCTTGCCCTCCCCCAACCTCATACGCCCCGATCACCACGTAG CGCCAAGGAGCCTGGGTCAG	3
amphK	TGACCCGGGCCCACCCCAACTACGTGCGAAACATCTGAGGAAG GTTCAACGGACG	4

rci Gene, coding for the shufflon recombinase

The 1155 bp gene *rci* (NC_019104.1), coding for the shufflon recombinase and described for *Salmonella enterica* subsp. *enterica* serovar Kentucky, can be found in the organisms episomal DNA, pR64 (pCS0010A).

Given the disparity between *Salmonella* spp. and *Streptomyces* spp. in terms of GC content and codon usage, the *rci* gene was codon optimized for use in *Streptomyces* spp. by GenScript Corp using their proprietary algorithm. This company then synthesized the gene. The optimized sequence had a GC content of 67.79%, versus the 48.83% original one, a *Streptomyces*-friendly codon-usage. This gene was cloned upstream of the genes to be shuffled and to avoid any risk of polymerase read-through, it was cloned in the 3'-5' direction.

Inducible promoter driving rci

The best-studied inducible promoter system in *Streptomyces* spp. is tipAp, which is induced by thiostrepton. As the inducer is itself an antibiotic, the use of tipAp implies cloning a thiostrepton-resistance cassette, *tsr*.

As this promoter is driving the expression of *rci*, it was cloned upstream of it and also in the 3' - 5' direction, to prevent any read-through towards the genes to be shuffled.

Recognition sequences placed between the genes to be shuffled

After the discovery of shufflons in 1987 by Komano *et al.* (25), some work was carried out by a few different groups in order to better understand the logic behind the recognition and inversion of sequences (23, 24, 27).

It has been previously determined (26, 27) that different recognition sites are inverted at different rates, higher for sequence "a", then "d", "b" and lower for sequence "c" (Table 3).

Table 3. List of shufflon recognition sequences and gene each will be shuffling

Recognition sequence name	Recognition sequence (5'-3')	SEQ ID NO:	Gene to shuffle	Fraction of rearranged molecules (%)
a	GTGCCAATCCGGTACGTGG	5	*momAVIII*	29
b	GTGCCAATCCGGTACCTGG	6	*tylG*	0.025
c	GTGCCAATCCGGTCGGTGG	7	*momAVI*	0.02
d	GTGCCAATCCGGTACTTGG	8	*amphK*	0.1

Regions of homology to the chromosomal loci for integration

The complete shufflon system could either be maintained as episomal DNA or integrated into the organism's chromosome. Due to the size of the full construct and vector backbone, of ca. 45 kbp, and to maintain enzyme stoichiometry, it was decided to integrate the shufflon pathway into the *pikII* gene of the pikromycin pathway. This way, the gene to be replaced would also be knocked out the same time that alternative genes are provided in the pathway.

Integration is to be achieved by a Campbell-like double crossover event. The regions of homology incorporated in the construct correspond to the first and last ca. 500 bp of the gene *pikAII*, which is the gene to be replaced with those in the shufflon system. Homology towards those regions will guarantee that the *pikAII* gene will be disrupted and that the shufflon system will take its place.

Constitutive promoter at the end of the shufflon system, for the prevention of polar effects

Gene disruption, such as the *pikAII* disruption by integration of the shufflon system, can cause polar effects, in which the transcription of the whole operon is affected.

In order to prevent this, the constitutive ermEp* promoter was cloned in the shufflon system, upstream of the 3' region of homology. By cloning the promoter at the 3' end of the shufflon system, one expects to prevent possible polar effects.

Gene synthesis (GenScript Corp)

The expression of the *rci* gene was improved due to the high discrepancy in terms of codon usage between *Streptomyces* spp. and *Salmonella* spp. GenScript's proprietary algorithm was used to improve protein expression of the gene.

As in any method for DNA parts assembly, higher numbers of fragments lead to lower probabilities obtaining the right construct. Therefore, it was sought to compress the number of parts as much as possible.

With this in mind, and given the 5' location of the *rci* gene and promoter, it was decided to synthesize the 5' region of homology towards *pikAII* together with the *rci* system and the small region of homology towards the backbone of the vector (YAC, pYes-1L) and to the DNA part downstream of *rci* as well.

The *Streptomyces* spp.-inducible promoter tipAp was cloned in the gene *tsr* to confer the organism with resistance to the inducer molecule thiostrepton, which is an antibiotic. This gene was directly synthesized by GenScript Corp., as it already originated from a *Streptomyces* sp. Similarly to what is traditionally done, the region cloned exceeded that of the *tsR* ORF to ensure successful transcription. Similar to the approach taken for the 5' extremity of this construct, the 3' region of homology towards *pikAII* and pYES-1L, and homology to the DNA fragment upstream was also synthesized with the *tsR* operon.

The backbone carrying the synthesized parts was pUC57, containing the parts released by plasmid digestion with NsiI and PstI (the *rci*-inclusive part) or Bsu36I and BglII (the *tsR*-inclusive part).

The digestions were run in 1% agarose gels at 120 v, and the parts were purified from the gels.

gBlock synthesis (IDT DNA)

Short linear synthetic DNA sequences were synthesized by Integrated DNA Technologies, Inc. (IDT DNA). These sequences were designed to be stitching double-stranded DNA blocks, homologous to both an upstream and downstream amplicon. These stitching blocks are crucial for the yeast-based assembly of fragments that do not share

homology with the neighboring DNA fragments, as they provide homology for correct assembly.

The gBlocks designed and used were the following: gBlock1 (links *tylG* and *momAVI,* while adding the shufflon recognition sequences to *momAVI),* gBlock 2 (links *momAVI* and the YAC backbone), and gBlock 3 (links *amphK* and *momAVIII,* while adding the shufflon recognition sequences to *momAVIII).*

PCR

Genes *amphK*, *tylG*, *momAVI* and *momAVIII* were obtained by PCR from the genomic DNA preps of *S. nodosus*, *S. fradiae* and *S. cinnamonensis* (*mon* genes), respectively.

A vast array of PCR conditions was tested in order to find the amplification conditions that allowed for gene amplification. The conditions that eventually led to the successful amplification of all parts are listed in Table 4.

Initially, attempts were made to amplify *momAVI* in its entirety (single amplicon) and *momAVIII* in two amplicons. Yet, extensive processing did not allow for successful amplification of fragments with the expected size. A different approach was then undertaken. Following this approach, *momAVI* was subdivided into two amplicons (instead of one) and *momAVIII* into three (instead of two) (Fig. 8), resulting in successful amplification.

Upon experimentation, several conditions were found to result in highly specific and efficient amplification (Table 4). From the array of available options, the conditions used for PCR scale-up of all amplicons (Fig. 7) are described in Table 4.

Table 4. PCR conditions that allowed for successful amplification of genes and gene parts

Gene	Amplicons	Polymerase	Additives	Primer annealing T (C)	Primer extension T (C)
tylG	-	Herculase II	-	58	72
momAVI	1	Herculase II	D2 + M2	58	72
	2	Herculase II	D2 + M2	58	72
momAVIII	1	Herculase II	D1	58	72
	2	Herculase II	D2 + M2	60	72
	3	Herculase II	D1.5 + M1.5	58	72
amphK	-	KOD	D2.5	56	68

In order to obtain by PCR the several amplicons used for the shufflon system, PCR experimentation was performed. Information about the annealing loci (Fig. 9) and sequences of the primers (Table 5) used is provided.

Table 5. Primers used for the amplifications of DNA pieces.

	PCR Primers Sequence (5'-3')	SEQ ID NO:
YAC1		
1f	CCTGGCCGGCCGGCGGAGCACACCCGG	51
1r	TCACGGCCTCTCCTCTCTCC	52
2f	GGAGAAGACATGCCGGGTACG	53
2r	TCACCGCGTGCCCCACTC	54
2.5f	CGACGGTGATTTCAGGTCCG	55
YAC2		
1f	CCGAGCGCAGCGGCGGCCGCGCTGATACCGCCGCcaattgTGACCCGGGCCCACCCCA AC	56
1r	GGCCTCGGCGGTCAGATCC	57
2f	CCACGTAGCGCCAAGGAGCC	58
2r	CGTACAGCGTCTCCAGGTCC	59
3f	CCGACGACGAGCCGATCGCG	60
3r	CGTTGCCGTGTCCGTTGCCCTGC	61
2.3f	GTGCGACGCCTCATCGACG	62
2.3r	CCGACACGTCGCCGAGACC	63
2.6f	GCGAGCAGGATCCGTACGG	64
2.6r	GCTCACCAGTCCCTGTGAGC	65

Assembling of parts (synthesized and PCRed) using an yeast-based recombination method

Several methods for the assembly of multiple DNA parts are available. Some of the most commonly used include the following: Golden Gate (47), which relies on the use of restriction enzymes whose recognition sequence and cut site differ; isothermal assembly (48), aka Gibson assembly, which is a one-pot 50 C reaction of exonuclease and ligase activity, allowing for the generation sticky homologous ends and its ligation; and yeast-based assembly (49), which relies on the organism's extraordinary ability to recombine homologous DNA at high efficiency.

Yet, the nature of the gene sequences that code natural product biosynthetic enzymes greatly limit the methods one can use. Experience shows that the few enzymes used for Golden-Gate based assembly cannot usually be selected in such assemblies, as they would also cut within the parts. This problem does not exist with Gibson-based assembly, but the high part number, and the repetitive and high-GC nature of the genes results in incorrect assembly more often. Yet, these are issues are not experienced when using yeast-based assembly.

Thus, for the assembly of the many, very large and high GC shufflon system parts, it was decided to use the yeast-based system. More precisely, the GeneArt® High-Order seamless Genetic Assembly (Life Technologies, Inc).

As already described, once the colonies of yeast carrying the assemblies were grown, PCRs for confirmation (Fig. 8) of the correct assembly of parts was carried out. Though not a great deal of information was provided, due to their dirtiness and lack of some expected PCR products, there was enough data to select a few transformants for additional studies.

Three seemingly correctly assembled yeast colonies were selected per construct. The *E. coli* transformants carrying YAC1 and YAC2 were mini-prepped using the BAC Zymo Research kit. The YAC preps were then used as template for confirmation PCRs with Herculase II, KOD and Kapa Hifi polymerases. A selection of three polymerases was used as a way to guarantee that all fragments would be successfully amplified. It was previously seen that different DNA polymerases amplified different DNA fragments with different success rates. Also in this case, by combining the results obtained from the different PCRs, a nearly 100% success in PCR was reached. The single PCR that did not work was not due to missing DNA fragments, but to troublesome PCR, since both the up and downstream PCRs did work, which proves that the fragment is indeed there.

Table 6. List of primers used for screening of correct assemblies and sequencing

Confirmation and sequencing primers	Primer sequences Forward/Reverse (5'-3')	SEQ ID NO:
YAC 1		
P1	GGTAGCGTGCATCCTGTTAGCC/ GTCATGGACGGTCCTTCGAGACC	9 10
P2	CCGATCTGCTCTTCGACCGG/ TCACGGCCTCTCCTCTCTCC	11 12
P3	CGTTTGATGAGGTGTGCGGC/ GATGTCCGCCAGCCAGAAGC	13 14
P4	GGAGAAGACATGCCGGGTACG/ CGTACGTCGCGAGGAGTGCG	15 16
P5	GGACCTGGGCTTCGACTCGC/ GCACTTGCCGATCTATGCGG	17 18
YAC 2		
P1	CTCGCTGGTCTCGGTCAGCC/ GGCCTCGGCGGTCAGATCC	19 20
P2	GCGATCAGTCCTGAGCGCG/ CGTATCACCGCAAGGACCTTGTGG	21 22
P3	GTGCGACGCCTCATCGACG/ CGTACAGCGTCTCCAGGTCC	23 24
P4	CGACGTACGCCTTCCAGCG/ CGCACCGAACGTGTCCCA	25 26
P5	GGTCGAGGAGGTGATGCGCG/ CGTTGCCGTGTCCGTTGCCCTGC	27 28
P6	GCACTTGCCGATCTATGCGG/ GCACTTGCCGATCTATGCGG	29 30

Construction a reporting suicide vector; for shufflon integration and double crossover selection and confirmation of gusA reporting capabilities

One of the challenging aspects of this project consists on the successful integration of the shufflon system into the chromosome of *S. venezuelae*. The goal is to achieve a clean integration, with loss of YAC backbone.

Failed attempts to obtain in a timely manner the reporter vector pKGLP2 led to the in house construction of a similar but ideal system - pSETSC3 - described below.

Very few reporter systems have been successfully used in *Streptomyces* spp. The widely used *lacZ* system cannot be used as a reporter in these organisms, due to their production of ß-galactosidases. Most commonly, the *neo* and *cat* genes (neomycin and chloramphenicol resistance-conferring genes) are preferred, but many *Streptomyces* spp. are also naturally resistant to these compounds. The spore pigment-conferring gene *whiE*, is an easy and visual reporter, but useless for strains that produce already pigmented spores. Furthermore, problems have been found when trying to express *gfp* in *Streptomyces* spp., given the difference in codon usage (28). A more recently described reporter system, using the *gusA* reporter gene, was preferred. *gusA*, which codes for ß - glucuronidase can be used as an efficient and visible reporter, by overlaying the bacterial colonies with the enzymatic substrate X - Gluc (5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, cyclohexylammonium salt) (50).

Streptomyces spp. have a potent native DNA restriction machinery, which cleaves foreign methylated DNA being transformed into this organism. Spores and mycelia transformation yield, for many species, a very low number - if any - of transformants. The most efficient way of introducing DNA into *Streptomyces* spp. is by means of conjugation using methylation-deficient *E. coli* ET12567 (CamR) either with or without plasmid pUZ8002 (KanR), a conjugative plasmid that provides the *tra* machinery for the initiation of DNA transfer, without being itself transmitted to the recipient strain. pSet152-gusA does not contain the *tra* machinery, therefore it was electroporated into *E. coli* ET12567 pUZ8002.

Bioactivity and SMILES of predicted novel structures

Possible gene combinations and predicted molecule bioactivity scores

Upon integration of the shufflon system and induction of shuffling events, the semi-synthetic novel Pikromycin pathway (Fig, 10) consists of genes *pikAI, pikAIII, pikAIV* and the heterologous genes: (1) *momAVI, amphK, tylG, momAVIII,* (3) *momAVI,* (4) *momAVI,*

amphK, (5) *momAVI, amphK, tylG*, (6) *amphK, tylG, momAVIII*, (7) *tylG*, (8) *momAVIII*, (9) *amphK*, (10) *tylG, momAVIII*, (11) *momAVI, momAVIII*, (12) *tylG, amphK*, (13) *momAVI, tylG*, (14) *amphK, momAVIII*, (15) *momAVI, tylG, momAVIII*, or (16) *momAVI, amphK, momAVIII*.

The different combination of genes (Fig. 11), and thus enzymes, are expected to result in novel molecules. Initially, the bioactivity of novel pathways was only determined for those where PikAII would be replaced by a single new enzyme (Fig. 5). This was due to the goal being gene selection from a pool of available options (Table 7). In Table 8 the bioactivity score is shown for all possible pathway combinations.

Bioactivity score is estimated using a web-based predictor (www.molinspiration.com/cgi-bin/properties). Briefly, it compares the properties of the submitted molecules to a large database of molecules of known activity. The score varies between -2 and 2, being the most frequent value for bioactivity 0.50, according to information provided on the website. Yet, these values are merely indicative, as they rely on the prediction of a structure that might not be indeed assembled by the pathway.

Table 7. List of genes preselected as potential DNA to be shuffled

Species	Molecule	Gene
S. halstedii	Halstoctacosanolide	*hlsE*
S. halstedii	Vicenistatin	*vinP2*
S. avermitilis	Oligomycin	*olmA5*
S. avermitilis	Oligomycin	*olmA6*
S. avermitilis	Oligomycin	*olmA3*
S. cinnamonensis	Monensin	*momAVIII*
S. hygroscopicus	Rapamycin	*rapC*
S. hygroscopicus	Geldanamycin	*gdmAI*
S. hygroscopicus	Herbamycin	*hbmall*
S. halstedii	Vicenistatin	*vinP3*
S. avermitilis	Avermectin	*aveA2*
S. nodosus	Amphotericin	*amphK*
S. avermitilis	Oligomycin	*olmA4*
S. violaceusniger	Nigericin	*nigAIX*
S. violaceusniger	Nigericin	*nigAII*
S. violaceusniger	Nigericin	*nigAVIII*
S. coelicolor	Rifampicin	*rifC*
S. pristinaespiralis	Pristinamycin	*snbC*
S. cinnamonensis	Monensin	*momAI*
S. coelicolor	Rifampicin	*rifE*
S. hygroscopicus	Geldanamycin	*gdmAII*
S. cinnamonensis	Monensin	*momAVI*
S. nodosus	Amphotericin	*amphB*

Species	Molecule	Gene
S. fradiae	Tylactone	*tylG*

Table 8. Predicted bioactivity scores of predicted molecules

Bioactivity	**Original**	**1**	**2**	**3**	**4**
GPCR ligand	-0.66	ND	0.15	0.09	-0.13
Ion channel modulator	-1.57	ND	0.02	-0.34	-0.81
Kinase inhibitor	-1.45	ND	-0.31	-0.44	-0.81
Nuclear receptor ligand	-1.03	ND	33	0.11	-0.32
Protease inhibitor	-0.32	ND	0.18	0.22	0.03
Enzyme inhibitor	-0.92	ND	0.2	0.06	-0.32
	5	**6**	**7**	**8**	**9**
GPCR ligand	-0.87	-1.32	0.09	0.15	0.14
Ion channel modulator	-1.87	-2.36	-0.34	0.02	-0.2
Kinase inhibitor	-1.71	-2.14	-0.44	-0.31	-0.36
Nuclear receptor ligand	-1.32	-1.73	0.11	0.33	0.22
Protease inhibitor	-0.46	-0.75	0.22	0.18	0.17
Enzyme inhibitor	-1.16	-1.52	0.06	0.2	0.09
	10	**11**	**12**	**13**	**14**
GPCR ligand	-0.62	-0.62	-0.17	-0.38	-0.4
Ion channel modulator	-1.54	-1.54	-0.82	-1.15	-1.22
Kinase inhibitor	-1.39	-1.39	-0.8	-1.06	-1.1
Nuclear receptor ligand	-0.91	-0.91	-0.31	-0.61	-0.6
Protease inhibitor	-0.27	-0.27	0.06	-0.08	-0.15
Enzyme inhibitor	-0.82	-0.82	-0.31	-0.55	-0.61
	15	**16**			
GPCR ligand	ND	-1.19			
Ion channel modulator	ND	-2.35			
Kinase inhibitor	ND	-2.15			
Nuclear receptor ligand	ND	-1.74			
Protease inhibitor	ND	-0.78			
Enzyme inhibitor	ND	-1.53			

Table 9. Predicted pathway SMILES

Number	Predicted SMILES
1	C(C)C(O0)C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O0)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CCC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CCC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
2	C(C)C(O0)C(C)C=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)C(C)C(=O)C(C)C(=O)O
3	C(C)C(O0)C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O
4	C(C)C(O0)C(C)C=CC(O)CC=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CCC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CCC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC=CC(O)C(C)C(=O)C(C)C(=O)O
5	C(C)C(O0)C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CCC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CCC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC=CC(O)C(C)C(=O)C(C)C(=O)O
6	C(C)C(O0)C(C)C=CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O0)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O)CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CC=CC(C)C(O0)C(C)))C(C)C=CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CC=CC(C)C(O)C(C)))C(C)C=CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
7	C(C)C(O0)C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O
8	C(C)C(O0)C(C)C=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)C(C)C(=O)C(C)C(=O)O
9	C(C)C(O0)C(C)C=CC=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CC=CC(C)C(O0)C(C)))C(C)C=CC=CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C=CC=CC(C)C(O)C(C)))C(C)C=CC=CC(O)C(C)C(=O)C(C)C(=O)O
10	C(C)C(O0)C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
11	C(C)C(O0)C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O

Number	Predicted SMILES
	)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
12	C(C)C(O0)C(C)C=CC=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O)CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC=CC(C)C(O0)C(C)))C(C)C=CC=CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)C=CC=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C(=O)O
13	C(C)C(O0)C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(=O)C(C)C(=O)O
14	C(C)C(O0)C(C)C=CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CC=CC(C)C(O0)C(C)))C(C)C=CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CC=CC(C)C(O)C(C)))C(C)C=CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
15	C(C)C(O0)C(C)C=CC(O)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)CC(O)CC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC(O)CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O
16	C(C)C(O0)C(C)C=CC(O)CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O0)CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O0)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O)C(C)C=CC(O)CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O.C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CCC(O)C=CC(C)C(O0)C(C)))C(C)C=CC(O)CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C0(=O).C(C)C(O(C(=O)C(C)C(=O)C(C)C(O)C(C)C(O)C=CCC(O)C=CC(C)C(O)C(C)))C(C)C=CC(O)CC=CC(O)C(C)C(O)C(C)C(=O)C(C)C(=O)O

Materials and Methods

Microorganisms and growth conditions

All organisms used in this Example are listed on Table 10, where information about growth conditions is also provided. Table 11 lists the plasmids and yeast artificial chromosomes(YACs) used and built for this project.

Medium ISP2 (International Streptomyces Project, formulae #2) consisted of 4.0 g/L of yeast extract, 10.0 g/L of malt extract and 4.0 g/L of dextrose. ISP2 plates for general purposes were made by adding 15 g/L of agar (Apex), while ISP2 plates for spore production were made with 18 g/L of yeast-grade agar (Sunrise). When growing *Streptomyces* spp. in liquid, sterile coils were added to the culture flasks, to avoid clumping. Organisms were stored at - 80 C in 30% glycerol in water (yeast or spores) or LB (bacteria).

LB medium (Luria Bertani, Miller) consisted of 10.0 g/L of tryptone, 5.0 g/L of yeast extract, 10.0 g/L of sodium chloride and 15.0 g/L of agar (Apex).

The yeast medium CSM-Trp (Complete Supplement Mixture minus tryptophan) consisted of 1.7 g/L of yeast nitrogen base (MPBio), 5.0 g/L of ammonium sulfate,20.0 g/L of dextrose and 20 g/L yeast-grade agar (Sunrise).

Table 10. List of microorganisms and growth conditions

Species	Strain	Temperature (C)	Medium
S. venezuelae	DSM 41110	30	ISP2
S. fradiae	DSM 41456	30	ISP2
S. cinnamonensis	DSM 1042	30	ISP2
S. nodosus	DSM 40109	30	ISP2
E. coli	MG1655	37	LB
E. coli	Top10	37	LB
S. cerevisiae	MaV203	30	CSM-Trp
E. coli	ET12567	37	LB + Chloramphenicol 34 µg/Ml

Table 11. List of plasmids

Plasmid	Features	Resistance	Strain	Source
pUZ8002	DNA transfer machinery	Kanamycin	Replicative in *E. coli*	(28)
pYes1L	YAC	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	GENEART® kit (Life Technologies)
pSET152	Streptomyces integrative plasmid (*attB*)	Apramycin	Integrative in *Streptomyces* spp., replicative in *E. coli*	(29)
pSET-gusA	pSET152 carrying *gusA*	Apramycin	Replicative in *E. coli;* suicide vector in *Streptomyces* spp.	This study
pSETSC1	pSET152 without integration machinery	Apramycin	Replicative in *E. coli;* suicide vector in *Streptomyces* spp.	This study
pSETSC2	pSETSC1 carrying a *cos* site	Apramycin	Replicative in *E. coli;* suicide vector in *Streptomyces* spp.	This study
pSETSC3	pSETSC2 carrying the *gusA* reporter	Apramycin	Replicative in *E. coli;* suicide vector in *Streptomyces* spp.	This study
pSETSC-Shuf	pSETSC3 carrying insert 1 and 2	Apramycin	Replicative in *E. coli;* suicide vector in *Streptomyces* spp.	This study
pYES1L-insert1	pYES1L carrying insert 1	Apramycin	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-insert2	pYES1L carrying insert 2	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pUC57-*tsR*	pUC57 carrying the *tsR*-inclusive synthetic DNA	Ampicillin	Replicative in *E. coli*	This study
pUC57-*rci*	pUC57 carrying the *rci*-inclusive synthetic DNA	Ampicillin	Replicative in *E. coli*	This study

Selection of a base pathway

In order to select a pathway onto which add gene diversity and enzyme variation by shufflon-based conditional expression, the following aspects were taken into account. The available pathway characterization, as a(deep knowledge of the pathway will facilitate the selection of genes to shuffle and help predict molecule outcome. Known enzyme promiscuity was also considered, as it has been previously determined that some enzymes are more prone to accept substrates other than their usual, than other enzymes. Given the fact that the shufflon system will be made of genes from different pathways and organisms, higher enzyme promiscuity will increase the odds of successful novel molecule assembly (30, 31). The molecule(s) assembled by the pathway were also examined because if a single pathway is capable of assembling several related molecules, there is an increase probability of the shufflon system resulting in several novel molecules also being assembled (32). Additional molecule diversity can be generated by tailoring enzymes, which adds post-assembly modifications to the assembled molecules. Sugar or lipid moieties can be added (33, 34).

Selection of genes to shuffle

A preselection of genes to shuffle, listed in Table 7, was made from the pathway information available online (gate.smallsoft.co.kr:8008/pks/mapsidb/, www.bio.nite.go.jp/pks/top, and smart.embl-heidelberg.de/), taking into account the factors described below.

Phylogenetic distance was examined through evolutionary relatedness (carried out by performing protein homology searches using BLAST and against the entire NCBI database). The top 100 hits, corresponding to the 100 sequences in the database with the highest homology to the query sequence, were used for phylogenetic comparisons. Unrooted trees displaying phylogenetic distance between the sequences were generated using NCBI's BLAST Tree View Neighbor Joining algorithm. The Maximum Sequence Difference allowed was 0.85 using the evolutionary distance model according to Grishin.

Due to the technical limitations associated with very large genes (natural product genes can be as large as 25,000 bp), gene size was restricted.

Bioactivity was determined using the freely available algorithms that predict molecule structure from the genetic information
(NP.searcher (35)) and molecule function (Molinspiration (36)).

The selection of a pathway and its gene to be replaced influenced the choice of heterologous genes selected to be incorporated into the biosynthetic pathway and shuffled.

Reliable access to the strains was a factor in determining the genes to be used or replaced. Only strains from reputable collections were used.

In silico design of full construct

The full shufflon system construct uses the following parts: heterologous genes, which will be shuffled, *rci* gene, which codes for the recombinase (shufflon) that inverts DNA sequences, inducible promoter driving *rci, r*ecognition sequences, which are placed between the genes to be shuffled, regions of homology to the chromosomal loci, which will allow for integration of the shufflon system into a desired locus, constitutive promoter at the end of the shufflon system, which will prevent the occurrence of polar effects, and unique restriction sites are also used, in order to recover the shufflon system from the carrier YAC and subsequently clone it in the target vector.

The several parts used for the system were obtained using three strategies: gene synthesis (GenScript Corp), gBlock synthesis (IDTDNA), and PCR.

The amplification of high GC and long templates can be challenging (37, 38), using specialized polymerases and conditions. A set of different DNA polymerases, specifically engineered for the amplification of long and/or high GC templates, were used for successful amplification of the parts, including: Phusion® Polymerase NEB®, Kapa Hifi Biosystems®, KOD Xtreme High GC Long Template EMD Milipore®, KOD XL EMD Milipore®, Herculase II Hot Start Agilent®, AccuTaq LA DNA Polymerase Sigma Aldrich®, and TaKaRa LA Taq Clonetech®.

A vast array of conditions was tested, where different buffers, and additives at different concentrations and combinations were used (Tables 12 and 13). This was done in order to find the amplification conditions that allowed for gene amplification.

Table 12. PCR mix protocols (25µL reactions)

Reagent	Phusion	Kapa	Herculase II	Robust
Buffer* (µL)	5	5	5	5
Additive A** (µL)	0 - 3	0 - 3	0 - 3	0 - 3
Additive B** (µL)	0 - 3	0 - 3	0 - 3	0 - 3
dNTP* (µL)	1	1	1	0.5
Primer (forward) (µL)	0.5	0.5	0.5	0.5
Primer (reverse) (µL)	0.5	0.5	0.5	0.5
DNA (template) (µL)	0.5	0.5	0.5	0.5

Enzyme (µL)	0.2	1	0.5	0.2
H_2O (µL)	13.3 - 17.3	12.5 - 16.5	13 - 17	13.8 - 17.8

Reagent	KOD	XL KOD	AccuTaq	TaKaRa
Buffer* (µL)	12.5	2.5	2.5	12.5
Additive A** (µL)	0 - 3	0 - 3	0	0
Additive B** (µL)	0 - 3	0 - 3	0	0
dNTP* (µL)	5	2.5	2.5	4
Primer (forward) (µL)	0.5	0.5	0.5	0.5
Primer (reverse) (µL)	0.5	0.5	0.5	0.5
DNA (template) (µL)	0.5	0.5	0.5	0.5
Enzyme (µL)	0.5	0.5	0.25	0.25
H_2O (µL)	1.5 - 5.5	14 - 18	18.25	6.75

* The buffer and dNTPs volume varies according to the stock provided with the individual kits. The final concentrations are always 1x for the buffers, 0.3 mM for each dNTP, 0.3 µM per primer and 100 ng DNA/reaction.

** The additive was added, per reaction, in the following fashion:

Table 13. List of type and volume of additives used for PCR

Volume/25 (µL)	Additive Combination
0	**b, d** or **m**
1	**b, d** or **m**
1.5	**b, d, m, b + m** or **d + m**
2	**b, d** or **m**
2.5	**b, d, m, b + m** or **d + m**
3	**b, d** or **m**

b = 5M Betaine
d = 100% DMSO
m = 2 mM $MgCl_2$

For the DNA polymerases used, the PCR cycling protocols used are indicated in Table 14.

Table 14. PCR Cycling Protocols

Step	Cycles (#)	Phusion		Cycles (#)	Kapa	
		T (C)	t (mm:ss)		T (C)	t (mm:ss)
Initial denaturation	1	98	0:30	1	98	3:00

Denaturation	28	98	0:10	28	98	0:20
Annealing		58	0:30		60	0:20
Extension		70	4:00		72	4:00
Final extension	1	70	10:00	1	72	5:00

Step	Cycles (#)	KOD		Cycles (#)	XL KOD	
		T (C)	t (mm:ss)		T (C)	t (mm:ss)
Initial denaturation	1	94	2:00	1	94	2:00
Denaturation	28	98	0:10	28	94	0:30
Annealing		58	0:30		56	0:05
Extension		70	3:00		72	3:00
Final extension		-		1	72	10:00

Step	Cycles (#)	Herculase II		Cycles (#)	Robust	
		T (C)	t (mm:ss)		T (C)	t (mm:ss)
Initial denaturation	1	95	2:00	1	95	5:00
Denaturation	28	95	0:20	28	95	0:15
Annealing		56	0:20		58	0:15
Extension		72	4:00		72	1:00
Final extension	1	72	3:00		-	

The PCR reaction volume was 25 µL. Scaled-up reaction were 4 of 50 µL. Reactions were run in 1% agarose gels, at 120 V for 1 - 1.5 h and the 1 kb NEB DNA ladder was used has size marker.

Assembling of parts (synthesized and PCRed) using an yeast-based recombination method

The PCRed and synthesized gene fragments were assembled using the GeneArt® High-Order Genetic Assembly kit (Life Technologies, Inc.).

This assembly kit has been devised to assemble up to 10 DNA fragments of up to 110 kb. The higher the number of fragments, the lower the probability of the yeast assembling the correct construct. Taking this into account, the final 13-fragment construct (12 + backbone) was subdivided in two fragments, for a total of 7 + 7 fragments (twice 6 + backbone) (Fig. 1):

YAC 1 - Linearized pYes1L; *rci*-containing fragment (contains XbaI site); gBlock 1; gBlock 2; *tylG* and *momAVI* (gene split in 2 pieces; contains MfeI site); and

YAC 2 - Linearized pYes1L; gBlock 3 (contains MfeI site); *amphK*; *momAVIII*

(gene split in 3 pieces) and *tsr*-containing fragment (contains the Bsu36I site).

In brief, 100 ng of each linearized fragment was mixed together in a total volume of <10 µL. One hundred microliters of competent yeast cells were then added to this DNA mixture, to which a PEG/Lithium Acetate solution was added afterwards.

After a 30 min incubation at 30 °C, ß-mercaptoetanol was added to the cell-DNA mixture, which was then heat-shocked at 42 °C.

After replacement of the transformation solution with NaCl, the cells were plated on yeast medium (CSM-Trp) for selection of yeast cells carrying the tryptophan-producing vector (and insert) and incubated at 30 °C.

Upon 3 days of incubation at 30 C, PCR was performed on the yeast colonies to identify transformants with the correct assembly. Several pairs of primers were used, each targeting a recombination locus (junctions).

A first screening of the transformants was performed by lysing 8 yeast colonies per construct and performing PCR using the lysate as template DNA.

The lysing was performed by resuspending the yeast colonies in 20 mM NaOH and boiling the suspensions for 10 minutes at 99 °C in a thermocycler.

One microliter of each lysate was used as template in the 25 µL PCR reactions, with Kapa Robust as DNA polymerase and no additives.

Three yeast colonies per assembly were selected to be moved to *E. coli*, based on the PCR screening. They were lysed using a proprietary lysis buffer and beads included with the GeneArt® kit. Briefly, one microliter of the lysate was electroporated into *E. coli* Top10 competent cells provided with the kit, and plated on Luria Bertani (LB) medium supplemented with spectinomycin (50 µg/mL) for selection of transformants.

Construction reporting suicide vector, for shufflon integration and double crossover selection

A reporter vector was designed to carry the shufflon system and to allow for its integration into the host organism, while reporting on its efficiency.

Thus, the vector was designed to include an apramycin resistance cassette, an origin of replication that replicates in *E. coli* but not in *Streptomyces* spp. (ColE1), a codon-optimized reporter gene *gusA* and a *cos* site for vector stability with large inserts.

The desired plasmid was derived from pSET152, an integrative vector. Prior to cloning in the parts of interest, the C31 and integrase-coding regions were removed, as the

goal was to integrate the vector's insert in a specific locus, which is not the attB site of the organism. The resulting plasmid was named pSETSC1.

The integration machinery was removed from pSET152 by PCRing out the backbone parts of interest by PCRing around it. This PCR product, which contained the resistance cassette, the origin of replication, and the *lacZ* counter selection machinery, was ligated into the cos site fragment (PCRed from pYES1L) by isothermal assembly. In order for this to be achieved, the primers used in the amplification of both pieces contained regions of homology to each other, with an annealing temperature of 50 C. Furthermore, the *cos* site amplicon also contained a NcoI cloning site, (pSETSC2) into which the codon-optimized *gusA* was subsequently cloned (pSETSC3, Fig. 2).

At the same time that the reporter suicide vector was being built, another vector was designed to contain the reporter gene alone (pSET152-gusA), which was used to determine whether *gusA* is expressed in the target organism and reports in a visual manner and to assess whether it reports even if integrated into the chromosome (in single copy).

All restriction enzyme-based DNA digestion was done in 50 µL reactions at 37 C, following the product protocol. DNA dephosphorylation, when used, was performed using Alkaline Phosphatase Calf Intestinal (CIP).

Construction of a gusA*-based reporter system in* Streptomyces *spp.*

In order to test this, a DNA amplicon carrying the ermEp* promoter and *gusA*, from pSETSC3 (Fig. 2), was PCRed using primers that added XbaI restriction sites at both extremities of the amplicon. Upon digestion of the fragment and the pSET152 recipient vector, these pieces were ligated together and transformed into *E. coli* . The gene *gusA* and the promoter ermEp* driving it were PCRed from the suicide reporter vector. The restriction site XbaI was added to both terminus of the amplicon by PCR, which upon digestion was cloned into the digested and dephosphorylated backbone of pSET152 by CIP, by overnight ligation at 16 C (NEB T4 DNA Ligase), resulting in plasmid pSET-gusA.

All enzymes (restriction and ligation) used were purchased from NEB Inc. The isothermal assembly mix was prepared in house.

Confirmation of the correct ligation was performed by sequencing the construct (Genewiz Inc.).

Moving PSET152-gusA into S. venezuelae

Briefly, 10 mL of an exponentially growing culture of the methylation-deficient *E. coli* ET12567, carrying the nontransmissible plasmid pUZ8002 and the transmissible plasmid pSet152-gusA was pelleted and resuspended in 1 mL of LB and mixed 1:1 with 500 μL of synchronally-germinating spores (50 C, 10 min) in Yeast Tryptone (YT) medium.

References for Example 1

1. Nikaido H. 2009. Multidrug resistance in bacteria. Annu. Rev. Biochem. 78:119.
2. Bronzwaer S. 2003. European antimicrobial resistance surveillance as part of a Community strategy. Groningen Rijksuniv.
3. Knothe H, Shah P, Krcmery V, Antal M, Mitsuhashi S. 1983. Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of
Klebsiella pneumoniae and Serratia marcescens. Infection 11:315–317.
4. Alanis AJ. 2005. Resistance to antibiotics: are we in the post-antibiotic era? Arch. Med. Res. 36:697–705.
5. McGowan JE. 1983. Antimicrobial Resistance in Hospital Organisms and Its Relation to Antibiotic Use. Clin. Infect. Dis. 5:1033–1048.
6. Levy SB, Marshall B. 2004. Antibacterial resistance worldwide: causes, challenges and responses. Nat. Med. 10:S122–9.
7. Centers for Disease Control and Prevention. 2013. Antibiotic resistance threats.
8. Cantón R, Morosini M-I. 2011. Emergence and spread of antibiotic resistance following exposure to antibiotics. FEMS Microbiol. Rev. 35:977–91.
9. U.S. Food and Drug Administration. 2014. Antibacterial Drug Development Task Force.
10. The White House. 2014. National Strategy for Combating Antibiotic Resistant Bacteria.
11. The White House, Office of the Press Secretary. 2014. Executive Order – Combating Antibiotic-Resistant Bacteria.
12. Executive Office of the President, President's Council of Advisors on Science and Technology. 2014. Report to the President on Combating Antibiotic Resistance.
13. Bérdy J. 2012. Thoughts and facts about antibiotics: where we are now and where we are heading. J. Antibiot. (Tokyo). 65:385–95.
14. Bergmann S, Schümann J, Scherlach K, Lange C, Brakhage A a, Hertweck C. 2007. Genomics-driven discovery of PKS-NRPS hybrid metabolites from Aspergillus nidulans. Nat. Chem. Biol. 3:213–7.
15. Allsop a E. 1998. New antibiotic discovery, novel screens, novel targets and impact of microbial genomics. Curr. Opin. Microbiol. 1:530–4.
16. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat. Rev. Drug Discov. 6:29–40.
17. Silver LL. 2011. Challenges of antibacterial discovery. Clin. Microbiol. Rev. 24:71–109.
18. Donadio S, Maffioli S, Monciardini P, Sosio M, Jabes D. 2010. Antibiotic discovery in the twenty-first century: current trends and future perspectives. J. Antibiot. (Tokyo). 63:423–30.
19. Fischbach MA, Walsh CT. 2009. Antibiotics for Emerging Pathogens. Science (80-.). 325:1089.
20. Komano T. 1999. Multiple Inversion Systems and Integrons.
21. Esposito D, Scocca JJ. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. Nucleic Acids Res. 25:3605–3614.
22. Siuti P, Yazbek J, Lu TK. 2013. Synthetic circuits integrating logic and memory in living cells. Nat. Biotechnol. 31:448–52.
23. Gyohda A, Funayama N, Komano T. 1997. Analysis of DNA inversions in the shufflon of plasmid R64. J. Bacteriol. 179:1867–1871.
24. Gyohda A, Furuya N, Kogure N, Komano T. 2002. Sequence-specific and nonspecific binding of the Rci protein to the asymmetric recombination sites of the R64 shufflon. J. Mol. Biol. 318:975–83.
25. Komano T, Kubo A, Nisioka T. 1987. shufflon: multi-inversion of four contiguous DNA segments of plasmid R64 crestes seven different open reading frames. Nucleic Acids Res. 15.
26. Gyohda A, Zhu S, Furuya N, Komano T. 2006. Asymmetry of shufflon-specific recombination sites in plasmid R64 inhibits recombination between direct sfx sequences. J. Biol. Chem. 281:20772–9.
27. Gyohda A, Komano T. 2000. Purification and Characterization of the R64 Purification and Characterization of the R64 shufflon-Specific Recombinase 182.
28. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical Streptomyces genetics. John Innes Foundation.

29. Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE. 1992. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116:43–49.
30. Zimmermann M, Fischbach M a. 2010. A family of pyrazinone natural products from a conserved nonribosomal peptide synthetase in Staphylococcus aureus. Chem. Biol. 17:925–30.
31. Meier JL, Burkart MD. 2009. The chemical biology of modular biosynthetic enzymes. Chem. Soc. Rev. 38:2012–45.
32. Mootz HD, Schwarzer D, Marahiel M a. 2002. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. Chembiochem 3:490–504.
33. Khosla C, Keasling J. 2003. Metabolic engineering for drug discovery and development. Nat. Rev. Drug Discov. 2.
34. Staunton J, Wilkinson B. 1997. Biosynthesis of Erythromycin and Rapamycin. Chem. Rev. 97:2611–2630.
35. Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. 2011. NRPSpredictor2— a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39:W362–W367.
36. Molinspiration.
37. Jensen M a, Fukushima M, Davis RW. 2010. DMSO and betaine greatly improve amplification of GC-rich constructs in de novo synthesis. PLoS One 5:e11024.
38. Mamedov T, Pienaar E. 2008. A fundamental study of the PCR amplification of GC-rich DNA templates. ... Biol. Chem. 32:452–457.
39. Kittendorf J, Sherman D. 2009. The methymycin/pikromycin pathway: a model for metabolic diversity in natural product biosynthesis. Bioorg. Med. Chem. 17:2137–2146.
40. Hansen D a, Rath CM, Eisman EB, Narayan ARH, Kittendorf JD, Mortison JD, Yoon YJ, Sherman DH. 2013. Biocatalytic synthesis of pikromycin, methymycin, neomethymycin, novamethymycin, and ketomethymycin. J. Am. Chem. Soc. 135:11232–8.
41. Xue Y, Sherman DH. 2001. Biosynthesis and combinatorial biosynthesis of pikromycin-related macrolides in Streptomyces venezuelae. Metab. Eng. 3:15–26.
42. Xue Y, Zhao L. 1998. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. Proc. ... 95:12111–12116.
43. Christiansen G, Philmus B, Hemscheidt T, Kurmayer R. 2011. Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity. J. Bacteriol. 193:3822–31.
44. Jenke-Kodama H, Dittmann E. 2009. Bioinformatic perspectives on NRPS/PKS megasynthases: advances and challenges. Nat. Prod. Rep. 26:874–83.
45. Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH. 2007. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol. Biol. 7:78.
46. Villiers B, Hollfelder F. 2011. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. Chem. Biol. 18:1290–9.
47. Engler C, Gruetzner R, Kandzia R, Marillonnet S. 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS One 4:e5553.
48. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth 6:343–345.
49. Struhl K, Stinchcomb DT, Scherer S, Davis RW. 1979. High-frequency transformation of yeast: autonomous replication of hybrid DNA molecules. Proc. Natl. Acad. Sci. 76:1035–1039.
50. Myronovskyi M, Welle E, Fedorenko V, Luzhetskyy A. 2011. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. Appl. Environ. Microbiol.77:5370–83.

Example 2

The experiments described below were conducted to develop and test the tools used in Example 1.

Testing the reporter gusA *in* S. venezuelae

The reporting capability of *gusA* in *Streptomyces* spp. was tested in *S. venezuelae*. The sequence of *gusA* (ID 946149) was obtained from NCBI. Since the original host is *E. coli*, it was codon optimized and synthesized by Genscript Corp. In order to test its activity in *S. venezuelae*, the optimized *gusA* was cloned into the integrative plasmid pSET152 via Isothermal Assembly [2]. The purified DNA was then electroporated into *E. coli*

MG1655 and plated on apramycin-supplemented LB (50 μg/mL). The colonies were screened for proper assembly and the selected clones were tested for β-glucuronidase activity. This testing was not used as a method for colony selection because the overlaying process leads to cell mixing.

Positive clones were re-streaked onto LB-Apramycin, incubated for 24 hours and overlaid with 1 mL of 0.5 mg/mL of 5-bromo-4-chloro-3-indolyl-beta-D-glucuronic acid, cyclohexylammonium sal (X-Gluc). The plates were checked for substrate breakdown into a blue pigment after 3 hours of incubation at room temperature. A strong blue/green color developed within this time frame. The hydrophobicity of the spore layer hinders the absorption of the solution, having the phenotype been stronger around individual colonies than the busiest sections of the plate.

It was thus confirmed that in *S. venezuelae* the codon-optimized *gusA* is an efficient reporter.

Recovery of inserts from YAC-1 and YAC-2, and assembly of full construct

In order to assemble the complete shufflon construct and move it to its dedicated integrative vector, 2 inserts were recovered from the YAC backbone . The recovery of the two inserts was performed by restriction digest of the two YACs. YAC 1 was digested with the restriction enzymes PmeI and XbaI, and YAC2 with PmeI and Bsu36I. The digestion step was allowed to occur for 10 hours at 37 C,
and following the product protocol. The digestions were gel-purified (0.7% agarose).

The pSETSC3 vector digested with Bsu36I and XbaI, and the two inserts, digested as aforementioned (Fig. 12), were ligated overnight at 16 C using NEB T4 DNA Ligase and according to the product specifications. The ligation reaction was subsequently purified and electroporated into *E. coli* MG1655.

A subset of the transformants obtained was midiprepped and PCRed using set of primers for the ligated loci (Fig 13). Initial attempts at performing colony-PCR for confirmation of proper assembly were unsuccessful.

Upon confirmation of correct assembly using the midiprepped DNA, two constructs were selected and cryopreserved.

The 2-part shufflon system was initially made up of several smaller parts that were assembled in yeast. The two parts were designed to include restriction sites that enable an easy recovery of the insert and removal of the Yeast Artificial Chromosome (YAC) backbone. Furthermore, they were designed to share a restriction site for assembling the 2-part shufflon into a complete one.

By digestion, gel purification, and 3-way ligation of the parts to the pSETSC3 backbone, it was possible to fully construct, *in vitro,* the pathway designed *in silico.*

Three-way ligations, particularly of very large inserts, are not particularly efficient. By using the information obtained from PCRs aimed at amplifying the ligated loci, it was possible to select isolates with the proper assembly (Fig. 13). As previously experienced, only the combined use of information from PCRs using a variety of polymerases allowed for such selection. Again, this was due to the very high GC content of the construct and the high repetitiveness of the coding DNA.

Rebuilding the shufflon system

Removing rci from the shufflon system implied, nonetheless, rebuilding it. Given that the assembly in yeast was performed in two independent YACs (1 and 2; see Example 1), only one was altered: the one containing *rci.*

Furthermore, given that all pieces of YAC1 would remain the same except for one, only one new piece was obtained (Fig. 15). This piece would connect the regions up and downstream of the *rci* locus (which previously connected them).

A new step of yeast-based assembly was performed and several yeast colonies were checked for successful assembly. Upon confirmation of a correct construct, named YAC3, it was moved to *E. coli* Top10, extracted, digested and ligated into pSETSC3 as already explained for the original pathway, resulting in pSETSC3-shufflon_b. As before, the colonies were checked for proper assembly, and correct 3-way ligations were obtained and stored at -80C.

Moving rci *to a temperature-sensitive replicative plasmid*

To address the concern that the tipAp promoter would be leaky in *S. venezuelae*, which would not allow for complete stop or control of Rci-mediated shuffling, it was decided to move this gene out of pSETSC3 and onto an independent plasmid.

This plasmid was derived from pKC1139 3, a replicative temperature-sensitive plasmid. The *rci* gene driven by the tipAp promoter was cloned in the restriction sites HindIII and BamHI, which were added to the insert by means of PCR (Fig. 16). Given that the selection marker for this plasmid was apramycin, which is also present in plasmid carrying the shufflon system and will only be removed if a double crossover event will be achieved, it was also decided to replace it with a hygromycin resistance cassette.

The hygromycin resistance cassette was amplified from plasmid pIJ10700 by PCR and cloned into restriction site xhoI, located in the middle of the gene coding for the apramycin resistance.

Testing rci-*based shuffling*

To address concerns that *rci* might not work in *S. venezuelae*, it was decided to pause the incursion into the use of *rci* as a tool to generate chemotype diversity. A new approach to test *rci*-mediated shuffling was then devised.

In this approach, four different genes coding for fluorescent proteins were chosen for the screening of the shuffling events and their frequency, in *S. venezuelae*. Given that the shufflon pathway of interest – the one carrying genes coding for polyketide synthases – is made of very large genes of very high GC content, the same type and length of DNA was used in between recognition sites.

Therefore, four different constructs were designed, with different inter-recognition sites distances: 1, 5, 10 and 15 kbp of high GC DNA. This inter-site DNA was PCR - amplified from *Streptomyces coelicolor* A3(2), a species of Streptomyces close enough to share many of the cumbersome aspects of the species from this genus (the very high GC content results in very tight DNA and potential inaccessibility of enzymes to the nucleotides). Yet, this strain is also not close enough to share high homology that could result in the integration of the construct through the DNA amplicons.

Using the NCBI Genome Graphic View, the genome of *S. coelicolor* A3(2) was scanned and 5, 10 and 15 kbp of regions not coding for antibiotics of genes that could potential affect the host phenotype were selected (Table 15).

These regions were then blasted against the whole chromosome of *S. venezuelae*, to assess the degree of similarity (Table 15), which was lower than 82% in up to 47% coverage of the sequences (i.e., very low).

Table 15. Length, coordinates and similarity of *S. coelicolor* DNA fragments to be flipped by Rci with the 4-color system

Piece	Length (kbp)	Locus (chromosome coordinates)	Coverage %	Similarity %
A	5	566133..571711	31	80
	10	565091..575044	24	80
	15	77298..93961	3	73
B	5	682108..687070	NA	NA
	10	676426..686815	5	73

Piece	Length (kbp)	Locus (chromosome coordinates)	Coverage %	Similarity %
	15	581427..598131	14	76
C	5	8319286..8324450	26	82
	10	8318514..8329225	40	82
	15	8317786..8333304	25	82
D	5	1295599..1300243	29	82
	10	1289613..1300079	28	82
	15	1284740..1299845	47	82

Coverage % - Stretch of DNA with some degree of similarity; Similarity% - How similar the covered stretches of DNA are; NA - No significant similarity found

Selection of fluorescent proteins

To this date, only green and red fluorescent proteins (FPs) have been developed or shown to be functional in Streptomyces spp [4,5]. Four genes coding for fluorescent proteins used in other prokaryotes were codon optimized and synthesized by Genscript Corp. The selection of these four FPs took into account their excitation and emission spectra, so that minimal overlap occurs (Table 16).

Table 16. Excitation and emission wavelengths for the selected FPs

Protein	Excitation (nm)	Emission (nm)
dsRed2	563	582
nirFP	605	670
ebfp2	380	448
mKate2	588	633

Testing the FPs in S. venezuelae

Upon codon optimization and synthesis of the four genes coding for the FPs, it was taken into account that optimization alone might not be the single barrier to the expression of these genes by *S. venezuelae*.

Therefore, it was decided to test these independently, before testing the 4-FP shufflon systems.

A constitutive expression vector, based on pSET152, was then designed and built (Fig. 18). The constitutive promoter ermEp* was cloned by Isothermal Assembly, having the plasmid backbone been amplified by PCR so as to remove the lacZ system. Thinking of maximizing its usefulness to the *Streptomyces* community of researchers, a multiple cloning

site (MCS) was cloned downstream of ermEp*. This MCS was designed to include 12 restriction sites with lower GC content, so as to increase their usefulness with high GC sequences by increasing the probability of uniqueness of restriction sites.

The terminator sequence from phage T4 was cloned downstream of the MCS and the final expression vector was designated pSETSCexp.

All for genes coding for the FPs have been successfully cloned into the MCS of pSETSCexp, in the sites BglII and SpeI, as confirmed by PCR (Fig. 19).

These constructs were subsequently moved to E.coli ET12456 pUZ8002 and conjugated into *S. venezuelae*, as described below.

Nonetheless, upon multiples attempt, no ex-conjugants were obtained. In order to troubleshoot this, different lots of mannitol- soya medium were used. The plates were prepared with soya flour or powder, and 10 or 20 mM of $MgCl_2$ was used. The *E. coli* cells carrying the constructs were also grown to initial, mid or late exponential phase, for conjugation.

Upon 8 attempts, no ex-conjugants were obtained, whereas the plasmid pSET-GusA used as control was very successfully integrated. It is believed that the levels of expression of the FPs, driven by ermEp*, were too high and thus toxic to *S. venezuelae* (Fig. 30).

Therefore, it was decided to rebuild these plasmids where the promoter driving expression is weaker. Promoter C4 6 was selected and ordered as a gBlock (IDTDNA) exactly as done for the pSETSCexp vector. The approach to cloning of this gBlock and genes coding for the several FPs was the exact same as for pSETSCexp. This new low-expression vector was designated pSETSCexp-low and or pSETSCexp-lowX, depending on the gene cloned in.

Obtaining the 5, 10, and 15 kb amplicons

Upon selection of *S. coelicolor* and the loci for amplification, PCR was performed. Given the knowledge gathered from the previous high GC PCR experiments, the first polymerase used was KOD Hot Start Polymerase, without any additives. This allowed for the amplification of three of the four 5 kbp amplicons, and one of the 10 kbp. None of the 15 kbps amplicons were obtained.

Following this, Herculase II with and without additives, KOD with 8% DMSO and $MgCl_2$, as well as Kapa Hifi were tested. None of the PCR reactions were successful. The PCR reactions were then further modified, as 8% betaine or DMSO, and/or 8% $MgCl_2$, were

used as additives for KOD-based PCRs. The remaining 5 kbp amplicon was obtained when using 8% DMSO, but no other amplicons were obtained.

Next, the same protocol was used for Herculase II, which resulted in the successful amplification of an additional 10 kbp amplicon, but no additional amplicons.

At this point, it was decided to follow a different strategy. Given that there were no real requirements in terms of amplicon boundaries, new primers were designed to the same overall locus, but with primers annealing slightly further up or downstream of the initial sites.

Following this approach the remaining 5 and 10 kb amplicons were successfully obtained. Nonetheless, only a single 15 kb piece was amplified.

Another new approach is in the process of being implemented, where the 15 kb pieces of *S. coelicolor* DNA are obtained by digestion of cosmids carrying its complete genome. The pieces will then be obtained by restriction digest of the cosmids. The homology required for yeast assembly will be provided by overhangs added by PCR to the 4 genes coding for the FPs.

Construction of a YAC-E. coli-Streptomyces integrative shuttle vector

For the 4-FPs shufflon pathway, the ultimate goal was to integrate it into the chromosome of *S. venezuelae*. This would enable a comparable circumstance to when integrating the antimicrobial shufflon pathway. Given that the 4-FPs pathway was to be assembled in yeast, it made sense to build a YAC that would also enable the integration of the pathway into *S. venezuelae*. As such, the near totality of pSET152-gusA plasmid was PCR-amplified. Its *E. coli* oriR was its only portion not amplified, as it pYES-1L already contained one. The primers also added overhangs to the amplicon, for recombination with pYES-1L in yeast. A gBlock containing ermEp* and further linking the two vectors were also used (Fig. 21).

Upon assembly in yeast the construct was colony-PCRed, for confirming proper assembly, and moved to *E.coli*, as previously described. This vector can be linearized by digestion with the restriction enzyme PmeI.

Building a vector carrying rci and tsR

Previously, a temperature sensitive vector carrying *rci* – pKCSC6 - was assembled. tsR was not cloned in this same vector. Nonetheless, the new shufflon 4-FPs systems did not carry either of these genes. As such, a new version of pKC1139 was designed to include both

rci and *tsR*. The selection marker was switched from apramycin to hygromycin. In the case of pKCSC5 this was mandatory, as pYes-SC4 is a C31/attP integrative plasmid carrying that same marker.

rci and *tsR* were cloned in the HindIII and XbaI unique restriction sites of pKC1139, in a three-way ligation where both inserts shared the KpnI restriction site. The hygromycin selection marker was cloned into the XhoI site, disrupting the Apramycin resistance cassette.

Assessing the presence of Rci recognition sites in the chromosome of S. venezuelae

Another concern was that, given the small size of the Rci recognition sites (20 bp), the *S. venezuelae* chromosome might contain regions that Rci could interpret as being recognition sites.

Using NCBI Blast, its whole genome was scanned for such sequences. No full homology was found The closest homology was 12 bp without any mismatches, and there was a single case of 18 bp with one mismatch, which corresponded to a siderophore biosynthetic gene (Fig. 22).

Nonetheless, this was performed for *S. venezuelae* ATCC 10712, not ATCC 15439 (DSM 41110) - the target organism. Only ATCC 10712 has been fully sequenced.

Testing the DNA flipping abilities or Rci, in E. coli

Two independent constructs were inoculated overnight without thiostrepton (50 μg/mL), the inducer. The cultures were then reinoculated with and without thiostrepton until the exponential phase was reached. At this point they were reinoculated with and without thiostrepton once again, and grown overnight. Twenty-five milliliters of culture were then midiprepped and the DNA was used as template for qPCR (Kapa SYBR Fast qPCR Master Mix) and regular PCR (Herculase II).

Primers were designed to anneal to the 3' or 5' of each gene, having all possible combinations of primer sets been used (Fig. 23). Primers were also designed to the middle of each gene to be flipped, to serve as reference for gene abundance. The amplicons were 100-150 bp long.

The flipping observed is listed in Fig. 23. It was particularly high for the recognition sequence b, which flipped ca. 6 kbp, and lowest for the recognition sequence c, which flipped ca. 5.5 kbp (Fig. 23). This differed from that reported previously [7], tested in the original shufflon system (plasmid R64) with ca. 250 bp between recognition sequences. Reports of flipping frequencies vary among studies [7–10].

Materials and methods

Table 17. List of strains used in this study

Species	Strain	Temperature (C)	Medium	Source
S. venezuelae	DSM 41110	30	ISP2	DSMZ
S. coelicolor	A3(2)	30	ISP2	USDA
E. coli	MG1655	37	LB	Lab collection
E. coli	ET12567	37	LB + Chloramphenicol 34 μg/mL	Lab collection
E. coli	Top10	37	LB	Geneart™

Table 18. List of plasmids sourced from other labs, used in this study

Plasmid	Features	Selection	Strain	Source
pUZ8002	DNA transfer machinery	Kanamycin	Replicative in *E. coli*	14
pSET152	Streptomyces integrative plasmid (*attB*)	Apramycin	Integrative in *Streptomyces* spp., replicative in *E. coli*	3
pKC1139	Shuttle vector with temperature-sensitive Streptomyces oriR	Apramycin	Replicative in *E. coli* and *Streptomyces spp.*	3
pIJ10700		Hygromycin	Replicative in *E. coli*	S. O'Rourke (unpublished)

Table 19. List of plasmids built for this study

Plasmid	Features	Selection	Strain	Source
pSET-gusA	pSET152 carrying *gusA*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pSETSC3	pSETSC2 carrying the *gusA* reporter	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pSETSC3-Shufflon	pSETSC3 carrying insert 1 and 2	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pYES1L-insert1	pYES1L carrying insert 1	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-insert2	pYES1L carrying insert 2	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-insert3	pYES1L carrying insert 3	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES-SC4	pYES1L carrying part of pSET-gusA	Apramycin Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*, integrative in *Streptomyces spp.*	This study
pKC1139-rci	pKC1139 carrying *rci*	Apramycin	Replicative in *E. coli* and *Streptomyces*	This study

			spp.	
pKC1139-rci-tsr	pKC1139 carrying *rci* and *tsR*	Apramycin	Replicative in *E. coli* and *Streptomyces spp.*	This study

Plasmid	Features	Selection	Strain	Source
pKC-SC5	pKC1139 carrying *rci* and *tsr*	Hygromycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pKC-SC6	pKC1139 carrying *rci*	Hygromycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pSETSCexp	pSET152 carrying the ermEp* promoter, a MCS and terminator	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-mKate2	Expresses *mKate2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-nirFP	Expresses *nirFP*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-dsRed2	Expresses *dsRed2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-ebfp2	Expresses *ebfp2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pYES-SC4-1	1 kbp – 4 FP system	Apramycin Spectinomycin Tryptophan	Replicative in *E. coli* and yeast; integrative in *Streptomyces* spp.	This study
pYES-SC4-5	5 kbp – 4 FP system	Apramycin Spectinomycin Tryptophan	Replicative in *E. coli* and yeast; integrative in *Streptomyces* spp.	This study
pYES-SC4-10	10 kbp – 4 FP system	Apramycin Spectinomycin Tryptophan	Replicative in *E. coli* and yeast; integrative in *Streptomyces* spp.	This study

Media Recipes

MS agar was created from 20 g/L agar, 20 g/L mannitol, 20 g/L soya flour, and 1L tap water. It was then autoclaved twice for 15 minutes at 115 C.

2xYT broth was created from 16 g/L tryptone, 1- g/L yeast extract, 5 g/L NaCl, and 1 L deionized water, followed by autoclaving for 20 minutes at 121 C.

References for Example 2

1. Myronovskyi, M., Welle, E., Fedorenko, V. & Luzhetskyy, A. Beta-glucuronidase as a sensitive and versatile reporter in actinomycetes. Appl. Environ. Microbiol. 77, 5370–83 (2011).

2. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth 6, 343–345 (2009).
3. Bierman, M. et al. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116, 43–49 (1992).
4. Sun, J., Kelemen, G. H., Fernández-Abalos, J. M. & Bibb, M. J. Green fluorescent protein as a reporter for spatial and temporal gene expression in Streptomyces coelicolor A3(2). Microbiology 145 (Pt 9, 2221–7 (1999).
5. Nguyen, K. D., Au-Young, S. H. & Nodwell, J. R. Monomeric red fluorescent protein as a reporter for macromolecular localization in Streptomyces coelicolor. Plasmid 58, 167–173 (2007).
6. Siegl, T., Tokovenko, B., Myronovskyi, M. & Luzhetskyy, A. Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. Metab. Eng. 19, 98–106 (2013).
7. Gyohda, A., Funayama, N. & Komano, T. Analysis of DNA inversions in the shufflon of plasmid R64. J. Bacteriol. 179, 1867–1871 (1997).
8. Gyohda, A., Zhu, S., Furuya, N. & Komano, T. Asymmetry of shufflon-specific recombination sites in plasmid R64 inhibits recombination between direct sfx sequences. J. Biol. Chem. 281, 20772–9 (2006).
9. Gyohda, A. & Komano, T. Purification and Characterization of the R64 Purification and Characterization of the R64 shufflon-Specific Recombinase. 182, (2000).
10. Gyohda, A., Furuya, N., Kogure, N. & Komano, T. Sequence-specific and non-specific binding of the Rci protein to the asymmetric recombination sites of the R64 shufflon. J. Mol. Biol. 318, 975–83 (2002).
11. Warth, L., Haug, I. & Altenbuchner, J. Characterization of the tyrosine recombinase MrpA encoded by the Streptomyces coelicolor A3(2) plasmid SCP2*. Arch. Microbiol. 193, 187–200 (2011).
12. Fedoryshyn, M., Welle, E., Bechthold, A. & Luzhetskyy, A. Functional expression of the Cre recombinase in actinomycetes. Appl. Microbiol. Biotechnol. 78, 1065–1070 (2008).
13. Herrmann, S. et al. Site-specific recombination strategies for engineering actinomycete. Appl. Environ. Microbiol. 78, 1804–1812 (2012).
14. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical streptomyces genetics. (John Innes Foundation, 2000).

Example 3

Testing the PFs in S. venezuelae

The selection and cloning of the new, codon-optimized genes encoding fluorescent proteins (FPs) was explained in Example 2. It was also explained that no exconjugants were obtained after over 8 attempts of conjugating the pSETSCexp plasmids carrying each of the FPs into *S. venezuelae*.

Possible causes for this phenotype include: toxicity associated with the overexpression of the FPs, mutation in the resistance cassette, which impedes selection, and mutation in the *int* gene, resulting in nonfunctional recombinase for integration of pSETSCexp into the *attB* site.

Initially, it was decided to build a lower-expression vector, but upon further Discussion, the pSETSCexp plasmids were rebuilt. Each FP, preceded by the ermEp* promoter and an RBS and followed by a terminator, was PCR-amplified.
The plasmid backbone was obtained by PCR, for assembly with each FP amplicon by Isothermal Assembly [1]. Upon assembly, the mixture was transformed into *E.coli* DHSα.

Upon confirmation of proper assembly, transformation into *E.coli* ET12456 pUZ8002 and conjugation into *S. venezuelae*, exconjugants were successfully obtained. It was thus concluded that the previous failure to obtain exconjugants resided on the integration machinery within the plasmid.

Each of the strains carrying individual FPs was grown in Tryptic Soy Broth supplemented with apramycin and incubated at 30C and for a period of up to 96 hours. To prevent clumping, a sterile stainless steel spring (Ace Glass Incorporated, USA) was added to each culture.

Aliquots of all cultures were taken after overnight growth and at 24, 72 and 96 hours, for analysis by Fluorescence Assisted Cell Sorting (FACS). The aliquots were diluted 1:10 (overnight and 24 hours) or 1:100 (72 and 96 hours) in Phosphate Buffered Saline (pH=7) and ran in a BDFACS LSR Fortessa cell analyzer (BO Biosciences, CA). For each sample, 50,000 cells were analyzed and gated using a forward scatter and side scatter.

After overnight and 24 hours of growth, no fluorescence was observed. The cultures were allowed to grow for a total period of 4 days. Aliquots were taken daily and analyzed through FACS. New and daily re-inoculated cultures were also ran, and a slight part of the population started fluorescing after the third day of incubation. Nonetheless, this represented less than 0.5 % of the total population, which for this purpose does not suffice.

E. coli's transcriptional machinery recognizes the ermEp* promoter used to drive the expression of the genes encoding the fluorescent proteins. As no fluorescence was observed when the genes were expressed in *S. venezuelae,* it was decided to check whether they were in *E. coli.* Again, no any fluorescence was detected by FACS. This suggests that the problem might reside in the genes themselves, codon-optimized for *S. venezuelae* but also expressible in *E. coli.* The issue could potentially be misfolding.

To assess the flipping abilities of Rci, it is imperative that, upon flipping, the transcribed fluorescent proteins glow, otherwise the flipping events will not be registered or recognized. As such, it was decided to determine the flipping events by qPCR.

In order to detect the flipping events by qPCR, several primers were designed, as illustrated (Fig. 24). The design of the primers enabled the use of a single set of primers for each of the shuffling regions, independent of the size of the sequence to be shuffled (1, 5 or10 kbp). Nonetheless, several sets of primers were designed and used to properly calculate the inversion events by qPCR. These sets included primers to standardize for DNA amount and DNA replication efficiency.

Assembly of the 5 and the 10 kbp amplicons

For the purpose of testing the flipping efficiency of Rci, it was decided to proceed only with the 1, 5 and 10 kbp inter-recognition sites assemblies (Fig. 25). On the basis of this decision lies the fact that, in the antibiotic-shufflon system, only one gene is over 6 kbp, and the technical impracticability of manipulating such large constructs - the total size would be above 75 kbp - and the conjugation issues arising thereof.

Initially, the 5 and 10 kbp 4-FP constructs - of 9 pieces each - were tentatively assembled in yeast. Upon two failed attempts, it was decided to subdivide the process into 2 steps, which would reduce the number of pieces to assemble per reaction to 5 from 9.

Facing failed attempts again, the suspicion of the failed assembly fell on the YAC-*E. coli-Streptomyces* integrative shuttle vector built for this purpose. A possible explanation is its low stability in yeast due to the great capacity of this organism to recombine DNA.

Thus, it was decided to assemble these pathways in the traditional pYES-lL YAC, in one step and in two steps. Three new amplicons were obtained due to homology to the YAC backbone at the 5' and 3': 5 kbp amplicon A, 10 kbp amplicon A, dsRed2 amplicon (now the 3' of the 2-step insert).

These amplicons were obtained via PCR with KOD (5 and 10 kbp pieces) or Hifi (dsRed2) polymerase without additives. These pieces were then combined in the following mixtures: pYes-lL, 5 or 10 kbp *A* amplicon, 5 or 10 kbp B amplicon, *ebfp2,* and *dsRed2* (with homology to pYes-lL).

Transformants were recovered and the YACs were moved into *E. coli.* Upon confirmation of proper assembly, they were digested with Pmel following the insert directions. This constituted the first step in the assembly process (Fig. 26). Once digested, this YAC constituted the backbone for the second step in the assembly.

The second step combined the following parts (Fig. 27): pYes-lst_step, 5 or 10 kbp C amplicon, 5 or 10 kbp D amplicon, *mKate*, and *nirFP* (with homology to pYes-1 L).

Colonies was checked for proper assembly of the whole construct. The yeast-based assembly was explained above.

The YACs obtained after this second step were move into *E. coli* and the culture was miniprepped using a kit specific for the purification of large-sized plasmids (ZRBAC DNA Miniprep Kit, Zymo Research).

The insert was designed to be popped out with the restriction enzymes Eco RV and Bsu361, both predicted to be unique cutters. Nonetheless, it was found that there were

errors/differences in the *S. coelicolor* A3(2) sequence deposited in the NCBI database, which resulted in the insert being cut into additional fragments. This is known to happen with sequences obtained from NCBI, as they correspond to the sequencing of a specific organism that might have accumulated differences from the one in use. Another issue is potential sequencing errors when elucidating the genomic makeup of this organism.

Upon digestion with each of the enzymes, it was found that the additional restriction sites corresponded to the recognition sites of Eco RV. This blunt end cutter was then replaced with Bsu361, which cuts in the YAC backbone and originated a sticky end.

In order to clone the insert into the blunt-end site of PmeI, in the recipient backbone, the sticky-ends of the 5 and 10 kbp inserts were first blunted using Quick Blunt Enzyme Mix (NEB). The total size of the final vectors was ca. 30 kbp for the 5-kbp 4-FP shufflon system and 50 kbp for the 10 kbp one.

Construction of an E coli-Streptomyces *integrative shuttle vector, pSET-SC4, for the 4-FP assemblies*

This vector was designed to replace the YAC-*E. coli-Streptomyces* integrative shuttle vector. It consisted of the pSET152 backbone (Fig. 28) to which a *cos* site , and the ermEp* promoter followed by recognition sequence a and the PmeI restriction site were added.

The promoter, recognition sequence, and restriction site were ordered as a gBlock from IDTDNA Corp and the *cos* site was obtained via PCR from pWEB-TNC using KAPA Hifi polymerase (no additives).

The 4-FP pathways were then cloned into this plasmid in the PmeI site and checked clones for the correct orientation.

Successful cloning in the proper orientation was achieved for the 10-kbp shuffling, but not for the 5-kbp one. It was then decided to assemble the insert into the backbone by Isothermal Assembly. For this, the backbone was PCR-amplified with primers with an overhang that matched the 5' (reverse primer) or 3' (forward primer) of the insert. Upon transformation into *E. coli* l0B and PCR screening, colonies carrying the right construct were found.

The 1, 5 and 10 kbp shuffling constructs were transformed into *E.coli* ET12456 pUZ8002 and conjugated into *S. venezuelae.* Exconjugants were successfully obtained and restreaked for purity.

Transforming rci *and* tsR *into the* S. venezuelae *strains*

Now that the integration of the multiple pathways in *S. venezuelae* was successful, the genes encoding the recombinase and the thiostrepton resistance had to be conjugated into the several strains.

Spores of *S. venezuelae* attB ::pSET-SC4-1kbp, -5kbp and -10kbp were used for conjugation with *E.coli* ET12456 pUZ8002 carrying pKC-SC5 (HygR). Upon several attempts, no exconjugants were obtained. It was determined that the hygromycin cassette was not a good selection marker in this strain of *S. venezuelae,* with this plasmid backbone. This conclusion was based on the fact that when using a version of pKC-SC5 with apramycin resistance (instead of hygromycin) and selecting for thiostrepton resistance (versus hygromycin resistance), exconjugants were indeed obtained. The selection was performed using thiostrepton (an antibiotic and the inducer of *rci* expression) and not apramycin because the integrated pSET-SC4-1 /5 /10 kbp plasmids contained already the apramycin resistance cassette.

Nonetheless, two different plasmids were built concomitantly (Figs. 29-30), as new tsr and rci expression vectors, with a gentamycin selection marker. One plasmid was pKC1139-rci+tsr (ApraR) where the gentamycin cassette (Fig. 30) was introduced in the single blunt-end cutter NcoI site in the apramycin cassette by T4 DNA-ligase mediated ligation. The ligation was transformed into *E. coli* DH5 α and plated onto LB with 10 μg/mL of gentamycin. Positive clones (pKC-SC7) were moved into the conjugation strain *E.coli* ET12456 pUZ8002.

Another plasmid also with a replicative origin of replication (but not temperature sensitive) was built (Fig. 29). This plasmid was based on pFX5832 and was generated using a version with gentamycin as selective marker, previously generated for unrelated purposes. An amplicon carrying the *tsr* and *rci* genes was PC Red from pKC1139-rci+tsr and the backbone of pFX583-GentR was amplified using primers that added overhangs homologous to the 5' and 3' of the *tsr* and *rci* amplicon. The two parts were joined by means of the Isothermal Assembly and transformed into *E. coli* DH5 α. Positive clones (pFX-SC6) were moved into the conjugation strain *E.coli* ET12456 pUZ8002.

The *rci* and *tsr,* carried by the appropriate plasmids, were conjugated into the *S. venezuelae* strains carrying the 1, 5 and 10 kbp 4-color shuffling pathways.

Calculation of Rci-*mediated flipping frequencies*

Given that the expression of the fluorescent proteins in *S. venezuelae* was not successful, as already explained, it was not possible to use them as a method to discriminate and quantify the DNA flipping frequencies, mediated by Rci. Instead, qPCR (KAPA SYBR® FAST Universal 2X qPCR Master Mix) was used in a LightCycler® 96 System (Roche, USA), with the downside that it is not possible to individualize flipping per sequence, ie, in a given sequence (cell) what flipping occurred. Instead, a general picture of the population is obtained.

In order to quantify the DNA inversion frequencies of each recognition site (a, b, c and d) in the whole population, several sets of primers were designed, as explained below. The melting temperature was ca. 60 C and the length 18-20 bp. The amplicon was ca. 100 bp in length.

Primers that amplify the middle region of each FP were used as reference for the relative amount of template, and sets of primers were designed to amplify DNA in the eventuality of an inversion, or lack thereof.

Cultures of *S. venezuelae* attB::pSET-SC4-1kbp and 10 kbp, with and without Rci were incubated for a period of 3 days, with 2 mL aliquots being taken on a daily basis, for qPCR measurements. A total of 25 - 50 ng of genomic DNA was used per reaction. The DNA was extracted using the ZR Fungal/Bacterial DNA MiniPrep Kit (Zymo Research) and the flipping rates were calculated as described elsewhere[3] and followed the Livak method[4]. The equations for calculation are shown below and in Fig. 31:

$$\%\text{Flipped} = 2^{\left(\frac{\text{flipped}}{\text{flipped + unflipped}}\right)} \times 100$$

This equation incorporates the results obtained when no Rci is present, so that the values shown correspond solely to measurement of interest. There was no flipping when Rci was absent, but there was non-specific binding, which can introduce error in the measurements - thus the values have to be incorporated into the equation.

The values were adjusted for the different DNA replication rates and primer annealing efficiency.

The original publication that describes the recombination ability determined the inversion frequencies as being 29, 0.025, 0.02 and 0.1 for the recognition sequences *a, b, c,* and *d* respectively *(a> d > b >* c). The DNA shuffled in the original context is ca. 400 bp long, representing the 3' end of the *pi IV* gene in some strains of *E. coli* and *Salmonella thyphimurium*[7].

For the shortest inter-recognition site, of ca. 1 kbp, the highest flipping frequency was found for sequences a, followed by d, c and b *(a > d > c > b)*. This is similar to the relative frequencies observed in the native organisms *(a > d > b >* c). Nonetheless, opposites results were obtained for the larger, 10 kbp, DNA pieces to be flipped by each pair of sites (Table 20), with *b* being the sequence that that resulted in the highest inversion rate, followed by *c, d* and *a.* These frequencies represent the average percentage of inverted sequences over the 4-day period.

Table 20. Percentage of inverted DNA, per pair of recognition sites and intersite distance during the duration of the screening.

Recognition sequence	Estimated % of flipping
a	1
b	32
c	0.09
d	9.6

Looking at each time point specifically, the recognition site *a* stands out as the overall best performer. The two recognition sites displaying the highest inversion rate were *a* (1 kbp) and *c* (10 kbp). The prevalence of *a* as the most efficient sequences was found also at 90 hours. The sites *d* were the best performers at 42 hours, while at 66 hours, both *a* and *d* performed well (Table 21).

The recognition sites that performed the best at inverting large DNA sequences (10 kpb) varied over the course of the analysis (Table 21), with *c* being the most efficient one at 18 hours, *a* at 42, *b* at 66, and *c* and *d* at 90 hours.

Table 21. Most efficient recognition sites in terms of percentage of inverted sequences, for each time point.

Sampling time (h)	Recognition sequence		
	1 kbp	5 kbp	10 kbp
18	*a*	*a, d*	*c*
42	*d*	*a, b*	*a*
66	*a, d*	*a*	*b*
90	*a*	*a, b*	*c, d*

In the case of sequence *a,* the percentage of inverted sequences displayed an opposite trend when comparing the short and the long flipping. The inversion of short, 1 kbp DNA

sequences, increased over time, whereas it dramatically decreased for the 10 kbp sequences after 42 hours (Fig. 32).

For recognition sequence *b,* the 1 kbp inversion events were more frequent in the being and at the end of the experiments, with an inflexion point situated between the 66 and 71 hours.

For both recognition sequences *c* and *d,* the same pattern of increased inversion frequency, with a maximum between the 42 and 66 hours, was observed for the short 1 kb sequences. In the particular case of the 10 kbp sequences, the highest inversion frequency of *c* was detected in the first time point, whereas the performance of Rci on *c* remained stable at higher values after the first time point (Fig. 32).

As already explained, these values were calculated for the whole population and do not give information on the inversion of a given segment in regards to other, in the same cell.

The possibility of performing digital PCR or qPCR using sequence-specific probes (instead of SYBR-based qPCR), is being considered. In the case of digital PCR, upon multiplexing, the flipping of each sequence in each DNA molecule could be determined. Thus, it would provide the same results as the FACS, with the additional benefit of being independent of the FP expression (required for recognizing the flipping when measure by FACS). The use of probes to perform qPCR would enable a more exact quantification of the flipping events, as in this case it would be independent of the length of the amplicon (which varies by only a few base pairs, but can lead to higher fluorescence levels due to the longer amplicon size).

Integration of the PKS-shufflon system in S. venezuelae

The PKS-shufflon system was designed to be integrated into the *pikAII* gene of the pikromycin pathway in *S. venezuelae,* to delete the vast majority of this gene. For this purpose, the pathway to integration was designed and built to be surrounded by regions of homology towards the chromosomal *pikAII.* By means of a double crossover event, the synthetic shufflon pathway should be integrated and the plasmid backbone lost. A single crossover event results in the integration of the whole plasmid and the backbone is not lost.

Thus far, seven attempts to conjugate the plasmid into *S. venezuelae* were made but no exconjugants were obtained.

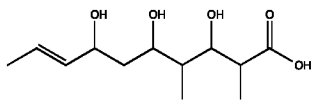
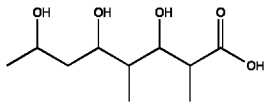
It was suspected that the integration, by single crossover, was very inefficient and thus no exconjugants were being obtained. This led to another backbone being built, for integration in the *attB* site of *S. venezuelae.* The basis of this vector was pSET152, to which

XbaI and Bsu36I restriction sites and a *cos* site obtained by PCR from pWEB-TNC (Fig. 33) were added by Isothermal Assembly.

Upon PCR-confirmation of the correct three-way assembly (Fig. 33), the new plasmid was obtained in large quantity and linearized using PmeI. To prevent recircularization without insert any insert, the phosphate groups were removed by incubation with the phosphatase CIP (NEB Biolabs).

The 2 parts of the synthetic PKS-shufflon previously obtained (Examples 1-2) were ligated into this vector by means of T4 DNA Polymerase (Example 1). Upon confirmation of proper 3-way assembly (Fig. 34), the plasmid was electroporated into *E.coli* ET12456 pUZ8002 and conjugated into *S. venezuelae.* Nonetheless, upon 4 attempts, no exconjugants were obtained. This indicates that the pathway with multiple heterologous *pks* genes is indeed functional in *S. venezuelae* and is leading to the assembly of molecules that are toxic for the host organism. It was also briefly considered that the issue might be the size of the plasmid (ca. 43 kbp), but since the integration of the 10 kbp 4-FP shufflon system of ca. 50 kbp was successful, the issue should indeed be the toxicity of the new molecules. This assumption is further supported by the predictive algorithm results for the possible molecules being assembled by the 4 genes cloned in an on position in the PKS-shufflon (Table 22).

Table 22. SMILES, structures and activity prediction for molecules being assembled by the genes incorporated in the PKS-shufflon pathway, to be integrated

Gene combination *	**SMILES**	**Structure**	**Activity prediction****	
1, 2, 3, 4 1, 2, 4 1, 3, 4 1, 4	CC=CC(O)CC(O)C(C)C(O)C(C)C(=O)O		GPCR ligand	0.28
			Ion channel modulator	0.38
			Kinase inhibitor	- 0.55
			Nuclear receptor ligand	0.69
			Protease inhibitor	0.18
			Enzyme inhibitor	0.87
2, 3, 4 2, 4	CC(O)CC(O)C(C)C(O)C(C)C(=O)O		GPCR ligand	-0.09
			Ion channel modulator	0.13
			Kinase inhibitor	-0.76
			Nuclear	0.28

Gene combination *	SMILES	Structure	Activity prediction**	
			receptor ligand	
			Protease inhibitor	-0.04
			Enzyme inhibitor	0.56

* When no molecule is predicted to be assembled, no data is provided
l-momAVI
2 -*amphK*
3 - *ty/G*
4-momAVIII
**Higher values correspond to higher probabilities of activity (-3<x~3; x>0.10 is considered good)

Materials

Table 23. List of bacterial strains used and/or developed during this study

Species	Strain	Temperature (C)	Medium	Source
S. venezuelae	DSM 41110	30	ISP2	DSMZ
E. coli	MG1655	37	LB	Lab collection
E. coli	ET12567	37	LB + Chloramphenicol 34 μg/mL	Lab collection
E. coli	Top10	37	LB	Geneart™
E. coli	DH5α	37	LB	Lab collection
E. coli	10B	37	LB	Lab collection
S. venezuelae attB::pSET-SC4-1kb	DSM 41110	30	ISP2, Apramycin	This study
S. venezuelae attB::pSET-SC4-5kb	DSM 41110	30	ISP2, Apramycin	This study
S. venezuelae attB::pSET-SC4-10kb	DSM 41110	30	ISP2, Apramycin	This study
S. venezuelae attB::pSET-SC4-1kb, pKC1139-rci-tsr	DSM 41110	30	ISP2, Apramycin, Thiostrepton	This study
S. venezuelae attB::pSET-SC4-10kb, pKC1139-rci-tsr	DSM 41110	30	ISP2, Apramycin, Thiostrepton	This study
S. venezuelae attB::pSET-SC4-1kb, pKC-SC7	DSM 41110	30	ISP2, Apramycin, Thiostrepton, Gentamycin	This study
S. venezuelae attB::pSET-SC4-5kb, pKC-SC7	DSM 41110	30	ISP2, Apramycin, Thiostrepton, Gentamycin	This study
S. venezuelae attB::pSET-SC4-10kb, pKC-SC7	DSM 41110	30	ISP2, Apramycin, Thiostrepton, Gentamycin	This study
S. venezuelae attB::pSET-SC4-	DSM 41110	30	ISP2, Apramycin, Thiostrepton,	This study

1kb, pFX-SC6			Gentamycin	
S. venezuelae attB::pSET-SC4-5kb, pFX-SC6	DSM 41110	30	ISP2, Apramycin, Thiostrepton, Gentamycin	This study
S. venezuelae attB::pSET-SC4-10kb, pFX-SC6	DSM 41110	30	ISP2, Apramycin, Thiostrepton, Gentamycin	This study

Table 24. List of plasmids created for the purpose of this research project

Plasmid	Features	Selection	Strain	Source
pSET-SC4	pSET152 carrying a *cos* site, the ErmEp* promoter and a PmeI site	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pSETSC3b	pSET152 carrying a *cos* site,the ErmEp* promoter and a XbaI and Bsu36I sites	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pSETSC3b-Shufflon	pSETSC3b carrying inserts 1 and 2	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp.	This study
pYES1L-5kb	pYES1L carrying insert 1	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-10kb	pYES1L carrying insert 2	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-5.1 kb	pYES1L carrying part 1 of the 5 kb shuffling system	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-10.1 kb	pYES1L carrying part 1 of the 10 kb shuffling system	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-5.2 kb	pYES1L carrying the full 5 kb shuffling system	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pYES1L-10.2 kb	pYES1L carrying the full10 kb shuffling system	Spectinomycin Tryptophan	Replicative in *S. cerevisiae* and *E. coli*	This study
pKC1139-rci	pKC1139 carrying *rci*	Apramycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pKC1139-rci-tsr	pKC1139 carrying *rci* and *tsR*	Apramycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pKC-SC5	pKC1139 carrying *rci*	Hygromycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pKC-SC7	pKC1139 carrying *rci* and *tsR*	Gentamycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pFX-SC6	pFX583-GentR carrying *rci* and *tsR*	Gentamycin	Replicative in *E. coli* and *Streptomyces spp.*	This study
pSETSCexp-mKate2 (repeat)	Expresses *mKate2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study

Plasmid	Features	Selection	Strain	Source
pSETSCexp-nirFP (repeat)	Expresses *nirFP*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-dsRed2 (repeat)	Expresses *dsRed2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSETSCexp-ebfp2 (repeat)	Expresses *ebfp2*	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSET-SC4-1kb	pSET-SC4 carrying the 1kb-shuffling system	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSET-SC4-5kb	pSET-SC4 carrying the 5kb-shuffling system	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study
pSET-SC4-10kb	pSET-SC4 carrying the 10kb-shuffling system	Apramycin	Replicative in *E. coli*; integrative in *Streptomyces* spp	This study

Table 25. List of independently developed published plasmids used during this study

Plasmid	Features	Selection	Strain	Source
pUZ8002	DNA transfer machinery	Kanamycin	Replicative in *E. coli*	14
pSET152	Streptomyces integrative plasmid (*attB*)	Apramycin	Integrative in *Streptomyces* spp., replicative in *E. coli*	3
pKC1139	Shuttle vector with temperature-sensitive Streptomyces oriR	Apramycin	Replicative in *E. coli* and *Streptomyces spp.*	3
pIJ10700		Hygromycin	Replicative in *E. coli*	S. O'Rourke (unpublished)

References for Example 3

1. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth 6, 343–345 (2009).
2. Lussier, F. X., Denis, F. & Shareck, F. Adaptation of the highly productive T7 expression system to Streptomyces lividans. Appl. Environ. Microbiol. 76, 967–970 (2010).
3. Mimee, M., Tucker, A. C., Voigt, C. A. & Lu, T. K. Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota. Cell Syst. 1, 62–71 (2015).
4. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–408 (2001).
5. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical streptomyces genetics. (John Innes Foundation, 2000).
6. Bierman, M. et al. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. Gene 116, 43–49 (1992).
7. Gyohda, A., Funayama, N. & Komano, T. Analysis of DNA inversions in the shufflon of plasmid R64. J. Bacteriol. 179, 1867–1871 (1997).

Example 4

Preparation of a cell-free translation and transcription system

Given the inability to integrate the PKS-shufflon pathway in *S. venezuelae*, it was decided to test whether a cell-free transcription and translation system could be used for flipping DNA. The parts used for the cell-free transcription and translation system include: DNA, which serves as template for transcription; S30-based reaction, which is the basis of the system and contains the S30 cell fraction of the cell extract (ribosomes and soluble proteins); and a control gene, *gusA* was selected as the gene to test if the system was indeed transcribing and translating genes.

Given that the system used to transcribe and translate genes uses exogenously-supplemented RNA polymerase, this was an opportunity to use the most convenient promoter-polymerase pair. The T7 promoter was used, given that the RNA T7 polymerase can be purchased.

Also, since all genomic DNA is removed from S30, any chromosomal genes of interest were added back to the system: *pikAI*, *III*, *IV* and *V* had to be PCR amplified from the chromosome of *S. venezuelae* DSM 41110.

The T7 promoter and a ribosomal binding site were added upstream of the initiation codon of each gene in one of 2 different ways: by amplifying the gene of interesting using forward primers carrying overhangs with the T7 promoter and an RBS, or by amplifying cosmid backbones with a reverse primer carrying overhangs with the T7 promoter and an RBS (Fig. 35).

The first option was used for genes *rci*, *pikAI, III, IV* and *V* (Fig. 36), as they were amplified from the chromosome of *S. venezuelae.* Given the size of pikAI, it was PCRed in 1 or 2 pieces (if 1 piece failed). The second option was used for the PKS-shufflon pathway, as its genes were organized in an operon and it was used to replace the ermEp* promoter with T7.

The PCR conditions initially used were Herculase II Polymerase without additives and 58 C as annealing temperature. As this method was not successful, the PCR protocol had to be modified for the amplification of these genes. The rci was successfully amplified and a T7 promoter and an RBS were added upstream of its initiation codon.

It was decided to perform a full-range analysis, where magnesium and/or DMSO were added to the reaction with KOD polymerase, in varying concentrations. Utilizing this approach, it was possible to find functional amplification conditions for each gene.

After learning that PCR-based transcription did not perform in a satisfactory way, it was decided to clone these genes/amplicons into vectors.

This was achieved by blunting the amplicons previously obtained by PCR: this step also provided the phosphate groups to the amplicons and allowed the dephosphorylation of the blunt-end digested cosmid backbone. Cosmids were used instead of plasmids due to the size of the amplicons.

The template DNA for testing the efficiency of the system - *gusA* driven by T7 - had been previously constructed for another purpose.

All plasmids were maxiprepped from 400 mL of culture using ZymoPure® MaxiPrep kit. The DNA extraction process was followed by ethanol precipitation for concentrating and further cleaning the DNA, which was then resuspended in sterile water and stored at -20 °C.

Testing the cell-free system

Two 5 µL aliquots were taken at 15 and 60 min. One aliquot was mixed with X-Gluc to assess color formation while the other was ran in a Bis-Tris 4-12% protein gel (NuPage) and stained with Coomassie Blue. No color development was observed for the aliquot mixed with X-Gluc and the only band found in the gel, with the expected weight of *gusA*, was also present in the control, both for the ermEp*- and T7-driven expressions. The Coomassie Blue staining will not allow proteins in small amount to be seen on a gel. Furthermore, the system's proteins could potentially be obscuring GusA, as they run at ca. the same height. The lack of color development with X-Gluc could be due to the enzymatic action inhibition cause by the S30 system.

New approaches to calculate the Rci-based DNA inversion

It is well known in the field that amplifying high GC DNA is troublesome. This is also well documented in Examples 1-3.

The qPCR reaction is a DNA amplification reaction. In the case of the Sybr® Green assay, the amount of DNA being accumulated as the amplification process takes place is quantified based on the amount of dye binding double-stranded DNA. This is a problem because it is extraordinarily difficult to obtain a single, specific amplicon, when amplifying high GC DNA. When performing PCR, there is more flexibility in terms of buffer content and amplification conditions, which is not the case with qPCR.

The approach initially taken to quantify the flipping efficiency, and given that an all-flipped version of the system was not available, was to use a gene of reference to estimate the Rci-mediated flipping efficiency (Example 3).

Standards DNA-Ct and qPCR

In this alternative approach to measure the Rci flipping efficiency, instead of relying on the internal standard and discount error, an "all-flipped" control was built. It was impractical to build versions of all 3 large pathways that contained the genes in a flipped (5'-3') orientation. Yet, the original (unflipped) pathways were built so that the same sets of primers could be used independently of the size of the DNA sequence to be flipped. Given that it is possible to predict the DNA sequence of the flipped version, it was possible to build a synthetic DNA sequence that contained the expected amplicons amplified, if the flipping indeed occurred (Figs. 37A-37C). These amplicons, and additional DNA up and downstream of each of them, were synthesized as a gBlock and cloned into pSET152. Upon integration into the attB locus, the culture was grown and the genomic DNA extracted. The mere gBlock was not used directly as control DNA, but instead was placed in the context of the host organism, to mimic the other constructs and the background associated with amplifying DNA from a genomic prep of high GC DNA.

Now that a positive control was available, it was possible to estimate the number of molecules of DNA that had in fact flipped. For this, a standard curve relating the Ct values of the qPCR reaction with the number of molecules flipped. The Ct value (Cycle threshold) represents the number of cycles for the fluorescence measured to rise above the baseline value. The higher this value is, the more cycles are required and thus the less abundant the template is.

For calculating the standard curve, the flipped and unflipped templates were added to the qPCR reaction in different concentrations, ranging from 102 to 107 molecules of DNA per well. The number of molecules of DNA was estimated using Thermo Scientific's DNA Copy Number Calculator. The fluorescence (Ct value) measured per DNA concentration (copies/well) was plotted (Fig. 67).

For determining the flipping efficiency of Rci, 3 independent experiments with replicates were performed. In each experiment, the 1, 5, and 10 kb flipping strains of *S. venezuelae*, carrying Rci or not were grown overnight in tryptic soy broth, 30 C, 250 rpm. Upon overnight growth, the cultures were reinoculated into TSB to an optical density of 0.04 at 600 nm. Samples (2 mL aliquots) were taken in early exponential phase (4 hours), mid

exponential phase (7 hours) and 20 hours of incubation, according to the growth curve previously determined (Fig. 40).

The DNA was purified using a Promega Maxwell 16 System Plant DNA Extraction Kit.

The qPCR reaction was performed using a Roche Light Cycler 480 II and 384-well plates. The reactions were prepared using a Kapa qPCR Universal Sybr® Green Kit (Kapa Biosystems) and following the manufacturer's instructions. The extension temperature used was 63C, in order to decrease the probability of non-specific amplification.

Table 26. Averaged percentage of flipped DNA molecules per sampling time and recognition sequence, using Sybr® Green-based qPCR.

Sampling phase	Recognition sequence A (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	2.18±1.54	4.72±1.54	6.43±4.44
Mid exponential	2.44±1.4	8.55±8.7	19.46±27.66
Stationary	4.27±1.27	4.48±2.74	2.46±1.53

Sampling phase	Recognition sequence B (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	2.21±1.93	2.95±1.11	5.96±1.48
Mid exponential	2.58±1.01	3.15±0.23	4.33±3.81
Stationary	3.21±1.4	3.88±2.51	2.74±0.52

Sampling phase	Recognition sequence C (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	33±26	48±39	44±9
Mid exponential	40±20	25±24	62±32
Stationary	38±13	63±32	36±16

Sampling phase	Recognition sequence D (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	0.01±0.02	0.001±0.0006	0.001±0.001
Mid exponential	0.02±0.03	0.03±0.05	19±34
Stationary	0.001±0.0006	0.003±0.002	0.007±0.01

Table 27. Flipping efficiency per recognition site, for each independent experiment and inter-site distance, using Sybr® Green-based qPCR.

Sampling phase	Independent experiment	Recognition sequence A (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	0.63	6.47	11.54
	2	4.06	18.39	51.40
	3	5.58	7.58	4.20
Mid exponential	1	3.70	3.68	3.53
	2	1.67	1.88	3.12
	3	3.04	3.45	1.81
Stationary	1	2.21	4.01	4.22
	2	1.60	5.38	3.87
	3	4.20	2.40	1.37

Sampling phase	Independent experiment	Recognition sequence B (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	0.003	4.20	4.72
	2	3.57	3.34	8.55
	3	4.75	5.94	2.33
Mid exponential	1	3.55	2.11	7.59
	2	2.62	2.90	3.26
	3	2.02	4.61	2.57
Stationary	1	3.08	2.55	5.53
	2	1.58	3.21	1.17
	3	2.87	1.09	3.33

Sampling phase	Independent experiment	Recognition sequence C (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	3.45	27.94	36.47
	2	62.13	49.25	98.34
	3	51.82	43.90	48.54
Mid exponential	1	44.28	92.75	53.90
	2	36.16	2.06	37.51
	3	34.38	20.13	18.23
Stationary	1	51.29	21.79	43.18
	2	21.38	23.37	49.75
	3	27.20	42.74	39.88

Sampling phase	Independent experiment	Recognition sequence D (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	0.03	1.44×10^{-03}	2.16×10^{-03}
	2	0.06	0.08	59.59
	3	0.001	9.99×10^{-04}	0.02
Mid exponential	1	3.59×10^{-04}	2.93×10^{-04}	3.96×10^{-04}
	2	2.39×10^{-03}	0.001	2.53×10^{-03}
	3	5.32×10^{-04}	0.003	2.08×10^{-03}
Stationary	1	7.05×10^{-04}	5.60×10^{-04}	3.00×10^{-04}
	2	8.82×10^{-04}	4.18×10^{-04}	2.38×10^{-03}
	3	1.77×10^{-03}	0.005	9.47×10^{-05}

The results were averaged and the number of molecules of DNA were calculated using the control logarithmic regression lines, which were calculated by plotting the number of molecules, in logarithmic scale, against the Ct values (Figs. 38-39). The percentages of flipping were calculated using the following fraction:

% Flipped sequences = # *ON* Amplicons / (# *ON* Amplicons + # *OFF* Amplicons) x 100

Despite the attempts to reduce non-specific amplification, there was still a considerable amount of non-specificity in the qPCR amplification. Furthermore, the use of a higher extension temperature worked well with the primers for the flipped amplicons, but was detrimental to the amplification of unflipped sequences.

This impaired the calculation of precise flipping efficiencies. The initial approach (Example 3) seemed to overestimate the flipping efficiency of Rci. This new approach seemed to underestimate such efficiencies. Three independent experiments were also performed, which led to higher standard deviations (Table 26).

PrimeTime probes and qPCR

Aiming to increase the specificity of the qPCR-based flipping quantification, it was decided to use sequence-specific labeled probes instead of Sybr® Green-based quantification (Fig. 41).

These synthetic probes - PrimeTime ZEN™ double-quenched probes for qPCR (IDT DNA Technologies) - hybridize onto the complementary DNA sequence. Once that region is amplified, the DNA polymerase hydrolyzes the FAM fluorophore located in the 5' end of the probe, releasing it. This causes a physical separation between the fluorophore and the internal ZEN™ and 3' Iowa Black® quenchers, which results in fluorescence between emitted (Fig. 41).

New sets of primers were also ordered along with the probes, for the flipped and unflipped amplicons. The PrimeTime qPCR Master Mix (IDT DNA Technologies) was also used, according to the manufacturer instructions. The 384 plates were run in a Roche LightCycler 480 II.

The measurements obtained for each sample were averaged and using the standard curves the number of molecules were extrapolated. They were then converted into percentage of flipped molecules over the number of flipped and unflipped molecules.

The flipping efficiencies obtained for each individual experiment varied, likely due to the fact that tipAp, the promoter driving the expression of *rci*, is very leaky. This results in an inability to start the assay with fully unflipped constructs.

Similarly to the previous results, the flipping efficiencies varied with the distance between recognition sites as well as with the recognition site used. The ranking of sites in terms of flipping efficiency, throughout the sampling times, was D>A>C>B, for the 1 kbp distance, C>D>A>B for 5 kb and D/A/C>B. Looking at the results per sampling time, C>D>A>B for the first time point, A>D>B>C for the second point and D>C>A>B for the

last one (Table 28). As previsouly observed, there was a great inter-independent experiment variation (Table 29), possibly due to the leakiness of tipAp, as explained above.

Table 28. Averaged percentage of flipped DNA molecules per sampling time and recognition sequence, using PrimeTime qPCR probes.

Sampling phase	Recognition sequence A (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	33±58	50±71	33±58
Mid exponential	50±71	0.03±0.04	99±2
Stationary	67±58	33±58	33±58

Sampling phase	Recognition sequence B (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	33±58	1.04±1.44	0.92±1.18
Mid exponential	2.5±3.8	36±55	65±56
Stationary	1.15±1.04	34±57	11±19

Sampling phase	Recognition sequence C (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	33±58	67±58	67±58
Mid exponential	33±58	0	33±58
Stationary	33±58	67±58	67±58

Sampling phase	Recognition sequence D (% of flipped sequences)		
	1 kb	5 kb	10 kb
Early exponential	67±58	33±58	33±58
Mid exponential	67±58	33±58	33±58
Stationary	100±0	33±58	100±0

Table 29. Flipping efficiency per recognition site, for each independent experiment and inter-site distance, using PrimeTime probes

Sampling phase	Independent experiment	Recognition sequence A (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	0.15	5.50×10^{-10}	2.52×10^{-03}
	2	3.70×10^{-04}	3.23×10^{-15}	96.13
	3	100.00	100.00	100.00
Mid exponential	1	100.00	-	0.00
	2	-	9.23×10^{-11}	100.00
	3	1.13×10^{-05}	9.44×10^{-07}	0.00
Stationary	1	0.00	100.00	100.00
	2	100.00	0.08	100.00
	3	100.00	0.01	6.10×10^{-13}

Sampling phase	Independent experiment	Recognition sequence B (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	1.78×10^{-03}	0.38	2.26
	2	6.94	8.07	98.94
	3	1.45	0.92	32.56
Mid exponential	1	100.00	0.01	5.50×10^{-04}
	2	0.22	2.13×10^{-03}	0.02
	3	2.03	0.06	0.05
Stationary	1	0.32	2.66	0.51
	2	0.36	100.00	95.95
	3	6.45×10^{-07}	100.00	0.16

Sampling phase	Independent experiment	Recognition sequence C (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	2.79×10^{-09}	1.82×10^{-07}	2.17×10^{-09}
	2	1.71×10^{-05}	6.11×10^{-06}	0.09
	3	5.83×10^{-15}	3.60×10^{-07}	100.00
Mid exponential	1	100.00	100.00	100.00
	2	5.70×10^{-23}	-	100.00
	3	100.00	100.00	100.00
Stationary	1	3.69×10^{-15}	100.00	100.00
	2	100.00	1.65×10^{-11}	5.20×10^{-10}
	3	4.38×10^{-12}	100.00	2.83×10^{-07}

Sampling phase	Independent experiment	Recognition sequence D (% of flipped sequences)		
		1 kb	5 kb	10 kb
Early exponential	1	5.43×10^{-10}	6.79×10^{-05}	6.37×10^{-09}
	2	100.00	4.15×10^{-05}	8.37×10^{-04}
	3	100.00	7.98×10^{-06}	100.00
Mid exponential	1	100.00	6.59×10^{-09}	100.00
	2	100.00	4.75×10^{-22}	0.00
	3	100.00	100.00	-
Stationary	1	100.00	100.00	5.10×10^{-23}
	2	3.15×10^{-05}	100.00	100.00
	3	100.00	7.17×10^{-22}	100.00

When comparing the three approaches undertaken thus far (Table 30), with the published characterization of Rci's flipping efficiency in *Enterobacteriaceae*, the second approach (Sybr® Green with DNA standards) led to results more similar to those published, in relative terms.

The third approach is the most reliable one, as probes - especially those used in this study - are the most precise method of quantification. For the smaller inter-recognition site distance, the relative results were similar to those of the native system (with distance smaller than 500 bp).

Table 30. Rci flipping efficiencies using three different methods: (2) Sybr® Green and a DNA standard, (3) qPCR double-quenched probes and a DNA standard

Described % of rearranged molecules, in the original Enterobacteriaceae native system:

Recognition sequence a>d>b>c
29>0.1>0.025>0.02

Approaches 2 and 3, with efficiencies per time point averaged together:

	1 kbp	5 kbp	10 kbp
2	**c > a/b > d** 33 > 2/2 > 0.01	**c > a > b > d** 48 > 5 > 3 > 0.001	**c > a/b > d** 44 > 6/6 > 0.001
3	**d > a > c > b** 78 > 50 > 33 > 12	**c > d > a > b** 44 > 33 > 28 > 24	**c/d > a > b** 56/56 > 55 > 26

Testing the expression of Sreptomyces-gfp in S. venezuelae

Given the GC-content associated difficulties in the both exact and precise measurement of Rci-mediated DNA flipping efficiencies, it was decided to test the expression of *gfp* in the target organism.

This *gfp* was modified for *S. coelicolor* A3(2) in 1999[1]. This process involved codon-optimization, given the codon usage bias of the high GC streptomycetes. Since then it has been successfully used in multiple *Streptomyces* spp, such as *S. lividans* [2] and *S. hygroscopicus* [3] . Nonetheless, the level of performance were never has high as found in other organisms such as *E. coli*.

If this fluorescent protein works in *S. venezuelae*, it will be possible to build plasmids carrying this gfp surrounded by Rci recognition sites and a downstream copy of *rci* . Yet, this will also depend on whether 100% of the population fluoresces or not, as if it does not, the flipping efficiencies will be underestimated. An option would be to extensively determine the percentage of the population that does not fluoresce, and include such percentage in the calculation of the flipping efficiencies (i.e., if on average only 90% of the population fluoresces, and if it is determined that 90% of the population flipped, the efficiency is in fact 100%).

Nonetheless, before embarking on the construction of these vectors, the gfp carried by pSET152 and driven by the promoter rpsLp(XC) a plasmid was integrated into the attB site of *S. venezuelae*. The strain carrying gfp was incubated in TSB supplemented with apramycin, whereas the wildtype, used as negative control, in TSB alone. Growth was promoted at 30C and 250 rpm. *E. coli* pZ8-gfp was used as a positive control and incubated in TSB supplemented with kanamycin, at 30C and 250 rpm. Aliquots were taken at 18 hours

for *S. venezuelae* and 6 hours for *E. coli*. The cultures were diluted 1:100 and 1:1000 in PBS and analyzed in a BD-FACS LSR Fortessa cell analyzer (BD Biosciences, CA). The fluorescence was measured using a FITC filter with a 490 nm excitation laser and a 520 emission filter.

Except for the positive control, none of the samples displayed fluorescence: *gfp* did not seem to work in *S. venezuelae* DSM 41110. Yet, it was decided to test this again, over a longer course of time and with multiple reinoculation steps: an overnight culture of *S. venezuelae* carrying *gpf* (and the respective controls) and reinoculated and grown overnight again. This second culture was then used to inoculate TSB (with the appropriate antibiotic, when needed). The reinoculation step was repeated 3 times (once a day, the last time point was 6 hours before analysis). Six hours after the third reinoculation step, 100 µL of each culture were mixed with 900 µL of PBS and tested for fluorescence by FACS. Interestingly, this time, fluorescence was observed, along with minimal background.

The fluorescence intensity varied with the incubation time and so did the percentage of the population that fluoresced. Six hours post-reinoculation, three populations of *S. venzuelae* were observed, two of which fluoresced with high intensity. A third population did not fluoresce. The oldest population (ca. 54 hours old) presented the highest percentage of non-fluorescent cells, possibly related with the aging population and sporulation. The positive control - *E. coli* expressing *gfp* - was not included in the *S. venezuelae* histograms as the cell morphology and the growth rate are very different, having served as an independent control of fluorescence. Interestingly this strain did not fluoresce at the last time point, in early stationary phase.

When fully characterizing the Rci-mediated DNA inversion independently of the previously built PKS-shufflon pathway and its characteristics, it is more straightforward to test each pair of recognition sites independently (instead of using the 4-color system).

Integration of the PKS-shufflon pathway in S. coelicolor A3(2)

As previously explained (Example 3), it was not possible to integrate the PKS-shufflon pathway in *S. venezuelae*, despite innumerable attempts.

This pathway was designed to interact with the *S. venezuelae* pikromycin pathway to generate new polyketides. As the PKS genes in the biosynthetic pathway were cloned in the 5'-3' orientation, it is suspected that new molecules are being made, which results in cell death. Given the death phenotype, it is impossible to test this *in vivo*. As such, it is hoped that the cell-free transcription and expression system will provide insight on this.

Nonetheless, it was decided to try to integrate the two version of this pathway - with or without *rci* - in S. *coelicolor* A3(2). This strain of Streptomyces does not carry the pikromycin pathway, despite carrying other natural product pathways. In order to be able to integrate it in this organism, the plasmid pSET-SC4 was used (Example 3), as it integrated into the attB site of the organism.

Shuffling in S. coelicolor A3(2)

Following the usual conjugation protocol (please see report#2), both pathways were successfully integrated into *S. coelicolor*. Interestingly, the strain containing the rci-less pathway (all genes *on*) did not produce prodigiosin (red pigment) or actinorhodin (blue pigment). On the other hand, the pathway carrying rci produced none, both or either molecules.

The pigment production phenotype was particularly interesting in the case of the white colonies, as it was possible to see red spots developing with a single colony and over time, in some of them.

It is thought that expression of the additional PKSs interferes with the proper assembly of actinorhodin and prodigiosin. When the recombinase was integrated together with the PKS-pathway, the resulting colonies displayed different phenotypes, in terms of pigment production. This further supports that Rci will be turning some of the genes *off*.

Materials and Methods

Preparation of reagents and S30 fraction for the cell-free system

The following reagents were prepared for this system: 0.1 M Tris-HCl pH 7.6, 10 mM $MgCl_2$, 10% Glycerol, 1 mM PMSF, 50 mM $CaCl_2$, 5 mM $CaCl_2$, 25 mM EGTA-KOH pH 7.0, 0.1 M Magnesium acetate, 5% TCA, 10% TCA, 1 M HEPES-KOH pH 7.5, 1 $MgCl_2$, 4 M NH_4Cl, 14.3 M B-mercaptoethanol, 0.55 M Dithiothreitol, 38 mM ATP pH 7.0, 50 mM Glycine-KOH, 88 mM CTP pH 7.0, 88 mM ATP pH 7.0, 88 mM UTP pH 7.0, 88 mM GTP pH 7.0, 0.42 M PEP pH 7.0, 55 mM Alanine, 55 mM Cysteine, 55 mM Aspartic acid, 55 mM Glutamic acid, 55 mM Phenylalanine, 55 mM Glycine, 55 mM Histidine, 55 mM Isoleucine, 55 mM Lysine, 55 mM Leucine, 55 mM Methionine, 55 mM Asparagine, 55 mM Pyrolysine, 55 mM Proline, 55 mM Glutamine, 55 mM Arginine, 55 mM Serine, 55 mM Threonine, 55 mM Selenocysteine, 55 mM Valine, 55 mM Tryptophan, 55 mM Tyrosine, 40% PEG 8000, 2000 U/mL Pyruvate kinase, 4.2 M Ammonium acetate, and 8.4 M Potassium acetate.

The sporulation medium, NEF, consisted of : 5 g glucose, 1 g yeast extract, 1 g casamino acids, 0.5 g beef extract, 20 g agar, and water to 1 L, pH 8.0 (1 M KOH).

The broth for liquid cultures, YEME, consisted of 3 g yeast extract, 5 g bacto-peptone, 3 g malt extract, 10 g glucose, 340 g sucrose, 5 g PEG 6000, and water to 1 L; 5mM (final) $MgCl_2 \cdot 6H_20$ after autoclaving.

TSB, tryptone soya broth (Oxoid) was also used.

S30 preparation protocol

Cells were grown in Tryptic Soy Broth with stainless steel springs – The original protocol asks for YEME supplemented with 5 g/L of PEG 6000. Yet, *S. venezuelae* was incapable of growing in this medium. Growing the cells in YEME without PEG 6000, and with stainless steel springs to prevent clumping was attempted. Nonetheless, no growth was observed. Next, growing the cells in TSB with springs was attempted. This medium had been previously used to grow *S. venezuelae* carrying the 4-color shufflon system.

The incubation of 1 L of medium was performed in 3 L baffled Fernbach flasks, but no growth was observed. It was hypothesized that the lack of growth could be due to over-oxygenation of the culture. The protocol for this system called for "vigorous shaking" but also stated said to grow the cells in "1 L in 2L [regular] flasks", which meant a lower air-to-surface ratio. It was thus decided to grow the spores in smaller flasks with less shaking: 4 x 500 mL in 4 x 1L flasks, at 170 rpm vs 250 rpm. This finally resulted in the sought-after mycelia.

The mycelium was harvested by centrifugation (7,000 rpm, 10 min, 4° C), washed twice with half a volume of washing buffer and once with half a volume of S30 buffer. The pellets were resuspended in S30 buffer and filtered through Whatman no. 1 paper, in a Buchner funnel. Upon dryness of the filter, it was placed in a 50 mL falcon tube to which S30 buffer with 10% glycerol was added. The tube was rotated to resuspend the mycelium from the filter.

A French-press located in a 4°C room was washed thoroughly with cold 30% ethanol, followed by cold sterile water and cold S30 buffer. After this, the sample was loaded and pressed at 12 x 103 psi into centrifuge tubes.

The A260 of the resulting extract - S30 - was measured and nuclease treatment was performed: 200 µL of S30 (40 - 60 A260 units) were mixed with 4.1 µL of $CaCl_2$ 50 mM and 2 µL of nuclease (150 units). After 30 minutes of incubation, the nuclease was inactivated by adding 2.6 µl of 25mM EGTA-KOH pH7.0.

The activity of the S30 fraction obtained was tested using *gusA*, a ß-glucuronidase gene that converts 5-bromo-4-chloro-3-indolyl glucuronide (X-Gluc) to a blue pigment (see Example 1): 40 µl S30 (2-3 A260 units), 32 µl S30 buffer, 8 µl 0.1 M magnesium, acetate (12mM final), 32 µl Synthesis mix, and 8 µl Plasmid DNA (8 µg) or water blank.

References for Example 4

1. Sun, J., Kelemen, G. H., Fernández-Abalos, J. M. & Bibb, M. J. Green fluorescent protein as a reporter for spatial and temporal gene expression in Streptomyces coelicolor A3(2). Microbiology 145 (Pt 9, 2221–7 (1999).
2. Vrancken, K. et al. pspA overexpression in Streptomyces lividans improves both Sec- and Tat-dependent protein secretion. Appl. Microbiol. Biotechnol. 73, 1150–7 (2007).
3. Kuščer, E. et al. Roles of rapH and rapG in positive regulation of rapamycin biosynthesis in Streptomyces hygroscopicus. J. Bacteriol. 189, 4756–4763 (2007).

Example 5

Production and analysis of the small molecules secreted to the growth medium

The PKS pathway integrated into the chromosome of *S. coelicolor* is made up of 4 genes, surrounded by sites (short DNA sequences) that are recognized by the recombinase Rci and flipped.

The strain carrying the pathway as well as Rci was grown on ISP2 medium, for recombination events to occur, and colonies displaying different phenotypes were picked and streaked on Supplemented Minimal Medium (SMMS)[1], for production of natural products. Strains carrying only Rci, only the pathway or neither, were also streaked on SMMS.

The plates were incubated at 30C for 5 days and the molecules were extracted by mixing the content of the plates (agar and cells) with 100% ethyl acetate. The mixture was left overnight at room temperature, with gentle shaking, to allow the molecules to transition from the agar to the organic solvent.

The next day the samples were centrifuged and the supernatants (organic solvent) collected. Aliquots were submitted for Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-ToF) analysis. The matrix used was alpha-cyano, having 1 µL been spotted and analyzed under positive ionization (performed at the Biopolymers and Proteomics Core Facility at the Koch Institute, MIT).

The molecules with mass values ranging between 150 and 800 Da detected by MALDI-ToF were analyzed using the software mMass[2] (Figure 1). The unique peaks detected in each test sample and not present in the control samples are listed on Table 1.

Given that all strains were grown in the same conditions and that multiple negative control strains were used, it can be stated that the mass values listed on Table 1 correspond to molecules produced by the pathway integrated, or in conjunction or as a consequence of the expression of such pathway. The random flipping of genes, mediated by Rci, leads to different sets of genes being expressed at a given time. With this, a larger diversity of molecules is expected to be assembled by the enzymes encoded in the pathway. Such diversity is reflected in the diversity of new masses found in each of the samples. The change in phenotype (production of pigmented antibiotics) is assumed to be associated with the production of new molecules, as a result of the gene shuffling. Thus, it was used for the selection of a few producer strains, which were analyzed by MALDI-ToF.

Table 31. List of unique mass values for each of the test samples. Values common to any of the 3 control samples were not included.

Unique mass values (Da)/strain				
1	2	3	4	5
224.0837	146.05	164.05	190.0624	212.1232
391.5259	155.1352	212.0677	191.091	331.45
392.55	173.05	379.4	212.1118	379.4
393.55	178.15	391.5874	229.1461	391.6
394.5847	179.0574	392.6	234.1499	392.6
442.55	189.9411	393.6	331.4636	393.6
474.45	205.15	394.6104	341.5563	394.6405
490.4678	207.15	407.6	379.4	406.6
496.4855	209.1937	408.6	381.6272	407.6
506.7	210.1552	409.6	391.6	408.6
507.65	212.0565	744.701	392.6	632.9
509.8539	218.15	753.9	393.6	633.8485
513.4	219.2096	766.65	394.65	
546.85	220.1934	791.8308	407.6016	
548.5405	221.25		408.6148	
564.05	223.15		409.65	
578.8	226.15			
599.2736	228.0627			
603.9796	231.0799			
606.45	234.1			
609.8783	246.1735			
612.2779	247.2			
623.7656	250.1			
626.2384	268.25			

Unique mass values (Da)/strain				
638.2	270.2188			
681.8	288.3295			
	333.3926			
	334.4			
	335.332			
	338.5507			
	341.55			
	358.6			
	379.4			
	381.6			
	383.6			
	648.3			

References for Example 5

1. Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. Practical streptomyces genetics. (John Innes Foundation, 2000).
2. Martin Strohalm. mMass - Open Source Mass Spectrometry Tool. (2013). at < www.mmass.org>

Sequences

Name	Sequence	SEQ ID NO:
RBS+tylG	CCGGCCGGCGGAGCACACCCGGCCGTCTCCGGCCCGGCCGCGGCCGGGCCGGAAGCCAT CCGCCGCCCACCCGGTACCGACCCCTCAAGCCCTTCAAGCCCTTCGACCCGTCCGATCAGT CAGTCCGGCGGTCCTCCACGACCGGTCCGGAATCGCCCCCACACGAGTCAGGAAGCACAC CATGGCCATGTCCGCCGAGAGGCTGACGGAGGCGCTGCGGACCTCGCTCAAGGAGGCCG AGCGGCTCCGGCGGCAGAACCGCGAACTGAGGGCCGCGCGGGACGCGGCGCGGGAGCCG ATCGCCGTCGTCGGCATGGCCTGCCGCTACCCGGGCGGTGTCACCGGCCCCGAGGAGCTG TGGGAGCTGGTGGCCGGAGGCCGGGACGCGATCGGGCCGTTCCCCGTGGACCGGGGCTG GGACGTGGCGTCGGTGTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGCCGGG AGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCGCCGC GCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAGGCGC TGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTACGTGG GCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAGGGCT ATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTCGGTC TCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTGCATC TGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTGGCG GTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCCGAC GGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTCGGC GTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGCCTG GTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAACGGC CCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAGGAC GTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCCGG GGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGCTC GCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAGAT GCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACCCC GCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGCAGCG GACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAATGCCCA TGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGCCGGAACCGGCCGGGGAGGCCCCCGG CGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGGTGGTCTCCGGACGCG ACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACCTCTCCCGCACCCCCG GGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCGCAGCCTTTGACCACCG CGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCCTGGACGCGTTGGCCGA AGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGACCGGGACGGCAGGATGGCCT TCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGGATGGCCCACGACCTGCATGCCGCCC ATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACGGACCGTCTCGACCCGCTGCTCGGCCG GCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGCTCGCCCGAAGCGGCACTCCTGGACCG GACCGAGTACACCCAGCCGGCGCTCTTCGCCGTCGAGGTGGCGCTCCACCGGCTGCTGGA	31

Name	Sequence	SEQ ID NO:
	GCACTGGGGGATGCGCCCCGACCTGCTGCTGGGGCACTCGGTGGGCGAACTGGCGGCCGC CCACGTCGCGGGTGTGCTCGATCTCGACGACGCCTGCGCGCTGGTGGCCGCCCGCGGCAG GCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATGGTCTCCGTGCGGGCCGGCGAGGACGA GGTCCGCGCACTGCTGGCCGGCCGCGAGGACGCCGTCTGCGTCGCCGCGGTGAACGGCCC CCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCG CCGGACGAGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCACGCCTTCCACTCACCCCTGA TGGACGGCATGCTCGCCGGATTCCGGGAGGTCGCCGCCGGCCTGCGCTACCGGGAACCGG AGCTGACGGTCGTCTCCACGGTCACGGGGCGGCCCGCCCGCCCCGGTGAACTCACCGGCC CCGACTACTGGGTGGCCCAGGTCCGTGAGCCCGTGCGCTTCGCGGACGCGGTCCGCACGG CACACCGCCTCGGAGCCCGCACCTTCCTGGAGACCGGCCCGGACGGCGTGCTGTGCGGCA TGGCAGAGGAGTGCCTGGAGGACGACACCGTGGCCCTGCTGCCGGCGATCCACAAGCCC GGCACCGCGCCGCACGGTCCGGCGGCTCCCGGCGCGCTGCGGGCGGCCGCCGCCGCGTAC GGCCGGGGCGCCCGGGTGGACTGGGCCGGGATGCACGCCGACGGCCCCGAGGGGCCGGC CCGCCGCGTCGAACTGCCCGTCCACGCCTTCCGGCACCGCCGCTACTGGCTCGCCCCGGG CCGCGCGGCGGACACCGACGACTGGATGTACCGGATCGGCTGGGACCGGCTGCCGGCTGT GACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTGATCCACCCCGACAGCCCGCGCTG CCGGGAGCTGTCCGGCCACGCCGAACGCGCGCTGCGCGCCGCGGGCGCGAGCCCCGTACC GCTGCCCGTGGACGCTCCGGCCGCCGACCGGGCGTCCTTCGCGGCACTGCTGCGCTCCGC CACCGGACCTGACACACGAGGTGACACAGCCGCGCCCGTGGCCGGTGTGCTGTCGCTGCT GTCCGAGGAGGATCGGCCCCATCGCCAGCACGCCCCGGTACCCGCCGGGGTCCTGGCGAC GCTGTCCCTGATGCAGGCTATGGAGGAGGAGGCGGTGGAGGCTCGCGTGTGGTGCGTCTC CCGCGCCGCGGTCGCCGCCGCCGACCGGGAACGGCCCGTCGGCGCGGGCGCCGCCCTGTG GGGGCTGGGGCGGGTGGCCGCCCTGGAACGCCCCACCCGGTGGGGCGGTCTCGTGGACCT GCCCGCCTCGCCCGGTGCGGCGCACTGGGCGGCCGCCGTGGAACGGCTCGCCGGTCCCGA GGACCAGATCGCCGTGCGCGCGTCCGGCAGTTGGGGCCGGCGCCTCACCAGGCTGCCGCG CGACGGCGGCGGCCGGACGGCCGCACCCGCGTACCGGCCGCGCGGCACGGTGCTCGTCA CCGGTGGCACCGGCGCGCTCGGCGGGCATCTCGCCCGCTGGCTCGCCGCGGCGGGCGCCG AACACCTGGCGCTCACCAGCCGCCGGGGCCCGGACGCGCCCGGCGCCGCCGGACTCGAG GCCGAACTCCTCCTCCTGGGCGCCAAGGTGACGTTCGCCGCCTGCGACACCGCCGACCGC GACGGCCTCGCCCGGGTCCTGCGGGCGATACCGGAGGACACCCCGCTCACCGCGGTGTTC CACGCCGCGGGCGTACCGCAGGTCACGCCGCTGTCCCGTACCTCGCCCGAGCACTTCGCC GACGTGTACGCGGGCAAGGCGGCGGGCGCCGCGCACCTGGACGAACTGACCCGCGAACT CGGCGCCGGACTCGACGCGTTCGTCCTCTACTCCTCCGGCGCCGGCGTCTGGGGCAGCGC CGGCCAGGGTGCCTACGCCGCCGCCAACGCCGCCCTGGACGCGCTCGCCCGGCGCCGTGC GGCGGACGGACTCCCCGCCACCTCCATCGCCTGGGGCGTGTGGGGCGGCGGCGGTATGGG GGCCGACGAGGCGGGCGCGGAGTATCTGGGCCGGCGCGGTATGCGCCCCATGGCACCGG TCTCCGCGCTCCGGGCGATGGCCACCGCCATCGCCTCCGGGGAACCCTGCCCCACCGTCA CCCACACCGACTGGGAGCGCTTCGGCGAGGGCTTCACCGCCTTCCGGCCCAGCCCTCTGA TCGCGGGGCTCGGCACGCCGGGCGGCGGCCGGGCGGCGGAGACCCCCGAGGAGGGGAAC GCCACCGCTGCGGCGGACCTCACCGCCCTGCCGCCCGCCGAACTCCGCACCGCGCTGCGC GAGCTGGTGCGAGCCCGGACCGCCGCGGCGCTCGGCCTCGACGACCCGGCCGAGGTCGCC GAGGGCGAACGGTTCCCCGCCATGGGCTTCGACTCCCTGGCCACCGTACGGCTGCGCCGC GGACTCGCCTCGGCCACGGGCCTCGACCTGCCCCCCGATCTGCTCTTCGACCGGGACACC CCGGCCGCGCTCGCCGCCCACCTGGCCGAACTGCTCGCCACCGCACGGGACCACGGACCC GGCGGCCCCGGGACCGGTGCCGCGCCGGCCGATGCCGGAAGCGGCCTGCCGGCCCTCTAC CGGGAGGCCGTCCGCACCGGCCGGGCCGCGGAAATGGCCGAACTGCTCGCCGCCGCTTCC CGGTTCCGCCCCGCCTTCGGGACGGCGGACCGGCAGCCGGTGGCCCTCGTGCCGCTGGCC GACGGCGCGGAGGACACCGGGCTCCCGCTGCTCGTGGGCTGCGCCGGGACGGCGGTGGC CTCCGGCCCGGTGGAGTTCACCGCCTTCGCCGGAGCGCTGGCGGACCTCCCGGCGGCGGC CCCGATGGCCGCGCTGCCGCAGCCCGGCTTTCTGCCGGGAGAACGAGTCCCGGCCACCCC GGAGGCATTGTTCGAGGCCCAGGCGGAAGCGCTGCTGCGCTACGCGGCCGGCCGGCCCTT CGTGCTGCTGGGGCACTCCGCCGGCGCCAACATGGCCCACGCCCTGACCCGTCATCTGGA GGCGAACGGTGGCGGCCCCGCAGGGCTGGTGCTCATGGACATCTACACCCCCGCCGACCC CGGCGCGATGGGCGTCTGGCGGAACGACATGTTCCAGTGGGTCTGGCGGCGCTCGGACAT CCCCCCGGACGACCACCGCCTCACGGCCATGGGCGCCTACCACCGGCTGCTTCTCGACTG GTCGCCCACCCCCGTCCGCGCCCCCGTACTGCATCTGCGCGCCGCGGAACCCATGGGCGA CTGGCCACCCGGGGACACCGGCTGGCAGTCCCACTGGGACGGCGCGCACACCACCGCCG GCATCCCCGGAAACCACTTCACGATGATGACCGAACACGCCTCCGCCGCCGCCCGGCTCG TGCACGGCTGGCTCGCGGAACGGACCCCGTCCGGGCAGGGCGGGTCACCGTCCCGCGCGG CGGGGAGAGAGGAGAGGCCGTGA	
RBS+momAVI	TCAGGAGTGGAGAAGACATGCCGGGTACGAACGACATGCCGGGTACCGAGGACAAGCTC CGCCACTACCTGAAGCGAGTGACCGCGGATCTCGGACAGACCCGTCAGCGCCTGCGCGAC GTGGAGGAGCGCCAGCGGGAACCGATCGCCATCGTCGCGATGGCCTGCCGCTACCCGGGC GGGGTGGCCTCCCCCGAGCAGCTGTGGGACCTGGTCGCCTCACGCGGCGACGCCATCGAG GAGTTCCCCGCCGACCGCGGCTGGGACGTGGCGGGCCTCTACCACCCCGACCCGGACCAC CCCGGCACGACCTATGTACGAGAGGCCGGATTCCTGCGGGACGCCGCCCGCTTCGACGCC GACTTCTTCGGCATCAACCCGCGCGAGGCGCTCGCCGCCGACCCGCAGCAACGGGTGCTC CTCGAAGTGTCGTGGGAACTGTTCGAGCGGGCGGGCATCGACCCCGCCACGCTCAAGGAC ACCCTCACCGGCGTGTACGCGGGGGTGTCCAGCCAGGACCACATGTCCGGGAGCCGGGTC CCGCCGGAGGTCGAGGGCTACGCCACCACGGGAACCCTCTCCAGCGTCATCTCCGGCCGC ATCGCCTACACCTTCGGCCTGGAGGGCCCGGCGGTGACGCTCGACACGGCGTGCTCGGCA TCGCTGGTCGCGATCCACCTCGCCTGCCAGGCCCTGCGCCAGGGCGACTGCGGCCTGGCG GTGGCGGGAGGCGTGACCGTACTGTCCACGCCGACGGCGTTCGTGGAGTTCTCACGCCAG CGCGGACTCGCACCGGACGGCCGCTGCAAGCCGTTCGCCGAGGCCGCCGACGGCACCGG ATTCTCCGAGGGCGTCGGCCTGATCCTCCTGGAACGCCTCTCCGACGCCCGCCGCAACGG	32

Name	Sequence	SEQ ID NO:
	ACATCAAGTACTCGGCGTCGTACGCGGATCGGCCGTCAACCAGGACGGCGCGAGCAACG GCCTGACCGCCCCGAACGACGTCGCCCAGGAACGCGTGATCCGCCAGGCCCTGACCAACG CCCGCGTCACCCCGGACGCCGTCGACGCCGTGGAGGCACACGGCACCGGCACCACGCTCG GCGACCCGATCGAGGGGAACGCACTCCTCGCGACGTACGGAAAGGACCGCCCCGCCGAC CGGCCGCTGTGGCTCGGCTCTGTGAAGTCGAACATCGGCCACACGCAGGCGGCTGCGGGC GTCGCAGGCGTCATCAAGATGGTGATGGCGATGCGCCACGGCGAGCTGCCCGCCTCCCTG CACATCGACCGGCCCACGCCCCACGTGGACTGGGAGGGCGGGGGAGTGCGGTTGCTCACC GATCCCGTGCCGTGGCCACGGGCCGACCGCCCCCGCCGCGCGGGGGTCTCCTCCTTCGGC ATCAGCGGCACCAACGCCCACCTGATCGTGGAACAGGCCCCCGCCCCGCCCGACACGGCC GACGACGCCCCGGAAGGCGCCGCAACCCCCGGCGCTTCCGACGGCCTCGTGGTGCCGTGG GTGGTGTCGGCCCGTAGTCCGCAGGCCCTGCGTGATCAGGCCCTGCGTCTGCGCGACTTTG CCGGTGACGCGTCCCGAGCGCCGCTCACCGACGTGGGCTGGTCTTTGCTGCGGTCGCGTG CGCTGTTCGAGCAGCGGGCGGTGGTGGCGGGGCGTGAGAGGGCTGAACTGCTGGCGGGG CTGGCTGCGTTGGCCGCTGGTGAGGAGCACCCGGCTGTGACGCGGTCCCGTGAGGAAGCG GCGGTTGCTGCGAGCGGTGATGTGGTGTGGCTGTTCAGTGGTCAGGGCAGTCAGTTGGTC GGTATGGGTGCTGGTTTGTATGAGCGGTTCCCGGTGTTTGCGGCTGCGTTTGATGAGGTGT GCGGCTTGCTGGAGGGGGAGCTGGGGGTTGGTTCGGGTGGGTTGCGGGAGGTGGTGTTCT GGGGCCCGCGGGAGCGGTTGGATCACACGGTGTGGGCGCAGGCGGGGTTGTTTGCGTTGC AGGTGGGGTTGGCCCGGTTGTGGGAGTCGGTCGGGGTGCGGCCGGATGTGGTGCTCGGGC ATTCGATCGGTGAGATCGCGGCCGCGCATGTGGCGGGGGTCTTTGATCTGGCGGATGCGT GTCGGGTGGTGGGGGCGCGGGCGCGTTTGATGGGTGGGTTGCCTGAGGGTGGGGCGATGT GTGCGGTGCAGGCCACGCCCGCCGAGCTGGCCGCGGATGTGGATGGCTCGTCCGTGAGTG TGGCGGCGGTCAACACACCTGACTCGACGGTGATTTCAGGTCCGTCGGGTGAGGTGGATC GGATTGCTGGGGTGTGGCGGGAGCGTGGGCGTAAGACGAAGGCGCTGAGCGTGAGTCAT GCTTTCCATTCGGCGTTGATGGAGCCGATGCTCGGGGAGTTCACGGAAGCGATACGAGGG GTCAAGTTCAGGCAGCCGTCGATCCCGCTCATGAGCAATGTCTCCGGAGAGCGGGCCGGC GAGGAGATCACATCCCCGGAGTACTGGGCGAGGCATGTACGCCAGACAGTGCTCTTCCAG CCCGGCGTCGCCCAAGTGGCCGCTGAGGCACGCGCGTTCGTCGAACTCGGCCCCGGCCCC GTACTGACCGCCGCCGCCCAGCACACCCTCGACCACATCACCGAGCCGGAAGGCCCCGAG CCGGTCGTCACCGCGTCCCTCCACCCCGACCGGCCGGACGACGTGGCCTTCGCGCACGCC ATGGCCGACCTCCACGTCGCCGGTATCAGCGTGGACTGGTCGGCGTACTTCCCTGACGAC CCCGCCCCCCGCACCGTCGACCTGCCCACCTACGCCTTCCAGGGGCGGCGCTTCTGGCTGG CGGACATCGCGGCGCCCGAGGCCGTGTCCTCGACGGACGGTGAGGAGGCCGGGTTCTGG GCCGCCGTCGAAGGTGCGGACTTCCAGGCGCTCTGCGACACCCTGCACCTCAAGGACGAC GAGCACCGCGCGGCTCTGGAGACGGTGTTCCCCGCGCTGTCCGCGTGGCGGCGCGAACGA CGTGAGCGGTCGATCGTCGATGCCTGGCGGTACCGGGTCGACTGGCGGCGCGTCGAGCTG CCGACACCCGTTCCGGGCGCCGGTACCGGTCCCGACGCCGACACGGGCCTCGGGGCGTGG CTGATCGTGGCTCCCACGCACGGGTCGGGTACTTGGCCGCAAGCCTGTGCCCGGGCGTTG GAGGAGGCGGGCGCGCCGGTACGTATCGTCGAGGCCGGCCCGCACGCCGACCGGGCGGA CATGGCGGACCTGGTCCAGGCATGGCGGGCAAGCTGTGCGGACGACACCACCCAGCTCG GAGGAGTGCTCTCCCTGCTGGCTCTCGCCGAGGCACCGGCCACCAGTTCCGACACCACTT CCCACACCAGTACCAGTTGCGGTACCGGCTCTCTCGCGTCCCACGGCCTCACCGGCACCTT GACGCTGCTGCACGGTCTGCTGGATGCGGGCGTCGAAGCGCCTCTCTGGTGTGCCACGCG CGGCGCCGTGTCGTGCGGCGACGCCGATCCGCTCGTCTCCCCGTCGCAGGCCCCGGTCTG GGGACTCGGACGCGTGGCCGCCCTGGAGCATCCGGAGTTGTGGGGCGGCCTGGTCGACCT GCCCGCCGACCCGGAGTCGCTCGACGCGAGCGCGTTGTATGCGGTTCTGCGCGGAGACGG CGGCGAGGATCAGGTCGCGCTGCGCCGGGGCGCGGTCCTCGGCCGTCGCCTGGTGCCCGA CGCAACCCCGGACGTGGCCCCCGGCTCGTCCCCGGACGTGTCCGGAGGCGCAGCCCATGC CGACGCGACCTCCGGGGAGTGGCAGCCGCATGGTGCCGTCCTCGTCACCGGAGGCGTCGG CCACCTGGCCGATCAGGTCGTACGGTGGCTCGCCGCGTCCGGCGCCGAACACGTCGTACT CCTGGACACGGGCCCCGCCAACAGCCGTGGTCCCGGCCGGAACGACGACCTCGCCGCGG AAGCCGCCGAACACGGCACCGAGCTGACGGTCCTGCGGTCCCTGAGCGAGCTGACAGAC GTATCCGTACGTCCCATACGGACCGTCATCCACACATCGCTGCCCGGCGAGCTCGCGCCG CTGGCCGAGGTCACCCCCGACGCGCTCGGCGCGGCCGTGTCCGCCGCCGCGCGGCTGAGC GAACTCCCCGGCATCGGGTCAGTGGAGACCGTGCTGTTCTTCTCCTCCGTGACGGCTTCGC TCGGCAGTAGGGAGCACGGCGCGTACGCCGCCGCCAACGCCTACCTCGACGCCCTGGCGC AACGGGCCGGTGCCGATGCTGCGAGCCCCCGGACGGTCTCGGTCGGGTGGGGCATCTGGG ATCTGCCGGACGACGGTGACGTGGCACGCGGCGCCGCCGGGCTGTCCCGGAGGCAGGGA CTCCCGCCGCTGGAACCGCAGTTGGCGCTCGGCGCCCTGCGCGCGGCGCTCGACGGGGGC AAGGGGCACACGCTGGTCGCCGACATCGAGTGGGAGCGGTTCGCGCCGCTGTTCACGCTG GCCAGGCCCACCCGGCTGCTCGACGGGATCCCCGCGGCCCAGCGGGTCCTCGACGCCTCC TCGGAGAGCGCCGAGGCCTCGGAGAACGCCTCGGCCCTCCGTCGCGAACTGACGGCCCTG CCCGTGCGGGAGCGGACCGGGGCACTTCTCGACCTGGTCCGCAAACAGGTGGCCGCCGTC CTGCGCTACGAGCCGGGCCAAGACGTGGCGCCCGAGAAGGCCTTCAAGGACCTGGGCTTC GACTCGCTCGTGGTCGTGGAGCTGCGCAACCGGCTGCGCGCCGCCACCGGGCTCCGGCTG CCCGCCACCCTGGTCTACGACTACCCCACACCCCGCACCCTCGCCGCACACCTGCTGGAC AGGGTGCTGCCCGACGGCGGCGCGGCAGAGCTCCCCGTGGCCGCCCACCTGGACGACCTG GAGGCGGCCCTCACCGACCTGCCGGCCGACGACCCCCGGCGCAAGGGCCTGGTCCGGCGT CTACAGACGCTGCTGTGGAAGCAGCCCGACGCCATGGGGGCGGCGGGCCCCGCCGACGA GGAGGAGCAAGCCGCGCCCGAGGACCTGTCGACCGCGAGCGCCGACGACATGTTCGCCC TGATCGACCGGGAGTGGGGCACGCGGTGA	
RBS+ amphK	TGACCCGGGCCCACCCCAACTACGTGCGAAACATCTGAGGAAGGTTCAACGGACGATGCC GGACGAAAGTAAGCTCGTCGATTACCTGAAGTGGGTCACGGCGGATCTCCACCAGACCCG CAGGCGCCTCCAGGAAGCCGAGTCGGGCCGCCACGAGCCCGTGGCGATCGTCGGCATGG CGTGCCGCTTCCCCGGCGGTGTGCGTTCCCCGGAGGACCTCTGGGAGATGCTCGCCGACG	33

Name	Sequence	SEQ ID NO:
	GCCGCGACGCCATCTCCGGGTTCCCCGCCGACCGCGGCTGGGACCTGGAGACGCTGGCCG GCGACGGAGCGGGCGGCAGCAGCACGCAGGAGGGAGGTTTCCTGCACGACGTGGCCGAC TTCGACCCGGGGTTCTTCGACATCTCGCCCCGCGAGGCGCTGGCCATGGACCCGCAGCAG CGGCTGCTCCTGGAGACCGCCTGGGAGGCCGTGGAGCGTGCCGGGATCGCCCCCGGCAGC CTCCGCGGCAGCCGCACCGGCGTGTTCGTCGGCACCAACTCGCAGGACTACGCCCACCTC GTCCTCGCCTCCGACGACGACATGGGCGGCTACGCGGGCAACGGCCTGGCCGCCAGCGTG ATGTCCGGCCGGCTGTCCTTTGCGCTCGGCTTCGAAGGACCCGCCGTCACCCTCGACACCG CGTGCTCCTCGGCCCTGGTCGCCCTGCACCTGGCCGCCCAGTCGGTGCGCTCCGGCGAGG CCGACCTCGCGCTGGCCGGCGGTGTCACCGTCATGACCACCTCGTCCAGCTTCGTGGGCTT CAGCCTCCAGGGCGGCCTCGCCACGGACGGCCGCTGCAAGGCGTTCGCCGACTCCGCCGA CGGCACCGGCTGGTCCGAGGGCGTCGGCATGATCCTCGTCGAACGGCTCTCCGAGGCCCG GCGCAAGGGCCACCCGGTGCTCGCCGTGCTGCGCGGCTCCGCCGTGAACCAGGACGGCGC CTCCAACGGCCTCAGCGCCCCCAACGGGCCCGCCCAGCAGCGCGTCATCCGCGACGCGCT GGCCTCCGCCGGGCTCTCGCCCGCCGACGTCGACGCCGTCGAGGCGCACGGCACCGGCAC CACCCTCGGCGACCCCATCGAGGCGCAGGCCCTGCTCGCCACCTACGGCCAGGACCGCGA CGCGAGCCGTCCGCTGCGCCTCGGCACGGTGAAGTCCAACATCGGCCACACCCAGGCCGC CGCCGGCGCCGCGGGCGTCATCAAGATGGTCCTGGCCCTGCGCCACGGGCTGCTGCCCCG CACCCTGCACATCGACGCGCCGTCCACACACGTCGACTGGGACGCAGGACACGTCAGCCT GCTCACCGAGGCCACCCCCTGGCCCGAGGGCGAGCAGACGCGCCGGGCCGGTGTGTCGTC CTTCGGCATCAGCGGCACCAACGCCCATGTCATCCTCGAAGAGGCCCCCGCGGCCGAGGA GGACACCGACGCCGAGCAGCGGCCCGAGCCGGTCGTCCCGGGCGCCGTGCCGTGGCCCGT GTCGGCCCGTACCGCCGACGCCCTCGACGCCCAGCTGGAGAAGGTGCGGCCGCTCGCCGC GTCCGGCGCCGACCCCGTCGCCGTCGGCCACGCCCTGGCCGTCACCCGCACCCCCTTCGA ACACCGGGCGCTGCTGGTCGCCGCGGACGGCAATCTCACCGAGGCCGCACGCGGCACCGT CCCCTCCGGCGACCGCCCCGGGCTGGCCGTGCTGTTCTCCGGGCAGGGCGCCCAGCGCCT CGGCATGGGCCGTGAACTCCACGCCCGCTTCCCGGTGTTCGCCGCCGCCCTGGACGAGAC CCTCGCCCTCCTCGACGAGCGCCTCGGCCACTCGCTGCGCGACGTCATCTGGGGCGAGGA CCCCGTGGCCCTCGACGACACCGGCCACACCCAGCCCGCGCTGTTCGCCGTGGAGGTCGC CCTCTACCGCCTGTTCGCCTCCTGGGGTCTGCGCCCCGACCACCTCGCCGGACACTCCGTC GGGGAGATCGCCGCCGTGCATGTGGCCGGCGCACTGTCCCTGGAGGACGCCTGCACCCTG GTTGCCGCCCGCGCCGGCCTGATGCGGGACCTGCCGGCCGGCGGAGCCATGGTCGCGCTG CGCGCCACCGAGGCCGAGGTGCGGCCGCTGCTCGACGAGCGGGTCTCGCTCGCCTCGGTC AACGGCCCCGATTCCGTGGTGGTCTCCGGCGCCGAGGACGCCGTCCTCGCCGTCGCCGAG CACTTCCGGAAGCAGGACCGGCGCACCACGCGGCTCTCCGTGAGCCACGCCTTCCACTCC CCGCTCGTCGACCCGATGCTCGACGCCTTCCGGGACGTCGTCGCGAAGCTCACCTTCGGC GAGCCGTCGGTACCCGTCGTCTCCACCCTCACCGGCGACGTCGTCGCCGCCGAGGAACTG GCCACGCCGCACTACTGGGTGTGCCACGCCCGGCAGGCCGTCCGCTTCGCCGACGCCGTG CGCACCCTCGTCGACGAGGGCGCCCGGACCTTCCTGGAGGTCGGCCCGGGCGGCGTGCTG TCCGCGCTCGTCGGTGAGAACACCCAGGAGGCCGGGGTCTCTGCCGTGCCCGCGCTGCGC AAGGACCAGCCGGAGGAGGCCTCCGTGCTCGCCGCCCTCGGCACCCTGTGGACCCGGGGC ACCGCGGCCGACTGGGACGCCGTGTTCGAGGGCACCCTGCCCGGCCGCCCGGAGTCCGTG GACCTGCCGACCTACGCCTTCCAGCGCGGCCGGTACTGGCCCACCGTGCGCGCCCGCTCC GGCGACCCCGCCGGACTGGGCCTCGGCGCCGCCGGGCACCCGCTGCTCAGCGCCACCGTC ACCCTCGCCGAGGCCGACGAGTCCGTCCTCACCGGACGGCTCTCGCCGCTGACCCACCCC TGGCTCACCGAGCACCGCGTCGACGGCCGGATCACCGTCCCCGGCACGGCCCTCGTGGAG ATCGCGGTCCGTGCGGGCGACGAGAACGGCACCCCCCGCCTGGACCGGCTCGACCTGCTC GCCCCGCTGACCCTCGGCGACCGCGACGCGGTGCTGCTCCAGGTGCGCGTCGGCCCCGAG GACGCCTCCGGCCGGCGGACGCTCTCCGTGCACGCCCGCCCGGCGACCGCGGACGACGCC CCCTGGACCACGTACGCCCGTGGTGTCCTCGCCCCCGACGAGGGCGCCGAGGACACCGCC GCGGACCTTGTGGAGTGGCCGCCCGCCGACGCCCGTCAGGTCCCGCTCACGGAGCTGGAG TCCGAAGGACGCGCCCTCGGCCCGCTCTTCAGCGGCCTCACCGGGGTGTGGCGGCACGAG GGCGAGGTGTTCGCCGAGGCCGAGCTGCCCGCAGGACCCGACAGCGGCTTCGGACTGCAC CCCGCGCTGCTGTCCACCGCCCTGCGCGCCGCCGCCGTCCTCGACGGCACCACGGCCGGC GAGCCCGCCGCCTTCGAGGGGCTCACCCTGCACGCCACCGGCGCCACCGCGCTGCGCGTA CGGCTGAGCACCACCGGGCCCGACACCGTCGACCTCACCGCCGTCGACCCCGCGGGCAAC CTCGTCCTGACCGCCGAGACCGTGCGTCTCGGCACCCCGGACGGCACCGCCGACACCCCG GCCGCCACCGGCCGGGGCGACCTGTTCGGGCTGAAGTGGGTGCCCGTCAAGGCCTCGGAG CGCGCCACCGGCACCCGCTGGGCCGTCGTCGGCTCCGACGAACTCGACCTCGGCTACGCC ATGCACCGCGCCGACGAGACCGTCACCGCCTACGCGGAGTCCCTCGGCGGCGCCATCGGC GACAGCGGGGTCGCCCCCGATGTGTTCCTGATCCCGCTCGCAGGCGAGAAGGACGCCGGA GCGGAGGGCGTGCACGCCCTCACCACCCGGGTCCTCGGCTACCTCCAGGAGTGGCTGTCC GAGCCCCGGCTCTCCGGCACCCGCCTGGTCTTCGTCACCCGCCGCGCCGTCGCCCTCGACG ACGAGGACGTCCTGGACCCGGCCGGCGCGGCCGTCTGGGGCCTGGTGCGCTCCGCGCAGA CCGAGAACCCCGGCAGCCTGCTGCTCGTCGACCTCGACGACACCTTCCTGTCCGCCGGAG TACTGCCCGACGTGCTGACGCTCGACGAGCAGCAGCTCGCGGTCCGCGACTACCAGGTCC GCGCCGCCCGGCTCGCCCGGCTGCCGCGCCCGGCCGACGACGCCCCCGCCGCCGACTGGA ACCCCGACGGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTCGGCGCCGCGCTCGCCC GGCACCTGGTCACCAGCCGGGGCGCCCGTCACCTGCTGCTGGCGAGCCGCCGCGGCCCGG ACGCGCCCGGCACGTCCGAACTGGTCGCCGAGCTGACCGGCCTGGGAGCCCAGGTGACCG TCTCCGCCTGCGACGTCGGCGACCGCGACGCGGTCGACACGCTCGTCGCGTCGGTGCCCG CCGAGCACCCGCTGACCGCCGTCGTGCACACCGCGGGTGTCCTGGACGACGCCCTGACCG GCTCGCTCACCCCGGAACAACTGGCTGGTGTGCTGCGCCCCAAGGCCGACGCGGCCCTGC ACCTGCACGAGGCGACCCTCGGCCAGGACCTCGCCGCGTTCGTGCTGTACTCCTCGATCTC CGGCGTCATCGGCGGCCCCGGCCAGGCCAACTACGCCGCCGCCAACGCCTCCCTGGACGC GCTCGCCCACCGGCGCAGGGCGGCCGGTCTGCCCGGTCTGTCCCTCGCCTGGGGGCCCTG	

Name	Sequence	SEQ ID NO:
	GGGCCGCGGCAGCGGCATGACCAGCCAGGTCAGCGACACCGACCTGGGGCGGATGGACC GCGGCGGCACCCCGCCGATGAGCCTCGAGGACGGCCTCGCCCTGTTCGACGCCGCCCTGG CCCGCACCGAGCCGATGGTCGTGCCCACCCGGATCAATGTCACCGGACTCCAGGTCCAGC AGACGCTGCCCGCGCTCTGGCGCGACCTGGTGCCGCGCCCCCGGCGCACCGCCGCCGCGG ACCGCTCGCCCAAGACCGTGCTCGACGGGCTGCGCACGCTCGACACCGCGGGCCGGGAG AAGCTGCTCACCGAGCTGGTCGTCGGCTTCACCGCGGGCCTGCTCGGCCACGCCGACCCC GCCGCCGTCGACCCCGAGCGCGGCTTCCTGGAGCTGGGCTTCGACTCGCTGGTCTCGGTC AGCCTGCGCAACCAGCTCGGCGAACTCCTCGGGCTGCGCCTGCCGACGTCGGTGGTCTTC GACAGCAAGACGCCGGTGAAGCTGGCCCGCCACCTCAACGAGGAACTGGGCGACCTGTC CGCCTCCGGCCCCGCGTCCGGCACCGCCGTCGCCGGCACCACGGTGCATCCCGACGACAC CCTGGTCGGCCTGTTCCACAACGCGGTGCGCGGCGGCAAGCTCGTCGAGGCGATGCGGAT GCTGAAGGCGGTCGCCAACACCCGGCCCACCTTCGAGACGCCCGCGGACCTGGAGGAGCT GTCCGAGGCCGTCACCCTGGCCACGGGGCCCGGCTCCCCGCGGCTGATCTTCGTCAGCGC GCCCGGCGCCACCGGCGGTGTCCACCAGTACGCGCGCATCGCCGCCCACTTCCGCGGCAA GCGCCATGTGTCGGCGATACCGCTGATGGGCTTCGCCCCCGGCGAGCTGCTGCCCGCCAC CAGTGAGGCGGCGGCCCGGATCGTCGCCGAGAGCGTCCTGATGGCGAGCGACGGCGAAC CGTTCGTGATGGTGGGCCACTCCACCGGCGGTTCCCTCGCCTACCTCGCGGCCGGGGTGCT GGAGGACACCTGGGGCGTGAAGCCGGAGGCGGTCGTCCTGCTCGACACGGCGTCCATCCG GTACAACCCCTCGGAGGGCAACAACCTCGACCAGACGACCCGCTTCTACCTCGCCGACAT CGACTCGCCGTCCGTGACGCTCAACAGCGCCCGGATGTCCGCGATGGCGCACTGGTTCAT GGCGATGACGGACATCGACGCGCCCGCCACGACCGCCCCCACCCTGCTCCTGCGCGCCAC ACAGGCCAACAACGGCTTCATGCTGGACACCTCCGCGGTGCCGGCGGACGTGGTGCGTGA CATCGAGGCCGACCATCTGTCCCTGGCCATGGAGCACTCGGATCTGACCGCCGAGGCCAT AGAGAACTGGCTCGCCGAACTTCCGGCCGGCGAGGCCTGA	
RBS+momAVIII	GCAGCGATCAGTCCTGAGCGCGATGAGCCCCGGCCACACCGGCCCCCTTGCCCTCCCCCA ACCTCATACGCCCCGATCACCACGTAGCGCCAAGGAGCCTGGGTCAGATGTCGAACGAGG AGAAGCTTCTCGATCACCTCAAGTGGGTCACCGCGGAGCTGCGCCAGGCCCGGCAACGGC TCCACGACAAGGAATCGACCGAGCCGGTCGCCATCGTCGGCATGGCCTGCCGCTACCCGG GCGGCGCCCGGTCGGCGGAGGACCTCTGGGAACTCGTGCGCGACGGGGGTGACGCGGTC GCGGGGTTCCCCGACGACCGGGGCTGGGACCTGGAGTCGCTGTATCACCCGGATCCGGAG CATCCGGCCACCAGCTATGTGCGGGACGGCGCCTTCCTGTACGACGCCGGCCATTTCGAC GCCGAGTTCTTCGGCATCAGCCCGCGCGAGGCCACGGCGATGGATCCGCAGCAGCGGCTG CTCCTGGAGACCGCGTGGGAGGCGATCGAGCACGCGGGCATGAACCCGCACGCGCTGAA GGGCAGCGACACCGGCGTCTTCACCGGAGTGAGCGCCCACGACTACCTGACGCTGATCAG CCAGACGGCCAGCGACGTCGAGGGGTACATCGGCACCGGCAACCTCGGCAGTGTGGTGTC GGGCCGGATCTCCTACACGGTGGGGCTCGAAGGTCCCGCGGTCACCGTCGACACGGCGTG CTCGTCGTCCCTGGTGGCGATCCATCTGGCAAGTCAGGCGCTGCGGCAGGGCGAGTGCTC GCTCGCGCTGGCGGGCGGTTCGACGGTGATGGCGACGCCGGGTTCGTTCACCGAGTTCTC CCGGCAGCGGGGGCTCGCGCCCGACGGGCGGTGCAAGCCGTTCGCGGCCGCCGCCGACG GCACCGGCTGGGGCGAGGGCGCCGGGGTGGTGGCGCTGGAGCTGCTCTCCGAAGCGCGG CGCCGCGGCCACAAGGTCCTTGCGGTGATACGGGGTTCGGCCACCAACCAGGACGGTACG AGCAACGGGCTCGCCGCCCCCAACGGCCCCTCCCAGGAACGCGTCATCCGCGCCGCCCTC GCCAACGCCCGCCTCTCCGCCGAGGACATCGACGCCGTAGAGGCGCACGGCACCGGCACC ACGCTCGGCGACCCCATCGAGGCGCAGGCCCTGATCGCCACGTACGGCCAGGGGCGCCCG GAGGACCGTCCGCTCTGGCTCGGCTCGGTGAAATCCAACATCGGCCACACGCAGGCCGCC GCCGGTGTCGCGGGCGTCATCAAGATGGTCATGGCGATGCGCAACGGTCTCCTCCCGACC TCCCTGCACATCGACGCCCCGTCACCGCACGTCCAGTGGGAACAGGGCAGCGTACGACTG CTCTCCGAGCCCGTGGACTGGCCGGCGGAGCGCACGCGGCGGGCAGGCATCTCGGCGTTC GGGATCTCCGGGACGAACGCGCACCTGATTCTGGAGGAGGCGCCGCCGGAAGAGGACGC GCCGGGCCCCGTGGCGGCTGAGCCGGGTGGGGTCGTGCCGTGGGTGGTGTCCGGGCGGAC GCCGGACGCGTTGCGTGAACAGGCGCGGCGGCTGGGCGAGTTCGCGGCCGGGCTGGCGG ATGCGTCGGTCTCCGAGGTGGGCTGGTCGCTGGCCACGACACGCGCGTTGTTCGATCAGC GGGCCGTGGTCGTGGGGCGGGACTTGGCTCAGGCTGGTGCCAGCCTGGAGGCATTGGCCG CGGGTGAGGCGTCGGCGGATGTGGTGGCCGGGGTGGCCGGTGATGTGGGTCCTGGGCCG GTGTTGGTGTTTCCGGGGCAGGGGTCGCAGTGGGTGGGCATGGGCGCCCAGCTCCTTGAC GAGTCGCCCGTCTTCGCGGCGCGGATCGCGGAGTGTGAGCAGGCGCTGTCGGCGCATGTG GACTGGTCGCTGAGTGATGTCTTGCGCGGGGACGGGAGCGAGCTGTCCCGGGTCGAGGTC GTGCAGCCCGTGCTGTGGGCGGTGATGGTCTCGCTGGCTGCGGTGTGGGCCGATTACGGC ATCACCCCGGCTGCCGTCATCGGGCACTCGCAGGGCGAGATGGCTGCCGCGTGTGTGGCG GGGGCACTGTCACTGGAGGATGCGGCACGGATCGTAGCCGTACGACAGCGACGCTTCGT CAGCTGCAAGGGCACGGCGACATGGCCTCGCTCAGCACCGGTGCCGAGCAGGCCGCAGA GCTGATCGGCGACCGGCCGGGCGTAGTCGTCGCGGCCGTCAACGGACCGTCGTCGACCGT GATTTCGGGCCCGCCGGAGCATGTGGCAGCCGTCGTCGCCGATGCGGAGGCGCAGGGACT CAGGGCCCGCGTCATCGACGTCAGGTACGCCTCCCACGGTCCCCAGATCGACCAGCTCCA CGACCTCCTCACCGACCGGCTCGCCGACATCCAGCCGACCACCACCGACGTGGCGTTCTA CTCGACGGTCACCGCAGAGCGCCTCGACGACACCACCGCCCTAGACACCGCCTACTGGGT CACCAACCTCCGCCAGCCCGTCCGCTTCGCCGACACCATCGAAGCGCTGCTGGCCGACGG CTACCGCCTGTTCATCGAGGCCAGCCCCCACCCCGTCCTCAACCTCGGCATCCAGGAGAC CATCGAGCAGCAGGCCGGTGCTGCGGGGACGGCCGTCACCATCCCCACCCTGCGTCGCGA CCACGGTGACACCACCCAGCTCACCCGCGCGGCCGCCCACGCCTTCACCGCCGGCGCCCC CGTCGACTGGCGGCGCTGGTTCCCGGCCGACCCCACCCCCCGTACCGTCGACCTCCCCACC TACGCCTTCCAGCACAAGCACTACTGGGTGGAGCCGCCCGCGGCGGTCGCAGCCGTGGGT GGTGGGCACGATCCGGTCGAGGCCCGGGTGTGGCAGGCGATCGAGGACCTGGACATCGA CGCCCTCGCCGGCAGTCTGGAGATCGAGGGGCAGGCGGAGAGCGTCGGAGCGCTGGAGT CCGCGCTGCCCGTCCTCTCGGCCTGGCGGCGTCGGCACCGCGAGCAGTCCACCGTCGACT	34

Name	Sequence	SEQ ID NO:
	CCTGGCGTTATCAGGTCACTTGGAAGCATCTGCCCGACGTGCCGGCGCCGGAGCTCAGCG GGGCCTGGCTGCTGCTCGTGCCCGCCGCGCACGCCGACCACCCGGCCGTCCTCGCGACCG CGCAGACGCTGACCGCCCATGGTGGCGAGGTGCGACGCCACGTGGTCGACGCACGTGCCA TGGAGCGTACGGAGTTGGCGCAGGAGCTGCGTGTCCTGATGGACGGGGCCGCGTTTGCCG GAGTCGTCAATCTGCTGGCCCTGGACGAGGAGCCGCATCCCGAGCACTCGGCCGTGCCCG CCGGACTCGCCGCGACGACCGCTCTCGTCCAGGCCCTCGCGGACAACGGCGCCGACATCG CCGTACGCACTCTCACGCAGGGGGCTGTTTCCACGAGCGCCGGCGACGCCCTCACCCACC CGGTGCAGGCTCAGGTGTGGGGGCTGGGGCGCGTCGCTGCGCTGGAGTATCCGCGGCTGT GGGGCGGGCTGGTCGATCTGCCCGCTCGTATCGACCATCAGACGCTGGCCCGGCTGGCCG CCGCGCTGGTTCCGCAGGACGAGGACCAGATCTCCATCCGGCCGTCCGGCGTCCATGCCC GTCGCCTTGCACACGCGCCCGCCAACACGGTCGGCAGCGGGCTTGGTTGGCGGCCCGACG GCACCACTCTCATCACCGGCGGGACCGGCGGCATCGGCGCCGTCCTCGCGCGGTGGCTCG CCCGTGCGGGCGCCCCGCACCTCCTCCTGACCAGCCGCCGCGGCCCCGACGCCCCGGGAG CACAGGAACTCGCCGCGGAACTGACGGAGTTGGGGGCCGCCGTCACCGTCACCGCCTGCG ACGTCGGCGACCGCGAGCAGGTGCGACGCCTCATCGACGATGTCCCCGCCGAGCACCCGC TGACCGCCGTCATCCACGCGGCCGGCGTGCCGAACTACATCGGTCTCGGCGACGTGTCGG GTGCCGAGCTGGACGAGGTGCTGCGTCCGAAGGCGCTCGCCGCTCACCATCTGCATGAAC TGACCCGGGAGTTGCCGCTCTCGGCGTTCGTGATGTTCTCGTCGGGCGCAGGCGTGTGGG GCAGTGGCCAGCAGGGCGCCTATGGTGCGGCCAACCACTTCCTCGACGGCCCTCGCCGAGC ACCGCCGCGCCGAGGGCCTGCCCGCCACCTCCATCGCCTGGGGGCCCTGGGCCGAGGCGG GCATGGCGGCGGACCAGGCCGCGTTGACGTTCTTCAGCCGCTTCGGCCTGCACCCGCTCA GCCCGGAGCTGTGCGTCAAGGCGCTGCAGCAGGCCCTGGACGCGGGTGAGACGACGCTG ACCGTGGCGAACTTCGACTGGGCGCAGTTCACGTCGACGTTCACCGCGCAGCGGCCCAGC CCGCTCCTCGCCGATCTGCCCGAGAACCGGCGGGCCAGCGCACCCGCGGCACAGCAGGA AGACGCCACGGAGGCATCGTCGCTCCAGCAGGAGCTGACGGAGGCGAAGCCGGCGCAGC AGCGGCAGTTGCTGCTGCAGCACGTGCGCTCCCAGGCGGCGGCCACGCTGGGGCACTCGG ACGTCGACGCGGTGCCCGCCACCAAGCCGTTCCAGGAGCTGGGCTTCGACTCGCTGACCG CGGTGGAGCTGCGCAACAGGCTGAACAAGAGCACCGGCCTGACACTGCCGACCACGGTC GTCTTCGACCACCCCACCCCCGACGCGCTCACCGACGTCCTGCGGGCCGAGCTGTCCGGT GACGCGGCGGCCTCCGCCGACCCGGTGCGGGCGGCCGGGGCCTCCAGGGGCGCCGCCGA CGACGAGCCGATCGCGATCGTCGGCATGGCCTGCCGCTATCCGGGCGACGTCCGCTCCGC CGAGGAGCTGTGGGATCTGGTCGCGGCCGGCAAGGACGCCATGGGGGCCTTCCCCGACG ACCGGGGCTGGGACCTGGAGACGCTGTACGACCCGGACCCGGAGAGCCGCGGCACCAGC TATGTGCGCGAAGGCGGGTTCCTCTACGACGCGGGCGACTTCGACGCCGGATTCTTCGGC ATCAGCCCCCGCGAGGCCGTCGCGATGGACCCGCAGCAGCGGCTGCTCCTGGAGACCGCG TGGGAGGCGATCGAACGCGCGGGCCTCGACCGGGAGACCCTCAAGGGCAGCGACGCCGG GGTGTTCACGGGCCTGACCATCTTCGACTACCTCGCGCTCGTCGGTGAACAGCCCACCGA GGTCGAGGGCTACATCGGCACCGGCAACCTCGGCTGTGTCGCCTCCGGCAGGGTGTCGTA CGTACTCGGCCTAGAAGGCCCCGCCATGACGATCGACACCGGCTGCTCTTCGTCCCTGGT GGCGATCCACCAGGCGGCGCACGCGCTGCGCCAGGGCGAGTGCTCGCTCGCTCTCGCGGG CGGCGCGACGGTGATGGCCACGCCGGGCTCGTTCGTCGAGTTCTCGCTGCAGCGCGGGCT CGCCAAGGACGGCCGGTGCAAGCCGTTCGCGGCCGCCGCCGACGGCACCGGCTGGGCCG AGGGAGTCGGCCTGGTCGTACTCGAACGGCTCTCGGAGGCCCGGCGCAACGGCCACAAC GTCCTGGCGGTGATCCGGGGTTCGGCCATCAACCAGGACGGCACGAGCAACGGGCTCACG GCACCCAACGGGCAGGCGCAGCAGCGGGTGATCCGGCAGGCACTCGCCAACGCGCGGCT CTCCGCCGAGGACGTCGACGCGGTGGAGGCGCACGGCACCGGCACCATGCTGGGCGACC CCATCGAAGCCAGCGCGCTCGTCGCCACCTACGGCAAGGAGCGGCCCGCGGACCGGCCG CTGTGGCTCGGCTCGATCAAGTCGAACATCGGGCACGCGCAGGCGTCGGCCGGTGTCGCC GGTGTGATCAAGATGGTCATGGCGCTGCGCAACGAACAGCTGCCCGCCTCCCTGCACATC GACGCGCCCACGCCGCACGTGGACTGGGACGGCAGTGGCGTCCGCCTGCTGTCCGAACCG GTCTCCTGGCCGCGCGGCGAACGCCCGCGCCGCGCCGGGGTGTCCGCCTTCGGCATCTCC GGCACCAACGCGCACCTCATCCTGGAACAGGCCCCGGACGCGCCGGAGCCCGTGACCGCT CCGGCGGAGGACGCCGCGGCGCCGGCCGGAGTGGTGCCCTGGGTGGTGTCGGCGCGCGG CGAGGAGGCGCTGCGGGCCCAGGCCCGCCTGCTGGCCGACCGCGCCACCGCCGACCCGC GGCTCGCGTCGCCGCTGGACGTGGGCTGGTCCCTGGTCAAGACCCGGTCGGTGTTCGAGA ACCGAGCCGTCGTCGTGGGCAAGGACCGCCAGACTCTCCTCGCCGGGCTACGGTCCCTGG CGGCGGGCGAGCCGTCACCGGACGTCGTCGAGGGCGCCGTACAGGGCGCCTCCGGCGCG GGTCCGGTGTTGGTGTTTCCGGGGCAGGGGTCGCAGTGGGTGGGCATGGGTGCCCAGCTC CTTGACGAGTCCCCCGTCTTCGCGGCGCGGATCGCGGAGTGTGAGCGGGCCCTGTCGGCG CATGTGGACTGGTCGCTGAGTGCGGTGTTGCGCGGGGACGGGAGTGAGCTGTCCCGGGTC GAGGTCGTGCAGCCCGTGTTGTGGGCGGTGATGGTCTCGCTGGCTTCGGTGTGGGCCGAT TACGGCATCACCCCGGCTGCCGTCATCGGGCACTCGCAGGGCGAGATGGCCGCCGCGTGT GTGGCGGGGGCACTGTCACTGGAGGATGCGGCGCGGATCGTGGCCGTACGCAGTGACGC GCTCCGTCAGCTCATGGGGCAGGGCGACATGGCGTCGTTGGGGGCCGGCTCGGAGCAGGT TGCTGAGCTCATCGGCGACCGGCCCGGCGTGTGTGTCGCTGCCGTCAACGGGCCCTCCTCT ACGGTCATTTCAGGGCCGCCGGAGCATGTGGCAGCCGTGGTCGCGGATGCGGAGGCGCG AGGCCTGCGCGCCCGCGTCATCGACGTCGGATACGCCTCCCACGGCCCCCAGATCGACCA GCTCCACGACCTCCTCACCGAGCGCCTGGCCGACATCCGGCCCACGACCACGGACGTCGC CTTCTACTCCACGGTCACCGCCGAACGCCTCGACGACACCACCACCCTCGACACGGATTA CTGGGTCACCAACCTCCGCCAGCCCGTCCGCTTCGCCGACACCATCGAAGCGCTGCTGGC CGACGGCTACCGCCTGTTCATCGAGGCCAGCCCCCACCCCGTCCTCAACCTCGGCATGGA GGAGACCATCGAGCGGGCCGACATGCCCGCCACCGTCGTGCCCACCCTGCGCCGCGACCA CGGCGACGCCGCGCAGCTCACCCGCGCGGCCGCCCAGGCCTTCGGCGCGGGGGCGGAGG TCGACTGGACGGGCTGGTTCCCGGCCGTGCCGCTGCCTCGGGTGGTGGATCTGCCGACGT ACGCCTTCCAGCGGGAGCGGTTCTGGCTGGAGGGGCGCAGGGGGCTCGCCGGGGACCCG	

Name	Sequence	SEQ ID NO:
	GCGGGGCTCGGGCTCGCGTCCGCGGGGCATCCGCTGCTCGGAGCCGCCGTGGAACTCGCG GACGGCGGCAGTCACCTGCTGACCGGCCGGATCTCTCCGCGGGACCAGGCGTGGCTGGCC GAGCACCGGGTCATGGACACGGTGCTGCTGCCCGGTTCGGCGTTCGTGGAGCTCGCGCTG CAGGCCGCGGTGCGGGCCGGCTGCGCGGAGTTGGCGGAGCTGACGCTGCACACTCCGCTC GCCTTCGGGGACGAGGGTGCGGGCGCGGTCGACGTGCAGGTGGTGGTCGGTTCCGTGGCC GAGGACGGGCGGCGTCCCGTGACCGTCCATTCGCGGCCCACGGGTGAGGGCGAGGAGGC CGTGTGGACCCGGCATGCCGCGGGCGTGGTCGCTCCCCCGGGGCCCGACGCCGGGGACGC CTCGTTCGGCGGGACGTGGCCGCCCCCGGGCGCCACACCGGTCGGCGAGCAGGATCCGTA CGGGGAACTCGCTTCGTACGGCTATGACTTCGGGCCCGGCTCACAGGGACTGGTGAGCGC GTGGCGGCTCGGGGACGACCTTTTCGCCGAGGTGGCGCTGCCCGAGGCGGAGAGCGGCA GGGCCGACCGCTACCAGGTGCACCCGGTGTTGCTCGACGCCACGCTGCACGCGCTGATCC TGGACGCGGTCACGTCGTCCGCCGACACCGACCAAGTGCTGCTGCCGTTCTCCTGGAGCG GGTTGCGGGTGCACGCGCCGGGCGCTGAGAAGTTGCGGGTACGTATCGCACGCACCGCGC CCGACCAGCTGGCCCTCACGGCCGTGGACGGGGGCGGAGGCGGGGAGCCGGTTCTCACG CTGGAGTCGCTCACGGTACGGCCGGTGGCCGCCCACCAGATCGCGGGCGCCCGTGCGGCG GACCGTGACGCGCTGTTCCGGCTCGTGTGGATGGAGGTCGCCGCGCGGGCCGAGGAGACG GGCGGCGGCGCCCCGCGTGCCGCGGTCCTCGCGCCGGTCGAGAGTGGCCCGATGGGCGGT ACGTCGGCCGGTGCGCTGGCCGACGCCTTGTCCGATGCGCTGGCCGCCGGCCCCGTGTGG GACACGTTCGGTGCGCTCCGGGACGGAGTGGCGGCTGGGGGCGAAGCGCCCGATGTCGT GCTCGCCGTGTGCGCCGCGCCCGGCGCAGGTGCCGGGGCCGTTGCGGATGCCGATGGCAG GGGCGGCGACCCGGCCGGGTACGCGCGGCTGGCCACCGTGTCCCTTCTGTCGCTGCTCAA GGAGTGGGTGGACGACCCGGCGTTCGCGGCGACCCGTCTCGTCGTCGTCACCCGAGGCGC GGTCGCCGCGCGGCCGGGTGAGACCGCCGGTGATCTGGCGGGCGCGTCGCTGTGGGGTCT GGTGCGCAGCGCGCAGGCCGAGAATCCCGGTCGCCTCACGCTGCTCGACGTGGACGGACT GGAGTCTTCTCCGGCCACGCTGACCGGCGTACTGGCCTCCGGCGAACCGGAATTGGCGCT GCGTGACGGGCGCGCCTACGTGCCACGGCTCGTGCGCGACGACGCGTCGGTGCGGCTGGT GCCGCCGGTCGGATCGCTCACGTGGCGGCTGGCTCGGTGCCAAGAGGCGGGCGGCGGAC AGCAGTTGTCCCTTGTCGACGCTCCCGAGGCCGGACGGGCGCTGGAGCCGCACGAGGTGC GGGTGGCGGTGCGGGCCGCGGCGCCGGGGCCGCTCACGGCGGGCCAGGTCGAGGGTGCC GGTGTGGTGACGGAAGTCGGCGGCGAGGTCGGTTCGGTGGCCGTGGGCGACCGGGTGAT GGGGTTGTTCGACGCGGTGGGCCCGGTGGCCGTCACCGATGCCGCGCTGCTTATGCCGGT TCCGGCGGGCTGGAGTTGGGCGCAGGCGGCCGGGTCCTTGGGGGCCTATGTGTCCGCGTA TCACGTGCTGGCGGATGTCGTGGCGCCGCGCGGCGGCGAGACTCTGCTCGTCGGGGAGGA AACCGGTTCCGTCGGCCGTGCCGTGCTGCGTCTTGCTCTTGCCGGGCGGTGGCGGGTCGA GGCCGTGGACGGTGCGTCGACGGCGGATGATTCGGGCGCCGAGCGCGCGGCCGACGTCA CCCTCCGCCACGAGGGGGCCCTGGTGGTCCACCGTGCGGGCGGTCGTCCGGACGAGGGAC AGGCTGTCGTGCCGCCCGAGCCCGGGCGCGTACGGGAAATCCTCGCGGAGCTGACCGAGC TGACCGAGCTTGCCGAGATCACGGAGTCGGCGGAGCCCGGGCTGCCCGCGGAACGGGGT GACAGCCGTGCTCTGACGCCGCTCGACATCACCGTGTGGGACATCCGGCAGGCGCCCGCC GCGATGGCGGCCCCGCCCTCGGCGGGCACGACCGTGTTCTCGCTGCCTCCCGCCTTCGACC CCGAGGGCACCGTGCTGGTCACCGGCGGCACCGGAGCACTCGGCTCGCTGACGGCCCGTC ACCTGGTGGAACGGTACGGAGCCAGGCACCTGTTGCTGTCCAGCAGGCGGGGAGCCGAC GCGCCGGGCGCACTCGAACTGGCCGCCGACCTCTCCGCGCTCGGCGCGCGAGTCACCTTC GCCGCGTGCGACCCGGGCGACCGGGACGAAGCCGCCGCCCTCCTCGCGGCGGTGCCCTCG GACCACCCGCTGACCGCCGTCTTCCACTGCGCGGGCACCGTGAACGACGCCGTGGTGCAG AACCTCACGGCCGAGCAGGTCGAGGAGGTGATGCGCGTGAAGGCGGACGCCGCGTGGCA CCTGCACGAGCTGACGCGGGACGCGGACCTGTCCGCGTTCGTCCTGTACTCCTCGGTCGCC GGGCTGCTCGGCGGCCCCGGCCAGGGCAGCTACACGGCCGCCAACGCCTTCTTGGACGCG CTGGCCCGGCACCGGCACGACGGCGGTGCGGCGGCGACCTCCCTGGCGTGGGGCTACTGG GAGCTGGCGAGCGGCATGTCGGGACGGCTCACCGACGCCGACCGGGCGCGCCATGCCCG CGCCGGCGTGGTCGGGCTCGGCGCCGACGAGGGACTCGCCCTCCTCGACGCGGCGTGGGC CGGCGGACTGCCCCTGTACGCGCCGGTCCGTCTGGACCTGGCCCGGATGCGCCGGCAGGC CCAGAGCCACCCCGCACCGGCGCTGCTGCGCGACCTGGTGCGCGGGGGGAGCAAGAGCG GCGGCGGTGCCGTGTCGGCGGGGGCGGCCGCGCTGCTCAAGTCGCTCGGCGCGATGTCCG ACCCCGAGCGGGAGGAGGCGCTGCTCGACCTGGTGTGCACCCATATCGCGGCCGTCCTCG GCTACGACGCGGCCACGCCCGTCAACGCGACGCAGGGGCTGCGGGAACTCGGCTTCGACT CCCTGACCGCGGTGGAGCTGCGCAACCGGCTCTCGGCCGCGACGGGGCTGAAGCTGCCCG CCACGTTCGTCTTCGACCATCCCAACCCGGCGGAGCTCGCCGCGCAGCTGCGGCAGGAGC TGGCCCCGCGCGCGGCGGATCCGCTCGCCGACGTCCTGGCGGAGTTCGAGCGCATCGAGG ACTCGCTGCTCTCGGTCTCCTCGAAGGACGGCTCGGCGCGGGCCGAACTGGCGGGGCGGC TGCGCGCGACGCTGGCCAGGCTCGACGCGCCGCAGGACACGGCCGGCGAGGTCGCGGTG GCCACTCGTACACGTATCCAGGACGCGTCGGCGGACGAGATCTTCGCGTTCATCGACCGG GACCTCGGCAGGGACGGCGCGAGCGGACAGGGCAACGGACAGCCCACCGGGCAGGGCA ACGGACACGGCAACGGCAACGGCAACGGCAACGGCAACGGACACGGTCAGGCAGTGGA GGGGCAGCGATGA	
rci_ NC_019104.1	ATGCCGTCTCCACGCATCCGTAAAATGTCCCTGTCACGCGCACTGGATAAGTACCTGAAA ACAGTTTCTGTTCACAAGAAAGGGCATCAACAGGAGTTTTACCGGAGCAATGTTATCAAG CGATATCCCATTGCCCTTCGGAATATGGACGAAATAACAACCGTTGATATTGCTACATAC AGAGACGTTCGTTTAGCAGAAATCAACCCCCGAACGGGTAAACCCATTACAGGTAATACT GTACGTCTTGAACTCGCCCTTCTGTCATCTCTGTTCAATATTGCTCGTGTTGAATGGGGAA CCTGTCGTACTAACCCGGTTGAACTGGTTCGCAAGCCGAAAGTATCCTCCGGACGAGATC GCCGGCTAACGTCTTCAGAAGAACGTCGCCTTTCTCGCTATTTCCGCGAAAAAAATCTGAT GTTGTATGTCATTTTCCATCTTGCCCTTGAGACAGCCATGCGGCAGGGCGAAATACTGGCC TTACGTTGGGAGCACATTGATTTGCGCCACGGTGTGGCTCATTTACCTGAAACCAAAAAC GGTCACTCACGGGATGTTCCTCTGTCCAGACGTGCCCGTAACTTTCTTCAGATGATGCCCG	35

Name	Sequence	SEQ ID NO:
	TTAATCTCCACGGCAATGTTTTTGATTACACCGCATCCGGCTTTAAAAATGCCTGGAGAAT AGCCACACAACGACTTCGCATCGAGGACCTGCATTTTCACGATCTACGGCATGAAGCAAT AAGCCGCTTCTTCGAACTGGGTAGCCTGAATGTAATGGAGATTGCTGCAATATCAGGACA TCGTTCCATGAATATGCTGAAACGGTATACTCATCTTCGTGCATGGCAACTGGTCAGTAAA CTTGATGCCCGCCGGCGGCAGACACAAAAAGTGGCAGCATGGTTTGTGCCGTATCCTGCC CATATCACGACCATCAATGAAGAAAATGGGCAGAAAGCGCATCGTATTGAGATCGGTGAT TTTGATAACCTTCACGTCACTGCCACAACGAAAGAGGAAGCAGTTCACCGCGCCAGTGAG GTTTTGTTGCGTACACTGGCCATTGCAGCACAGAAAGGCGAACGTGTCCCATCTCCCGGA GCGTTACCTGTTAACGACCCTGACTACATTATGATTTGCCCTCTGAACCCGGGCAGCACCC CGCTGTAA	
optimized_rci	ATGCCGTCCCCCCGCATCCGCAAGATGAGCCTCTCCCGCGCCCTCGACAAGTACCTGAAG ACCGTGAGCGTCCACAAGAAGGGCCACCAGCAGGAGTTCTACCGGTCCAACGTCATCAAG CGCTACCCGATCGCCCTGCGGAACATGGACGAGATCACCACCGTCGACATCGCCACCTAC CGGGACGTGCGCCTGGCGGAGATCAACCCGCGCACCGGCAAGCCCATCACCGGCAACAC CGTCCGGCTGGAGCTCGCCCTGCTCTCCTCGCTGTTCAACATCGCGCGCGTGGAGTGGGGC ACCTGCCGGACCAACCCGGTCGAGCTGGTGCGCAAGCCCAAGGTCAGCTCCGGCCGGGAC CGCCGGCTCACCTCGTCCGAGGAGCGCCGGCTGTCCCGGTACTTCCGCGAGAAGAACCTG ATGCTCTACGTGATCTTCCACCTGGCCCTGGAGACCGCGATGCGCCAGGGCGAGATCCTG GCCCTCCGCTGGGAGCACATCGACCTGCGCCACGGCGTCGCGCACCTGCCGGAGACCAAG AACGGCCACAGCCGCGACGTGCCCCTGTCCCGCCGGGCCCGGAACTTCCTCCAGATGATG CCCGTCAACCTGCACGGCAACGTGTTCGACTACACCGCCAGCGGCTTCAAGAACGCCTGG CGCATCGCGACCCAGCGGCTGCGCATCGAGGACCTGCACTTCCACGACCTGCGCCACGAG GCCATCTCGCGGTTCTTCGAGCTGGGCAGCCTCAACGTCATGGAGATCGCCGCGATCTCC GGCCACCGCTCGATGAACATGCTCAAGCGGTACACCCACCTGCGCGCCTGGCAGCTGGTG TCGAAGCTCGACGCGCGTCGCCGCCAGACCCAGAAGGTCGCCGCGTGGTTCGTGCCGTAC CCCGCCCACATCACCACCATCGACGAGGAGAACGGCCAGAAGGCGCACCGCATCGAGAT CGGCGACTTCGACAACCTGCACGTCACCGCCACCACCAAGGAAGAGGCCGTGCACCGCGC GTCCGAGGTGCTGCTCCGGACCCTCGCCATCGCCGCGCAGAAGGGCGAGCGGGTCCCGTC GCCCGGCGCGCTGCCGGTCAACGACCCCGACTACATCATGATCTGCCCCCTGAACCCCGG CTCCACCCCGCTGTGA	36
tipAp(RBS)	CCCGGGCTGAGGGAGCCGACGGCACGCGGCGGCTCACGGCGTGGCACGCGGAACGTCCG GGCTTGCACCTCACGTCACGTGAGGAGGCAGCGTGGACGGCGTCAGAGAAGGGAGCGGA A	37
upstream_homology	CGAAGAGAAGTACCTCGACTACCTGCGTCGTGCCACGGCGGACCTCCACGAGGCCCGTGG CCGCCTCCGCGAGCTGGAGGCGAAGGCGGGCGAGCCGGTGGCGATCGTCGGCATGGCCT GCCGCCTGCCCGGCGGCGTCGCCTCGCCCGAGGACCTGTGGCGGCTGGTGGCCGGCGGCG AGGACGCGATCTCGGAGTTCCCCCAGGACCGCGGCTGGGACGTGGAGGGCCTGTACGACC CGAACCCGGAGGCCACGGGCAAGAGTTACGCCCGCGAGGCCGGATTCCTGTACGAGGCG GGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCG CAGCAGCGTCTCCTCCTGGAGGCCTCCTGGGAGGCGTTCGAGCACGCCGGGATCCCGGCG GCCACCGCGCGCGGCACCTCGGTCGGCGTCTTCACCGGCGTGATGTACCACGACTACGCC ACCCGTCT	38
downstream_homology	CTGCTCAGCGGGCTCACCCGCGGATCGCGGGTCGGCGGCGCGCCGGTCAACCAGCGCAGG GCAGCCGCCGGAGGCGCGGGCGAGGCGGACACGGACCTCGGCGGGCGGCTCGCCGCGAT GACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCACGTGGCGACCGT CCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGCGACACCGGTTT CGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACCGGGCTGCGGCT GCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGGCACCTGCTCGA CGAACTCGCCACGGCCGCGGGCGGGTCCTGGGCGGAAGGCACCGGGTCCGGAGACACGG CCTCGGCGACCGATCGGCAGACCACGGCGGCCCTCGCCGAACTCGACCGGCTGGAAGGC GTGCTCGCCTCCCTCGCG	39
pYES-1L	CCTCGCCGCAGTTAATTAAAGTCAGTGAGCGAGGAAGCGCGTAACTATAACGGTCCTAAG GTAGCGAATCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATA GATCGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGTAATAACCTATTTCTTA GCATTTTTGACGAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTCTCCACACCTCC GCTTACATCAACACCAATAACGCCATTTAATCTAAGCGCATCACCAACATTTTCTGGCGTC AGTCCACCAGCTAACATAAAATGTAAGCTTTCGGGGCTCTCTTGCCTTCCAACCCAGTCAG AAATCGAGTTCCAATCCAAAAGTTCACCTGTCCCACCTGCTTCTGAATCAAACAAGGGAA TAAACGAATGAGGTTTCTGTGAAGCTGCACTGAGTAGTATGTTGCAGTCTTTTGGAAATAC GAGTCTTTTAATAACTGGCAAACCGAGGAACTCTTGGTATTCTTGCCACGACTCATCTCCA TGCAGTTGGACGATATCAATGCCGTAATCATTGACCAGAGCCAAAACATCCTCCTTAAGT TGATTACGAAACACGCCAACCAAGTATTTCGGAGTGCCTGAACTATTTTTATATGCTTTTA CAAGACTTGAAATTTTCCTTGCAATAACCGGGTCAATTGTTCTCTTTCTATTGGGCACACA TATAATACCCAGCAAGTCAGCATCGGAATCTAGAGCACATTCTGCGGCCTCTGTGCTCTGC AAGCCGCAAACTTTCACCAATGGACCAGAACTACCTGTGAAATTAATAACAGACATACTC CAAGCTGCCTTTGTGTGCTTAATCACGTATACTCACGTGCTCAATAGTCACCAATGCCCTC CCTCTTGGCCCTCTCCTTTTCTTTTTTCGACCGAATTAATTCTTAATCGGCAAAAAAAGAA AAGCTCCGGATCAAGATTGTACGTAAGGTGACAAGCTATTTTTCAATAAAGAATATCTTC CACTACTGCCATCTGGCGTCATAACTGCAAAGTACACATATATTACGATGCTGTTCTATTA AATGCTTCCTATATTATATATATATAGTAATGTCGTGATCTATGGTGCACTCTCAGTACAATC TGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCC TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCT GCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGA TACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGGATCGCTTGCC TGTAACTTACACGCGCCTCGTATCTTTTAATGATGGAATAATTTGGGAATTTACTCTGTGT	40

Name	Sequence	SEQ ID NO:
	TTATTTATTTTTATGTTTTGTATTTGGATTTTAGAAAGTAAATAAAGAAGGTAGAAGAGTT ACGGAATGAAGAAAAAAAAATAAACAAAGGTTTAAAAAATTTCAACAAAAAGCGTACTT TACATATATATTTATTAGACAAGAAAAGCAGATTAAATAGATATACATTCGATTAACGAT AAGTAAAATGTAAAATCACAGGATTTTCGTGTGTGGTCTTCTACACAGACAAGGTGAAAC AATTCGGCATTAATACCTGAGAGCAGGAAGAGCAAGATAAAAGGTAGTATTTGTTGGCGA TCCCCCTAGAGTCTTTTACATCTTCGGAAAACAAAAACTATTTTTTCTTTAATTTCTTTTTT TACTTTCTATTTTTAATTTATATATTTATATTAAAAAATTTAAATTATAATTATTTTTATAG CACGTGATGAAAAGGACCCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTT GTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAAT GCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATT CCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAA AAGATGCTGAAGATCAGTTGGGACGCGTAGTCTAGACCAGCCAGGACAGAAATGCCTCG ACTTCGCTGCTACCCAAGGTTGCCGGGTGACGCACACCGTGGAAACGGATGAAGGCACGA ACCCAGTGGACATAAGCCTGTTCGGTTCGTAAGCTGTAATGCAAGTAGCGTATGCGCTCA CGCAACTGGTCCAGAACCTTGACCGAACGCAGCGGTGGTAACGGCGCAGTGGCGGTTTTC ATGGCTTGTTATGACTGTTTTTTTGGGGTACAGTCTATGCCTCGGGCATCCAAGCAGCAAG CGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGG GCAGTCGCCCTAAAACAAAGTTAAACATTATGAGGGAAGCGGTGATCGCCGAAGTATCG ACTCAACTATCAGAGGTAGTTGGCGCCATCGAGCGCCATCTCGAACCGACGTTGCTGGCC GTACATTTGTACGGCTCCGCAGTGGATGGCGGCCTGAAGCCACACAGTGATATTGATTTG CTGGTTACGGTGACCGTAAGGCTTGATGAAACAACGCGGCGAGCTTTGATCAACGACCTT TTGGAAACTTCGGCTTCCCCTGGAGAGAGCGAGATTCTCCGCGCTGTAGAAGTCACCATT GTTGTGCACGACGACATCATTCCGTGGCGTTATCCAGCTAAGCGCGAACTGCAATTTGGA GAATGGCAGCGCAATGACATTCTTGCAGGTATCTTCGAGCCAGCCACGATCGACATTGAT CTGGCTATCTTGCTGACAAAAGCAAGAGAACATAGCGTTGCCTTGGTAGGTCCAGCGGCG GAGGAACTCTTTGATCCGGTTCCTGAACAGGATCTATTTGAGGCGCTAAATGAAACCTTA ACGCTATGGAACTCGCCGCCCGACTGGGCTGGCGATGAGCGAAATGTAGTGCTTACGTTG TCCCGCATTTGGTACAGCGCAGTAACCGGCAAAATCGCGCCGAAGGATGTCGCTGCCGGC TGGGCAATGGAGCGCCTGCCGGCCCAGTATCAGCCCGTCATACTTGAAGCTAGACAGGCT TATCTTGGACAAGAAGAAGATCGCTTGGCCTCGCGCGCAGATCAGTTGGAAGAATTTGTC CACTACGTGAAAGGCGAGATCACCAAGGTAGTCGGCAAATAACCCTCGAGCATTCAAGG CGCCTTGATTATTTGACGTGGTTTGATGGCCTCCACGCACGTTGTGATATGTAGATGATAA TCATTATCACTTTACGGGTCCTTTCCGGTGATCCGACAGGTTACGGGGCGGCGACCTCGCG GGTTTTCGCTATTTATGAAAATTTTCCGGTTTAAGGCGTTTCCGTTCTTCTTCGTCATAACT TAATGTTTTTATTTAAAATACCTCGCGAGTGGCAACACTGAAAATACCCATGGAGCGGCG TAACCGTCGCACAGGAAGGACAGAGAAAGCGCGGATCTGGGAAGTGACGGACAGAACGG TCAGGACCTGGATTGGGGAGGCGGTTGCCGCCGCTGCTGCTGACGGTGTGACGTTCTCTG TTCCGGTCACACCACATACGTTCCGCCATTCCTATGCGATGCACATGCTGTATGCCGGTAT ACCGCTGAAAGTTCTGCAAAGCCTGATGGGACATAAGTCCATCAGTTCAACGGAGGTCTA CACGAAGGTTTTTGCGCTGGATGTGGCTGCCCGGCACCGGGTGCAGTTTGCGATGCCGGA GTCTGATGCGGTTGCGATGCTGAAACAATTATCCTGAGAATAAATGCCTTGGCCTTTATAT GGAAATGTGGAACTGAGTGGATATGCTGTTTTTGTCTGTTAAACAGAGAAGCTGGCTGTT ATCCACTGAGAAGCGAACGAAACAGTCGGGAAAATCTCCCATTATCGTAGAGATCCGCAT TATTAATCTCAGGAGCCTGTGTAGCGTTTATAGGAAGTAGTGTTCTGTCATGATGCCTGCA AGCGGTAACGAAAACGATTTGAATATGCCTTCAGGAACAATAGAAATCTTCGTGCGGTGT TACGTTGAAGTGGAGCGGATTATGTCAGCAATGGACAGAACAACCTAATGAACACAGAA CCATGATGTGGTCTGTCCTTTTACAGCCAGTAGTGCTCGCCGCAGTCGAGCGACAGGGCG AAGCCCTCGAGTGAGCGAGGAAGCACCAGGGAACAGCACTTATATATTCTGCTTACACAC GATGCCTGAAAAAACTTCCCTTGGGGTTATCCACTTATCCACGGGGATATTTTTATAATTA TTTTTTTTATAGTTTTTAGATCTTCTTTTTTAGAGCGCCTTGTAGGCCTTTATCCATGCTGGT TCTAGAGAAGGTGTTGTGACAAATTGCCCTTTCAGTGTGACAAATCACCCTCAAATGACA GTCCTGTCTGTGACAAATTGCCCTTAACCCTGTGACAAATTGCCCTCAGAAGAAGCTGTTT TTTCACAAAGTTATCCCTGCTTATTGACTCTTTTTTATTTAGTGTGACAATCTAAAAACTTG GCACACTTCACATGGATCTGTCATGGCGGAAACAGCGGTTATCAATCACAAGAAACGTAA AAATAGCCCGCGAATCGTCCAGTCAAACGACCTCACTGAGGCGGCATATAGTCTCTCCCG GGATCAAAAACGTATGCTGTATCTGTTCGTTGACCAGATCAGAAAATCTGATGGCACCCT ACAGGAACATGACGGTATCTGCGAGATCCATGTTGCTAAATATGCTGAAATATTCGGATT GACCTCTGCGGAAGCCAGTAAGGATATACGGCAGGCATTGAAGAGTTTCGCGGGGAAGG AAGTGGTTTTTTATCGCCCTGAAGAGGATGCCGGCGATGAAAAAGGCTATGAATCTTTTC CTTGGTTTATCAAACGTGCGCACAGTCCATCCAGAGGGCTTTACAGTGTACATATCAACCC ATATCTCATTCCCTTCTTTATCGGGTTACAGAACCGGTTTACGCAGTTTCGGCTTAGTGAA ACAAAAGAAATCACCAATCCGTATGCCATGCGTTTATACGAATCCCTGTGTCAGTATCGT AAGCCGGATGGCTCAGGCATCGTCTCTCTGAAAATCGACTGGATCATAGAGCGTTACCAG CTGCCTCAAAGTTACCAGCGTATGCCTGACTTCCGCCGCCGCTTCCTGCAGGTCTGTGTTA ATGAGATCAACAGCAGAACTCCAATGCGCCTCTCATACATTGAGAAAAAGAAAGGCCGC CAGACGACTCATATCGTATTTTCCTTCCGCGATATCACTTCCATGACGACAGGATAGTCTG AGGGTTATCTGTCACAGATTTGGGGGGTGGTTCGTCACATTTGTTCTGACCTACTGAGGGTA ATTTGTCACAGTTTTGCTGTTTCCTTCAGCCTGCATGGATTTTCTCATACTTTTTGAACTGT AATTTTTAAGGAAGCCAAATTTGAGGGCAGTTTGTCACAGTTGATTTCCTTCTCTTTCCCTT CGTCATGTGACCTGATATCGGGGGGTTAGTTTGTCATCATTGATGAGGGTTGATTATCACAG TTTATTACTCTGAATTGGCTATCCGCGTGTGTACCTCTACCTGGAGTTTTTCCCACGGTGGA TATTTCTTCTTGCGCTGAGCGTAAGAGCTATCTGACAGAACAGTTCTTCTTTGCTTCCTCGC CAGTTCGCTCGCTATGCTCGGTTACACGGCTGCGGCGAGCGCTAGTGATAATAAGTGACT GAGGTATGTGCTCTTCTTATCTCCTTTTGTAGTGTTGCTCTTATTTTAAACAACTTTGCGGT TTTTTGATGACTTTGCGATTTTGTTGTTGCTTTGCAGTAAATTGCAAGATTTAATAAAAAA	

Name	Sequence	SEQ ID NO:
	ACGCAAAGCAATGATTAAAGGATGTTCAGAATGAAACTCATGGAAACACTTAACCAGTGC ATAAACGCTGGTCATGAAATGACGAAGGCTATCGCCATTGCACAGTTTAATGATGACAGC CCGGAGGCGAGGAAAATAACCCGGCGCTGGAGAATAGGTGAAGCAGCGGATTTAGTTGG GGTTTCTTCTCAGGCTATCAGAGATGCCGAGAAAGCAGGGCGACTACCGCACCCGGATAT GGAAATTCGAGGACGGGTTGAGCAACGTGTTGGTTATACAATTGAACAAATTAATCATAT GCGTGATGTGTTTGGTACGCGATTGCGACGTGCTGAAGACGTATTTCCACCGGTGATCGG GGTTGCTGCCCATAAAGGTGGCGTTTACAAAACCTCAGTTTCTGTTCATCTTGCTCAGGAT CTGGCTCTGAAGGGGCTACGTGTTTTGCTCGTGGAAGGTAACGACCCCCAGGGAACAGCC TCAATGTATCACGGATGGGTACCAGATCTTCATATTCATGCAGAAGACACTCTCCTGCCTT TCTATCTTGGGGAAAAGGACGATGTCACTTATGCAATAAAGCCCACTTGCTGGCCGGGGC TTGACATTATTCCTTCCTGTCTGGCTCTGCACCGTATTGAAACTGAGTTAATGGGCAAATT TGATGAAGGTAAACTGCCCACCGATCCACACCTGATGCTCCGACTGGCCATTGAAACTGT TGCTCATGACTATGATGTCATAGTTATTGACAGCGCGCCTAACCTGGGTATCGGCACGATT AATGTCGTATGTGCTGCTGATGTGCTGATTGTTCCCACGCCTGCTGAGTTGTTTGACTACA CCTCCGCACTGCAGTTTTTCGATATGCTTCGTGATCTGCTCAAGAACGTTGATCTTAAAGG GTTCGAGCCTGATGTACGTATTTTGCTTACCAAATACAGCAATAGTAATGGCTCTCAGTCC CCGTGGATGGAGGAGCAAATTCGGGATGCCTGGGGAAGCATGGTTCTAAAAAATGTTGTA CGTGAAACGGATGAAGTTGGTAAAGGTCAGATCCGGATGAGAACTGTTTTTGAACAGGCC ATTGATCAACGCTCCTCAACTGGTGCCTGGAGAAATGCTCTTTCTATTTGGGAACCTGTCT GCAATGAAATTTTCGATCGCCTGATTAAACCACGCTGGGAGATTAGATAATGAAGCGTGC GCCTGTTATTCCAAAACATACGCTCAATACTCAACCGGTTGAAGATACTTCGTTATCGACA CCAGCTGCCCCGATGGTGGATTCGTTAATTGCGCGCGTAGGAGTAATGGCTCGCGGTAAT GCCATTACTTTGCCTGTATGTGGTCGGGATGTGAAGTTTACTCTTGAAGTGCTCCGGGGTG ATAGTGTTGAGAAGACCTCTCGGGTATGGCCAGGTAATGAACGTGACCAGGAGCTGCTTA CTGAGGACGCACTGGATGATCTCATCCCTTCTTTTCTACTGACTGGTCAACAGACACCGGC GTTCGGTCGAAGAGTATCTGGTGTCATAGAAATTGCCGATGGGAGTCGCCGTCGTAAAGC TGCTGCACTTACCGAAAGTGATTATCGTGTTCTGGTTGGCGAGCTGGATGATGAGCAGAT GGCTGCATTATCCAGATTGGGTAACGATTATCGCCCAACAAGTGCTTATGAACGTGGTCA GCGTTATGCAAGCCGATTGCAGAATGAATTTGCTGGAAATATTTCTGCGCTGGCTGATGC GGAAAATATTTCACGTAAGATTATTACCCGCTGTATCAACACCGCCAAATTGCCTAAATC AGTTGTTGCTCTTTTTTCTCACCCCGGTGAACTATCTGCCCGGTCAGGTGATGCACTTCAA AAAGCCTTTACAGATAAAGAGGAATTACTTAAGCAGCAGGCATCTAACCTTCATGAGCAG AAAAAAGCTGGGGTGATATTTGAAGCTGAAGAAGTTATCACTCTTTTAACTTCTGTGCTTA AAACGTCATCTGCATCAAGAACTAGTTTAAGCTCACGACATCAGTTTGCTCCTGGAGCGA CAGTATTGTATAAGGGCGATAAAATGGTGCTTAACCTGGACAGGTCTCGTGTTCCAACTG AGTGTATAGAGAAAATTGAGGCCATTCTTAAGGAACTTGAAAAGCCAGCACCCTGATGCG ACCACGTTTTAGTCTACGTTTATCTGTCTTTACTTAATGTCCTTTGCTACAGGCCAGAAAGC ATAACTGGCCTGAATATTCTCTCTGGGCCCACTGTTCCACTTGTATCGTCGGACTGATAAT CAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTC CCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATCGTCGGTC TGATAATCAGACTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGAC CATGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGGTCCCACTCGTATC GTCGGTCTGATTATTAGTCTGGAACCACGGTCCCACTCGTATCGTCAGTCTGATTATTAGT CTGGGACCACGGTCCCACTCGTATCGTCGGTCTGATTATTAGTCTGGGACCACGATCCCAC TCGTGTTGTCGGTCTGATTATCGGTCTGGGACCACGGTCCCACTTGTATTGTCGATCAGAC TATCAGCGTGAGACTACGATTCCATCAATGCCTGTCAAGGGCAAGTATTGACATGTCGTC GTAACCTGTAGAACGGAGTAACCTCGGTGTGCGGTTGTATGCCTGCTGTGGATTGCTGCTG TGTCCTGCTTATCCACAACATTTTGCGCACGGTTATGTGGACAAAATACCTGGTTACCCAG GCCGTGCCGGCACGTTAACCGGGCTGCATCCGATGCAAGTGTGTCGCTGTCGACGGCCTC CTCACCCGGTCACGTGAGCTCATTTAACCCACTCCACAAAAAGGCTCAACAGGTTGGTGG TTCTCACCACCAAAAGCACCACACCCCACGCAAAAACAAGTTTTTGCTGATTTTTCTTTAT AAATAGAGTGTTATGAAAAATTAGTTTCTCTTACTCTCTTTATGATATTTAAAAAAGCGGT GTCGGCGCGGCTACAACAACGCGCCGACACCGTTTTGTAGGGGTGGTACTGACTATTTTT ATAAAAAACATTATTTTATATTAGGGGTGCTGCTAGCGGCGCGGTGTGTTTTTTTATAGGA TACCGCTAGGGGCGCTGCTAGCGGTGCGTCCCTGTTTGCATTATGAATTAGTTACGCTAGG GATAACAGGGTAATATAGAACCCGAACGACCGAGCGCAGCGGCGGCCGCGCTGATACCG CCGC	
tsr_operon	CGGCCACGACACCCCCATCGGCATCGCGTGGCGGGCCCGATTCGCGTCCGCCCCGGGGAG GAGACGGGCGTCCGCGTCGAGGGCGAGGCGGGCTTCCACCCGTAGGCGATCCGGCGAGC ACGACACCCGCAGATTGGCTCGACGCAGCCCAGAAATGTATGATCAAGGCGAATACTTCA TATGCGGGGATCGACCGCGCGGGTCCCGGACGGGGAAGAGCGGGGAGCTTTGCCAGAGA GCGACGACTTCCCCTTGCGTTGGTGATTGCCGGTCAGGGCAGCCATCCGCCATCGTCGCGT AGGGTGTCACACCCCAGGAATCGCGTCACTGAACACAGCAGCCGGTAGGACGACCATGA CTGAGTTGGACACCATCGCAAATCCGTCCGATCCCGCGGTGCAGCGGATCATCGATGTCA CCAAGCCGTCGCGATCCAACATAAAGACAACGTTGATCGAGGACGTCGAGCCCCTCATGC ACAGCATCGCGGCCGGGGTGGAGTTCATCGAGGTCTACGGCAGCGACAGCAGTCCTTTTC CATCTGAGTTGCTGGATCTGTGCGGGCGGCAGAACATACCGGTCCGCCTCATCGACTCCTC GATCGTCAACCAGTTGTTCAAGGGGGAGCGGAAGGCCAAGACATTCGGCATCGCCCGCGT CCCTCGCCCGGCCAGGTTCGGCGATATCGCGAGCCGGCGTGGGGACGTCGTCGTTCTCGA CGGGGTGAAGATCGTCGGGAACATCGGCGCGATAGTACGCACGTCGCTCGCGCTCGGAGC GTCGGGGATCATCCTGGTCGACAGTGACATCACCAGCATCGCGGACCGGCGTCTCCAAAG GGCCAGCCGAGGTTACGTCTTCTCCCTTCCCGTCGTTCTCTCCGGTCGCGAGGAGGCCATC GCCTTCATTCGGGACAGCGGTATGCAGCTGATGACGCTCAAGGCGGATGGCGACATTTCC GTGAAGGAACTCGGGGACAATCCGGATCGGCTGGCCTTGCTGTTCGGCAGCGAAAAGGGT GGGCCTTCCGACCTGTTCGAGGAGGCGTCTTCCGCCTCGGTTTCCATCCCCATGATGAGCC	41

Name	Sequence	SEQ ID NO:
	AGACCGAGTCTCTCAACGTTTCCGTTTCCCTCGGAATCGCGCTGCACGAGAGGATCGACA GGAATCTCGCGGCCAACCGATAAGCGCCTCTGTTCCTCGGACGCTCGGTTCCTCGACCTCG ATTCGTCAGTGATGATCACCTCACACGGCAGCGATCACCACTGACATATCGAGGTCAACG GTCGTGGTCCGGGCGGGCACTCCTCGAAGGCGCGGCCGACGCCCTTGAACGACTCGATGG CGCTGACGCCTTCCGCGATCGCCTCGACTGTGAAAACGCTCGTGGACGTCGGCGGGTAGC CGATGAGTTCCATCGCCTTGCGCATCCCCTTGAGCATGGAGAACGCGTCCACCGGGTAGT CACGAAGGTACTTGCGAAGGCTTCGCGCCGTCCGAGACCGTGAGCCGGCGCAAGCCGGTG TGGCGCCGGTCCCGGATCC	
gBlock_1	ACCCCGTCCGGGCAGGGCGGGTCACCGTCCCGCGCGGCGGGGAGAGAGGAGAGGCCGTG ACCAGGTACCGGATTGGCACTTATCCGGTGCCAATCCGGTCGGTGGTCAGGAGTGGAGAA GACATGCCGGGTACGAACGACATGCCGGGTACCGAGGACAAGCTCCGCCACTACCTGAA GCGA	42
gBlock_2	CCGACGAGGAGGAGCAAGCCGCGCCCGAGGACCTGTCGACCGCGAGCGCCGACGACATG TTCGCCCTGATCGACCGGGAGTGGGGCACGCGGTGACCACCGACCGGATTGGCACGTAGG CCAGTGCCAATCCGGTACTTGGCAATTGCCTCGCCGCAGTTAATTAAAGTCAGTGAGCGA GGAAGCGC	43
gBlock_3	GCTTCATGCTGGACACCTCCGCGGTGCCGGCGGACGTGGTGCGTGACATCGAGGCCGACC ATCTGTCCCTGGCCATGGAGCACTCGGATCTGACCGCCGAGGCCATAGAGAACTGGCTCG CCGAACTTCCGGCCGGCGAGGCCTGACCAAGTACCGGATTGGCACTGGCCTACGTGCCAA TCCGGTACGTGGGCAGCGATCAGTCCTGAGCGCGATGAGCCCCGGCCACACCGGCCCCCT TGCCCTCCCCCAACCTCATACGCCCCGATCACCACGTAGCGCCAAGGAGCCTGGGTCAGA TGTCGAACGAGGAGAAGCTTCTCGATCACCTCAAGTGGGTCACCGCGGAGCTGCGCCAGG CCCGGCAACGGCTCCACGACAAGGAATCGACCGAGCCGGTC	44
gusA- NC_000913.3	ATGTTACGTCCTGTAGAAACCCCAACCCGTGAAATCAAAAAACTCGACGGCCTGTGGGCA TTCAGTCTGGATCGCGAAAACTGTGGAATTGATCAGCGTTGGTGGGAAAGCGCGTTACAA GAAAGCCGGGCAATTGCTGTGCCAGGCAGTTTTAACGATCAGTTCGCCGATGCAGATATT CGTAATTATGCGGGCAACGTCTGGTATCAGCGCGAAGTCTTTATACCGAAAGGTTGGGCA GGCCAGCGTATCGTGCTGCGTTTCGATGCGGTCACTCATTACGGCAAAGTGTGGGTCAAT AATCAGGAAGTGATGGAGCATCAGGGCGGCTATACGCCATTTGAAGCCGATGTCACGCCG TATGTTATTGCCGGGAAAAGTGTACGTATCACCGTTTGTGTGAACAACGAACTGAACTGG CAGACTATCCCGCCGGGAATGGTGATTACCGACGAAAACGGCAAGAAAAAGCAGTCTTA CTTCCATGATTTCTTTAACTATGCCGGGATCCATCGCAGCGTAATGCTCTACACCACGCCG AACACCTGGGTGGACGATATCACCGTGGTGACGCATGTCGCGCAAGACTGTAACCACGCG TCTGTTGACTGGCAGGTGGTGGCCAATGGTGATGTCAGCGTTGAACTGCGTGATGCGGAT CAACAGGTGGTTGCAACTGGACAAGGCACTAGCGGGACTTTGCAAGTGGTGAATCCGCAC CTCTGGCAACCGGGTGAAGGTTATCTCTATGAACTGTGCGTCACAGCCAAAAGCCAGACA GAGTGTGATATCTACCCGCTTCGCGTCGGCATCCGGTCAGTGGCAGTGAAGGGCGAACAG TTCCTGATTAACCACAAACCGTTCTACTTTACTGGCTTTGGTCGTCATGAAGATGCGGACT TGCGTGGCAAAGGATTCGATAACGTGCTGATGGTGCACGACCACGCATTAATGGACTGGA TTGGGGCCAACTCCTACCGTACCTCGCATTACCCTTACGCTGAAGAGATGCTCGACTGGGC AGATGAACATGGCATCGTGGTGATTGATGAAACTGCTGCTGTCGGCTTTAACCTCTCTTTA GGCATTGGTTTCGAAGCGGGCAACAAGCCGAAAGAACTGTACAGCGAAGAGGCAGTCAA CGGGGAAACTCAGCAAGCGCACTTACAGGCGATTAAAGAGCTGATAGCGCGTGACAAAA ACCACCCAAGCGTGGTGATGTGGAGTATTGCCAACGAACCGGATACCCGTCCGCAAGGTG CACGGGAATATTTCGCGCCACTGGCGGAAGCAACGCGTAAACTCGACCCGACGCGTCCGA TCACCTGCGTCAATGTAATGTTCTGCGACGCTCACACCGATACCATCAGCGATCTCTTTGA TGTGCTGTGCCTGAACCGTTATTACGGATGGTATGTCCAAAGCGGCGATTTGGAAACGGC AGAGAAGGTACTGGAAAAAGAACTTCTGGCCTGGCAGGAGAAACTGCATCAGCCGATTA TCATCACCGAATACGGCGTGGATACGTTAGCCGGGCTGCACTCAATGTACACCGACATGT GGAGTGAAGAGTATCAGTGTGCATGGCTGGATATGTATCACCGCGTCTTTGATCGCGTCA GCGCCGTCGTCGGTGAACAGGTATGGAATTTCGCCGATTTTGCGACCTCGCAAGGCATAT TGCGCGTTGGCGGTAACAAGAAAGGGATCTTCACTCGCGACCGCAAACCGAAGTCGGCG GCTTTTCTGCTGCAAAAACGCTGGACTGGCATGAACTTCGGTGAAAAACCGCAGCAGGGA GGCAAACAATGA	45
gusA-optimized	ATGCTGCGCCCGGTGGAAACCCCGACCCGCGAAATCAAGAAGCTCGACGGCCTGTGGGC GTTCAGCCTCGACCGCGAAAACTGCGGCATCGACCAGCGGTGGTGGGAGAGCGCGCTGC AGGAATCGCGCGCGATCGCCGTCCCGGGCTCGTTCAACGACCAGTTCGCGGACGCCGACA TCCGGAACTACGCCGGCAACGTCTGGTACCAGCGCGAGGTGTTCATCCCGAAGGGCTGGG CGGGCCAGCGGATCGTCCTGCGCTTCGACGCCGTGACCCACTACGGCAAGGTCTGGGTGA ACAACCAGGAAGTCATGGAACACCAGGGCGGCTACACCCCGTTCGAGGCGGACGTCACG CCGTACGTGATCGCCGGCAAGTCCGTCCGGATCACCGTCTGCGTGAACAACGAGCTGAAC TGGCAGACCATCCCGCCGGGCATGGTGATCACGGACGAGAACGGCAAGAAGAAGCAGTC CTACTTCCACGACTTCTTCAACTACGCGGGCATCCACCGCAGCGTCATGCTGTACACCACG CCGAACACCTGGGTCGACGACATCACCGTCGTGACGCACGTGGCGCAGGACTGCAACCAC GCCAGCGTCGACTGGCAGGTCGTGGCCAACGGCGACGTCTCGGTGGAGCTGCGGGACGC GGACCAGCAGGTCGTGGCCACCGGCCAGGGCACCTCGGGCACGCTGCAGGTCGTGAACC CGCACCTCTGGCAGCCGGGCGAGGGCTACCTGTACGAACTCTGCGTCACCGCGAAGTCGC AGACGGAGTGCGACATCTACCCGCTGCGGGTGGGCATCCGCTCCGTCGCCGTGAAGGGCG AGCAGTTCCTCATCAACCACAAGCCGTTCTACTTCACCGGCTTCGGCCGGCACGAGGACG CGGACCTGCGCGGCAAGGGCTTCGACAACGTCCTGATGGTGCACGACCACGCGCTCATGG ACTGGATCGGCGCCAACTCGTACCGGACCTCGCACTACCCGTACGCGGAGGAAATGCTGG ACTGGGCCGACGAGCACGGCATCGTCGTGATCGACGAAACGGCGGCCGTCGGCTTCAACC TGAGCCTCGGCATCGGCTTCGAGGCGGGCAACAAGCCGAAGGAACTCTACTCGGAGGAA GCCGTGAACGGCGAGACCCAGCAGGCGCACCTGCAGGCCATCAAGGAACTCATCGCGCG GGACAAGAACCACCCGTCCGTCGTGATGTGGAGCATCGCCAACGAGCCGGACACCCGGC	46

Name	Sequence	SEQ ID NO:
	CGCAGGGCGCGCGCGAGTACTTCGCCCCGCTGGCGGAAGCCACGCGGAAGCTCGACCCG ACCCGCCCGATCACGTGCGTCAACGTGATGTTCTGCGACGCGCACACCGACACGATCTCG GACCTGTTCGACGTGCTGTGCCTCAACCGGTACTACGGCTGGTACGTGCAGTCCGGCGAC CTGGAGACCGCGGAAAAGGTGCTCGAGAAGGAACTGCTCGCCTGGCAGGAGAAGCTGCA CCAGCCGATCATCATCACCGAATACGGCGTGGACACGCTGGCCGGCCTCCACTCGATGTA CACGGACATGTGGTCCGAGGAATACCAGTGCGCGTGGCTGGACATGTACCACCGGGTCTT CGACCGCGTGTCCGCCGTCGTGGGCGAGCAGGTCTGGAACTTCGCGGACTTCGCCACCAG CCAGGGCATCCTCCGGGTGGGCGGCAACAAGAAGGGCATCTTCACGCGGGACCGCAAGC CGAAGAGCGCGGCCTTCCTGCTCCAGAAGCGCTGGACCGGCATGAACTTCGGCGAGAAGC CGCAGCAGGGCGGCAAGCAGTGA	
insert_1	GAACGACCGAGCGCAGCGGCGGCCGCGCTGATACCGCCGCtctagaCGAAGAGAAGTACCTC GACTACCTGCGTCGTGCCACGGCGGACCTCCACGAGGCCCGTGGCCGCCTCCGCGAGCTG GAGGCGAAGGCGGGCGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGCGG CGTCGCCTCGCCCGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCGGA GTTCCCCCAGGACCGCGGCTGGGACGTGGAGGGCCTGTACGACCCGAACCCGGAGGCCA CGGGCAAGAGTTACGCCCGCGAGGCCGGATTCCTGTACGAGGCGGGCGAGTTCGACGCC GACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGTCTCCTCC TGGAGGCCTCCTGGGAGGCGTTCGAGCACGCCGGGATCCCGGCGGCCACCGCGCGCGGC ACCTCGGTCGGCGTCTTCACCGGCGTGATGTACCACGACTACGCCACCCGTCTTCACAGCG GGGTGGAGCCGGGGTTCAGGGGGCAGATCATGATGTAGTCGGGGTCGTTGACCGGCAGC GCGCCGGGCGACGGGACCCGCTCGCCCTTCTGCGCGGCGATGGCGAGGGTCCGGAGCAG CACCTCGGACGCGCGGTGCACGGCCTCTTCCTTGGTGGTGGCGGTGACGTGCAGGTTGTC GAAGTCGCCGATCTCGATGCGGTGCGCCTTCTGGCCGTTCTCCTCGTCGATGGTGGTGATG TGGGCGGGGTACGGCACGAACCACGCGGCGACCTTCTGGGTCTGGCGGCGACGCGCGTCG AGCTTCGACACCAGCTGCCAGGCGCGCAGGTGGGTGTACCGCTTGAGCATGTTCATCGAG CGGTGGCCGGAGATCGCGGCGATCTCCATGACGTTGAGGCTGCCCAGCTCGAAGAACCGC GAGATGGCCTCGTGGCGCAGGTCGTGGAAGTGCAGGTCCTCGATGCGCAGCCGCTGGGTC GCGATGCGCCAGGCGTTCTTGAAGCCGCTGGCGGTGTAGTCGAACACGTTGCCGTGCAGG TTGACGGGCATCATCTGGAGGAAGTTCCGGGCCCGGCGGGACAGGGGCACGTCGCGGCT GTGGCCGTTCTTGGTCTCCGGCAGGTGCGCGACGCCGTGGCGCAGGTCGATGTGCTCCCA GCGGAGGGCCAGGATCTCGCCCTGGCGCATCGCGGTCTCCAGGGCCAGGTGGAAGATCAC GTAGAGCATCAGGTTCTTCTCGCGGAAGTACCGGGACAGCCGGCGCTCCTCGGACGAGGT GAGCCGGCGGTCCCGGCCGGAGCTGACCTTGGGCTTGCGCACCAGCTCGACCGGGTTGGT CCGGCAGGTGCCCCACTCCACGCGCGCGATGTTGAACAGCGAGGAGAGCAGGGCGAGCT CCAGCCGGACGGTGTTGCCGGTGATGGGCTTGCCGGTGCGCGGGTTGATCTCCGCCAGGC GCACGTCCCGGTAGGTGGCGATGTCGACGGTGGTGATCTCGTCCATGTTCCGCAGGGCGA TCGGGTAGCGCTTGATGACGTTGGACCGGTAGAACTCCTGCTGGTGGCCCTTCTTGTGGAC GCTCACGGTCTTCAGGTACTTGTCGAGGGCGCGGGAGAGGCTCATCTTGCGGATGCGGGG GGACGGCATttccgctcccttctctgacgccgtccacgctgcctcctcacgtgacgtgaggtgcaagcccggacgttccgcgtgccacgcc gtgagccgccgcgtgccgtcggctccctcagcccgggttatccgactgttacggtagcgtgcatcctgttagccaatctgtcagtcagttgtcaggtc attaatgttatgtcattgacaatgatcatgttGATGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCCAG CGtggcctacGTGCCAATCCGGTACCTGGCCGGCCGGCGGAGCACACCCGGCCGTCTCCGGCC CGGCCGCGGCCGGGCCGGAAGCCATCCGCCGCCCACCCGGTACCGACCCCTCAAGCCCTT CAAGCCCTTCGACCCGTCCGATCAGTCAGTCCGGCGGTCCTCCACGACCGGTCCGGAATC GCCCCCACACGAGTCAGGAAGCACACCATGGCCATGTCCGCCGAGAGGCTGACGGAGGC GCTGCGGACCTCGCTCAAGGAGGCCGAGCGGCTCCGGCGGCAGAACCGCGAACTGAGGG CCGCGCGGGACGCGGCGCGGGAGCCGATCGCCGTCGTCGGCATGGCCTGCCGCTACCCGG GCGGTGTCACCGGCCCCGAGGAGCTGTGGGAGCTGGTGGCCGGAGGCCGGGACGCGATC GGGCCGTTCCCCGTGGACCGGGGCTGGGACGTGGCGTCGGTGTACGACCCGGATCCCGAG TCGAAGGGCACCACGTACTGCCGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGAC GCCGCCTTCTTCGGCATCTCGCCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTG CTGCTGGAGGTGTCCTGGGAGGCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGC GGCAGCCGCGGTGGTGTCTACGTGGGCGCCGCGCACGGCTCGTACGCCTCCGATCCCCGG CTGGTGCCCGAGGGCTCGGAGGGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCC GGCCGCATCTCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGC TCCTCCTCGCTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGG CTCGCGCTGGCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCC CGGCAGAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGG CACCGGCTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCG CGCGGGGCACACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTC CAACGGGCTGACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGC CGACGCCGGGCTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCC GGCTCGGCGACCCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCG GCGACCACCCGCTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCG CCGGTGTCGGCGGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCA CCCTCCACGCCGACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGC TCACCTCCGAGGTGCCGTGGCAGCGGACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCT TCGGCGTCGGCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGCCCCGCCCGCGC CGGAACCGGCCGGGGAGGCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCC CTGGCCTGGGTGGTCTCCGGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTC CGCGACCACCTCTCCCGCACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCG CCACGCGCGCAGCCTTTGACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCG CCGCCGCCCTGGACGCGTTGGCCGAAGGACGCGACGGTCCGGCGGTGGTGCGCGGAGTCC GCGACCGGGACGGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCCGGG ATGGCCCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTGACG	47

Name	Sequence	SEQ ID NO:
	GACCGTCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCTGGACGCCCGACCCGGC TCGCCCGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCGGCGCTCTTCGCCGTC GAGGTGGCGCTCCACCGGCTGCTGGAGCACTGGGGGATGCGCCCCGACCTGCTGCTGGGG CACTCGGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTGCTCGATCTCGACGACGCC TGCGCGCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTGCCGCCCGGCGGCGCGATG GTCTCCGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTGGCCGGCCGCGAGGACGC CGTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGC GGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGCACCAGGCGGCTCCGCG TCGCGCACGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCCGGATTCCGGGAGGTCG CCGCCGGCCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCCACGGTCACGGGGCGGC CCGCCCGCCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCCCAGGTCCGTGAGCCCG TGCGCTTCGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCCCGCACCTTCCTGGAGA CCGGCCCGGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTGGAGGACGACACCGTG GCCCTGCTGCCGGCGATCCACAAGCCCGGCACCGCGCCGCACGGTCCGGCGGCTCCCGGC GCGCTGCGGGCGGCCGCCGCCGCGTACGGCCGGGGCGCCCGGGTGGACTGGGCCGGGAT GCACGCCGACGGCCCCGAGGGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTTCCG GCACCGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCG GATCGGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCT GGTGATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGCCGAACGCGCGCT GCGCGCCGCGGGCGCGAGCCCCGTACCGCTGCCCGTGGACGCTCCGGCCGCCGACCGGGC GTCCTTCGCGGCACTGCTGCGCTCCGCCACCGGACCTGACACACGAGGTGACACAGCCGC GCCCGTGGCCGGTGTGCTGTCGCTGCTGTCCGAGGAGGATCGGCCCCATCGCCAGCACGC CCCGGTACCCGCCGGGGTCCTGGCGACGCTGTCCCTGATGCAGGCTATGGAGGAGGAGGC GGTGGAGGCTCGCGTGTGGTGCGTCTCCCGCGCCGCGGTCGCCGCCGCCGACCGGGAACG GCCCGTCGGCGCGGGCGCCGCCCTGTGGGGGCTGGGGCGGGTGGCCGCCCTGGAACGCCC CACCCGGTGGGGCGGTCTCGTGGACCTGCCCGCCTCGCCCGGTGCGGCGCACTGGGCGGC CGCCGTGGAACGGCTCGCCGGTCCCGAGGACCAGATCGCCGTGCGCGCGTCCGGCAGTTG GGGCCGGCGCCTCACCAGGCTGCCGCGCGACGGCGGCGGCCGGACGGCCGCACCCGCGT ACCGGCCGCGCGGCACGGTGCTCGTCACCGGTGGCACCGGCGCGCTCGGCGGGCATCTCG CCCGCTGGCTCGCCGCGGCGGGCGCCGAACACCTGGCGCTCACCAGCCGCCGGGGCCCGG ACGCGCCCGGCGCCGCCGGACTCGAGGCCGAACTCCTCCTCCTGGGCGCCAAGGTGACGT TCGCCGCCTGCGACACCGCCGACCGCGACGGCCTCGCCCGGGTCCTGCGGGCGATACCGG AGGACACCCCGCTCACCGCGGTGTTCCACGCCGCGGGCGTACCGCAGGTCACGCCGCTGT CCCGTACCTCGCCCGAGCACTTCGCCGACGTGTACGCGGGCAAGGCGGCGGGCGCCGCGC ACCTGGACGAACTGACCCGCGAACTCGGCGCCGGACTCGACGCGTTCGTCCTCTACTCCT CCGGCGCCGGCGTCTGGGGCAGCGCCGGCCAGGGTGCCTACGCCGCCGCCAACGCCGCCC TGGACGCGCTCGCCCGGCGCCGTGCGGCGGACGGACTCCCCGCCACCTCCATCGCCTGGG GCGTGTGGGGCGGCGGCGGTATGGGGGCCGACGAGGCGGGCGCGGAGTATCTGGGCCGG CGCGGTATGCGCCCCATGGCACCGGTCTCCGCGCTCCGGGCGATGGCCACCGCCATCGCC TCCGGGGAACCCTGCCCCACCGTCACCCACACCGACTGGGAGCGCTTCGGCGAGGGCTTC ACCGCCTTCCGGCCCAGCCCTCTGATCGCGGGGCTCGGCACGCCGGGCGGCGGCCGGGCG GCGGAGACCCCCGAGGAGGGGAACGCCACCGCTGCGGCGGACCTCACCGCCCTGCCGCC CGCCGAACTCCGCACCGCGCTGCGCGAGCTGGTGCGAGCCCGGACCGCCGCGGCGCTCGG CCTCGACGACCCGGCCGAGGTCGCCGAGGGCGAACGGTTCCCCGCCATGGGCTTCGACTC CCTGGCCACCGTACGGCTGCGCCGCGGACTCGCCTCGGCCACGGGCCTCGACCTGCCCCC CGATCTGCTCTTCGACCGGGACACCCCGGCCGCGCTCGCCGCCCACCTGGCCGAACTGCT CGCCACCGCACGGGACCACGGACCCGGCGGCCCCGGGACCGGTGCCGCGCCGGCCGATG CCGGAAGCGGCCTGCCGGCCCTCTACCGGGAGGCCGTCCGCACCGGCCGGGCCGCGGAA ATGGCCGAACTGCTCGCCGCCGCTTCCCGGTTCCGCCCCGCCTTCGGGACGGCGGACCGG CAGCCGGTGGCCCTCGTGCCGCTGGCCGACGGCGCGGAGGACACCGGGCTCCCGCTGCTC GTGGGCTGCGCCGGGACGGCGGTGGCCTCCGGCCCGGTGGAGTTCACCGCCTTCGCCGGA GCGCTGGCGGACCTCCCGGCGGCGGCCCCGATGGCCGCGCTGCCGCAGCCCGGCTTTCTG CCGGGAGAACGAGTCCCGGCCACCCCGGAGGCATTGTTCGAGGCCCAGGCGGAAGCGCT GCTGCGCTACGCGGCCGGCCGGCCCTTCGTGCTGCTGGGGCACTCCGCCGGCGCCAACAT GGCCCACGCCCTGACCCGTCATCTGGAGGCGAACGGTGGCGGCCCCGCAGGGCTGGTGCT CATGGACATCTACACCCCCGCCGACCCCGGCGCGATGGGCGTCTGGCGGAACGACATGTT CCAGTGGGTCTGGCGGCGCTCGGACATCCCCCCGGACGACCACCGCCTCACGGCCATGGG CGCCTACCACCGGCTGCTTCTCGACTGGTCGCCCACCCCCGTCCGCGCCCCCGTACTGCAT CTGCGCGCCCGCGGAACCCATGGGCGACTGGCCACCCGGGGACACCGGCTGGCAGTCCCAC TGGGACGGCGCGCACACCACCGCCGGCATCCCCGGAAACCACTTCACGATGATGACCGAA CACGCCTCCGCCGCCGCCCGGCTCGTGCACGGCTGGCTCGCGGAACGGACCCCGTCCGGG CAGGGCGGGTCACCGTCCCGCGCGGCGGGGAGAGAGGAGAGGCCGTGACCAGGTACCGG ATTGGCACttatccgGTGCCAATCCGGTCGGTGGTCAGGAGTGGAGAAGACATGCCGGGTACG AACGACATGCCGGGTACCGAGGACAAGCTCCGCCACTACCTGAAGCGAGTGACCGCGGA TCTCGGACAGACCCGTCAGCGCCTGCGCGACGTGGAGGAGCGCCAGCGGGAACCGATCG CCATCGTCGCGATGGCCTGCCGCTACCCGGGCGGGGTGGCCTCCCCCGAGCAGCTGTGGG ACCTGGTCGCCTCACGCGGCGACGCCATCGAGGAGTTCCCCGCCGACCGCGGCTGGGACG TGGCGGGCCTCTACCACCCCGACCCGGACCACCCCGGCACGACCTATGTACGAGAGGCCG GATTCCTGCGGGACGCCGCCCGCTTCGACGCCGACTTCTTCGGCATCAACCCGCGCGAGG CGCTCGCCGCCGACCCGCAGCAACGGGTGCTCCTCGAAGTGTCGTGGGAACTGTTCGAGC GGGCGGGCATCGACCCCGCCACGCTCAAGGACACCCTCACCGGCGTGTACGCGGGGGTGT CCAGCCAGGACCACATGTCCGGGAGCCGGGTCCCGCCGGAGGTCGAGGGCTACGCCACC ACGGGAACCCTCTCCAGCGTCATCTCCGGCCGCATCGCCTACACCTTCGGCCTGGAGGGC CCGGCGGTGACGCTCGACACGGCGTGCTCGGCATCGCTGGTCGCGATCCACCTCGCCTGC CAGGCCCTGCGCCAGGGCGACTGCGGCCTGGCGGTGGCGGGAGGCGTGACCGTACTGTCC	

Name	Sequence	SEQ ID NO:
	ACGCCGACGGCGTTCGTGGAGTTCTCACGCCAGCGCGGACTCGCACCGGACGGCCGCTGC AAGCCGTTCGCCGAGGCCGCCGACGGCACCGGATTCTCCGAGGGCGTCGGCCTGATCCTC CTGGAACGCCTCTCCGACGCCCGCCGCAACGGACATCAAGTACTCGGCGTCGTACGCGGA TCGGCCGTCAACCAGGACGGCGCGAGCAACGGCCTGACCGCCCCGAACGACGTCGCCCA GGAACGCGTGATCCGCCAGGCCCTGACCAACGCCCGCGTCACCCCGGACGCCGTCGACGC CGTGGAGGCACACGGCACCGGCACCACGCTCGGCGACCCGATCGAGGGGAACGCACTCC TCGCGACGTACGGAAAGGACCGCCCCGCCGACCGGCCGCTGTGGCTCGGCTCTGTGAAGT CGAACATCGGCCACACGCAGGCGGCTGCGGGCGTCGCAGGCGTCATCAAGATGGTGATG GCGATGCGCCACGGCGAGCTGCCCGCCTCCCTGCACATCGACCGGCCCACGCCCCACGTG GACTGGGAGGGCGGGGGAGTGCGGTTGCTCACCGATCCCGTGCCGTGGCCACGGGCCGA CCGCCCCCGCCGCGCGGGGGTCTCCTCCTTCGGCATCAGCGGCACCAACGCCCACCTGAT CGTGGAACAGGCCCCCGCCCCGCCCGACACGGCCGACGACGCCCCGGAAGGCGCCGCAA CCCCCGGCGCTTCCGACGGCCTCGTGGTGCCGTGGGTGGTGTCGGCCCGTAGTCCGCAGG CCCTGCGTGATCAGGCCCTGCGTCTGCGCGACTTTGCCGGTGACGCGTCCCGAGCGCCGCT CACCGACGTGGGCTGGTCTTTGCTGCGGTCGCGTGCGCTGTTCGAGCAGCGGGCGGTGGT GGCGGGGCGTGAGAGGGCTGAACTGCTGGCGGGGCTGGCTGCGTTGGCCGCTGGTGAGG AGCACCCGGCTGTGACGCGGTCCCGTGAGGAAGCGGCGGTTGCTGCGAGCGGTGATGTGG TGTGGCTGTTCAGTGGTCAGGGCAGTCAGTTGGTCGGTATGGGTGCTGGTTTGTATGAGCG GTTCCCGGTGTTTGCGGCTGCGTTTGATGAGGTGTGCGGCTTGCTGGAGGGGGAGCTGGG GGTTGGTTCGGGTGGGTTGCGGGAGGTGGTGTTCTGGGGCCCGCGGGAGCGGTTGGATCA CACGGTGTGGGCGCAGGCGGGGTTGTTTGCGTTGCAGGTGGGGTTGGCCCGGTTGTGGGA GTCGGTCGGGGTGCGGCCGGATGTGGTGCTCGGGCATTCGATCGGTGAGATCGCGGCCGC GCATGTGGCGGGGGTCTTTGATCTGGCGGATGCGTGTCGGGTGGTGGGGGCGCGGGCGCG TTTGATGGGTGGGTTGCCTGAGGGTGGGGCGATGTGTGCGGTGCAGGCCACGCCCGCCGA GCTGGCCGCGGATGTGGATGGCTCGTCCGTGAGTGTGGCGGCGGTCAACACACCTGACTC GACGGTGATTTCAGGTCCGTCGGGTGAGGTGGATCGGATTGCTGGGGTGTGGCGGGAGCG TGGGCGTAAGACGAAGGCGCTGAGCGTGAGTCATGCTTTCCATTCGGCGTTGATGGAGCC GATGCTCGGGGAGTTCACGGAAGCGATACGAGGGGTCAAGTTCAGGCAGCCGTCGATCCC GCTCATGAGCAATGTCTCCGGAGAGCGGGCCGGCGAGGAGATCACATCCCCGGAGTACTG GGCGAGGCATGTACGCCAGACAGTGCTCTTCCAGCCCGGCGTCGCCCAAGTGGCCGCTGA GGCACGCGCGTTCGTCGAACTCGGCCCCGGCCCCGTACTGACCGCCGCCGCCCAGCACAC CCTCGACCACATCACCGAGCCGGAAGGCCCCGAGCCGGTCGTCACCGCGTCCCTCCACCC CGACCGGCCGGACGACGTGGCCTTCGCGCACGCCATGGCCGACCTCCACGTCGCCGGTAT CAGCGTGGACTGGTCGGCGTACTTCCCTGACGACCCCGCCCCCCGCACCGTCGACCTGCC CACCTACGCCTTCCAGGGGCGGCGCTTCTGGCTGGCGGACATCGCGGCGCCCGAGGCCGT GTCCTCGACGGACGGTGAGGAGGCCGGGTTCTGGGCCGCCGTCGAAGGTGCGGACTTCCA GGCGCTCTGCGACACCCTGCACCTCAAGGACGACGAGCACCGCGCGGCTCTGGAGACGGT GTTCCCCGCGCTGTCCGCGTGGCGGCGCGAACGACGTGAGCGGTCGATCGTCGATGCCTG GCGGTACCGGGTCGACTGGCGGCGCGTCGAGCTGCCGACACCCGTTCCGGGCGCCGGTAC CGGTCCCGACGCCGACACGGGCCTCGGGGCGTGGCTGATCGTGGCTCCCACGCACGGGTC GGGTACTTGGCCGCAAGCCTGTGCCCGGGCGTTGGAGGAGGCGGGCGCGCCGGTACGTAT CGTCGAGGCCGGCCCGCACGCCGACCGGGCGGACATGGCGGACCTGGTCCAGGCATGGC GGGCAAGCTGTGCGGACGACACCACCCAGCTCGGAGGAGTGCTCTCCCTGCTGGCTCTCG CCGAGGCACCGGCCACCAGTTCCGACACCACTTCCCACACCAGTACCAGTTGCGGTACCG GCTCTCTCGCGTCCCACGGCCTCACCGGCACCTTGACGCTGCTGCACGGTCTGCTGGATGC GGGCGTCGAAGCGCCTCTCTGGTGTGCCACGCGCGGCGCCGTGTCGTGCGGCGACGCCGA TCCGCTCGTCTCCCCGTCGCAGGCCCCGGTCTGGGGACTCGGACGCGTGGCCGCCCTGGA GCATCCGGAGTTGTGGGGCGGCCTGGTCGACCTGCCCGCCGACCCGGAGTCGCTCGACGC GAGCGCGTTGTATGCGGTTCTGCGCGGAGACGGCGGCGAGGATCAGGTCGCGCTGCGCCG GGGCGCGGTCCTCGGCCGTCGCCTGGTGCCCGACGCAACCCCGGACGTGGCCCCCGGCTC GTCCCCGGACGTGTCCGGAGGCGCAGCCCATGCCGACGCGACCTCCGGGGAGTGGCAGCC GCATGGTGCCGTCCTCGTCACCGGAGGCGTCGGCCACCTGGCCGATCAGGTCGTACGGTG GCTCGCCGCGTCCGGCGCCGAACACGTCGTACTCCTGGACACGGGCCCCGCCAACAGCCG TGGTCCCGGCCGGAACGACGACCTCGCCGCGGAAGCCGCCGAACACGGCACCGAGCTGA CGGTCCTGCGGTCCCTGAGCGAGCTGACAGACGTATCCGTACGTCCCATACGGACCGTCA TCCACACATCGCTGCCCGGCGAGCTCGCGCCGCTGGCCGAGGTCACCCCCGACGCGCTCG GCGCGGCCGTGTCCGCCGCCGCGCGGCTGAGCGAACTCCCCGGCATCGGGTCAGTGGAGA CCGTGCTGTTCTTCTCCTCCGTGACGGCTTCGCTCGGCAGTAGGGAGCACGGCGCGTACGC CGCCGCCAACGCCTACCTCGACGCCCTGGCGCAACGGGCCGGTGCCGATGCTGCGAGCCC CCGGACGGTCTCGGTCGGGTGGGGCATCTGGGATCTGCCGGACGACGGTGACGTGGCACG CGGCGCCGCCGGGCTGTCCCGGAGGCAGGGACTCCCGCCGCTGGAACCGCAGTTGGCGCT CGGCGCCCTGCGCGCGGCGCTCGACGGGGGCAAGGGGCACACGCTGGTCGCCGACATCG AGTGGGAGCGGTTCGCGCCGCTGTTCACGCTGGCCAGGCCCACCCGGCTGCTCGACGGGA TCCCCGCGGCCCAGCGGGTCCTCGACGCCTCCTCGGAGAGCGCCGAGGCCTCGGAGAACG CCTCGGCCCTCCGTCGCGAACTGACGGCCCTGCCCGTGCGGGAGCGGACCGGGGCACTTC TCGACCTGGTCCGCAAACAGGTGGCCGCCGTCCTGCGCTACGAGCCGGGCCAAGACGTGG CGCCCGAGAAGGCCTTCAAGGACCTGGGCTTCGACTCGCTCGTGGTCGTGGAGCTGCGCA ACCGGCTGCGCGCCGCCACCGGGCTCCGGCTGCCCGCCACCCTGGTCTACGACTACCCCA CACCCCGCACCCTCGCCGCACACCTGCTGGACAGGGTGCTGCCCGACGGCGGCGCGGCAG AGCTCCCCGTGGCCGCCCACCTGGACGACCTGGAGGCGGCCCTCACCGACCTGCCGGCCG ACGACCCCCGGCGCAAGGGCCTGGTCCGGCGTCTACAGACGCTGCTGTGGAAGCAGCCCG ACGCCATGGGGGCGGCGGGCCCCGCCGACGAGGAGGAGCAAGCCGCGCCCGAGGACCTG TCGACCGCGAGCGCCGACGACATGTTCGCCCTGATCGACCGGGAGTGGGGCACGCGGTGA CCACCGACCGGATTGGCACgtaggccaGTGCCAATCCGGTACTTGGcaattgCCTCGCCGCAGTTA ATTAAAGTCAGTGAGCGAGGAAGCGC	

Name	Sequence	SEQ ID NO:
insert_2	GAACGACCGAGCGCAGCGGCGGCCGCGCTGATACCGCCGCcaattgTGACCCGGGCCCACCC CAACTACGTGCGAAACATCTGAGGAAGGTTCAACGGACGATGCCGGACGAAAGTAAGCT CGTCGATTACCTGAAGTGGGTCACGGCGGATCTCCACCAGACCCGCAGGCGCCTCCAGGA AGCCGAGTCGGGCCGCCACGAGCCCGTGGCGATCGTCGGCATGGCGTGCCGCTTCCCCGG CGGTGTGCGTTCCCCGGAGGACCTCTGGGAGATGCTCGCCGACGGCCGCGACGCCATCTC CGGGTTCCCCGCCGACCGCGGCTGGGACCTGGAGACGCTGGCCGGCGACGGAGCGGGCG GCAGCAGCACGCAGGAGGGAGGTTTCCTGCACGACGTGGCCGACTTCGACCCGGGGTTCT TCGACATCTCGCCCCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGCTGCTCCTGGAGA CCGCCTGGGAGGCCGTGGAGCGTGCCGGGATCGCCCCCGGCAGCCTCCGCGGCAGCCGCA CCGGCGTGTTCGTCGGCACCAACTCGCAGGACTACGCCCACCTCGTCCTCGCCTCCGACG ACGACATGGGCGGCTACGCGGGCAACGGCCTGGCCGCCAGCGTGATGTCCGGCCGGCTGT CCTTTGCGCTCGGCTTCGAAGGACCCGCCGTCACCCTCGACACCGCGTGCTCCTCGGCCCT GGTCGCCCTGCACCTGGCCGCCCAGTCGGTGCGCTCCGGCGAGGCCGACCTCGCGCTGGC CGGCGGTGTCACCGTCATGACCACCTCGTCCAGCTTCGTGGGCTTCAGCCTCCAGGGCGG CCTCGCCACGGACGGCCGCTGCAAGGCGTTCGCCGACTCCGCCGACGGCACCGGCTGGTC CGAGGGCGTCGGCATGATCCTCGTCGAACGGCTCTCCGAGGCCCGGCGCAAGGGCCACCC GGTGCTCGCCGTGCTGCGCGGCTCCGCCGTGAACCAGGACGGCGCCTCCAACGGCCTCAG CGCCCCCAACGGGCCCGCCCAGCAGCGCGTCATCCGCGACGCGCTGGCCTCCGCCGGGCT CTCGCCCGCCGACGTCGACGCCGTCGAGGCGCACGGCACCGGCACCACCCTCGGCGACCC CATCGAGGCGCAGGCCCTGCTCGCCACCTACGGCCAGGACCGCGACGCGAGCCGTCCGCT GCGCCTCGGCACGGTGAAGTCCAACATCGGCCACACCCAGGCCGCCGCCGGCGCCGCGG GCGTCATCAAGATGGTCCTGGCCCTGCGCCACGGGCTGCTGCCCCGCACCCTGCACATCG ACGCGCCGTCCACACACGTCGACTGGGACGCAGGACACGTCAGCCTGCTCACCGAGGCCA CCCCCTGGCCCGAGGGCGAGCAGACGCGCCGGGCCGGTGTGTCGTCCTTCGGCATCAGCG GCACCAACGCCCATGTCATCCTCGAAGAGGCCCCCGCGGCCGAGGAGGACACCGACGCC GAGCAGCGGCCCGAGCCGGTCGTCCCGGGCGCCGTGCCGTGGCCCGTGTCGGCCCGTACC GCCGACGCCCTCGACGCCCAGCTGGAGAAGGTGCGGCCGCTCGCCGCGTCCGGCGCCGAC CCCGTCGCCGTCGGCCACGCCCTGGCCGTCACCCGCACCCCCTTCGAACACCGGGCGCTG CTGGTCGCCGCGGACGGCAATCTCACCGAGGCCGCACGCGGCACCGTCCCCTCCGGCGAC CGCCCCGGGCTGGCCGTGCTGTTCTCCGGGCAGGGCGCCCAGCGCCTCGGCATGGGCCGT GAACTCCACGCCCGCTTCCCGGTGTTCGCCGCCGCCCTGGACGAGACCCTCGCCCTCCTCG ACGAGCGCCTCGGCCACTCGCTGCGCGACGTCATCTGGGGCGAGGACCCCGTGGCCCTCG ACGACACCGGCCACACCCAGCCCGCGCTGTTCGCCGTGGAGGTCGCCCTCTACCGCCTGT TCGCCTCCTGGGGTCTGCGCCCCGACCACCTCGCCGGACACTCCGTCGGGGAGATCGCCG CCGTGCATGTGGCCGGCGCACTGTCCCTGGAGGACGCCTGCACCCTGGTTGCCGCCCGCG CCGGCCTGATGCGGGACCTGCCGGCCGGCGGAGCCATGGTCGCGCTGCGCGCCACCGAGG CCGAGGTGCGGCCGCTGCTCGACGAGCGGGTCTCGCTCGCCTCGGTCAACGGCCCCGATT CCGTGGTGGTCTCCGGCGCCGAGGACGCCGTCCTCGCCGTCGCCGAGCACTTCCGGAAGC AGGACCGGCGCACCACGCGGCTCTCCGTGAGCCACGCCTTCCACTCCCCGCTCGTCGACC CGATGCTCGACGCCTTCCGGGACGTCGTCGCGAAGCTCACCTTCGGCGAGCCGTCGGTAC CCGTCGTCTCCACCCTCACCGGCGACGTCGTCGCCGCCGAGGAACTGGCCACGCCGCACT ACTGGGTGTGCCACGCCCGGCAGGCCGTCCGCTTCGCCGACGCCGTGCGCACCCTCGTCG ACGAGGGCGCCCGGACCTTCCTGGAGGTCGGCCCGGGCGGCGTGCTGTCCGCGCTCGTCG GTGAGAACACCCAGGAGGCCGGGGTCTCTGCCGTGCCCGCGCTGCGCAAGGACCAGCCG GAGGAGGCCTCCGTGCTCGCCGCCCTCGGCACCCTGTGGACCCGGGGCACCGCGGCCGAC TGGGACGCCGTGTTCGAGGGCACCCTGCCCGGCCGCCCGGAGTCCGTGGACCTGCCGACC TACGCCTTCCAGCGCGGCCGGTACTGGCCCACCGTGCGCGCCCGCTCCGGCGACCCCGCC GGACTGGGCCTCGGCGCCGCCGGGCACCCGCTGCTCAGCGCCACCGTCACCCTCGCCGAG GCCGACGAGTCCGTCCTCACCGGACGGCTCTCGCCGCTGACCCACCCCTGGCTCACCGAG CACCGCGTCGACGGCCGGATCACCGTCCCCGGCACGGCCCTCGTGGAGATCGCGGTCCGT GCGGGCGACGAGAACGGCACCCCCCGCCTGGACCGGCTCGACCTGCTCGCCCCGCTGACC CTCGGCGACCGCGACGCGGTGCTGCTCCAGGTGCGCGTCGGCCCCGAGGACGCCTCCGGC CGGCGGACGCTCTCCGTGCACGCCCGCCCGGCGACCGCGGACGACGCCCCCTGGACCACG TACGCCCGTGGTGTCCTCGCCCCCGACGAGGGCGCCGAGGACACCGCCGCGGACCTTGTG GAGTGGCCGCCCGCCGACGCCCGTCAGGTCCCGCTCACGGAGCTGGAGTCCGAAGGACGC GCCCTCGGCCCGCTCTTCAGCGGCCTCACCGGGGTGTGGCGGCACGAGGGCGAGGTGTTC GCCGAGGCCGAGCTGCCCGCAGGACCCGACAGCGGCTTCGGACTGCACCCCGCGCTGCTG TCCACCGCCCTGCGCGCCGCCGCCGTCCTCGACGGCACCACGGCCGGCGAGCCCGCCGCC TTCGAGGGGCTCACCCTGCACGCCACCGGCGCCACCGCGCTGCGCGTACGGCTGAGCACC ACCGGGCCCGACACCGTCGACCTCACCGCCGTCGACCCCGCGGGCAACCTCGTCCTGACC GCCGAGACCGTGCGTCTCGGCACCCCGGACGGCACCGCCGACACCCCGGCCGCCACCGGC CGGGGCGACCTGTTCGGGCTGAAGTGGGTGCCCGTCAAGGCCTCGGAGCGCGCCACCGGC ACCCGCTGGGCCGTCGTCGGCTCCGACGAACTCGACCTCGGCTACGCCATGCACCGCGCC GACGAGACCGTCACCGCCTACGCGGAGTCCCTCGGCGGCGCCATCGGCGACAGCGGGGTC GCCCCCGATGTGTTCCTGATCCCGCTCGCAGGCGAGAAGGACGCCGGAGCGGAGGGCGTG CACGCCCTCACCACCCGGGTCCTCGGCTACCTCCAGGAGTGGCTGTCCGAGCCCCGGCTCT CCGGCACCCGCCTGGTCTTCGTCACCCGCCGCGCCGTCGCCCTCGACGACGAGGACGTCC TGGACCCGGCCGGCGCGGCCGTCTGGGGCCTGGTGCGCTCCGCGCAGACCGAGAACCCCG GCAGCCTGCTGCTCGTCGACCTCGACGACACCTTCCTGTCCGCCGGAGTACTGCCCGACGT GCTGACGCTCGACGAGCAGCAGCTCGCGGTCCGCGACTACCAGGTCCGCGCCGCCCGGCT CGCCCGGCTGCCGCGCCCGGCCGACGACGCCCCCGCCGCCGACTGGAACCCCGACGGCAC CGTCCTGATCACCGGCGGCACCGGCGGCCTCGGCGCCGCGCTCGCCCGGCACCTGGTCAC CAGCCGGGGCGCCCGTCACCTGCTGCTGGCGAGCCGCCGCGGCCCGGACGCGCCCGGCAC GTCCGAACTGGTCGCCGAGCTGACCGGCCTGGGAGCCCAGGTGACCGTCTCCGCCTGCGA CGTCGGCGACCGCGACGCGGTCGACACGCTCGTCGCGTCGGTGCCCGCCGAGCACCCGCT	48

Name	Sequence	SEQ ID NO:
	GACCGCCGTCGTGCACACCGCGGGTGTCCTGGACGACGCCCTGACCGGCTCGCTCACCCC GGAACAACTGGCTGGTGTGCTGCGCCCCAAGGCCGACGCGGCCCTGCACCTGCACGAGGC GACCCTCGGCCAGGACCTCGCCGCGTTCGTGCTGTACTCCTCGATCTCCGGCGTCATCGGC GGCCCCGGCCAGGCCAACTACGCCGCCGCCAACGCCTCCCTGGACGCGCTCGCCCACCGG CGCAGGGCGGCCGGTCTGCCCGGTCTGTCCCTCGCCTGGGGGCCCTGGGGCCGCGGCAGC GGCATGACCAGCCAGGTCAGCGACACCGACCTGGGGCGGATGGACCGCGGCGGCACCCC GCCGATGAGCCTCGAGGACGGCCTCGCCCTGTTCGACGCCGCCCTGGCCCGCACCGAGCC GATGGTCGTGCCCACCCGGATCAATGTCACCGGACTCCAGGTCCAGCAGACGCTGCCCGC GCTCTGGCGCGACCTGGTGCCGCGCCCCCGGCGCACCGCCGCCGCGGACCGCTCGCCCAA GACCGTGCTCGACGGGCTGCGCACGCTCGACACCGCGGGCCGGGAGAAGCTGCTCACCG AGCTGGTCGTCGGCTTCACCGCGGGCCTGCTCGGCCACGCCGACCCCGCCGCCGTCGACC CCGAGCGCGGCTTCCTGGAGCTGGGCTTCGACTCGCTGGTCTCGGTCAGCCTGCGCAACC AGCTCGGCGAACTCCTCGGGCTGCGCCTGCCGACGTCGGTGGTCTTCGACAGCAAGACGC CGGTGAAGCTGGCCCGCCACCTCAACGAGGAACTGGGCGACCTGTCCGCCTCCGGCCCCG CGTCCGGCACCGCCGTCGCCGGCACCACGGTGCATCCCGACGACACCCTGGTCGGCCTGT TCCACAACGCGGTGCGCGGCGGCAAGCTCGTCGAGGCGATGCGGATGCTGAAGGCGGTC GCCAACACCCGGCCCACCTTCGAGACGCCCGCGGACCTGGAGGAGCTGTCCGAGGCCGTC ACCCTGGCCACGGGGCCCGGCTCCCCGCGGCTGATCTTCGTCAGCGCGCCCGGCGCCACC GGCGGTGTCCACCAGTACGCGCGCATCGCCGCCCACTTCCGCGGCAAGCGCCATGTGTCG GCGATACCGCTGATGGGCTTCGCCCCCGGCGAGCTGCTGCCCGCCACCAGTGAGGCGGCG GCCCGGATCGTCGCCGAGAGCGTCCTGATGGCGAGCGACGGCGAACCGTTCGTGATGGTG GGCCACTCCACCGGCGGTTCCCTCGCCTACCTCGCGGCCGGGGTGCTGGAGGACACCTGG GGCGTGAAGCCGGAGGCGGTCGTCCTGCTCGACACGGCGTCCATCCGGTACAACCCCTCG GAGGGCAACAACCTCGACCAGACGACCCGCTTCTACCTCGCCGACATCGACTCGCCGTCC GTGACGCTCAACAGCGCCCGGATGTCCGCGATGGCGCACTGGTTCATGGCGATGACGGAC ATCGACGCGCCCGCCACGACCGCCCCCACCCTGCTCCTGCGCGCCACACAGGCCAACAAC GGCTTCATGCTGGACACCTCCGCGGTGCCGGCGGACGTGGTGCGTGACATCGAGGCCGAC CATCTGTCCCTGGCCATGGAGCACTCGGATCTGACCGCCGAGGCCATAGAGAACTGGCTC GCCGAACTTCCGGCCGGCGAGGCCTGACCAAGTACCGGATTGGCACtggcctacGTGCCAATC CGGTACGTGGGCAGCGATCAGTCCTGAGCGCGATGAGCCCCGGCCACACCGGCCCCCTTG CCCTCCCCCAACCTCATACGCCCCGATCACCACGTAGCGCCAAGGAGCCTGGGTCAGATG TCGAACGAGGAGAAGCTTCTCGATCACCTCAAGTGGGTCACCGCGGAGCTGCGCCAGGCC CGGCAACGGCTCCACGACAAGGAATCGACCGAGCCGGTCGCCATCGTCGGCATGGCCTGC CGCTACCCGGGCGGCGCCCGGTCGGCGGAGGACCTCTGGGAACTCGTGCGCGACGGGGG TGACGCGGTCGCGGGGTTCCCCGACGACCGGGGCTGGGACCTGGAGTCGCTGTATCACCC GGATCCGGAGCATCCGGCCACCAGCTATGTGCGGGACGGCGCCTTCCTGTACGACGCCGG CCATTTCGACGCCGAGTTCTTCGGCATCAGCCCGCGCGAGGCCACGGCGATGGATCCGCA GCAGCGGCTGCTCCTGGAGACCGCGTGGGAGGCGATCGAGCACGCGGGCATGAACCCGC ACGCGCTGAAGGGCAGCGACACCGGCGTCTTCACCGGAGTGAGCGCCCACGACTACCTGA CGCTGATCAGCCAGACGGCCAGCGACGTCGAGGGGTACATCGGCACCGGCAACCTCGGC AGTGTGGTGTCGGGCCGGATCTCCTACACGGTGGGGCTCGAAGGTCCCGCGGTCACCGTC GACACGGCGTGCTCGTCGTCCCTGGTGGCGATCCATCTGGCAAGTCAGGCGCTGCGGCAG GGCGAGTGCTCGCTCGCGCTGGCGGGCGGTTCGACGGTGATGGCGACGCCGGGTTCGTTC ACCGAGTTCTCCCGGCAGCGGGGGCTCGCGCCCGACGGGCGGTGCAAGCCGTTCGCGGCC GCCGCCGACGGCACCGGCTGGGGCGAGGGCGCCGGGGTGGTGGCGCTGGAGCTGCTCTC CGAAGCGCGGCGCCGCGGCCACAAGGTCCTTGCGGTGATACGGGGTTCGGCCACCAACCA GGACGGTACGAGCAACGGGCTCGCCGCCCCCAACGGCCCCTCCCAGGAACGCGTCATCCG CGCCGCCCTCGCCAACGCCCGCCTCTCCGCCGAGGACATCGACGCCGTAGAGGCGCACGG CACCGGCACCACGCTCGGCGACCCCATCGAGGCGCAGGCCCTGATCGCCACGTACGGCCA GGGGCGCCCGGAGGACCGTCCGCTCTGGCTCGGCTCGGTGAAATCCAACATCGGCCACAC GCAGGCCGCCGCCGGTGTCGCGGGCGTCATCAAGATGGTCATGGCGATGCGCAACGGTCT CCTCCCGACCTCCCTGCACATCGACGCCCCGTCACCGCACGTCCAGTGGGAACAGGGCAG CGTACGACTGCTCTCCGAGCCCGTGGACTGGCCGGCGGAGCGCACGCGGCGGGCAGGCAT CTCGGCGTTCGGGATCTCCGGGACGAACGCGCACCTGATTCTGGAGGAGGCGCCGCCGGA AGAGGACGCGCCGGGCCCCGTGGCGGCTGAGCCGGGTGGGGTCGTGCCGTGGGTGGTGT CCGGGCGGACGCCGGACGCGTTGCGTGAACAGGCGCGGCGGCTGGGCGAGTTCGCGGCC GGGCTGGCGGATGCGTCGGTCTCCGAGGTGGGCTGGTCGCTGGCCACGACACGCGCGTTG TTCGATCAGCGGGCCGTGGTCGTGGGGCGGGACTTGGCTCAGGCTGGTGCCAGCCTGGAG GCATTGGCCGCGGGTGAGGCGTCGGCGGATGTGGTGGCCGGGGTGGCCGGTGATGTGGGT CCTGGGCCGGTGTTGGTGTTTCCGGGGCAGGGGTCGCAGTGGGTGGGCATGGGCGCCCAG CTCCTTGACGAGTCGCCCGTCTTCGCGGCGCGGATCGCGGAGTGTGAGCAGGCGCTGTCG GCGCATGTGGACTGGTCGCTGAGTGATGTCTTGCGCGGGGACGGGAGCGAGCTGTCCCGG GTCGAGGTCGTGCAGCCCGTGCTGTGGGCGGTGATGGTCTCGCTGGCTGCGGTGTGGGCC GATTACGGCATCACCCCGGCTGCCGTCATCGGGCACTCGCAGGGCGAGATGGCTGCCGCG TGTGTGGCGGGGGCACTGTCACTGGAGGATGCGGCACGGATCGTAGCCGTACGCAGCGAC GCGCTTCGTCAGCTGCAAGGGCACGGCGACATGGCCTCGCTCAGCACCGGTGCCGAGCAG GCCGCAGAGCTGATCGGCGACCGGCCGGGCGTAGTCGTCGCGGCCGTCAACGGACCGTCG TCGACCGTGATTTCGGGCCCGCCGGAGCATGTGGCAGCCGTCGTCGCCGATGCGGAGGCG CAGGGACTCAGGGCCCGCGTCATCGACGTCAGGTACGCCTCCCACGGTCCCCAGATCGAC CAGCTCCACGACCTCCTCACCGACCGGCTCGCCGACATCCAGCCGACCACCACCGACGTG GCGTTCTACTCGACGGTCACCGCAGAGCGCCTCGACGACACCACCGCCCTAGACACCGCC TACTGGGTCACCAACCTCCGCCAGCCCGTCCGCTTCGCCGACACCATCGAAGCGCTGCTG GCCGACGGCTACCGCCTGTTCATCGAGGCCAGCCCCCACCCCGTCCTCAACCTCGGCATCC AGGAGACCATCGAGCAGCAGGCCGGTGCTGCGGGGACGGCCGTCACCATCCCCACCCTGC GTCGCGACCACGGTGACACCACCCAGCTCACCCGCGCGGCCGCCCACGCCTTCACCGCCG	

Name	Sequence	SEQ ID NO:
	GCGCCCCCGTCGACTGGCGGCGCTGGTTCCCGGCCGACCCCACCCCCCGTACCGTCGACC TCCCCACCTACGCCTTCCAGCACAAGCACTACTGGGTGGAGCCGCCCGCGGCGGTCGCAG CCGTGGGTGGTGGGCACGATCCGGTCGAGGCCCGGGTGTGGCAGGCGATCGAGGACCTG GACATCGACGCCCTCGCCGGCAGTCTGGAGATCGAGGGGCAGGCGGAGAGCGTCGGAGC GCTGGAGTCCGCGCTGCCCGTCCTCTCGGCCTGGCGGCGTCGGCACCGCGAGCAGTCCAC CGTCGACTCCTGGCGTTATCAGGTCACTTGGAAGCATCTGCCCGACGTGCCGGCGCCGGA GCTCAGCGGGGCCTGGCTGCTGCTCGTGCCCGCCGCGCACGCCGACCACCCGGCCGTCCT CGCGACCGCGCAGACGCTGACCGCCCATGGTGGCGAGGTGCGACGCCACGTGGTCGACG CACGTGCCATGGAGCGTACGGAGTTGGCGCAGGAGCTGCGTGTCCTGATGGACGGGGCCG CGTTTGCCGGAGTCGTCAATCTGCTGGCCCTGGACGAGGAGCCGCATCCCGAGCACTCGG CCGTGCCCGCCGGACTCGCCGCGACGACCGCTCTCGTCCAGGCCCTCGCGGACAACGGCG CCGACATCGCCGTACGCACTCTCACGCAGGGGGCTGTTTCCACGAGCGCCGGCGACGCCC TCACCCACCCGGTGCAGGCTCAGGTGTGGGGGCTGGGGCGCGTCGCTGCGCTGGAGTATC CGCGGCTGTGGGGCGGGCTGGTCGATCTGCCCGCTCGTATCGACCATCAGACGCTGGCCC GGCTGGCCGCCGCGCTGGTTCCGCAGGACGAGGACCAGATCTCCATCCGGCCGTCCGGCG TCCATGCCCGTCGCCTTGCACACGCGCCCGCCAACACGGTCGGCAGCGGGCTTGGTTGGC GGCCCGACGGCACCACTCTCATCACCGGCGGGACCGGCGGCATCGGCGCCGTCCTCGCGC GGTGGCTCGCCCGTGCGGGCGCCCCGCACCTCCTCCTGACCAGCCGCCGCGGCCCCGACG CCCCGGGAGCACAGGAACTCGCCGCGGAACTGACGGAGTTGGGGGCCGCCGTCACCGTC ACCGCCTGCGACGTCGGCGACCGCGAGCAGGTGCGACGCCTCATCGACGATGTCCCCGCC GAGCACCCGCTGACCGCCGTCATCCACGCGGCCGGCGTGCCGAACTACATCGGTCTCGGC GACGTGTCGGGTGCCGAGCTGGACGAGGTGCTGCGTCCGAAGGCGCTCGCCGCTCACCAT CTGCATGAACTGACCCGGGAGTTGCCGCTCTCGGCGTTCGTGATGTTCTCGTCGGGCGCAG GCGTGTGGGGCAGTGGCCAGCAGGGCGCCTATGGTGCGGCCAACCACTTCCTCGACGCCC TCGCCGAGCACCGCCGCGCCGAGGGCCTGCCCGCCACCTCCATCGCCTGGGGGCCCTGGG CCGAGGCGGGCATGGCGGCGGACCAGGCCGCGTTGACGTTCTTCAGCCGCTTCGGCCTGC ACCCGCTCAGCCCGGAGCTGTGCGTCAAGGCGCTGCAGCAGGCCCTGGACGCGGGTGAG ACGACGCTGACCGTGGCGAACTTCGACTGGGCGCAGTTCACGTCGACGTTCACCGCGCAG CGGCCCAGCCCGCTCCTCGCCGATCTGCCCGAGAACCGGCGGGCCAGCGCACCCGCGGCA CAGCAGGAAGACGCCACGGAGGCATCGTCGCTCCAGCAGGAGCTGACGGAGGCGAAGCC GGCGCAGCAGCGGCAGTTGCTGCTGCAGCACGTGCGCTCCCAGGCGGCGGCCACGCTGGG GCACTCGGACGTCGACGCGGTGCCCGCCACCAAGCCGTTCCAGGAGCTGGGCTTCGACTC GCTGACCGCGGTGGAGCTGCGCAACAGGCTGAACAAGAGCACCGGCCTGACACTGCCGA CCACGGTCGTCTTCGACCACCCCACCCCCGACGCGCTCACCGACGTCCTGCGGGCCGAGC TGTCCGGTGACGCGGCGGCCTCCGCCGACCCGGTGCGGGCGGCCGGGGCCTCCAGGGGCG CCGCCGACGACGAGCCGATCGCGATCGTCGGCATGGCCTGCCGCTATCCGGGCGACGTCC GCTCCGCCGAGGAGCTGTGGGATCTGGTCGCGGCCGGCAAGGACGCCATGGGGGCCTTCC CCGACGACCGGGGCTGGGACCTGGAGACGCTGTACGACCCGGACCCGGAGAGCCGCGGC ACCAGCTATGTGCGCGAAGGCGGGTTCCTCTACGACGCGGGCGACTTCGACGCCGGATTC TTCGGCATCAGCCCCCGCGAGGCCGTCGCGATGGACCCGCAGCAGCGGCTGCTCCTGGAG ACCGCGTGGGAGGCGATCGAACGCGCGGGCCTCGACCGGGAGACCCTCAAGGGCAGCGA CGCCGGGGTGTTCACGGGCCTGACCATCTTCGACTACCTCGCGCTCGTCGGTGAACAGCC CACCGAGGTCGAGGGCTACATCGGCACCGGCAACCTCGGCTGTGTCGCCTCCGGCAGGGT GTCGTACGTACTCGGCCTAGAAGGCCCCGCCATGACGATCGACACCGGCTGCTCTTCGTC CCTGGTGGCGATCCACCAGGCGGCGCACGCGCTGCGCCAGGGCGAGTGCTCGCTCGCTCT CGCGGGCGGCGCGACGGTGATGGCCACGCCGGGCTCGTTCGTCGAGTTCTCGCTGCAGCG CGGGCTCGCCAAGGACGGCCGGTGCAAGCCGTTCGCGGCCGCCGCCGACGGCACCGGCT GGGCCGAGGGAGTCGGCCTGGTCGTACTCGAACGGCTCTCGGAGGCCCGGCGCAACGGC CACAACGTCCTGGCGGTGATCCGGGGTTCGGCCATCAACCAGGACGGCACGAGCAACGG GCTCACGGCACCCAACGGGCAGGCGCAGCAGCGGGTGATCCGGCAGGCACTCGCCAACG CGCGGCTCTCCGCCGAGGACGTCGACGCGGTGGAGGCGCACGGCACCGGCACCATGCTG GGCGACCCCATCGAAGCCAGCGCGCTCGTCGCCACCTACGGCAAGGAGCGGCCCGCGGA CCGGCCGCTGTGGCTCGGCTCGATCAAGTCGAACATCGGGCACGCGCAGGCGTCGGCCGG TGTCGCCGGTGTGATCAAGATGGTCATGGCGCTGCGCAACGAACAGCTGCCCGCCTCCCT GCACATCGACGCGCCCACGCCGCACGTGGACTGGGACGGCAGTGGCGTCCGCCTGCTGTC CGAACCGGTCTCCTGGCCGCGCGGCGAACGCCCGCGCCGCGCCGGGGTGTCCGCCTTCGG CATCTCCGGCACCAACGCGCACCTCATCCTGGAACAGGCCCCGGACGCGCCGGAGCCCGT GACCGCTCCGGCGGAGGACGCCGCGGCGCCGGCCGGAGTGGTGCCCTGGGTGGTGTCGG CGCGCGGCGAGGAGGCGCTGCGGGCCCAGGCCCGCCTGCTGGCCGACCGCGCCACCGCC GACCCGCGGCTCGCGTCGCCGCTGGACGTGGGCTGGTCCCTGGTCAAGACCCGGTCGGTG TTCGAGAACCGAGCCGTCGTCGTGGGCAAGGACCGCCAGACTCTCCTCGCCGGGCTACGG TCCCTGGCGGCGGGCGAGCCGTCACCGGACGTCGTCGAGGGCGCCGTACAGGGCGCCTCC GGCGCGGGTCCGGTGTTGGTGTTTCCGGGGCAGGGGTCGCAGTGGGTGGGCATGGGTGCC CAGCTCCTTGACGAGTCCCCCGTCTTCGCGGCGCGGATCGCGGAGTGTGAGCGGGCCCTG TCGGCGCATGTGGACTGGTCGCTGAGTGCGGTGTTGCGCGGGGACGGGAGTGAGCTGTCC CGGGTCGAGGTCGTGCAGCCCGTGTTGTGGGCGGTGATGGTCTCGCTGGCTTCGGTGTGG GCCGATTACGGCATCACCCCGGCTGCCGTCATCGGGCACTCGCAGGGCGAGATGGCCGCC GCGTGTGTGGCGGGGGCACTGTCACTGGAGGATGCGGCGCGGATCGTGGCCGTACGCAGT GACGCGCTCCGTCAGCTCATGGGGCAGGGCGACATGGCGTCGTTGGGGGCCGGCTCGGAG CAGGTTGCTGAGCTCATCGGCGACCGGCCCGGCGTGTGTGTCGCTGCCGTCAACGGGCCC TCCTCTACGGTCATTTCAGGGCCGCCGGAGCATGTGGCAGCCGTGGTCGCGGATGCGGAG GCGCGAGGCCTGCGCGCCCGCGTCATCGACGTCGGATACGCCTCCCACGGCCCCCAGATC GACCAGCTCCACGACCTCCTCACCGAGCGCCTGGCCGACATCCGGCCCACGACCACGGAC GTCGCCTTCTACTCCACGGTCACCGCCGAACGCCTCGACGACACCACCACCCTCGACACG GATTACTGGGTCACCAACCTCCGCCAGCCCGTCCGCTTCGCCGACACCATCGAAGCGCTG	

Name	Sequence	SEQ ID NO:
	CTGGCCGACGGCTACCGCCTGTTCATCGAGGCCAGCCCCCACCCCGTCCTCAACCTCGGC ATGGAGGAGACCATCGAGCGGGCCGACATGCCCGCCACCGTCGTGCCCACCCTGCGCCGC GACCACGGCGACGCCGCGCAGCTCACCCGCGCGGCCGCCCAGGCCTTCGGCGCGGGGGC GGAGGTCGACTGGACGGGCTGGTTCCCGGCCGTGCCGCTGCCTCGGGTGGTGGATCTGCC GACGTACGCCTTCCAGCGGGAGCGGTTCTGGCTGGAGGGGCGCAGGGGGCTCGCCGGGG ACCCGGCGGGGCTCGGGCTCGCGTCCGCGGGGCATCCGCTGCTCGGAGCCGCCGTGGAAC TCGCGGACGGCGGCAGTCACCTGCTGACCGGCCGGATCTCTCCGCGGGACCAGGCGTGGC TGGCCGAGCACCGGGTCATGGACACGGTGCTGCTGCCCGGTTCGGCGTTCGTGGAGCTCG CGCTGCAGGCCGCGGTGCGGGCCGGCTGCGCGGAGTTGGCGGAGCTGACGCTGCACACTC CGCTCGCCTTCGGGGACGAGGGTGCGGGCGCGGTCGACGTGCAGGTGGTGGTCGGTTCCG TGGCCGAGGACGGGCGGCGTCCCGTGACCGTCCATTCGCGGCCCACGGGTGAGGGCGAG GAGGCCGTGTGGACCCGGCATGCCGCGGGCGTGGTCGCTCCCCCGGGGCCCGACGCCGGG GACGCCTCGTTCGGCGGGACGTGGCCGCCCCCGGGCGCCACACCGGTCGGCGAGCAGGAT CCGTACGGGGAACTCGCTTCGTACGGCTATGACTTCGGGCCCGGCTCACAGGGACTGGTG AGCGCGTGGCGGCTCGGGGACGACCTTTTCGCCGAGGTGGCGCTGCCCGAGGCGGAGAG CGGCAGGGCCGACCGCTACCAGGTGCACCCGGTGTTGCTCGACGCCACGCTGCACGCGCT GATCCTGGACGCGGTCACGTCGTCCGCCGACACCGACCAAGTGCTGCTGCCGTTCTCCTG GAGCGGGTTGCGGGTGCACGCGCCGGGCGCTGAGAAGTTGCGGGTACGTATCGCACGCA CCGCGCCCGACCAGCTGGCCCTCACGGCCGTGGACGGGGGCGGAGGCGGGGAGCCGGTT CTCACGCTGGAGTCGCTCACGGTACGGCCGGTGGCCGCCCACCAGATCGCGGGCGCCCGT GCGGCGGACCGTGACGCGCTGTTCCGGCTCGTGTGGATGGAGGTCGCCGCGCGGGCCGAG GAGACGGGCGGCGGCGCCCCGCGTGCCGCGGTCCTCGCGCCGGTCGAGAGTGGCCCGAT GGGCGGTACGTCGGCCGGTGCGCTGGCCGACGCCTTGTCCGATGCGCTGGCCGCCGGCCC CGTGTGGGACACGTTCGGTGCGCTCCGGGACGGAGTGGCGGCTGGGGGCGAAGCGCCCG ATGTCGTGCTCGCCGTGTGCGCCGCGCCCGGCGCAGGTGCCGGGGCCGTTGCGGATGCCG ATGGCAGGGGCGGCGACCCGGCCGGGTACGCGCGGCTGGCCACCGTGTCCCTTCTGTCGC TGCTCAAGGAGTGGGTGGACGACCCGGCGTTCGCGGCGACCCGTCTCGTCGTCGTCACCC GAGGCGCGGTCGCCGCGCGGCCGGGTGAGACCGCCGGTGATCTGGCGGGCGCGTCGCTGT GGGGTCTGGTGCGCAGCGCGCAGGCCGAGAATCCCGGTCGCCTCACGCTGCTCGACGTGG ACGGACTGGAGTCTTCTCCGGCCACGCTGACCGGCGTACTGGCCTCCGGCGAACCGGAAT TGGCGCTGCGTGACGGGCGCGCCTACGTGCCACGGCTCGTGCGCGACGACGCGTCGGTGC GGCTGGTGCCGCCGGTCGGATCGCTCACGTGGCGGCTGGCTCGGTGCCAAGAGGCGGGCG GCGGACAGCAGTTGTCCCTTGTCGACGCTCCCGAGGCCGGACGGGCGCTGGAGCCGCACG AGGTGCGGGTGGCGGTGCGGGCCGCGGCGCCGGGGCCGCTCACGGCGGGCCAGGTCGAG GGTGCCGGTGTGGTGACGGAAGTCGGCGGCGAGGTCGGTTCGGTGGCCGTGGGCGACCG GGTGATGGGGTTGTTCGACGCGGTGGGCCCGGTGGCCGTCACCGATGCCGCGCTGCTTAT GCCGGTTCCGGCGGGCTGGAGTTGGGCGCAGGCGGCCGGGTCCTTGGGGGCCTATGTGTC CGCGTATCACGTGCTGGCGGATGTCGTGGCGCCGCGCGGCGGCGAGACTCTGCTCGTCGG GGAGGAAACCGGTTCCGTCGGCCGTGCCGTGCTGCGTCTTGCTCTTGCCGGGCGGTGGCG GGTCGAGGCCGTGGACGGTGCGTCGACGGCGGATGATTCGGGCGCCGAGCGCGCGGCCG ACGTCACCCTCCGCCACGAGGGGGCCCTGGTGGTCCACCGTGCGGGCGGTCGTCCGGACG AGGGACAGGCTGTCGTGCCGCCCGAGCCCGGGCGCGTACGGGAAATCCTCGCGGAGCTG ACCGAGCTGACCGAGCTTGCCGAGATCACGGAGTCGGCGGAGCCCGGGCTGCCCGCGGA ACGGGGTGACAGCCGTGCTCTGACGCCGCTCGACATCACCGTGTGGGACATCCGGCAGGC GCCCGCCGCGATGGCGGCCCCGCCCTCGGCGGGCACGACCGTGTTCTCGCTGCCTCCCGC CTTCGACCCCGAGGGCACCGTGCTGGTCACCGGCGGCACCGGAGCACTCGGCTCGCTGAC GGCCCGTCACCTGGTGGAACGGTACGGAGCCAGGCACCTGTTGCTGTCCAGCAGGCGGGG AGCCGACGCGCCGGGCGCACTCGAACTGGCCGCCGACCTCTCCGCGCTCGGCGCGCGAGT CACCTTCGCCGCGTGCGACCCGGGCGACCGGGACGAAGCCGCCGCCCTCCTCGCGGCGGT GCCCTCGGACCACCCGCTGACCGCCGTCTTCCACTGCGCGGGCACCGTGAACGACGCCGT GGTGCAGAACCTCACGGCCGAGCAGGTCGAGGAGGTGATGCGCGTGAAGGCGGACGCCG CGTGGCACCTGCACGAGCTGACGCGGGACGCGGACCTGTCCGCGTTCGTCCTGTACTCCT CGGTCGCCGGGCTGCTCGGCGGCCCCGGCCAGGGCAGCTACACGGCCGCCAACGCCTTCT TGGACGCGCTGGCCCGGCACCGGCACGACGGCGGTGCGGCGGCGACCTCCCTGGCGTGG GGCTACTGGGAGCTGGCGAGCGGCATGTCGGGACGGCTCACCGACGCCGACCGGGCGCG CCATGCCCGCGCCGGCGTGGTCGGGCTCGGCGCCGACGAGGGACTCGCCCTCCTCGACGC GGCGTGGGCCGGCGGACTGCCCCTGTACGCGCCGGTCCGTCTGGACCTGGCCCGGATGCG CCGGCAGGCCCAGAGCCACCCCGCACCGGCGCTGCTGCGCGACCTGGTGCGCGGGGGGA GCAAGAGCGGCGGCGGTGCCGTGTCGGCGGGGGCGGCCGCGCTGCTCAAGTCGCTCGGC GCGATGTCCGACCCCGAGCGGGAGGAGGCGCTGCTCGACCTGGTGTGCACCCATATCGCG GCCGTCCTCGGCTACGACGCGGCCACGCCCGTCAACGCGACGCAGGGGCTGCGGGAACTC GGCTTCGACTCCCTGACCGCGGTGGAGCTGCGCAACCGGCTCTCGGCCGCGACGGGGCTG AAGCTGCCCGCCACGTTCGTCTTCGACCATCCCAACCCGGCGGAGCTCGCCGCGCAGCTG CGGCAGGAGCTGGCCCCGCGCGCGGCGGATCCGCTCGCCGACGTCCTGGCGGAGTTCGAG CGCATCGAGGACTCGCTGCTCTCGGTCTCCTCGAAGGACGGCTCGGCGCGGGCCGAACTG GCGGGGCGGCTGCGCGCGACGCTGGCCAGGCTCGACGCGCCGCAGGACACGGCCGGCGA GGTCGCGGTGGCCACTCGTACACGTATCCAGGACGCGTCGGCGGACGAGATCTTCGCGTT CATCGACCGGGACCTCGGCAGGGACGGCGCGAGCGGACAGGGCAACGGACAGCCCACCG GGCAGGGCAACGGACACGGCAACGGCAACGGCAACGGCAACGGCAACGGACACGGTCA GGCAGTGGAGGGGCAGCGATGACCACGTACCGGATTGGCACgtaggccaggacagcaacggccacgaca cccccatcggcatcgcgtggcgggcccgattcgcgtccgccccggggaggagacgggcgtccgcgtcgagggcgaggcgggcttccacccgt aggcgatccggcgagcacgacacccgcagattggctcgacgcagcccagaaatgtatgatcaaggcgaatacttcatatgcggggatcgaccgc gcgggtcccggacggggaagagcggggagctttgccagagagcgacgacttccccttgcgttggtgattgccggtcagggcagccatccgccat cgtcgcgtagggtgtcacaccccaggaatcgcgtcactgaacacagcagccggtaggacgaccatgactgagttggacaccatcgcaaatccgtc cgatcccgcggtgcagcggatcatcgatgtcaccaagccgtcgcgatccaacataaagacaacgttgatcgaggacgtcgagcccctcatgcaca	

Name	Sequence	SEQ ID NO:
	gcatcgcggccggggtggagttcatcgaggtctacggcagcgacagcagtccttttccatctgagttgctggatctgtgcgggcggcagaacatac cggtccgcctcatcgactcctcgatcgtcaaccagttgttcaaggggagcggaaggccaagacattcggcatcgcccgcgtccctcgcccggcc aggttcggcgatatcgcgagccggcgtggggacgtcgtcgttctcgacggggtgaagatcgtcgggaacatcggcgcgatagtacgcacgtcgct cgcgctcggagcgtcggggatcatcctggtcgacagtgacatcaccagcatcgcggaccggcgtctccaaagggcagccgaggttacgtcttct ccctttcccgtcgttctctccggtcgcgaggaggccatcgccttcattcgggacagcggtatgcagctgatgacgctcaaggcggatggcgacatttc cgtgaaggaactcggggacaatccggatcggctggccttgctgttcggcagcgaaaagggtgggccttccgacctgttcgaggaggcgtcttccg cctcggtttccatccccatgatgagccagaccgagtctctcaacgtttccgtttccctcggaatcgcgctgcacgagaggatcgacaggaatctcgcg gccaaccgataagcgcctctgttcctcggacgctcggttcctcgacctcgattcgtcagtgatgatcacctcacacggcagcgatcaccactgacata tcgaggtcaacggtcgtggtccgggcgggcactcctcgaaggcgcggccgacgcccttgaacgactcgatggcgctgacgccttccgcgatcgc ctcgactgtgaaaacgctcgtggacgtcggcgggtagccgatgagttccatcgccttgcgcatcccccttgagcatggagaacgcgtccaccgggta gtcacgaaggtacttgcgaaggcttcgcgccgtccgagaccgtgagccggcgcaagccggtgtggcgccggtcccggatccacatgctggtGA TGCTGTTGTGGGCACAATCGTGCCGGTTGGTAGGATCCAGCGCTGCTCAGCGGGCTCACC CGCGGATCGCGGGTCGGCGGCGCGCCGGTCAACCAGCGCAGGGCAGCCGCCGGAGGCGC GGGCGAGGCGGACACGGACCTCGGCGGGCGGCTCGCCGCGATGACACCGGACGACCGGG TCGCGCACCTGCGGGACCTCGTCCGTACGCACGTGGCGACCGTCCTGGGACACGGCACCC CGAGCCGGGTGGACCTGGAGCGGGCCTTCCGCGACACCGGTTTCGACTCGCTCACCGCCG TCGAACTCCGCAACCGTCTCAACGCCGCGACCGGGCTGCGGCTGCCGGCCACGCTGGTCT TCGACCACCCCACCCCGGGGGAGCTCGCCGGGCACCTGCTCGACGAACTCGCCACGGCCG CGGGCGGGTCCTGGGCGGAAGGCACCGGGTCCGGAGACACGGCCTCGGCGACCGATCGG CAGACCACGGCGGCCCTCGCCGAACTCGACCGGCTGGAAGGCGTGCTCGCCTCCCTCGCG gatatcCCTCGCCGCAGTTAATTAAAGTCAGTGAGCGAGGAAGCGCGTAACTATAACGGTCC TAAGGTAGCG	
pSETSC3	atctacgtctgtcgagaagtttctgatcgaaaagttcgacagcgtctccgacctgatgcagctctcgcagggcgaagaatctcgtgctttcagcttcgat gtaggagggcgtggatatgtcctgagggtaaatagctgcgccgatggtttctacaaagatcgttatgttgatcggcactttgcatcggccgcgctccc gattccggaagtgcttgacattgggaatttatgcggtgtgaaatacccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattca ggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgg gtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttgggctgcaggtcgactctagaggatccgcggccgcgcg cgatatcgaattcgtaatcatgtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaag cctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaat cggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagc ggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggcc aggaaccgtaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcga aacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgc ctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaa ccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactg gtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatc tgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagc agcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattt tggtcatgagattatcaaaaggatcttcacctagatcctttggttcatgtgcagctccatcagcaaaaggggatgataagtttatcaccaccgactattt gcaacagtgccgttgatcgtgctatgatcgactgatgtcatcagcggtggagtgcaatgtcgtgcaatacgaatggcgaaaagccgagctcatcggt cagcttctcaaccttggggttacccccggcggtgtgctgctggtccacagctccttccgtagcgtccggcccctcgaagatgggccacttggactgat cgaggccctgcgtgctgcgctgggtccgggagggacgctcgtcatgcccctcgtggtcaggtctggacgacgagccgttcgatcctgccacgtcgc ccgttacaccggaccttggagttgtctctgacacattctggcgcctgccaaatgtaaagcgcagcgcccatccatttgcctttgcggcagcggggcca caggcagagcagatcatctctgatccattgcccctgccacctcactcgcctgcaagcccggtcgcccgtgtccatgaactcgatgggcaggtacttct cctcggcgtgggacacgatgccaacacgacgctgcatcttgccgagttgatggcaaaggttccctatggggtgccgagacactgcaccattcttcag gatggcaagttggtacgcgtcgattatctcgagaatgaccactgctgtgagcgctttgccttggcggacaggtggctcaaggagaagagccttcaga aggaaggtccagtcggtcatgcctttgctcggttgatccgctcccgcgacattgtggcgacagccctgggtcaactgggccgagatccgttgatcttc ctgcatccgccagaggcgggatgcgaagaatgcgatgccgctcgccagtcgattggctgagctcatgagcggagaacgagatgacgttggaggg gcaaggtcgcgctgattgctggggcaacacgtggagcggatcggggattgtctttcttcagctcgctgatgatatgctgacgctcaatgccgtttggc ctccgactaacgaaaatcccgcatttggacggctgatccgattggcacggcggacggcgaatggcggagcagacgctcgtccgggggcaatgag atatgaaaaagcctgaactcaccgcgacgtatcgggccctggccagctagctagagtcgacctgcaggtcccggggatcggtcttgccttgctcgt cggtgatgtacttcaccagctccgcgaagtcgctcttcttgatggagcgcatggggacgtgcttggcaatcacgcgcaccccccggccgttttagcg gctaaaaaagtcatggctctgccctcgggcggaccacgcccatcatgaccttgccaagctcgtcctgcttctcttcgatcttcgccagcagggcgag gatcgtggcatcaccgaaccgcgccgtgcgcgggtcgtcggtgagccagagtttcagcaggccgcccaggcggcccaggtcgccattgatgcgg gccagctcgcggacgtgctcatagtccacgacgcccgtgattttgtagccctggccgacggccagcaggtaggccgacaggctcatgccggccg ccgccgccttttcctcaatcgctcttcgttcgtctggaaggcagtacaccttgataggtgggctgcccttcctggttggcttggtttcatcagccatccgct tgccctcatctgttacgccggcggtagccggccagcctcgcagagcaggattcccgttgagcaccgccaggtgcgaataagggacagtgaagaa ggaacacccgctcgcgggtgggcctacttcacctatcctgcccggctgacgccgttggatacaccaaggaaagtctacacgaacccttttggcaaaa tcctgtatatcgtgcgaaaaaggatggatataccgaaaaaatcgctataatgaccccgaagcagggttatgcagcggaaaagatccgtcgacctgca ggcatgcaagctctagcgattccagacgtcccgaaggcgtggcgcggcttcccctgtccggagcaatcgccctgggtgggttacacgacgcccct ctatggcccgtactgacggacacaccgaagccccggcggcaaccctcagcggatgcccccggggcttcacgttttccccaggtcagaagcggttttcg gTATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAGCGCTATATGCGTT GATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCA GTCCTGCTCGCTTCGCTACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCG TCCTGTGGATCTGCCTCGCTGGCCTGCCGCAGTTCTTCAACCTCCCGGCGCAGCTTTTCGTT CTCAATTTCAGCATCCCTTTCGGCATACCATTTTATGACGGCGGCAGAGTCATAAAGCACC TCATTACCCTTGCCACCGCCTCGCAGAACGGGCATTCCCTGTTCCTGCCAGTTCTGAATGG TACGGATACTCGCACCGAAAATGTCAGCCAGCTGCTTTTTGTTGACTTCCATTGTTCATTC CACGGACAAAAACAGAGAAAGGAAACGACAGAGGCCAAAAAGCTCGCTTTCAGCACCTG TCGTTTCCTTTCTTTTCAGAGGGTATTTTAAATAAAAACATTAAGTTATGACGAAGAAGAA CGGAAACGCCTTAAACCGGAAAATTTTCATAAATAGCGAAAACCCGCGAGGTCGCCGCCC CGTAACAAGGCGGATCGCCGGAAAGGACCCGCAAATGATAATAATTATCAATTGCATACT ATCGACGGCACTGCTGCCAGATAACACCACCGGGGAAACATTCCATCATGATGGCCGTGC GGACATAGGAAGCCAGTTCATCCATCGCTTTCTTGTCTGCTGCCATTTGCTTTGTGACATC CAGCGCCGCACATTCAGCAGCGTTTTTCAGCGCGTTTTCGATCAACGTTTCAATGTTGGTA	49

Name	Sequence	SEQ ID NO:
	TCAACACCAGGTTTAACTTTGAACTTATCGGCACTGACGGTTACCTTGTTCTGCGCTGGCT CATCACGCTGGATACCAAGGCTGATGTTGTAGATATTGGTCACCGGCTGAGGTGTTTCGAT TGCCGCTGCGTGGATAGCACCATTccatggGATGCTGTTGTGGGCACAATCGTGCCGGTTGGT AGGATCCAGCGagaagggagcggaaATGCTGCGCCCGGTGGAAACCCCGACCCGCGAAATCAAG AAGCTCGACGGCCTGTGGGCGTTCAGCCTCGACCGCGAAAACTGCGGCATCGACCAGCGG TGGTGGGAGAGCGCGCTGCAGGAATCGCGCGCGATCGCCGTCCCGGGCTCGTTCAACGAC CAGTTCGCGGACGCCGACATCCGGAACTACGCCGGCAACGTCTGGTACCAGCGCGAGGTG TTCATCCCGAAGGGCTGGGCGGGCCAGCGGATCGTCCTGCGCTTCGACGCCGTGACCCAC TACGGCAAGGTCTGGGTGAACAACCAGGAAGTCATGGAACACCAGGGCGGCTACACCCC GTTCGAGGCGGACGTCACGCCGTACGTGATCGCCGGCAAGTCCGTCCGGATCACCGTCTG CGTGAACAACGAGCTGAACTGGCAGACCATCCCGCCGGGCATGGTGATCACGGACGAGA ACGGCAAGAAGAAGCAGTCCTACTTCCACGACTTCTTCAACTACGCGGGCATCCACCGCA GCGTCATGCTGTACACCACGCCGAACACCTGGGTCGACGACATCACCGTCGTGACGCACG TGGCGCAGGACTGCAACCACGCCAGCGTCGACTGGCAGGTCGTGGCCAACGGCGACGTCT CGGTGGAGCTGCGGGACGCGGACCAGCAGGTCGTGGCCACCGGCCAGGGCACCTCGGGC ACGCTGCAGGTCGTGAACCCGCACCTCTGGCAGCCGGGCGAGGGCTACCTGTACGAACTC TGCGTCACCGCGAAGTCGCAGACGGAGTGCGACATCTACCCGCTGCGGGTGGGCATCCGC TCCGTCGCCGTGAAGGGCGAGCAGTTCCTCATCAACCACAAGCCGTTCTACTTCACCGGCT TCGGCCGGCACGAGGACGCGGACCTGCGCGGCAAGGGCTTCGACAACGTCCTGATGGTGC ACGACCACGCGCTCATGGACTGGATCGGCGCCAACTCGTACCGGACCTCGCACTACCCGT ACGCGGAGGAAATGCTGGACTGGGCCGACGAGCACGGCATCGTCGTGATCGACGAAACG GCGGCCGTCGGCTTCAACCTGAGCCTCGGCATCGGCTTCGAGGCGGGCAACAAGCCGAAG GAACTCTACTCGGAGGAAGCCGTGAACGGCGAGACCCAGCAGGCGCACCTGCAGGCCAT CAAGGAACTCATCGCGCGGGACAAGAACCACCCGTCCGTCGTGATGTGGAGCATCGCCAA CGAGCCGGACACCCGGCCGCAGGGCGCGCGCGAGTACTTCGCCCCGCTGGCGGAAGCCA CGCGGAAGCTCGACCCGACCCGCCCGATCACGTGCGTCAACGTGATGTTCTGCGACGCGC ACACCGACACGATCTCGGACCTGTTCGACGTGCTGTGCCTCAACCGGTACTACGGCTGGT ACGTGCAGTCCGGCGACCTGGAGACCGCGGAAAAGGTGCTCGAGAAGGAACTGCTCGCC TGGCAGGAGAAGCTGCACCAGCCGATCATCATCACCGAATACGGCGTGGACACGCTGGCC GGCCTCCACTCGATGTACACGGACATGTGGTCCGAGGAATACCAGTGCGCGTGGCTGGAC ATGTACCACCGGGTCTTCGACCGCGTGTCCGCCGTCGTGGGCGAGCAGGTCTGGAACTTC GCGGACTTCGCCACCAGCCAGGGCATCCTCCGGGTGGGCGGCAACAAGAAGGGCATCTTC ACGCGGGACCGCAAGCCGAAGAGCGCGGCCTTCCTGCTCCAGAAGCGCTGGACCGGCAT GAACTTCGGCGAGAAGCCGCAGCAGGGCGGCAAGCAGTGATTGGGGACCCTAGAGGTCC CCTTTTTTATTTTTTGGGGACCCTAGAGGTCCCCTTTTTTATTTTccatggcgagacacccgggaagcctg	
pSET152-gusA	atctacgtctgtcgagaagtttctgatcgaaaagttcgacagcgtctccgacctgatgcagctctcgcagggcgaagaatctcgtgctttcagcttcgat gtaggagggcgtggatatgtcctgcgggtaaatagctgcgccgatggtttctacaaagatcgttatgttgatcggcactttgcatcggccgcgctccc gattccggaagtgcttgacattggggaatttatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattcgccattca ggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgtgctgcaaggcgattaagttgg gtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagcttgggctgcaggtcgactctagaGATGCTGTTGTG GGCACAATCGTGCCGGTTGGTAGGATCCAGCGagaagggagcggaaATGCTGCGCCCGGTGGAA ACCCCGACCCGCGAAATCAAGAAGCTCGACGGCCTGTGGGCGTTCAGCCTCGACCGCGAA AACTGCGGCATCGACCAGCGGTGGTGGGAGAGCGCGCTGCAGGAATCGCGCGCGATCGC CGTCCCGGGCTCGTTCAACGACCAGTTCGCGGACGCCGACATCCGGAACTACGCCGGCAA CGTCTGGTACCAGCGCGAGGTGTTCATCCCGAAGGGCTGGGCGGGCCAGCGGATCGTCCT GCGCTTCGACGCCGTGACCCACTACGGCAAGGTCTGGGTGAACAACCAGGAAGTCATGGA ACACCAGGGCGGCTACACCCCGTTCGAGGCGGACGTCACGCCGTACGTGATCGCCGGCAA GTCCGTCCGGATCACCGTCTGCGTGAACAACGAGCTGAACTGGCAGACCATCCCGCCGGG CATGGTGATCACGGACGAGAACGGCAAGAAGAAGCAGTCCTACTTCCACGACTTCTTCAA CTACGCGGGCATCCACCGCAGCGTCATGCTGTACACCACGCCGAACACCTGGGTCGACGA CATCACCGTCGTGACGCACGTGGCGCAGGACTGCAACCACGCCAGCGTCGACTGGCAGGT CGTGGCCAACGGCGACGTCTCGGTGGAGCTGCGGGACGCGGACCAGCAGGTCGTGGCCA CCGGCCAGGGCACCTCGGGCACGCTGCAGGTCGTGAACCCGCACCTCTGGCAGCCGGGCG AGGGCTACCTGTACGAACTCTGCGTCACCGCGAAGTCGCAGACGGAGTGCGACATCTACC CGCTGCGGGTGGGCATCCGCTCCGTCGCCGTGAAGGGCGAGCAGTTCCTCATCAACCACA AGCCGTTCTACTTCACCGGCTTCGGCCGGCACGAGGACGCGGACCTGCGCGGCAAGGGCT TCGACAACGTCCTGATGGTGCACGACCACGCGCTCATGGACTGGATCGGCGCCAACTCGT ACCGGACCTCGCACTACCCGTACGCGGAGGAAATGCTGGACTGGGCCGACGAGCACGGC ATCGTCGTGATCGACGAAACGGCGGCCGTCGGCTTCAACCTGAGCCTCGGCATCGGCTTC GAGGCGGGCAACAAGCCGAAGGAACTCTACTCGGAGGAAGCCGTGAACGGCGAGACCCA GCAGGCGCACCTGCAGGCCATCAAGGAACTCATCGCGCGGGACAAGAACCACCCGTCCG TCGTGATGTGGAGCATCGCCAACGAGCCGGACACCCGGCCGCAGGGCGCGCGCGAGTAC TTCGCCCCGCTGGCGGAAGCCACGCGGAAGCTCGACCCGACCCGCCCGATCACGTGCGTC AACGTGATGTTCTGCGACGCGCACACCGACACGATCTCGGACCTGTTCGACGTGCTGTGC CTCAACCGGTACTACGGCTGGTACGTGCAGTCCGGCGACCTGGAGACCGCGGAAAAGGTG CTCGAGAAGGAACTGCTCGCCTGGCAGGAGAAGCTGCACCAGCCGATCATCATCACCGAA TACGGCGTGGACACGCTGGCCGGCCTCCACTCGATGTACACGGACATGTGGTCCGAGGAA TACCAGTGCGCGTGGCTGGACATGTACCACCGGGTCTTCGACCGCGTGTCCGCCGTCGTG GGCGAGCAGGTCTGGAACTTCGCGGACTTCGCCACCAGCCAGGGCATCCTCCGGGTGGGC GGCAACAAGAAGGGCATCTTCACGCGGGACCGCAAGCCGAAGAGCGCGGCCTTCCTGCT CCAGAAGCGCTGGACCGGCATGAACTTCGGCGAGAAGCCGCAGCAGGGCGGCAAGCAGT GAtctagaggatccgcggccgcgcgcgatatcgaattcgtaatcatgtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaaca tacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcggga aacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgct gcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaaca	50

Name	Sequence	SEQ ID NO:
	tgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaa aatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccga ccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctatagctcacgctgtaggtatctcagttcggtgtaggtcgtt cgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacg acttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacgg ctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgc tggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagt ggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttggttcatgtgcagctccatcagcaaaaggg gatgataagtttatcaccaccgactatttgcaacagtgccgttgatcgtgctatgatcgactgatgtcatcagcggtggagtgcaatgtcgtgcaatacg aatggcgaaaagccgagctcatcggtcagcttctcaaccttggggttaccccccggcggtgtgctgctggtccacagctccttccgtagcgtccggcc cctcgaagatgggccacttggactgatcgaggccctgcgtgctgcgctgggtccgggagggacgctcgtcatgccctcgtggtcaggtctggacg acgagccgttcgatcctgccacgtcgcccgttacaccggaccttggagttgtctctgacacattctggcgcctgccaaatgtaaagcgcagcgcccat ccatttgccttttgcggcagcggggccacaggcagagcagatcatctctgatccattgcccctgccacctcactcgcctgcaagcccggtcgcccgtg tccatgaactcgatgggcaggtacttctcctcggcgtgggacacgatgccaacacgacgctgcatcttgccgagttgatggcaaaggttccctatgg ggtgccgagacactgcaccattcttcaggatggcaagttggtacgcgtcgattatctcgagaatgaccactgctgtgagcgctttgccttggcggaca ggtggctcaaggagaagagccttcagaaggaaggtccagtcggtcatgcctttgctccggttgatccgctcccgcgacattgtggcgacagccctgg gtcaactgggccgagatccgttgatcttcctgcatccgccagaggcgggatgcgaagaatgcgatgccgctcgccagtcgattggctgagctcatg agcggagaacgagatgacgttggaggggcaaggtcgcgctgattgctggggcaacacgtggagcggatcggggattgtctttcttcagctcgctg atgatatgctgacgctcaatgccgtttggcctccgactaacgaaaatcccgcatttggacggctgatccgattggcacggcggacggcgaatggcg gagcagacgctcgtccgggggcaatgagatatgaaaaagcctgaactcaccgcgacgtatcgggccctggccagctagctagagtcgacctgca ggtccccggggatcggtcttgccttgctcgtcggtgatgtacttcaccagctccgcgaagtcgctcttcttgatggagcgcatggggacgtgcttggc aatcacgcgcacccccgggccgttttagcggctaaaaaagtcatggctctgccctcgggcggaccacgcccatcatgaccttgccaagctcgtcctg cttctcttcgatcttcgccagcagggcgaggatcgtggcatcaccgaacccgcgccgtgcgcgggtcgtcggtgagccagagtttcagcaggccgc ccaggcggcccaggtcgccattgatgcgggccagctcgcggacgtgctcatagtccacgacgcccgtgattttgtagccctggccgacggccagc aggtaggccgacaggctcatgccggccgccgccgcctttttcctcaatcgctcttcgttcgtctggaaggcagtacaccttgataggtgggctgcccctt cctggttggcttggtttcatcagccatccgcttgccctcatctgttacgccggcggtagccggccagcctcgcagagcaggattcccgttgagcaccg ccaggtgcgaataagggacagtgaagaaggaacacccgctcgcgggtgggcctacttcacctatcctgcccggctgacgccgttggatacaccaa ggaaagtctacacgaaccctttggcaaaatcctgtatatcgtgcgaaaaaggatggatataccgaaaaaatcgctataatgacccccgaagcagggtt atgcagcggaaaagatccgtcgacctgcaggcatgcaagctctagcgattccagacgtcccgaaggcgtggcgcggcttccccgtgccggagca atcgcccctgggtgggttacacgacgcccctctatggcccgtactgacggacacaccgaagccccggcggcaaccctcagcggatgccccggggc ttcacgttttcccaggtcagaagcggttttcgggagtagtgccccaactggggtaacctttgagttctctcagttgggggcgtagggtcgccgacatga cacaaggggttgtgaccggggtggacacgtacgcgggtgcttacgaccgtcagtcgcgcgagcgcgagaatttcgagcgcagcaagcccagcga cacagcgtagcgccaacgaagacaaggcggccgaccttcagcgcgaagtcgagcgcgacgggggccggttcaggttcgtcgggcatttcagcg aagcgccgggcacgtcggcgttcgggacggcggagcgcccggagttcgaacgcatcctgaacgaatgccgcgccgggcggctcaacatgatca ttgtctatgacgtgtcgcgcttctcgcgcctgaaggtcatggacgcgattccgattgtctcggaattgctcgccctgggcgtgacgattgtttccactca ggaaggcgtcttccggcagggaaacgtcatggacctgattcacctgattatgcggctcgacgcgtcgcacaaagaatcttcgctgaagtcggcgaa gattctcgacacgaagaaccttcagcgcgaattgggcgggtacgtcggcgggaaggcgccttacggcttcgagcttgtttcggagacgaaggagat cacgcgcaacggccgaatggtcaatgtcgtcatcaacaagcttgcgcactcgaccactcccctttaccggacccttcgagttcgagcccgacgtaatc cggtggtggtggcgtgagatcaagacgcacaaacaccttcccttcaagccgggcagtcaagccgccattcacccgggcagcatcacggggctttg taagcgcatggacgctgacgccgtgccgacccggggcgagacgattgggaagaagaccgcttcaagcgcctgggacccggcaaccgttatgcg aatccttcgggacccgcgtattgcgggcttcgccgctgaggtgatctacaagaagaagccggacggcacgccgaccacgaagattgagggttacc gcattcagcgcgacccgatcacgctccggccggtcgagcttgattgcggaccgatcatcgagcccgctgagtggtatgagcttcaggcgtggttgg acggcaggggcgcggcaaggggctttcccgggggcaagccattctgtccgccatggacaagctgtactgcgagtgtggcgccgtcatgacttcg aagcgcggggaagaatcgatcaaggactcttaccgctgccgtcgccggaaggtggtcgacccgtccgcacctgggcagcacgaaggcacgtgc aacgtcagcatggcggcactcgacaagttcgttgcggaacgcatcttcaacaagatcaggcacgccgaaggcgacgaagagacgttggcgcttct gtgggaagccgcccgacgcttcggcaagctcactgaggcgcctgagaagagcggcgaacgggcgaaccttgttgcggagcgcgccgacgccc tgaacgcccttgaagagctgtacgaagaccgcgcggcaggcgcgtacgacggacccgttggcaggaagcacttccggaagcaacaggcagcg ctgacgctccggcagcaaggggcggaagagcggcttgccgaacttgaagccgccgaagccccgaagcttccccttgaccaatggttccccgaag acgccgacgctgacccgaccggccctaagtcgtggtgggggcgcgcgtcagtagacgacaagcgcgtgttcgtcgggctcttcgtagacaagatc gttgtcacgaagtcgactacgggcaggggggcagggaacgcccatcgagaagcgcgcttcgatcacgtgggcgaagccgccgaccgacgacga cgaagacgacgcccaggacggcacggaagacgtagcggcgtagcgagacacccgggaagcctg	

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, *i.e.*, to mean including but not limited to. Only the transitional phrases "consisting of" and "consisting essentially of" shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

What is claimed is:

CLAIMS

1. An engineered nucleic acid, comprising:

(a) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest; and

(b) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs.

2. The engineered nucleic acid of claim 1, wherein the at least two heterologous genes encode proteins selected from the group consisting of: enzymes, regulatory proteins and transport proteins.

3. The engineered nucleic acid of claim 1 further comprising (c) a promoter operably linked to a gene encoding a shufflon recombinase.

4. The engineered nucleic acid of claim 3, wherein the shufflon recombinase is a *Salmonella enterica* shufflon recombinase.

5. The engineered nucleic acid of claim 3, wherein the gene encoding the shufflon recombinase is codon-optimized for expression in a host cell of interest.

6. The engineered nucleic acid construct of claim 1, wherein the gene encoding the shufflon recombinase comprise a nucleotide sequence as set forth in SEQ ID NO: 35 or SEQ ID NO: 36.

7. The engineered nucleic acid of claim 1 further comprising (d) nucleotide sequences homologous to a chromosomal locus of a host cell of interest.

8. The engineered nucleic acid of claim 5, wherein the host cell of interest is a bacterial cell.

9. The engineered nucleic acid of claim 8, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp*. and *Lactobacillus spp*.

10. The engineered nucleic acid of claim 9, wherein the bacterial cell is a *Streptomyces* spp. cell.

11. The engineered nucleic acid of claim 10, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

12. The engineered nucleic acid of claim 3, wherein the gene encoding the shufflon recombinase is oriented in a 3' to 5' direction.

13. The engineered nucleic acid of claim 3, wherein the promoter of (a) and/or (c) is an inducible promoter.

14. The engineered nucleic acid of claim 1, wherein the engineered nucleic acid construct further comprises an antibiotic resistance gene.

15. The engineered nucleic acid of claim 1, wherein the RRSs recombine, in the presence of shufflon recombinase activity, at different rates relative to each other.

16. The engineered nucleic acid of claim 15, wherein the RRSs comprises a nucleotide sequence selected from sequences set forth as SEQ ID NO: 5-8.

17. The engineered nucleic acid of claim 1, wherein at least two of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

18. The engineered nucleic acid of claim 1, wherein the promoter of (a) is operably linked to at least three heterologous genes of a biosynthetic pathway of interest.

19. The engineered nucleic acid of claim 18, wherein at least three of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

20. The engineered nucleic acid of claim 18, wherein the promoter of (a) is operably linked to at least four heterologous genes of a biosynthetic pathway of interest.

21. The engineered nucleic acid of claim 20, wherein at least four of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

22. The engineered nucleic acid of claim 1, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 5' to 3' direction.

23. The engineered nucleic acid of claim 1, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 3' to 5' direction.

24. A cell comprising the engineered construct of claim 1.

25. The cell of claim 24, wherein the cell is bacterial cell.

26. The cell of claim 25, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

27. The cell of claim 26, wherein the bacterial cell is a *Streptomyces* spp. cell.

28. The cell of claim 27, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

29. The cell of claim 24, wherein the cell comprises a deletion or modification in a gene of a biosynthetic pathway of interest.

30. The cell of claim 29, wherein the biosynthetic pathway of interest is a pikromycin biosynthetic pathway.

31. The cell of claim 30, wherein the cell comprises a deletion or modification in a *pikAI-V* , *pikBI-VIII*, *pikBR* gene, *pikC* gene or *pikD* gene.

32. The cell of claim 31, wherein the cell comprises a deletion in a *pikAII* gene.

33. The cell of claim 24, wherein the engineered nucleic acid is present on an episomal vector or integrated into a chromosome of the cell.

34. A cell comprising:

(a) an engineered nucleic acid comprising a promoter operably linked to a gene encoding a shufflon recombinase; and

(b) an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs.

35. The cell of claim 34, wherein the at least two heterologous genes encode proteins selected from the group consisting of: enzymes, regulatory proteins and transport proteins.

36. The cell of claim 34, wherein the shufflon recombinase is a *Salmonella enterica* shufflon recombinase.

37. The cell of claim 34, wherein the gene encoding the shufflon recombinase is codon-optimized for expression in the cell.

38. The cell of claim 34, wherein the gene encoding the shufflon recombinase comprise a nucleotide sequence as set forth in SEQ ID NO: 35 or SEQ ID NO: 36.

39. The cell of claim 348, wherein the promoter of (a) and/or (b) is an inducible promoter.

40. The cell of claim 34, wherein the engineered nucleic acid construct of (a) and/or (b) further comprises an antibiotic resistance gene.

41. The cell of claim 34, wherein the RRSs recombine, in the presence of shufflon recombinase activity, at different rates relative to each other.

42. The cell of claim 41, wherein the RRSs comprises a nucleotide sequence selected from sequences set forth as SEQ ID NO: 5-8.

43. The cell of claim 34, wherein at least two of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

44. The cell of claim 34, wherein the promoter of (b) is operably linked to at least three heterologous genes of a biosynthetic pathway of interest.

45. The cell of claim 44, wherein at least three of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

46. The cell of claim 44, wherein the promoter of (b) is operably linked to at least four heterologous genes of a biosynthetic pathway of interest.

47. The cell of claim 46, wherein at least four of the heterologous genes are selected from *S. fradiae tylG*, *S. cinnamonensis momVI*, *S. cinnamonensis momVIII* and *S. nodosus amphK*.

48. The cell of claim 34, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 5' to 3' direction.

49. The cell of claim 34, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein each gene operably linked to the promoter of (a) is oriented in a 3' to 5' direction.

50. The cell of claim 34, wherein the cell is bacterial cell.

51. The cell of claim 50, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

52. The cell of claim 51, wherein the bacterial cell is a *Streptomyces* spp. cell.

53. The cell of claim 52, wherein the *Streptomyces* spp. cell is a *Streptomyces venezuelae* cell or a *Streptomyces coelicolor* cell.

54. The cell of claim 34, wherein the cell comprises a deletion or modification in a gene of a biosynthetic pathway of interest.

55. The cell of claim 54, wherein the biosynthetic pathway of interest is a pikromycin biosynthetic pathway.

56. The cell of claim 55, wherein the cell comprises a deletion or modification in a *pikAI-V* , *pikBI-VIII*, *pikBR* gene, *pikC* gene or *pikD* gene.

57. The cell of claim 56, wherein the cell comprises a deletion in a *pikAII* gene.

58. The cell of claim 34, wherein the engineered nucleic acid of (a) and/or (b) is present on an episomal vector or integrated into a chromosome of the cell.

59. A cell-free composition, comprising

(a) an engineered nucleic acid comprising a promoter operably linked to a gene encoding a shufflon recombinase;

(b) an engineered nucleic acid comprising (i) a promoter operably linked to at least two heterologous genes of a biosynthetic pathway of interest, and (ii) at least two different pairs of shufflon recombinase recognition sequences (RRSs), wherein at least one of the genes is located between at least one of the pairs of shufflon RRSs;

(c) a cell extract; and

(d) a polymerase.

60. The cell-free composition of claim 59, wherein the cell extract comprises at least one component selected from the group consisting of: ribosomes, amino acids, NTPs, phosphoenolpyruvate, pyruvate kinase, polyethylene glycol, ammonium acetate, potassium acetate and folinic acid.

61. The cell-free composition of claim 59 further comprising an engineered nucleic acid encoding additional proteins of the biosynthetic pathway of interest, or further comprising additional proteins of the biosynthetic pathway of interest.

62. The cell-free composition of claim 59, wherein the cell extract comprises a S30 cell fraction obtained from a bacterial cell.

63. The cell-free composition of claim 62, wherein the bacterial cell belongs to a genus selected from the group consisting of: *Yersinia* spp., *Escherichia* spp., *Klebsiella* spp., *Bordetella* spp., *Neisseria* spp., *Aeromonas* spp., *Franciesella* 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., *Bacteroides* spp., *Prevotella* spp., *Clostridium* spp., *Bifidobacterium spp.* and *Lactobacillus spp.*

64. The cell-free composition of claim 63, wherein the bacterial cell is a *Escherichia* spp. cell or a *Streptomyces* spp. cell.

65. The cell-free composition of claim 59, wherein the polymerase is a RNA polymerase.

66. The cell-free composition of claim 59, wherein the RNA polymerase is a RNA T7 polymerase.

67. A method of producing molecules, comprising incubating the cell-free composition of any one of claims 59-66 under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the biosynthetic pathway of interest.

68. A method of producing molecules, comprising:

culturing the cells of any one of claims 24-58 under conditions that result in expression of a shufflon recombinase, recombination of RSSs, expression of at least two heterologous genes, and production of molecules produced by the biosynthetic pathway of interest.

69. The method of claim 68 further comprising isolating the molecules.

70. The method of claim 69, wherein the molecules comprise antibiotics.

71. A cell, comprising:

(a) an engineered nucleic acid, comprising a promoter operably linked to at least one heterologous gene encoding an antimicrobial peptide or an antimicrobial protein located between a pair of shufflon recombinase recognition sites (RRSs); and

(b) a nucleic acid comprising a gene encoding a shufflon recombinase.

72. The cell of claim 71, wherein the promoter of (a) is oriented in the 5' to 3' direction, and wherein at least one heterologous gene encoding an antimicrobial peptide or an antimicrobial protein is oriented in a 3' to 5' direction.

73. The cell of claim 71 or 72, wherein the cell is a bacterial cell.

74. The cell of claim 73, wherein the bacterial cell is an *Escherichia* spp. cell.

75. The cell of claim 73, wherein the bacterial cell is an *Salmonella* spp. cell.

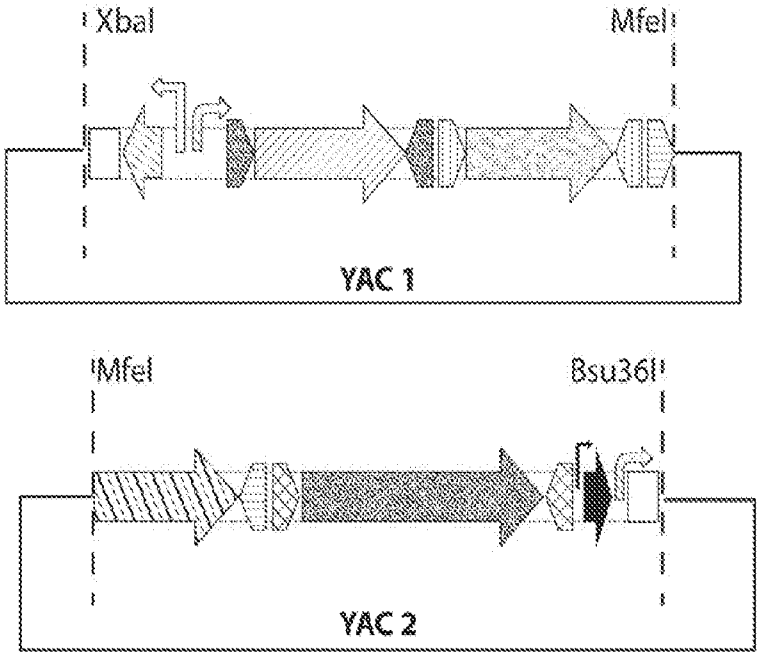


Fig. 1

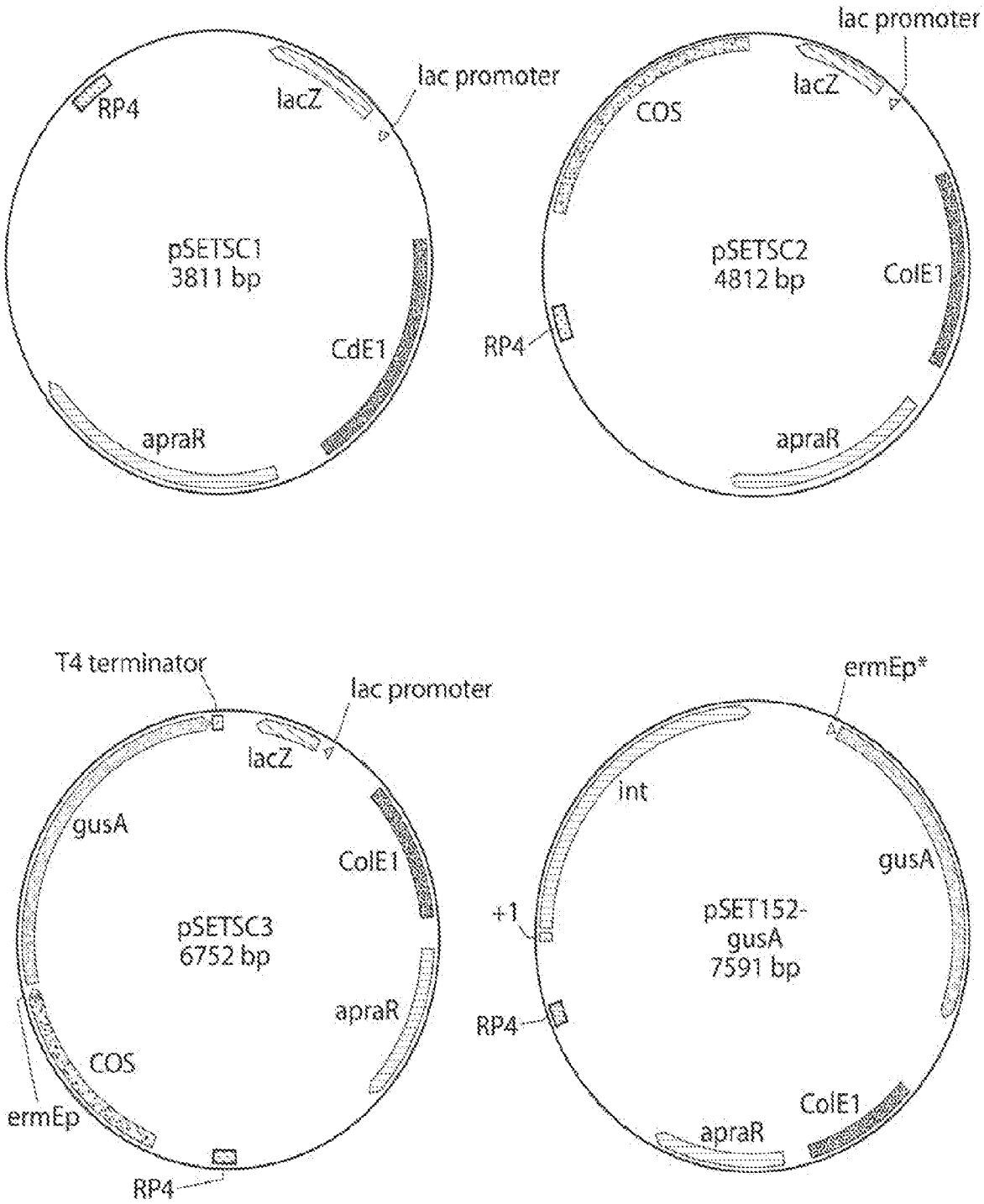

RP4
lacZ
lac promoter
pSETSC1
3811 bp
CdE1
apraR
lac promoter
lacZ
COS
pSETSC2
4812 bp
ColE1
RP4
apraR
T4 terminator
lac promoter
lacZ
gusA
ColE1
pSETSC3
6752 bp
apraR
COS
ermEp
RP4
ermEp*
int
gusA
+1
pSET152-
gusA
7591 bp
RP4
ColE1
apraR

Fig. 2

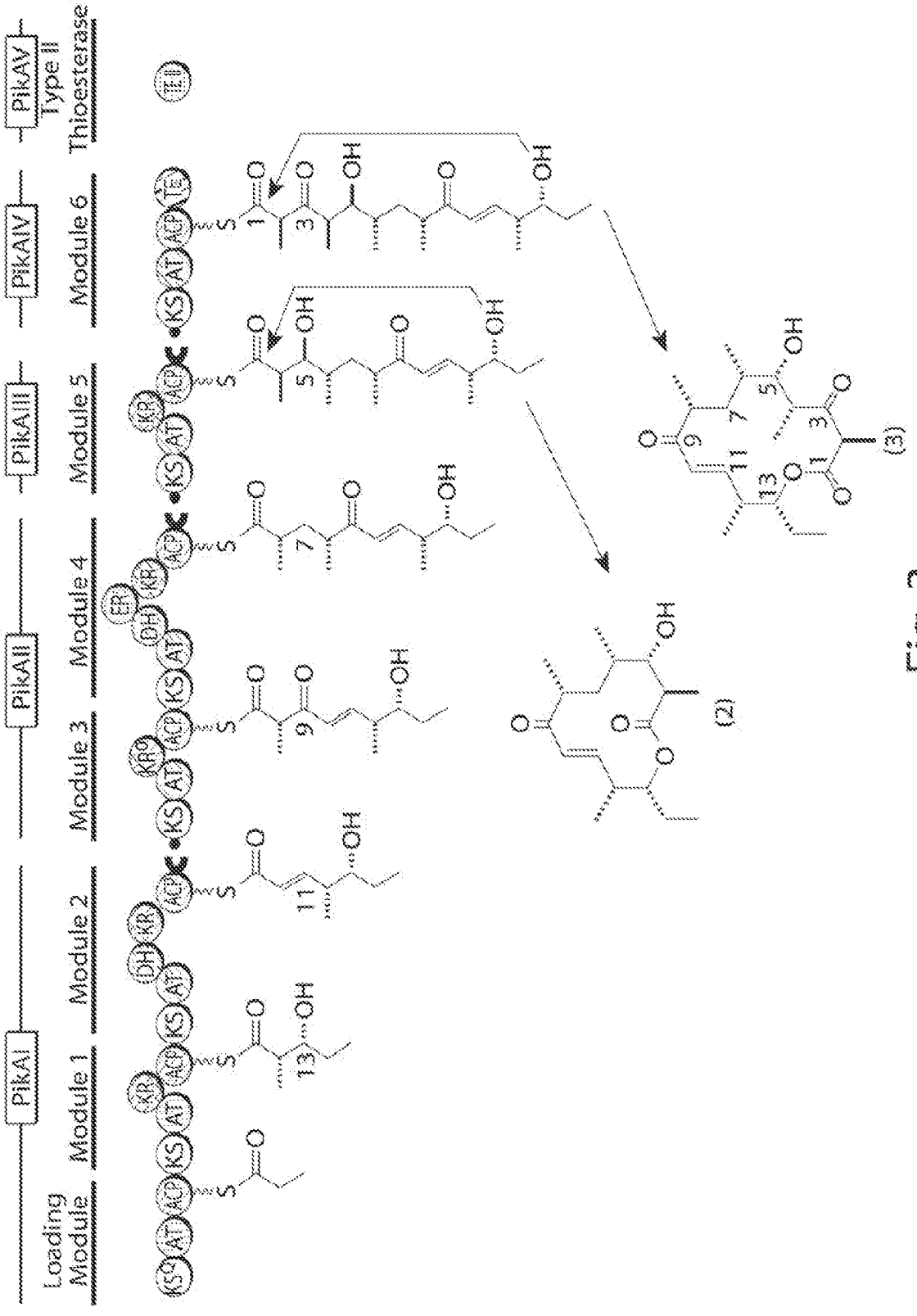

PikAI
PikAII
PikAIII
PikAIV
PikAV
Type II
Thioesterase
Loading Module
Module 1
Module 2
Module 3
Module 4
Module 5
Module 6
(2)
(3)

Fig. 3

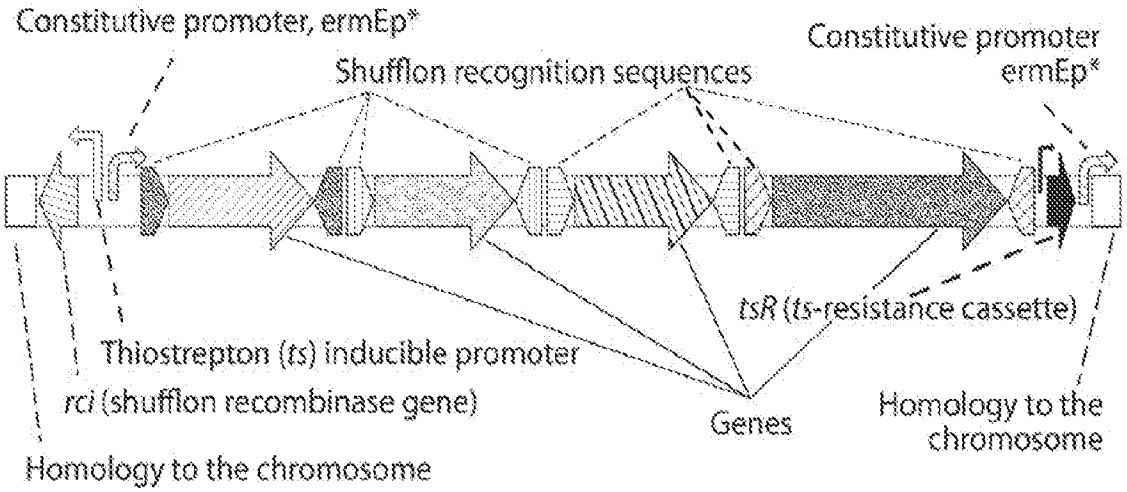


Fig. 4

GPCR ligand	0.09	0.09	0.14	0.15
Ion channel modulator	-0.34	-0.34	-0.20	0.02
Kinase inhibitor	-0.44	-0.44	-0.36	-0.31
Nuclear receptor ligand	0.11	0.11	0.22	0.33
Protease inhibitor	0.22	0.22	0.17	0.18
Enzyme inhibitor	0.06	0.06	0.09	0.20



Fig. 5

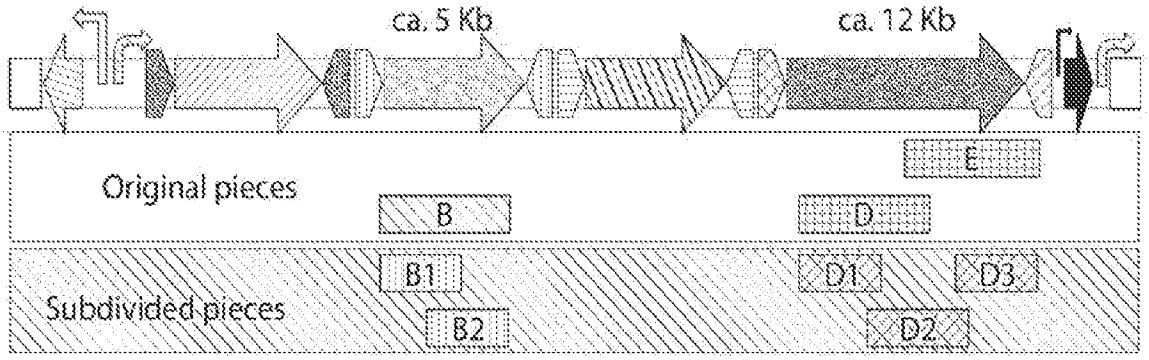


Fig. 6

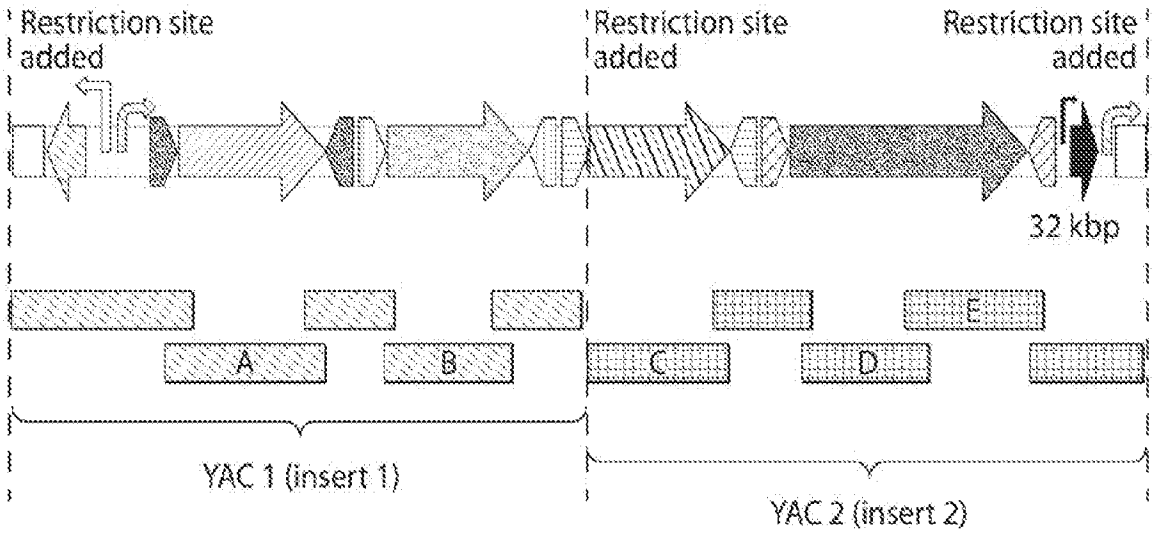

Restriction site added
Restriction site added
Restriction site added
32 kbp
A
B
C
D
E
YAC 1 (insert 1)
YAC 2 (insert 2)

Fig. 7

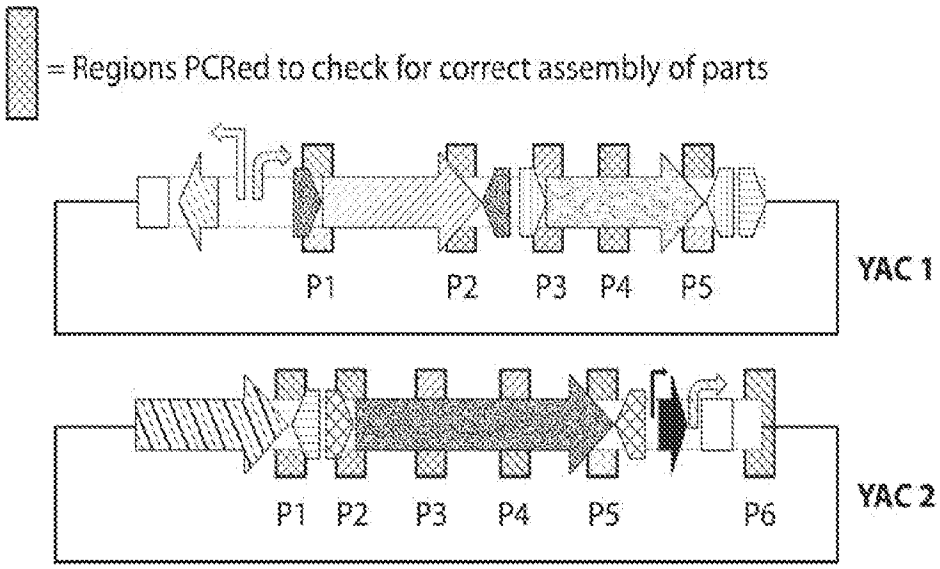

= Regions PCRed to check for correct assembly of parts
P1
P2
P3
P4
P5
YAC 1
P1
P2
P3
P4
P5
P6
YAC 2

Fig. 8

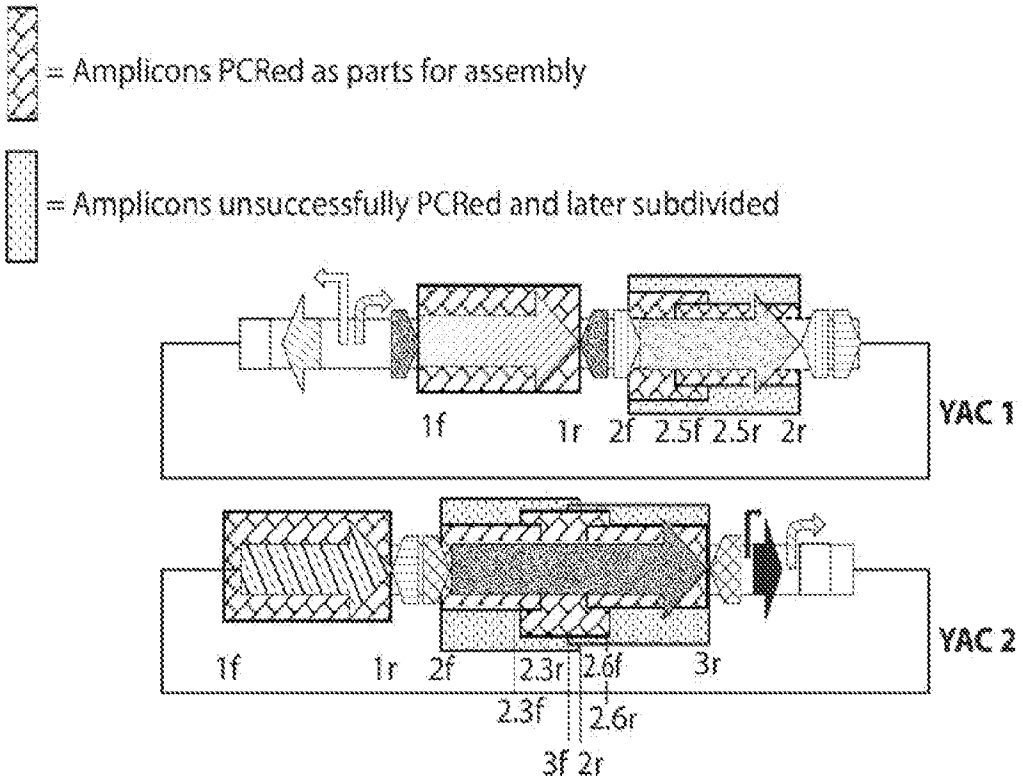

= Amplicons PCRed as parts for assembly
= Amplicons unsuccessfully PCRed and later subdivided
1f
1r
2f
2.5f
2.5r
2r
YAC 1
1f
1r
2f
2.3r
2.6f
3r
2.3f
2.6r
3f
2r
YAC 2

Fig. 9

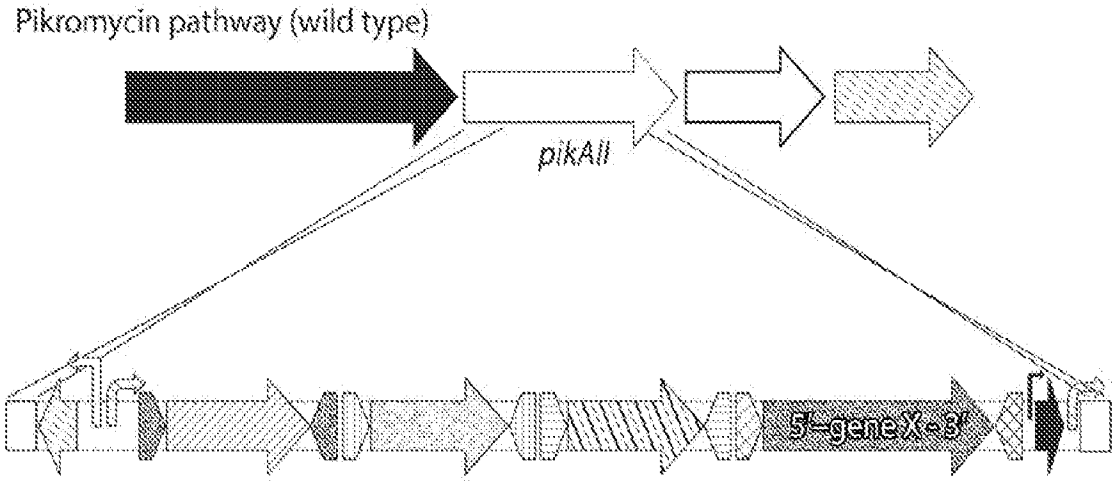

Pikromycin pathway (wild type)
pikAII
5'-gene X-3'

Fig. 10

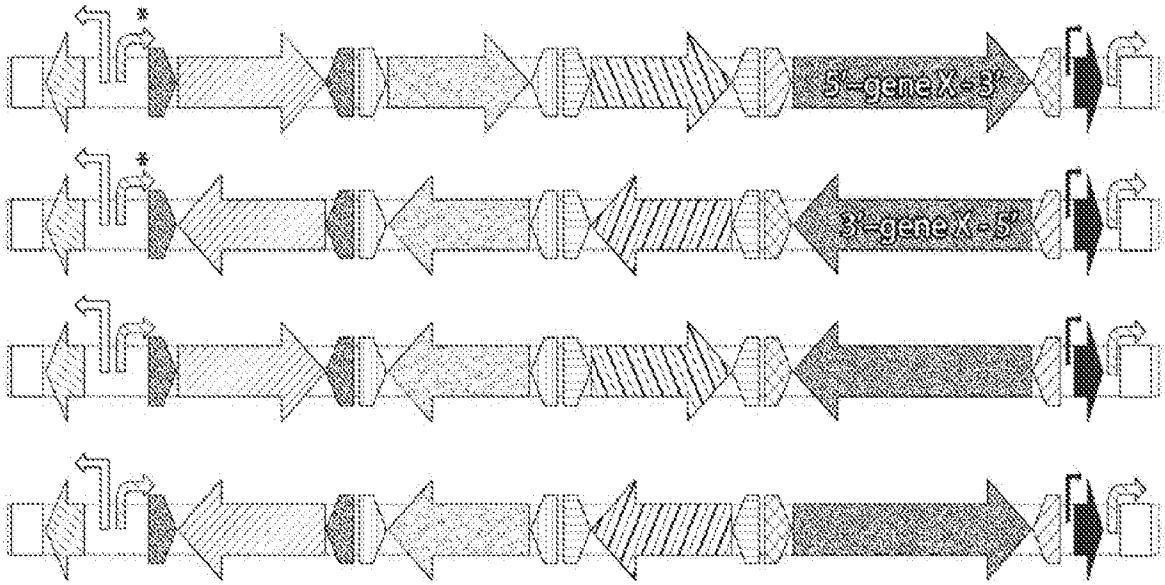


Fig. 11

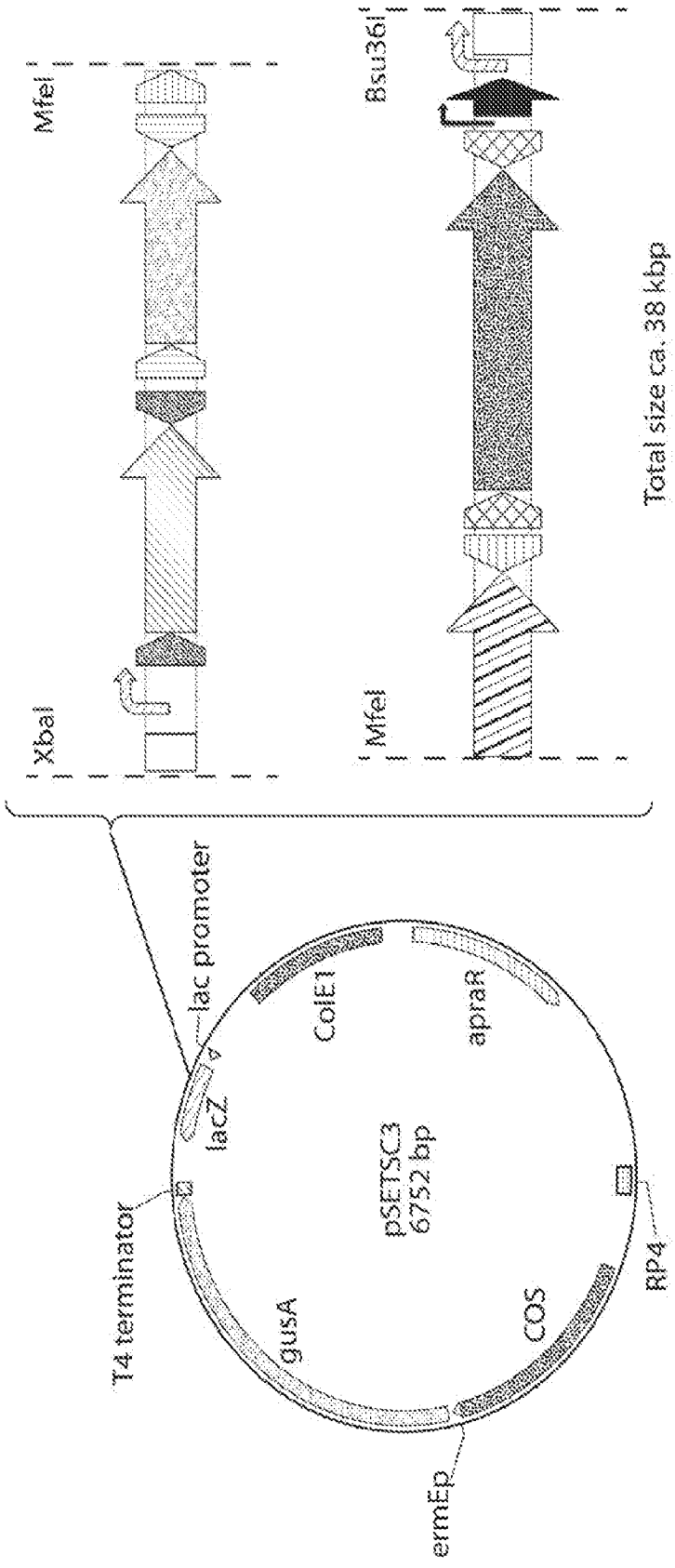

MfeI
Bsu36I
XbaI
MfeI
Total size ca. 38 kbp
lac promoter
ColE1
apraR
lacZ
pSETSC3
6752 bp
T4 terminator
gusA
COS
RP4
ermEp

Fig. 12

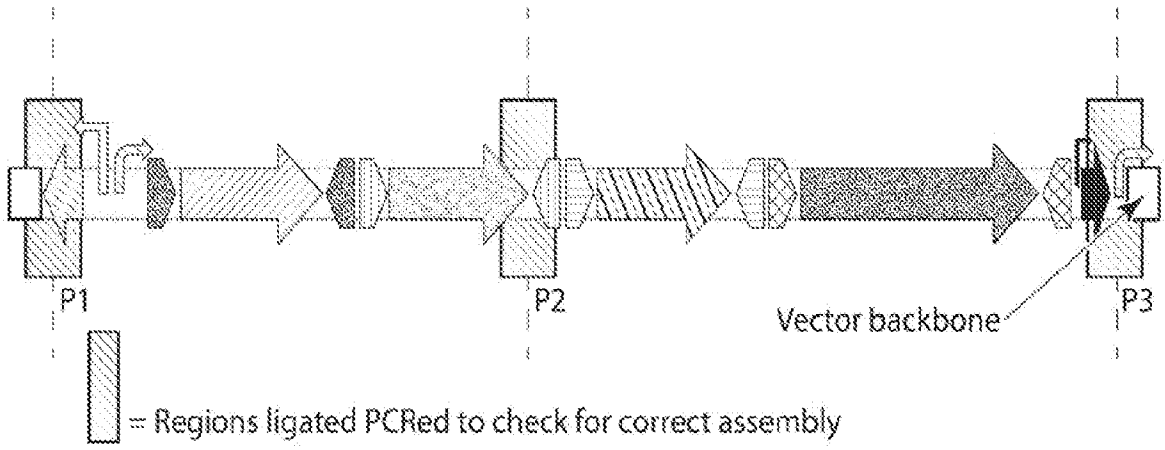

P1
P2
Vector backbone
P3
= Regions ligated PCRed to check for correct assembly

Fig. 13

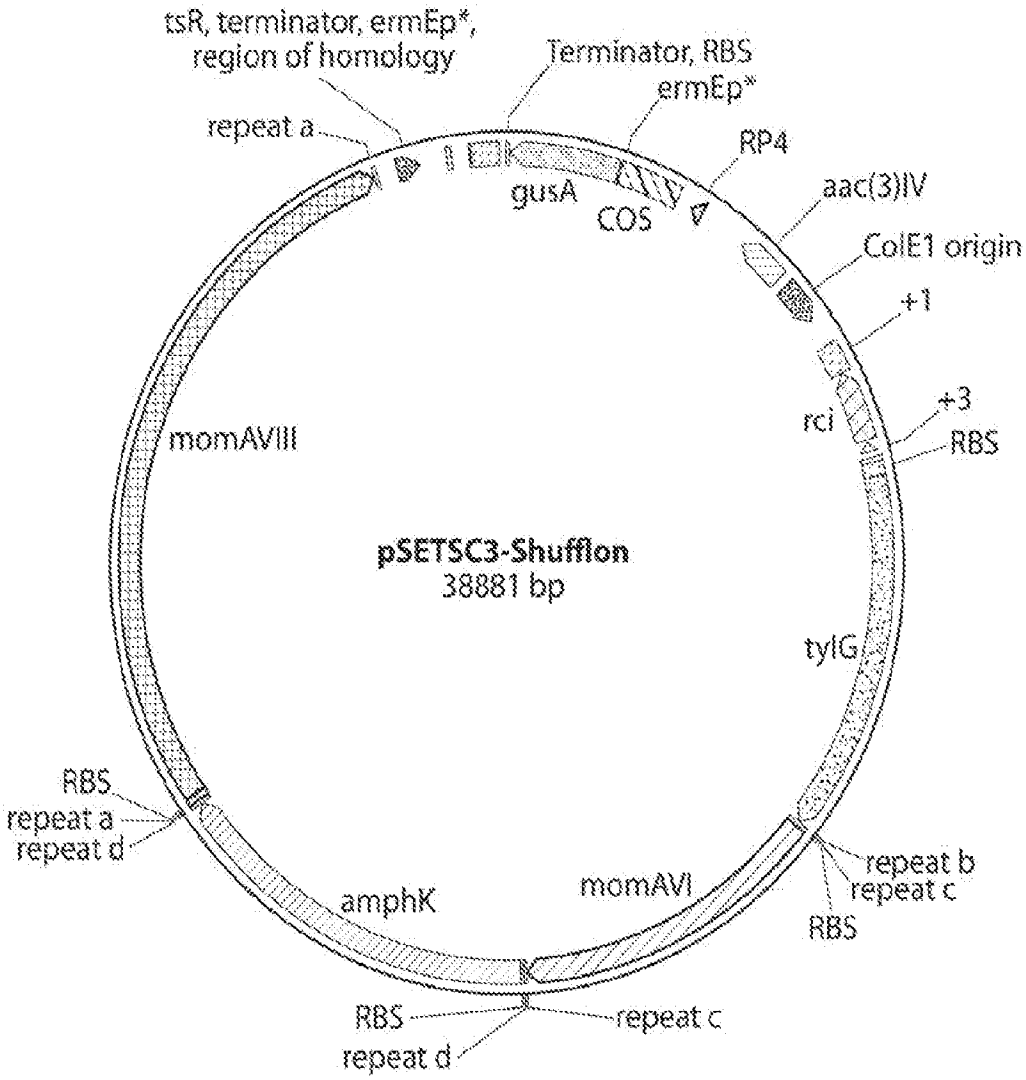

tsR, terminator, ermEp*, region of homology
Terminator, RBS
ermEp*
repeat a
RP4
gusA
COS
aac(3)IV
ColE1 origin
+1
+3
RBS
rci
momAVIII
pSETSC3-Shufflon
38881 bp
tylG
RBS
repeat a
repeat d
amphK
momAVI
repeat b
repeat c
RBS
RBS
repeat d
repeat c

Fig. 14

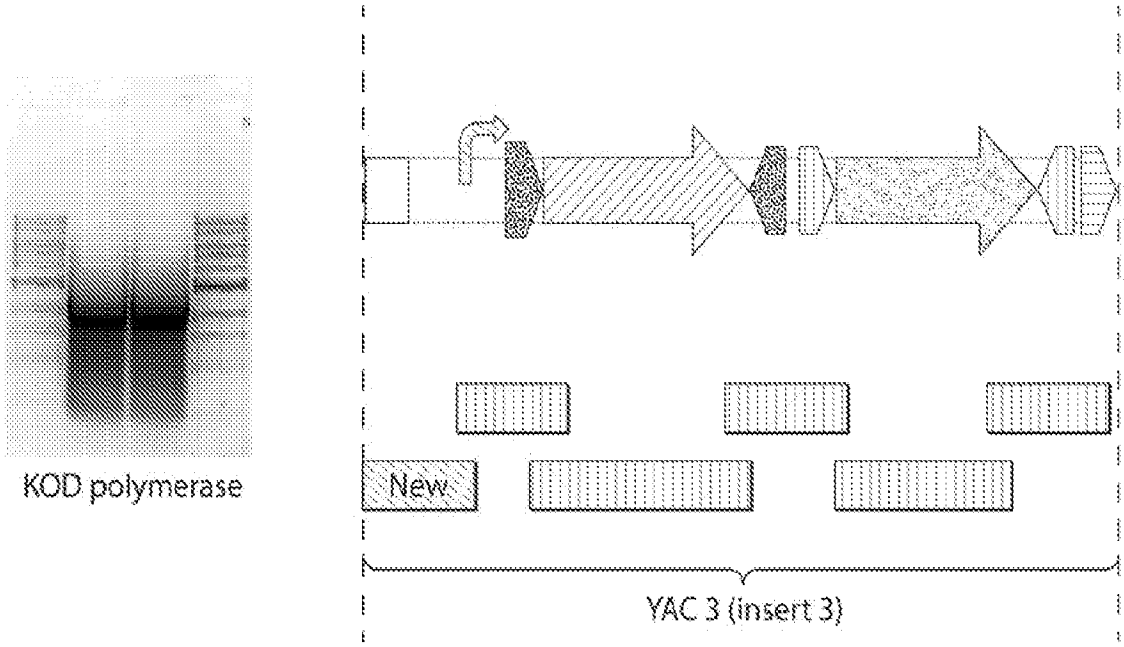

KOD polymerase
New
YAC 3 (insert 3)

Fig. 15

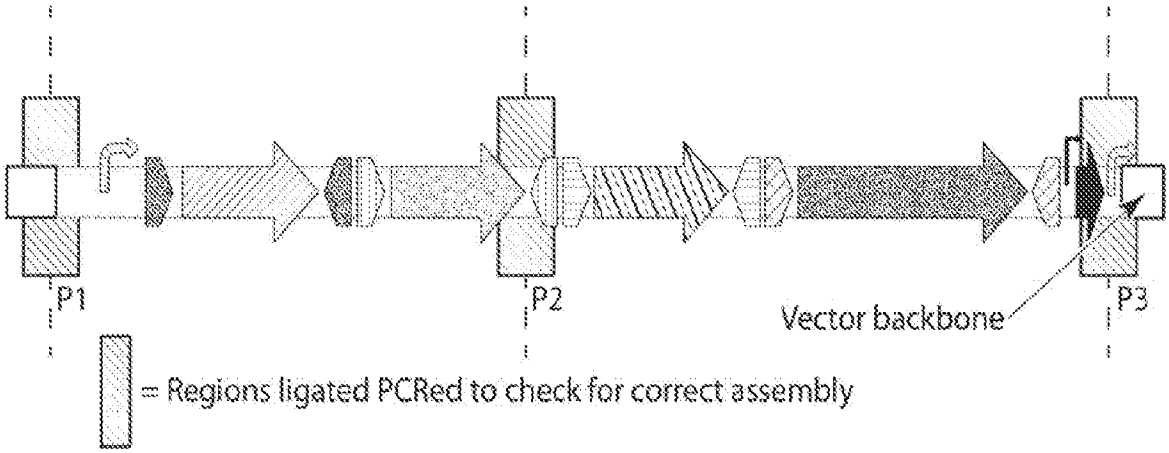

P1
P2
P3
Vector backbone
= Regions ligated PCRed to check for correct assembly

Fig. 16

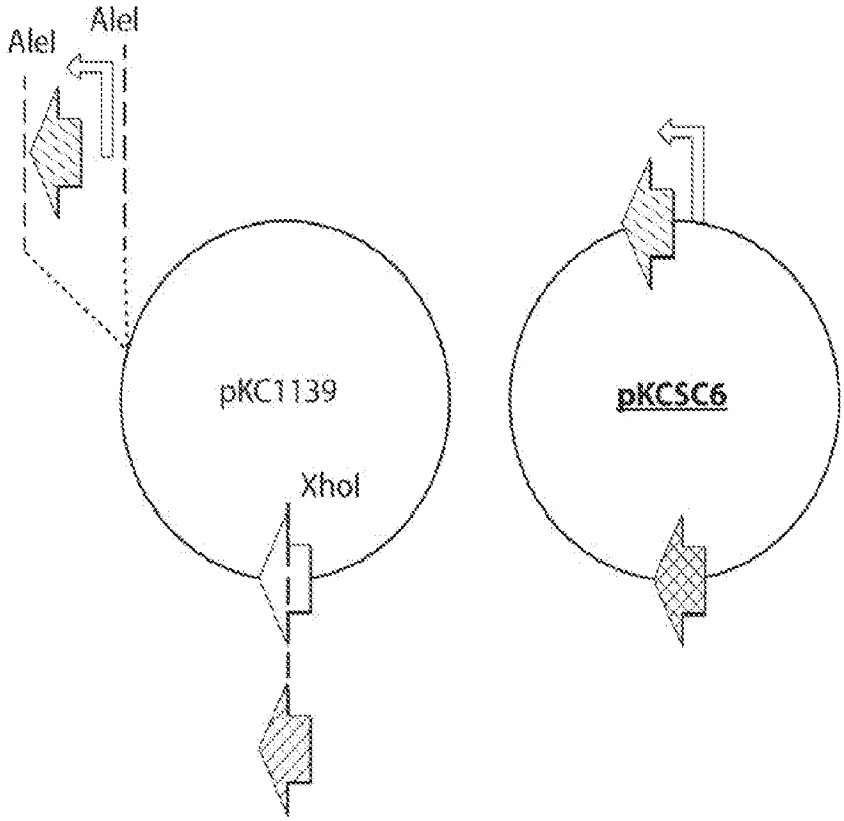

AleI
AleI
pKC1139
XhoI
pKCSC6

Fig. 17

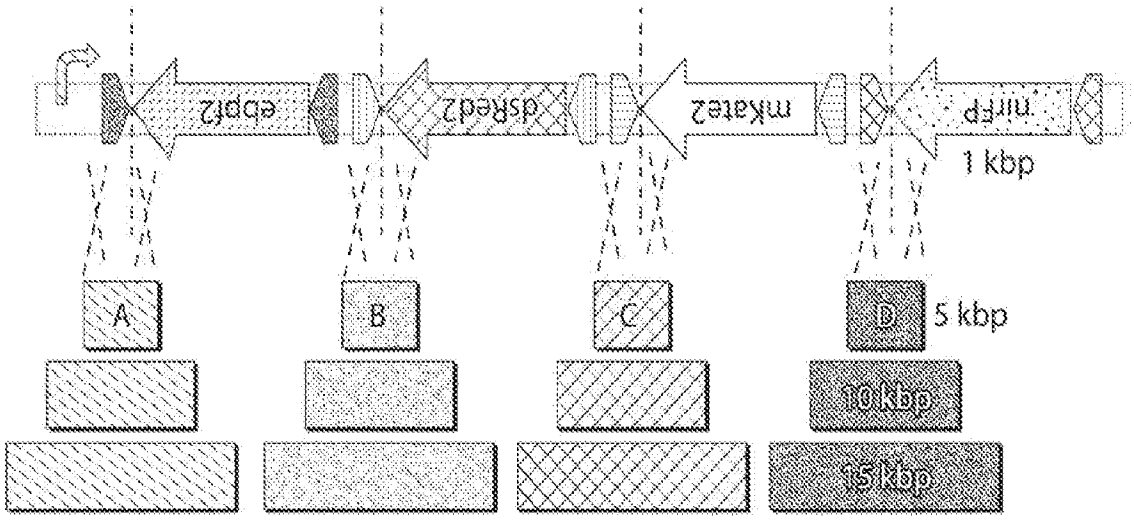

ebpR2
dsRed2
mKate2
nirFP
1 kbp
A
B
C
D
5 kbp
10 Kbp
15 Kbp

Fig. 18

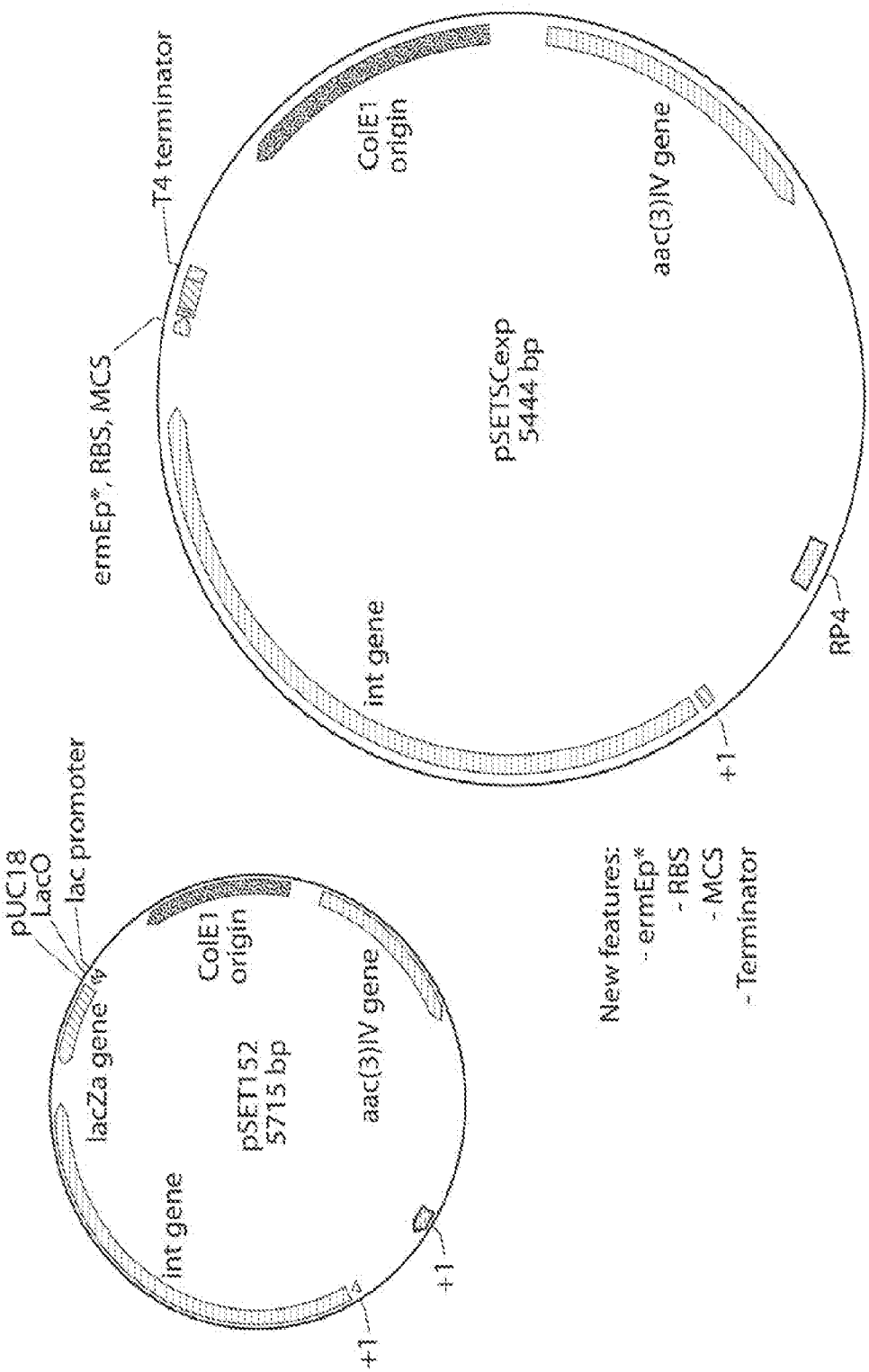

T4 terminator
ermEp*, RBS, MCS
ColE1 origin
aac(3)IV gene
pSETSCexp 5444 bp
int gene
RP4
+1
pUC18 LacO
lac promoter
lacZa gene
ColE1 origin
pSET152 5715 bp
aac(3)IV gene
int gene
+1
+1
New features:
- ermEp*
- RBS
- MCS
- Terminator

Fig. 19

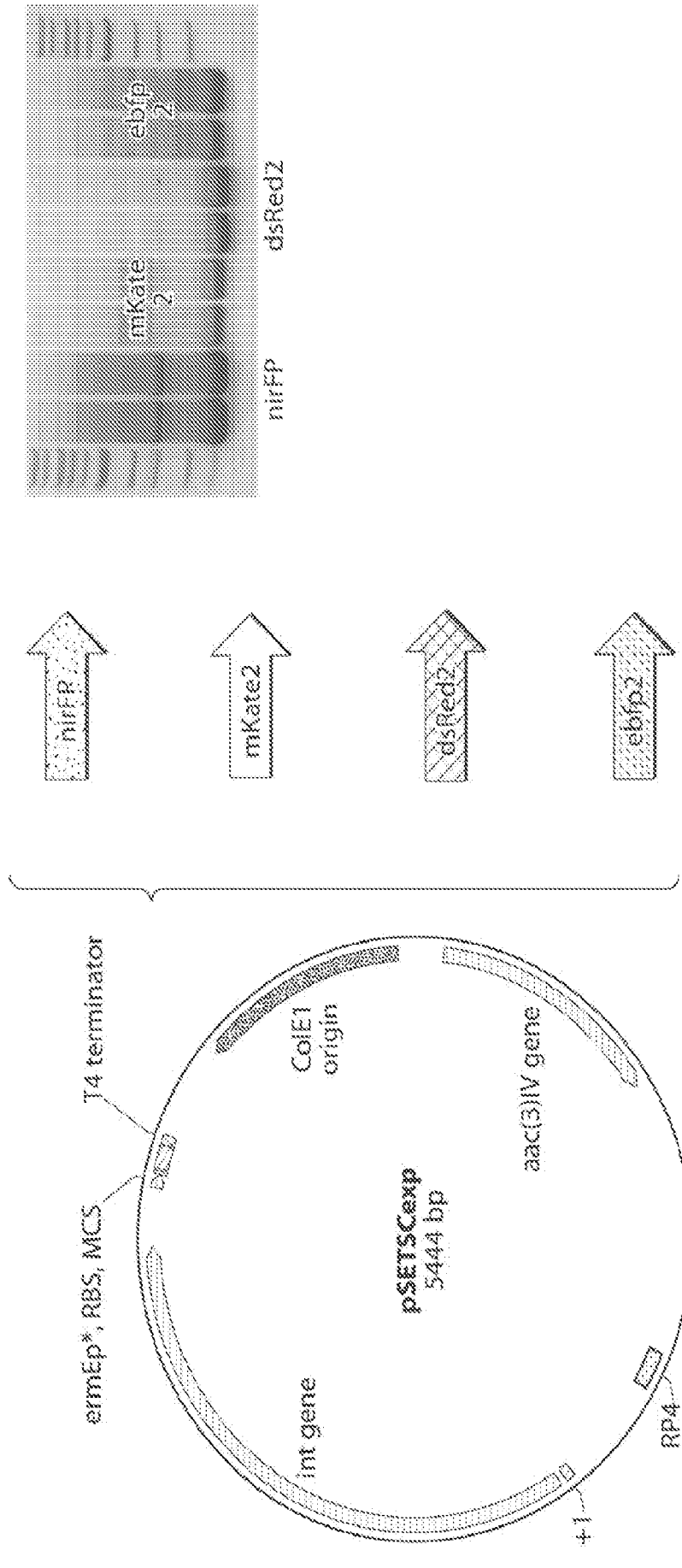

nirFP
mKate2
dsRed2
ebfp2
nirFP
mKate2
dsRed2
ebfp2
T4 terminator
ermEp*, RBS, MCS
ColE1 origin
aac(3)IV gene
pSETSCexp
5444 bp
int gene
RP4
+1

Fig. 20

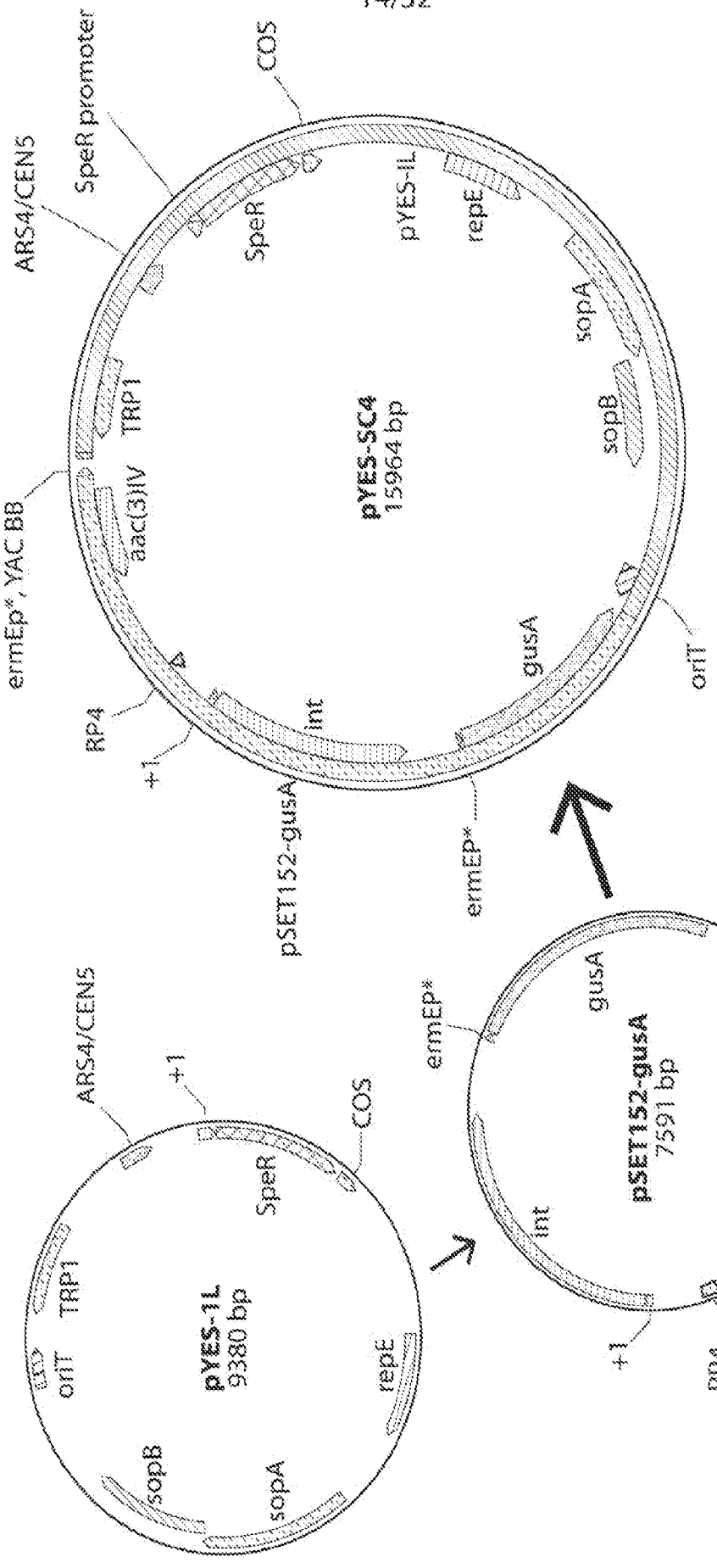

pYES-SC4
15964 bp
ARS4/CEN5
SpeR promoter
COS
SpeR
pYES-1L
repE
sopA
sopB
TRP1
aac(3)IV
ermEp*, YAC BB
RP4
+1
int
gusA
oriT
pSET152-gusA
ermEP*
pYES-1L
9380 bp
ARS4/CEN5
+1
COS
SpeR
TRP1
oriT
repE
sopB
sopA
pSET152-gusA
7591 bp
ermEP*
gusA
ColE1
apraR
int
+1
RP4

Fig. 21

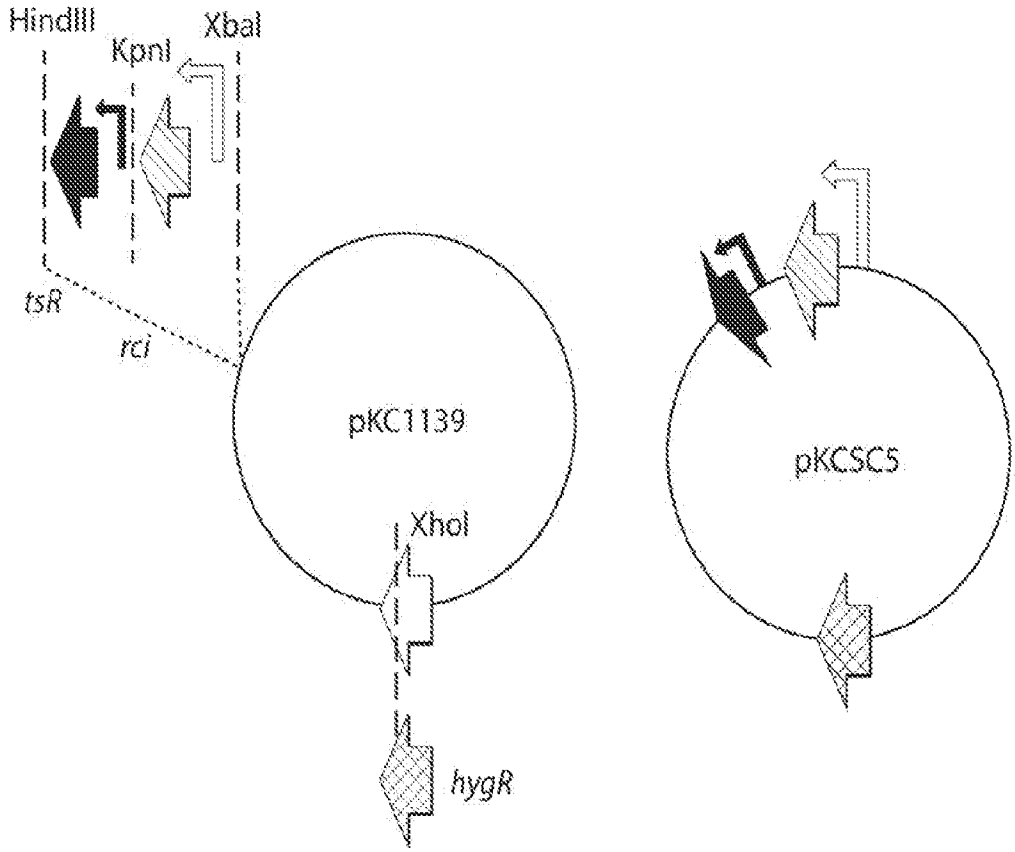
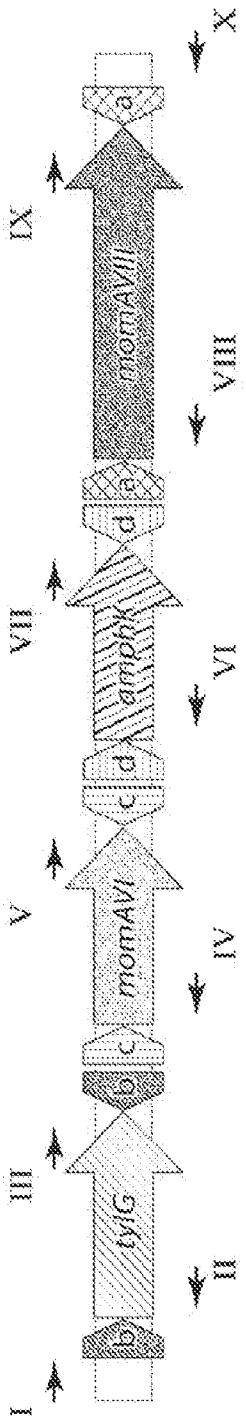
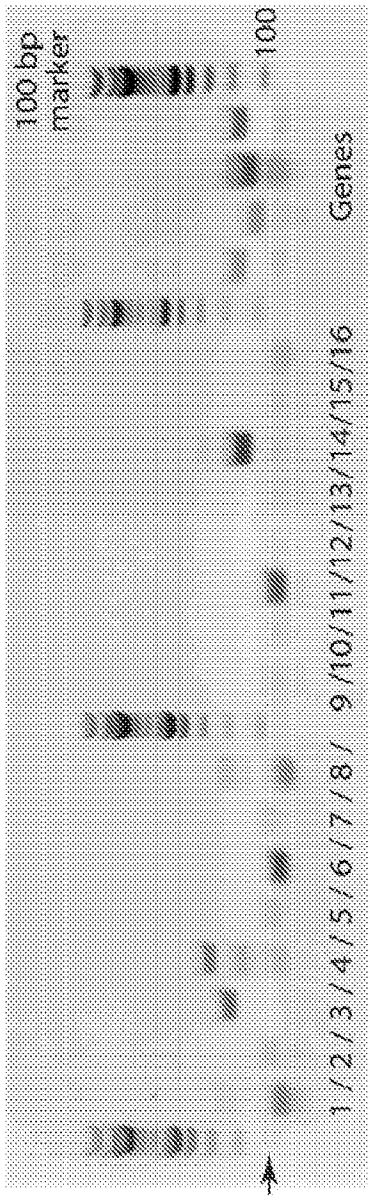


Fig. 22





1 - I + II	9 - II + V
2 - III + IV	10 - II + IV
3 - V + VI	11 - V + VII
4 - VII + VIII	12 - IV + VI
5 - IX + X	13 - IV + VII
6 - I + III	14 - VII + IX
7 - VIII + X	15 - VI + VII
8 - III + V	16 - VI + IX

Recognition sequence	Estimated % of flipping
a	1
b	32
c	0.09
d	9.6

Fig. 23

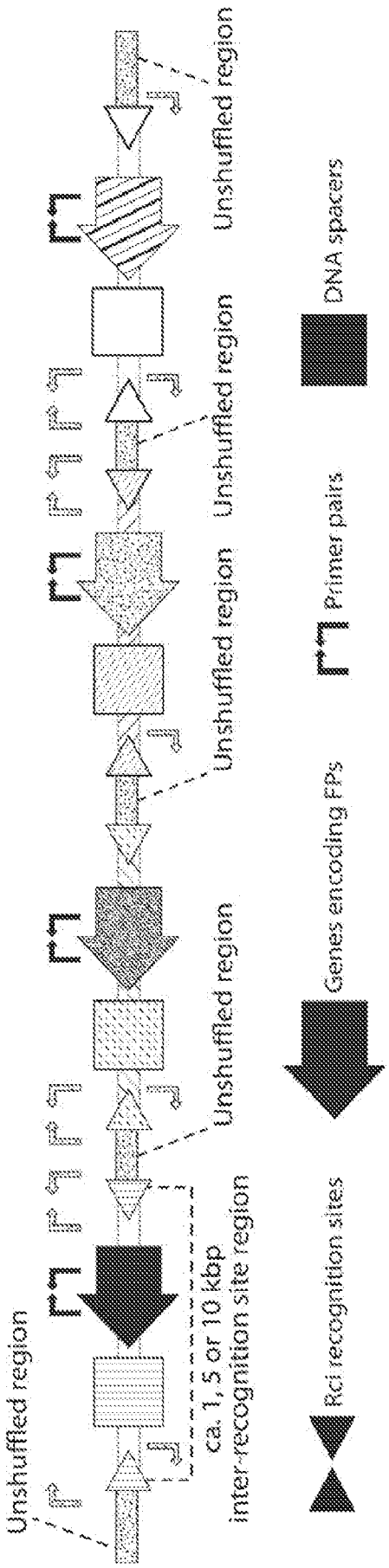

Unshuffled region
ca. 1, 5 or 10 kbp
inter-recognition site region
Unshuffled region
Unshuffled region
Unshuffled region
Unshuffled region
Unshuffled region
Rci recognition sites
Genes encoding FPs
Primer pairs
DNA spacers

Fig. 24

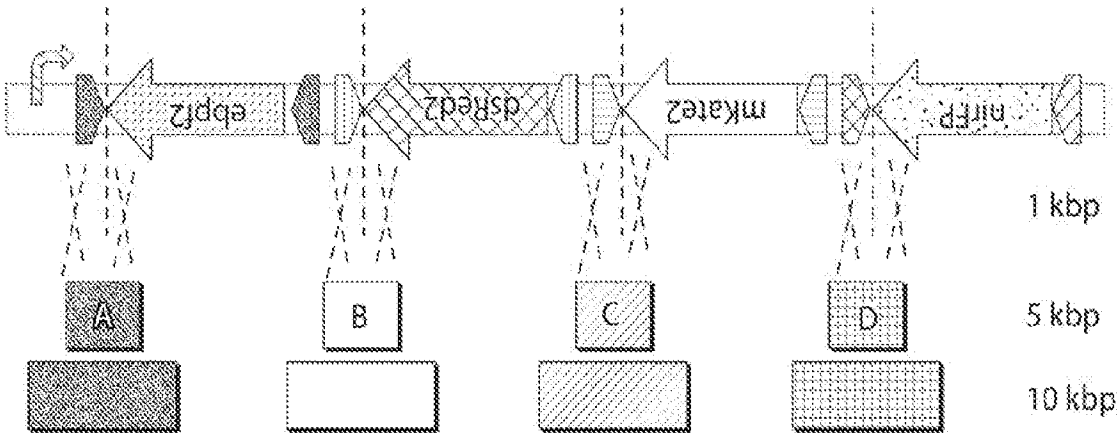


Fig. 25

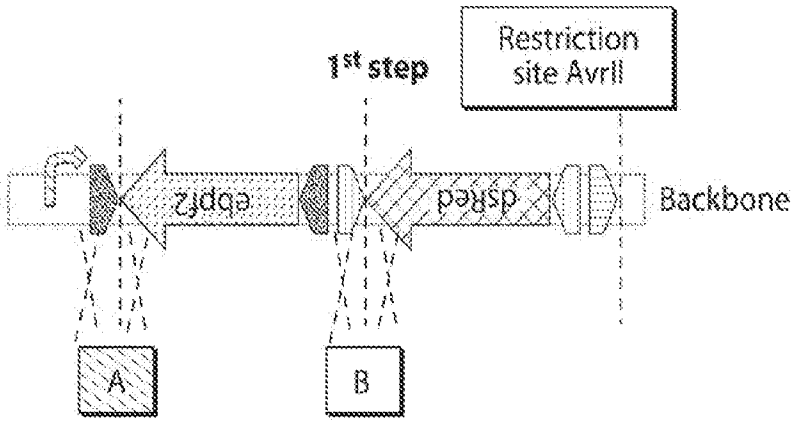


Fig. 26

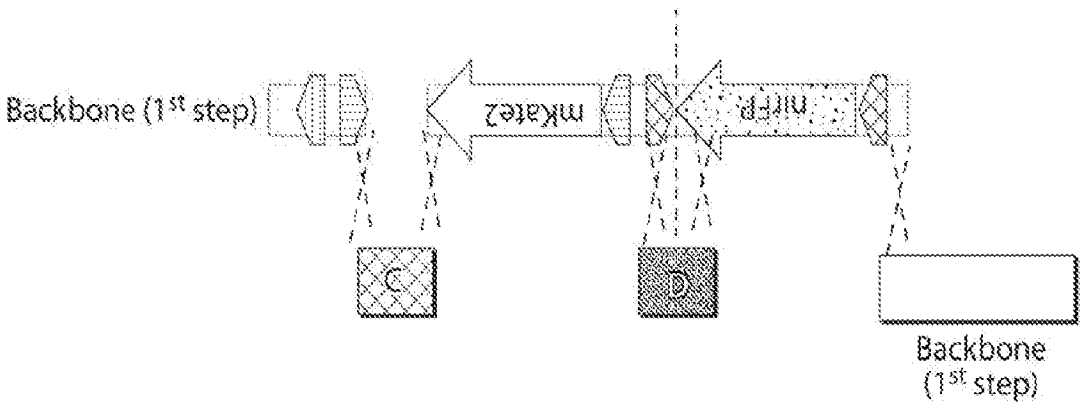


Fig. 27

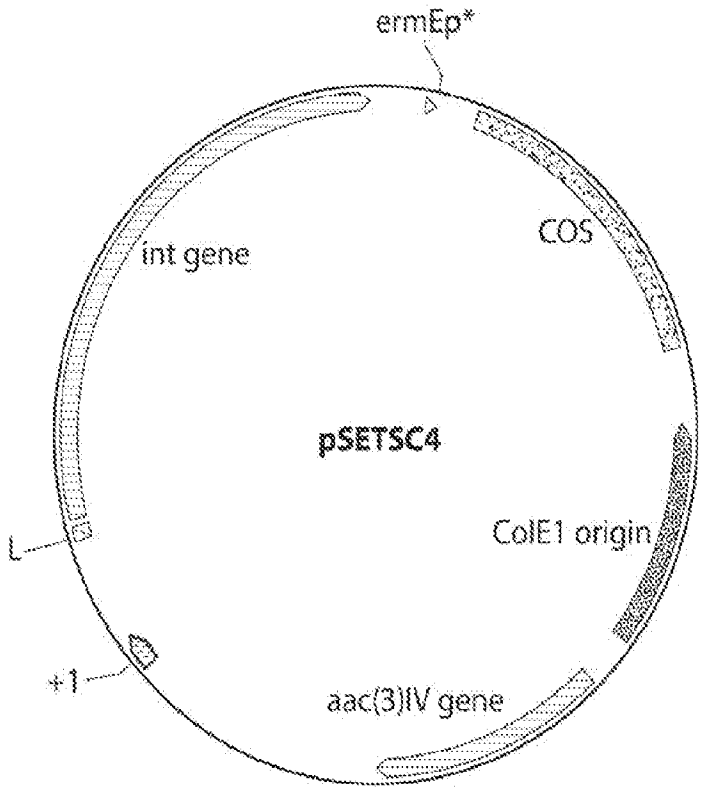

ermEp*
int gene
COS
pSETSC4
ColE1 origin
L
+1
aac(3)IV gene

Fig. 28

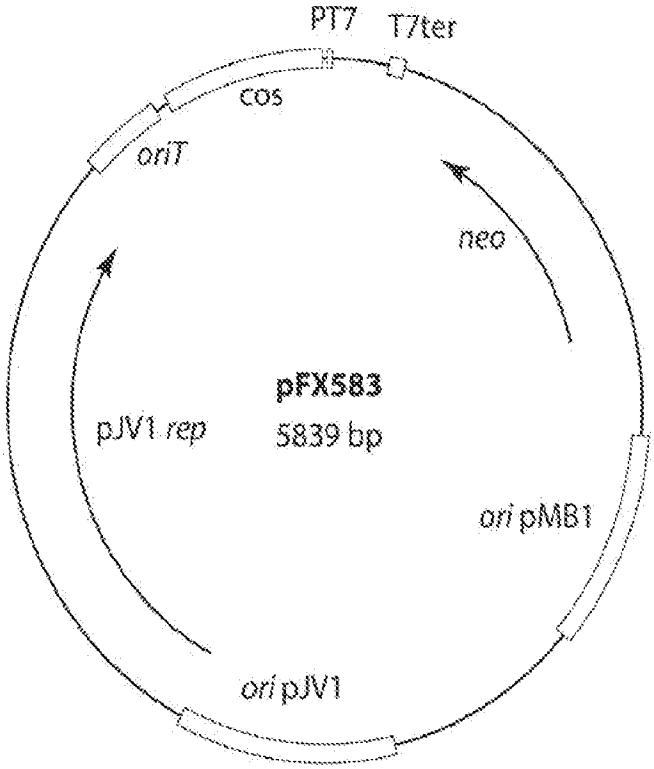

PT7
T7ter
cos
oriT
neo
pFX583
pJV1 rep
5839 bp
ori pMB1
ori pJV1

Fig. 29

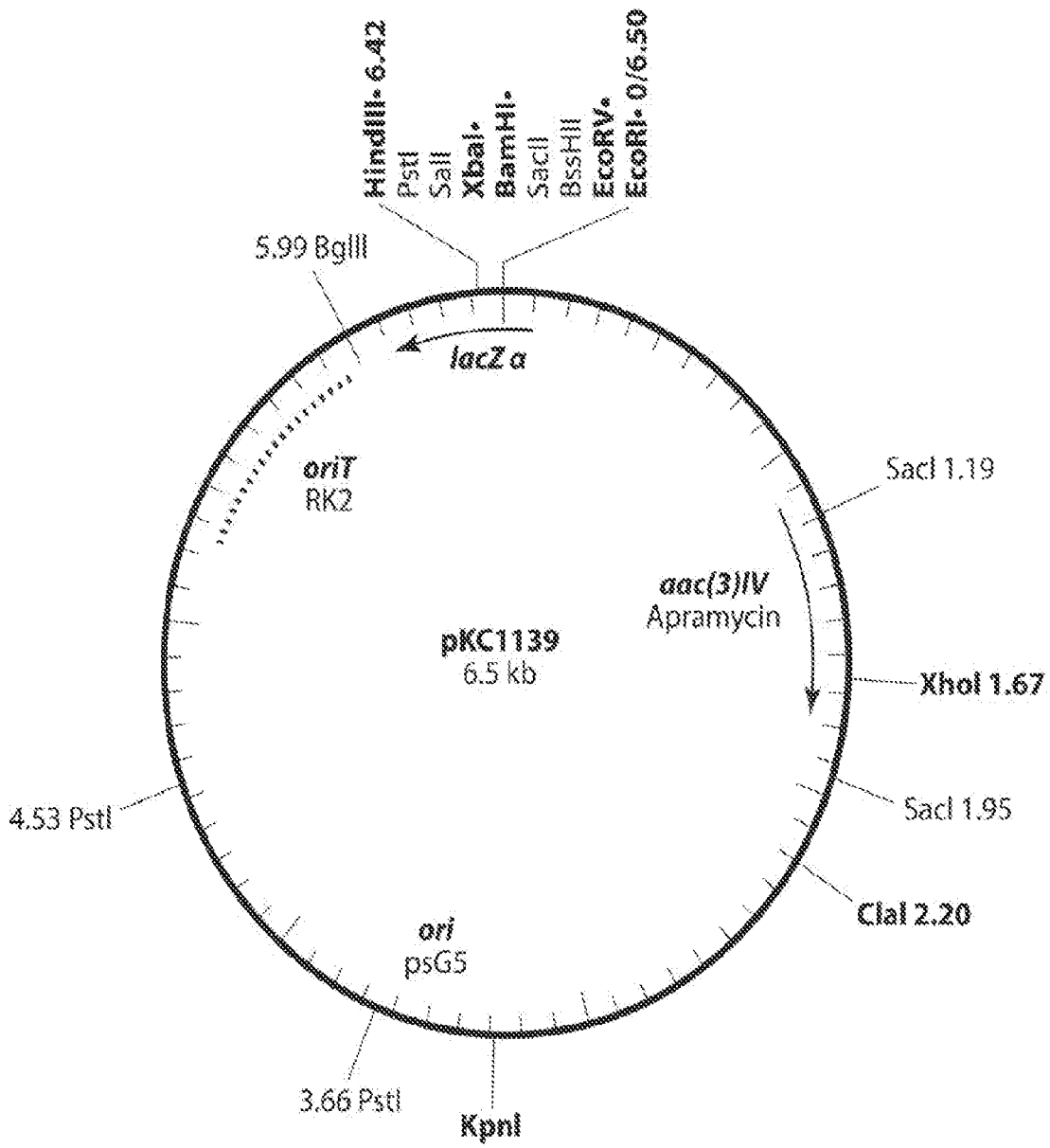

HindIII• 6.42
PstI
SalI
XbaI•
BamHI•
SacII
BssHII
EcoRV•
EcoRI• 0/6.50
5.99 BglII
lacZ α
oriT
RK2
SacI 1.19
aac(3)IV
Apramycin
pKC1139
6.5 kb
XhoI 1.67
SacI 1.95
ClaI 2.20
4.53 PstI
ori
psG5
3.66 PstI
KpnI

Fig. 30

$$\%\text{Flipped} = 2^{\left[\frac{[(CT(flip,Rci)-CT(FP,Rci))-(CT(flip,\ no\ Rci)-CT(FP,\ no\ Rci))]flipped}{[(CT(flip,Rci)-CT(FP,Rci))-(CT(flip,\ no\ Rci)-CT(FP,\ no\ Rci))]flipped + [(CT(unflip,Rci)-CT(FP,Rci))-(CT(unflip,\ no\ Rci)-CT(FP,\ no\ Rci))]unflipped}\right] \times 100}$$

Legend:

CT (flip/unflip, no Rci) = CT value of flipped/unflipped amplicon of interest in the Rci-less sample

CT (FP, no Rci) = CT value of control amplicon in the Rci-less sample

CT (flip/unflip, Rci) = CT value of the flipped/unflipped amplicon in the samples with Rci

CT (FP, Rci) = CT value of control amplicon in the samples with Rci

Fig. 31

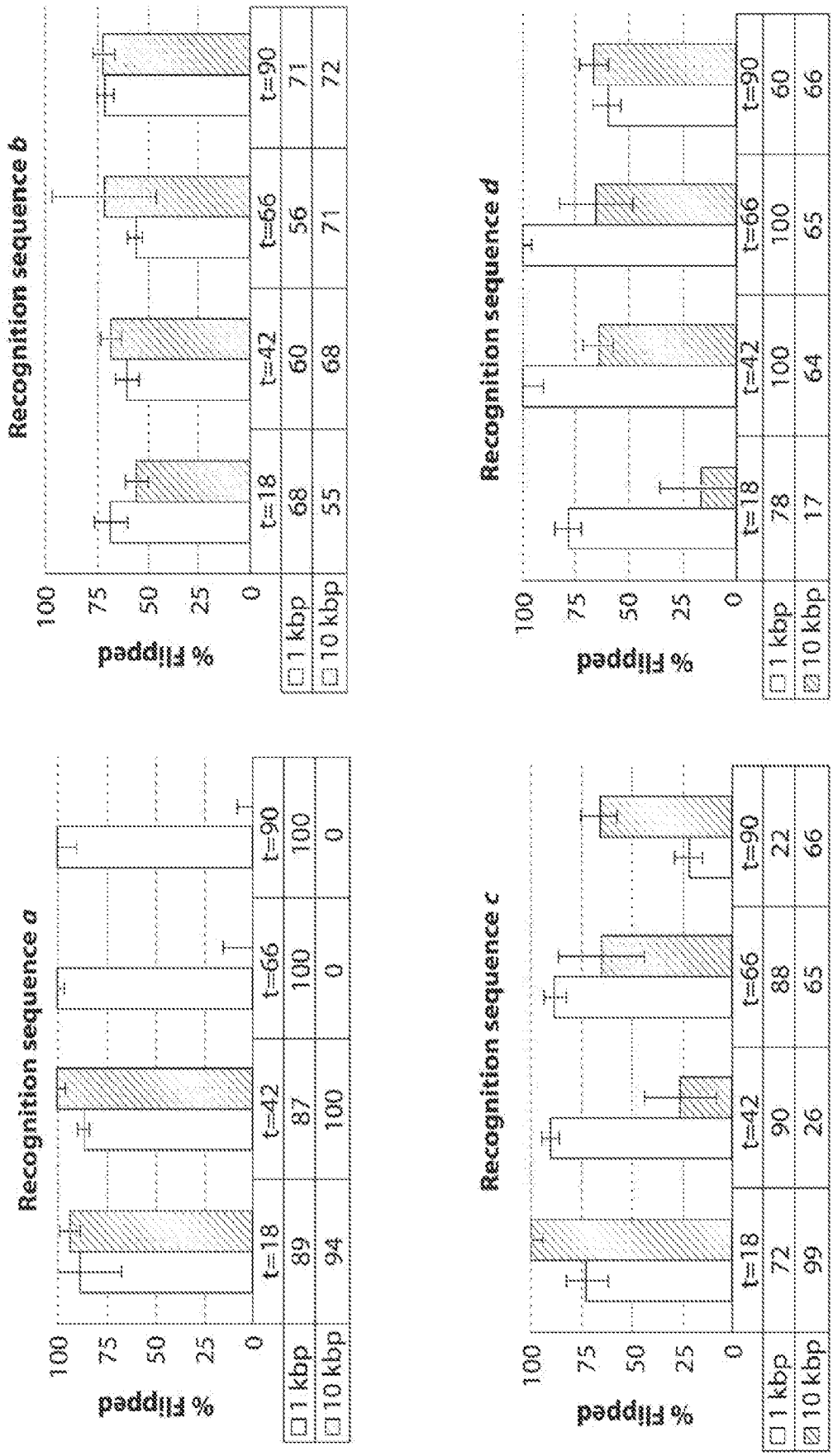

Recognition sequence a
% flipped
100
75
50
25
0
t=18
t=42
t=66
t=90
1 kbp
89
87
100
100
10 kbp
94
100
0
0
Recognition sequence b
% flipped
100
75
50
25
0
t=18
t=42
t=66
t=90
1 kbp
68
60
56
71
10 kbp
55
68
71
72
Recognition sequence c
% flipped
100
75
50
25
0
t=18
t=42
t=66
t=90
1 kbp
72
90
88
22
10 kbp
99
26
65
66
Recognition sequence d
% flipped
100
75
50
25
0
t=18
t=42
t=66
t=90
1 kbp
78
100
100
60
10 kbp
17
64
65
66

Fig. 32

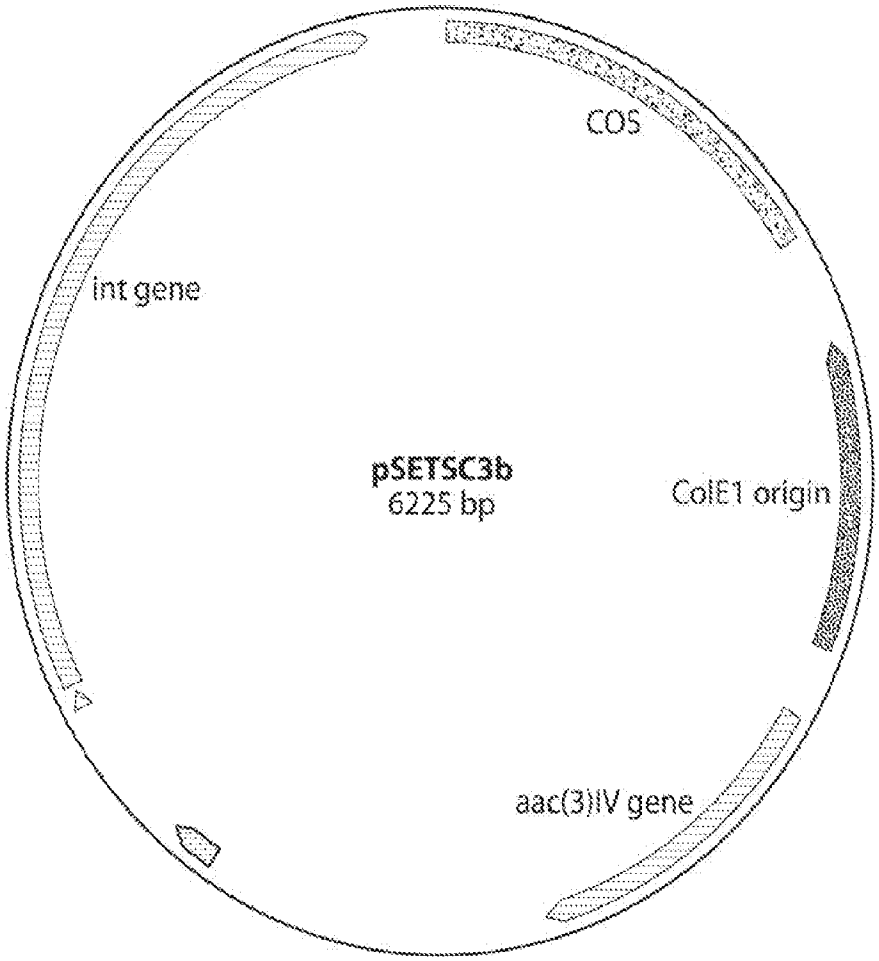


Fig. 33

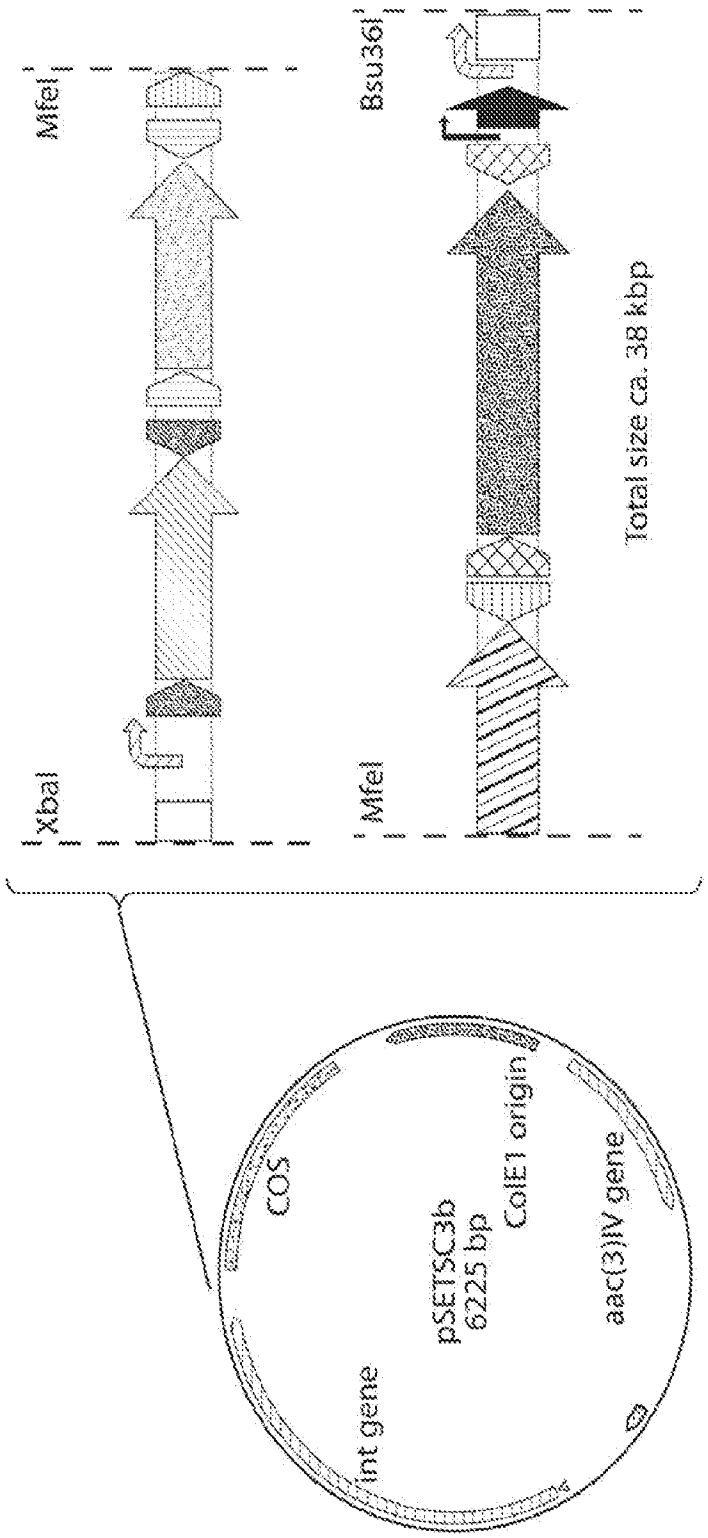

MfeI
Bsu36I
XbaI
MfeI
Total size ca. 38 kbp
COS
pSETSC3b
6225 bp
ColE1 origin
aac(3)IV gene
int gene

Fig. 34

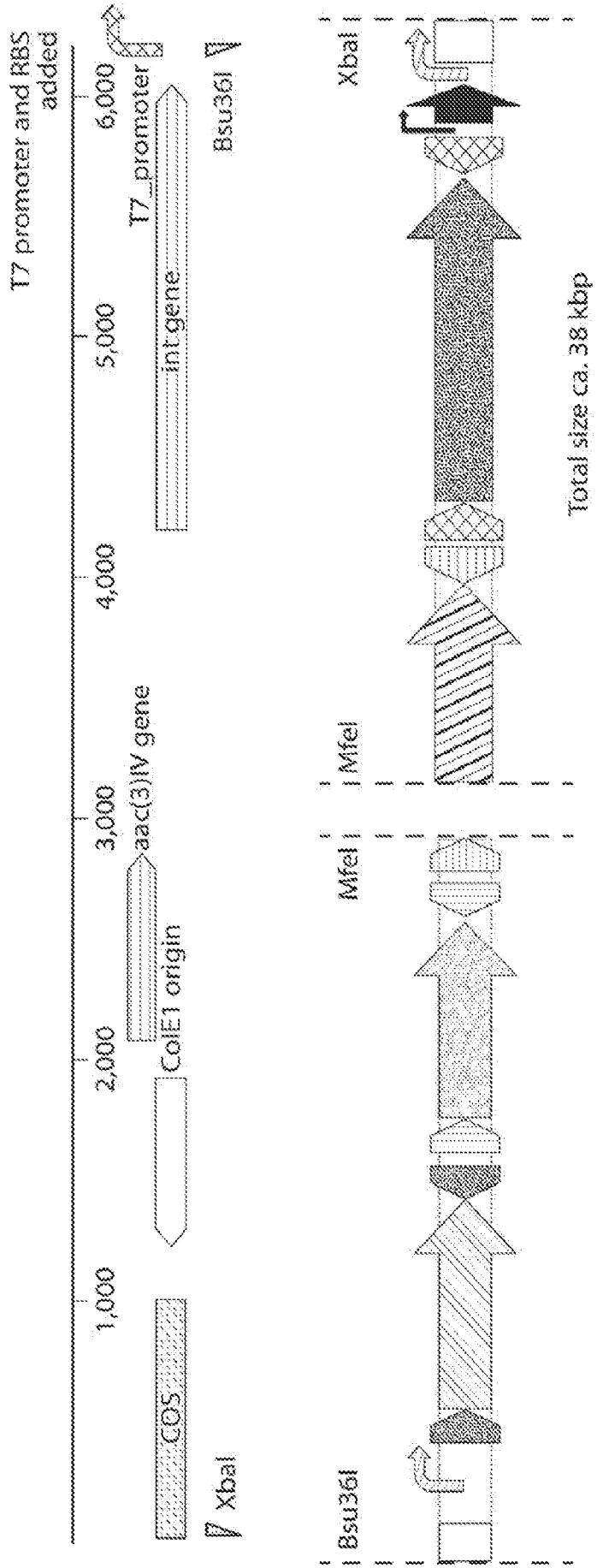

T7 promoter and RBS added
1,000
2,000
3,000
4,000
5,000
6,000
COS
ColE1 origin
aac(3)IV gene
int gene
T7_promoter
XbaI
Bsu36I
MfeI
XbaI
Bsu36I
MfeI
Total size ca. 38 kbp

Fig. 35

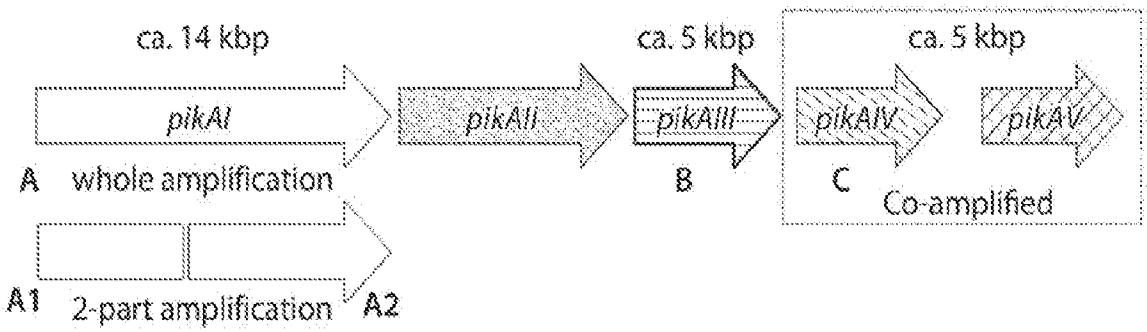

ca. 14 kbp
ca. 5 kbp
ca. 5 kbp
pikAI
pikAII
pikAIII
pikAIV
pikAV
A
whole amplification
B
C
Co-amplified
A1
2-part amplification
A2

Fig. 36

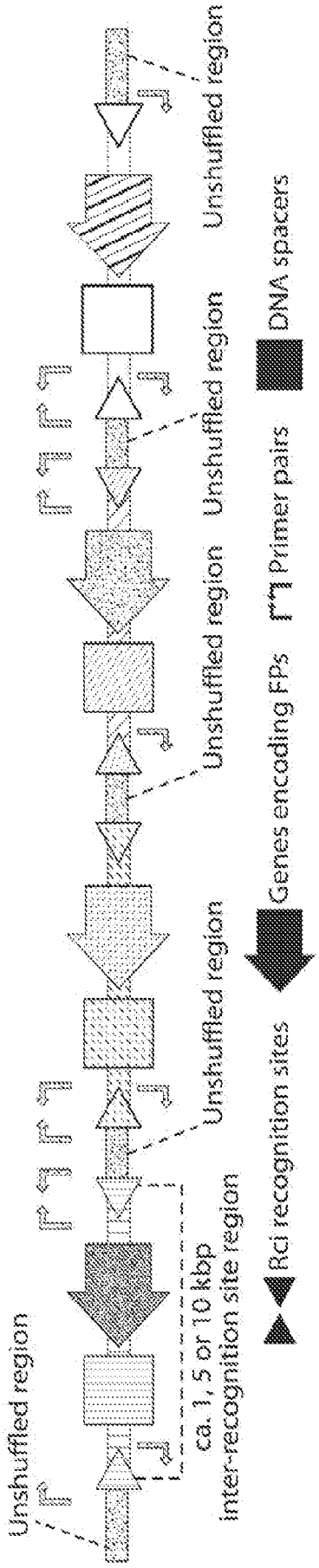


Fig. 37A

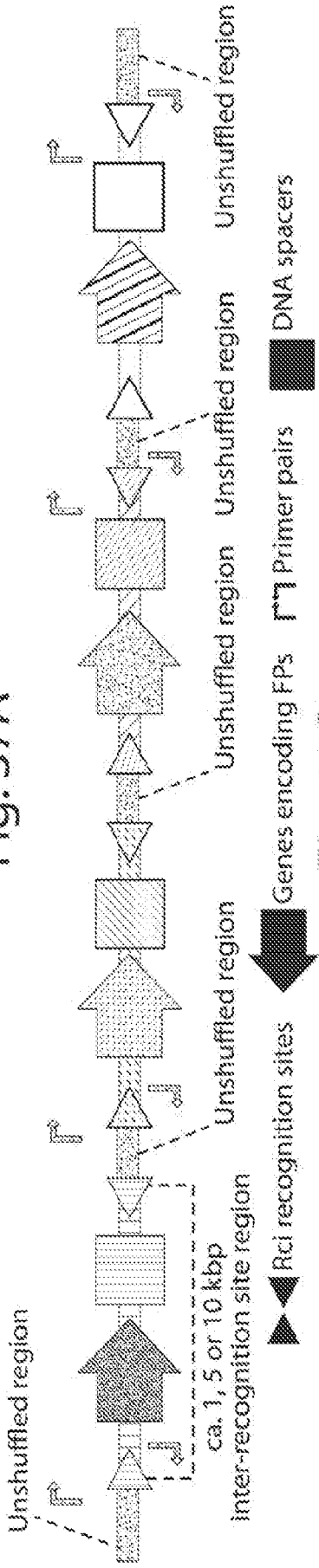


Fig. 37B

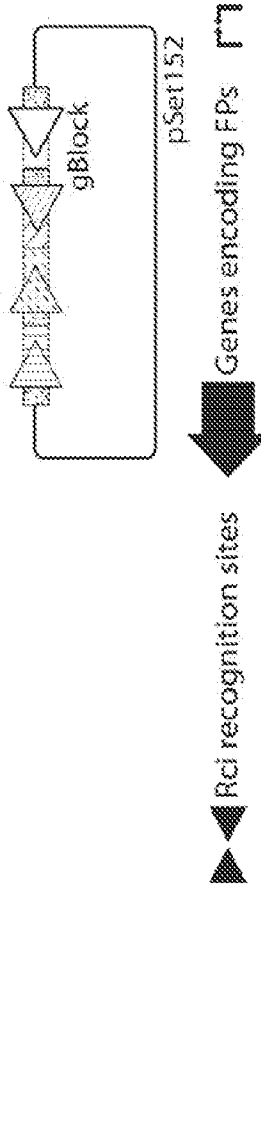


Fig. 37C

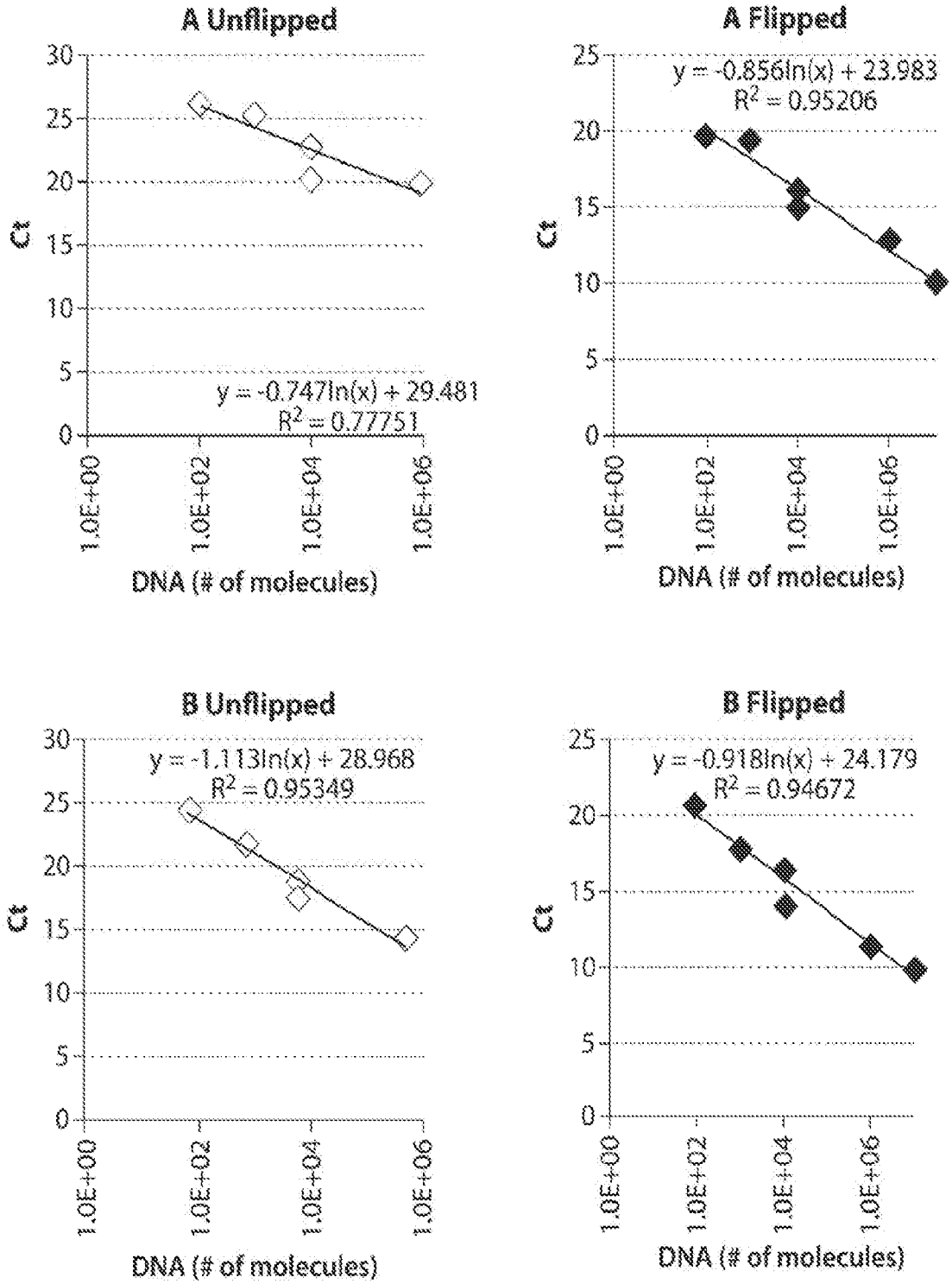

A Unflipped
y = -0.747ln(x) + 29.481
R² = 0.77751
Ct
DNA (# of molecules)
A Flipped
y = -0.856ln(x) + 23.983
R² = 0.95206
Ct
DNA (# of molecules)
B Unflipped
y = -1.113ln(x) + 28.968
R² = 0.95349
Ct
DNA (# of molecules)
B Flipped
y = -0.918ln(x) + 24.179
R² = 0.94672
Ct
DNA (# of molecules)

Fig. 38

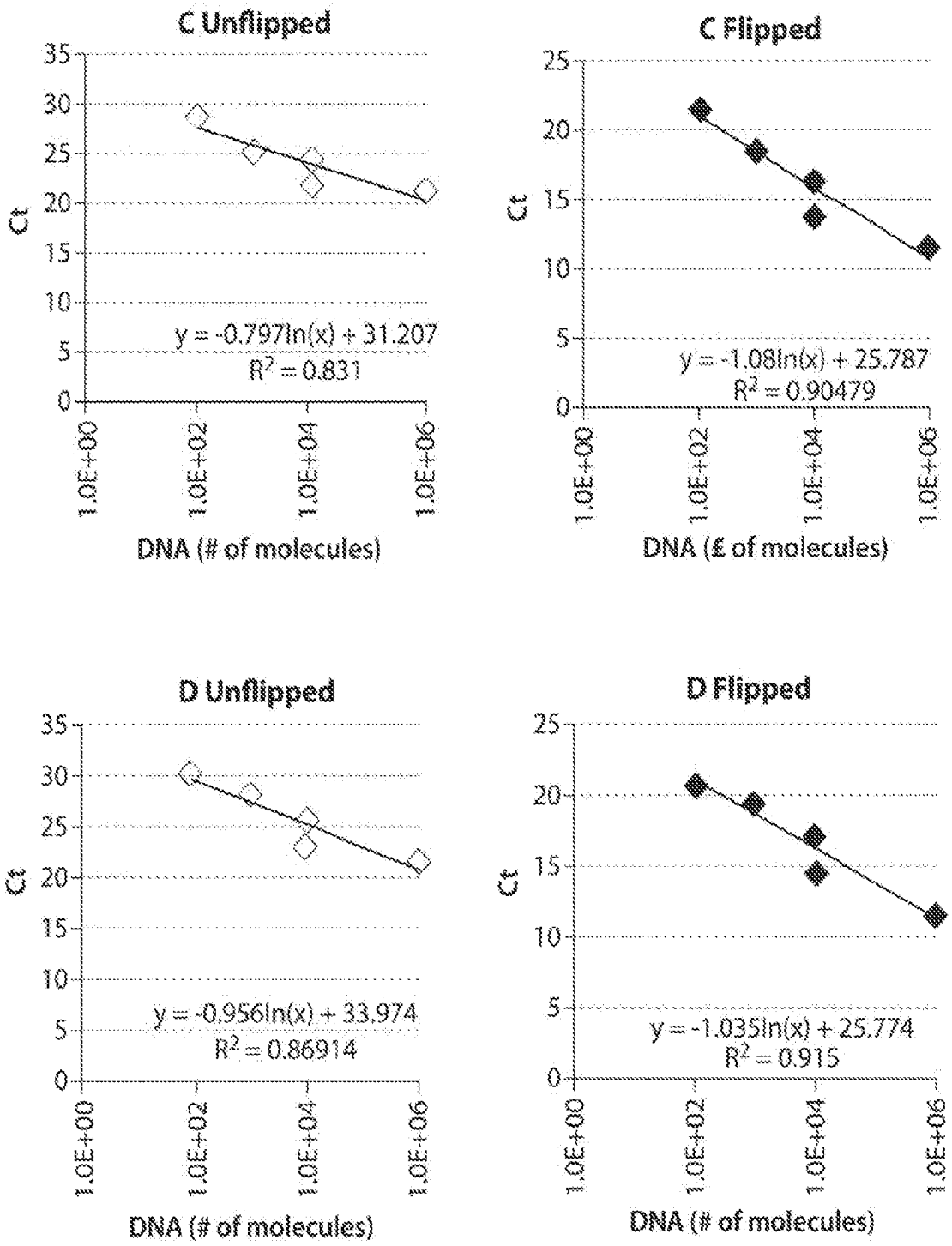

C Unflipped
Ct
DNA (# of molecules)
y = -0.797ln(x) + 31.207
R2 = 0.831
C Flipped
Ct
DNA (£ of molecules)
y = -1.08ln(x) + 25.787
R2 = 0.90479
D Unflipped
Ct
DNA (# of molecules)
y = -0.956ln(x) + 33.974
R2 = 0.86914
D Flipped
Ct
DNA (# of molecules)
y = -1.035ln(x) + 25.774
R2 = 0.915

Fig. 39

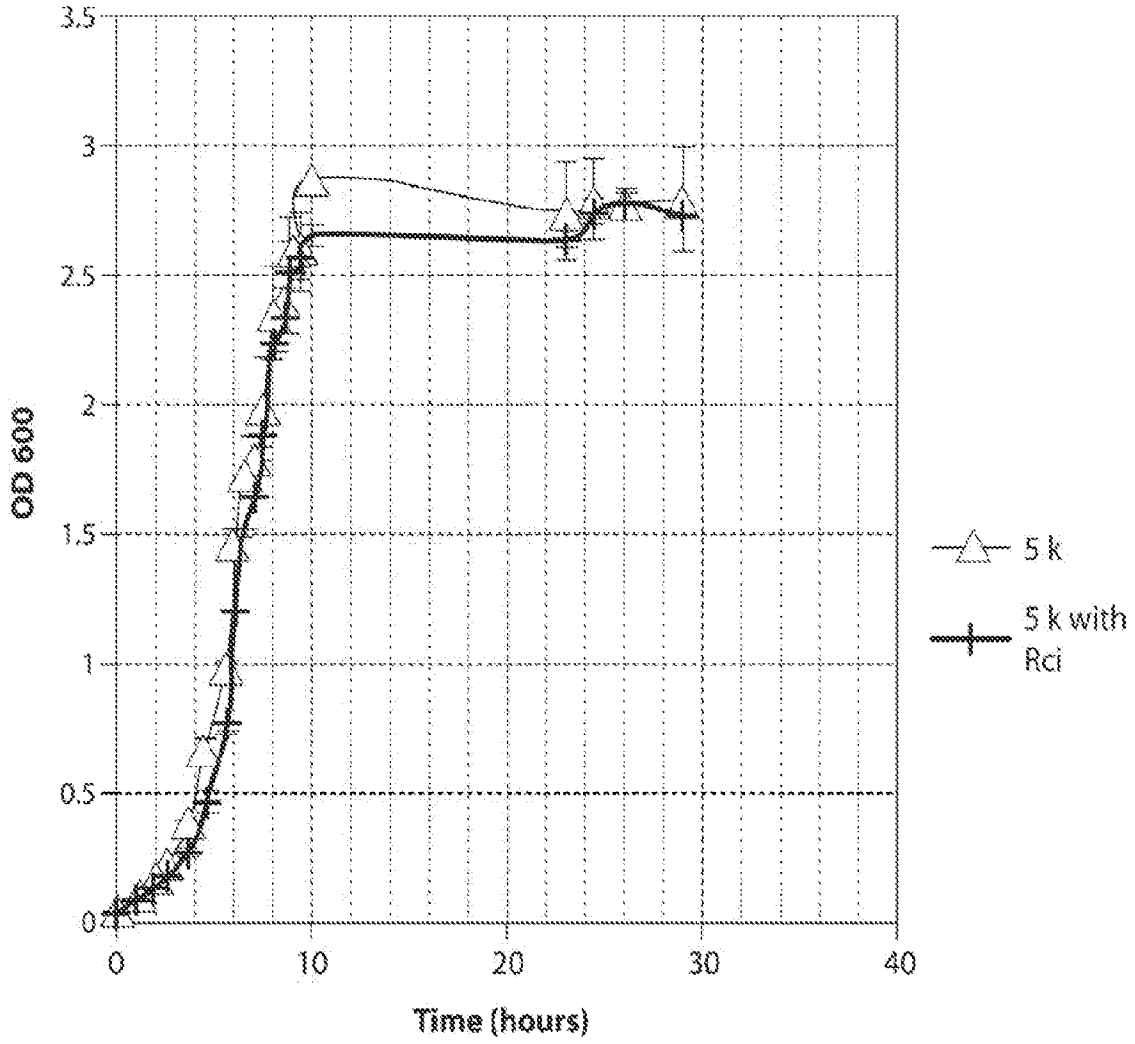


Fig. 40

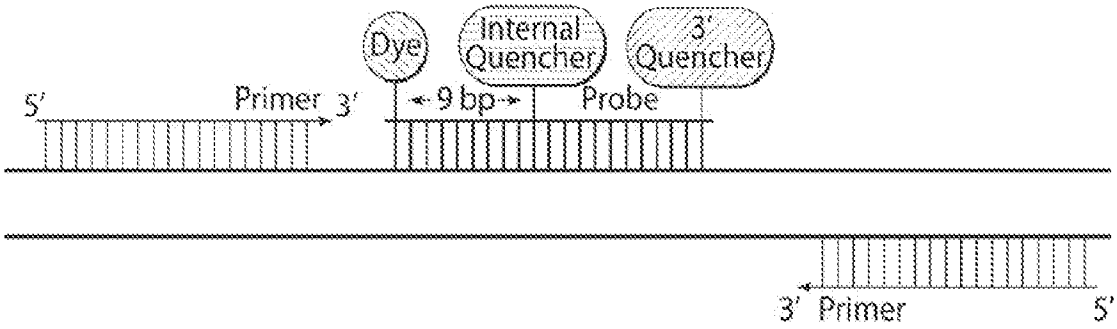

Dye
Internal Quencher
3' Quencher
5'
Primer 3'
9 bp
Probe
3' Primer 5'

Fig. 41

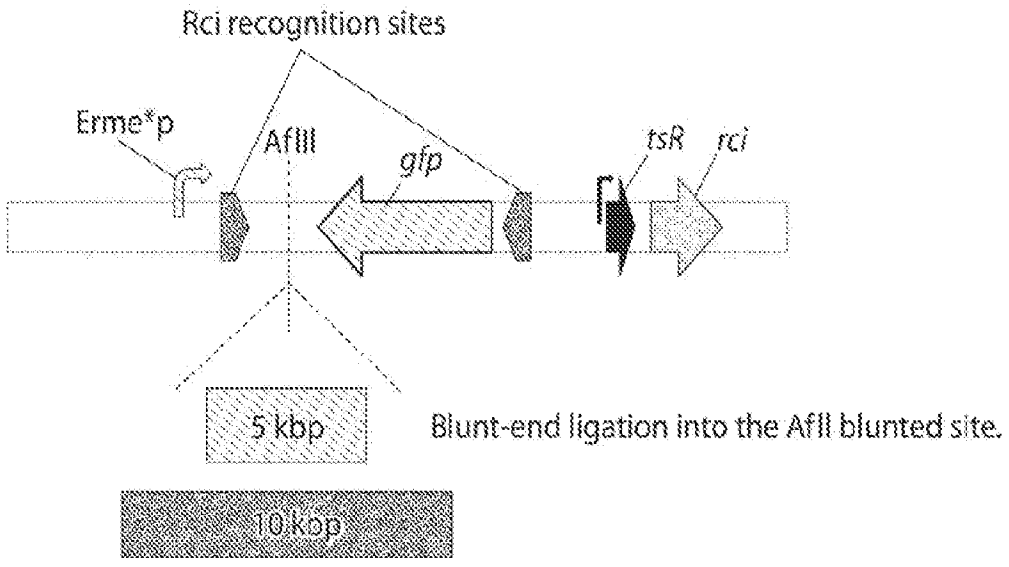

Rci recognition sites
Erme*p
AflII
gfp
tsR
rci
5 kbp
Blunt-end ligation into the AflI blunted site.
10 kbp

Fig. 42

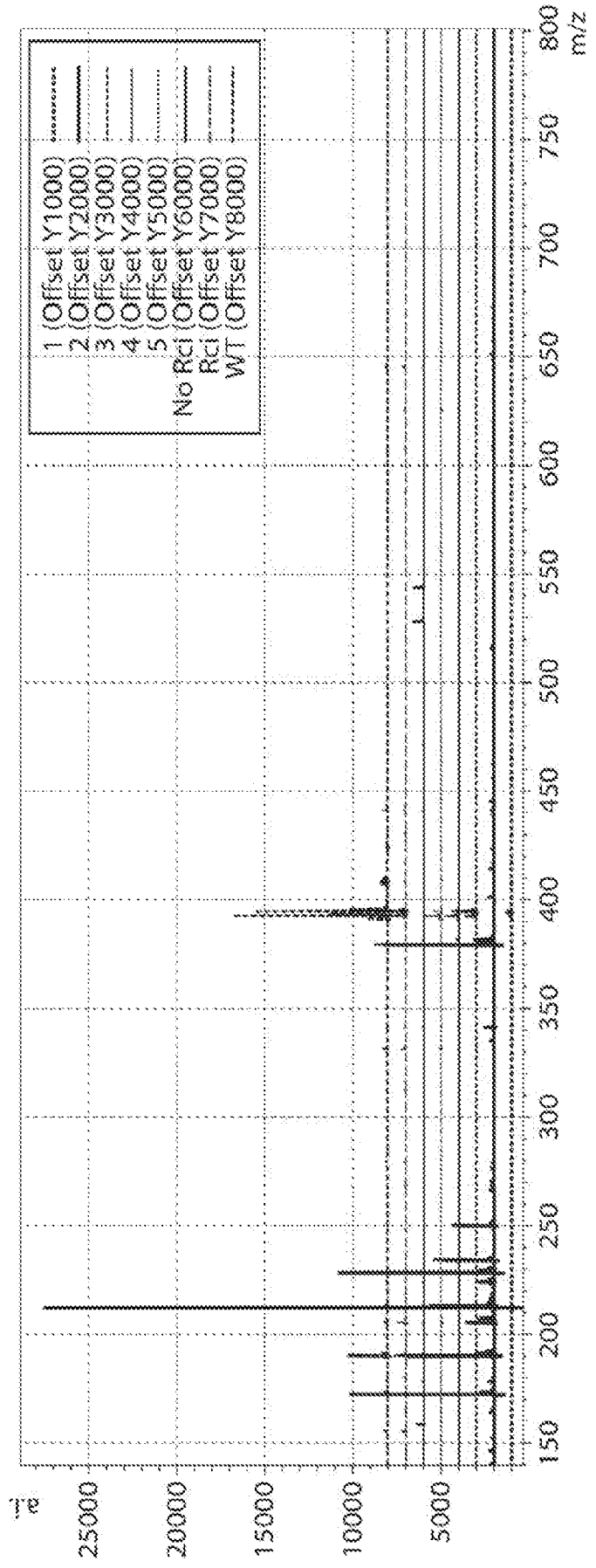

1 (Offset Y1000)
2 (Offset Y2000)
3 (Offset Y3000)
4 (Offset Y4000)
5 (Offset Y5000)
No Rci (Offset Y6000)
Rci (Offset Y7000)
WT (Offset Y8000)
m/z
150
200
250
300
350
400
450
500
550
600
650
700
750
800
5000
10000
15000
20000
25000

Fig. 43

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/50533

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☒ forming part of the international application as filed:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:
GenCore ver 6.4.1 SEQ ID NOs: 5-8, 35, 36

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/50533

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.: 68-70
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:

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 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.:

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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 16/50533

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/70, 15/76, 15/31, 1/21; C12Q 1/68 (2016.01)
CPC - C12N 15/70, 15/76; 15/63, 2800/30
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C12N 15/70, 15/76, 15/31, 1/21; C12Q 1/68 (2016.01)
CPC: C12N 15/70, 15/76, 15/63, 2800/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC: C12N 15/70, 15/76, 15/63, 2800/30 (text search)
USPC: 435/471, 252.33, 252.35, 252.8, 6.1; 536/23.2 (text search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Electronic data bases: PatBase; Google Patents; Google Scholar
Search terms: Recombinase, shufflon recombinase Rci protein (e.g. from Salmonella enterica, E. coli), shufflon recombinase recognition sequences (RRSs), R64 plasmid, genes of biosynthetic pathway, inversion systems, biosynthetic gene cluster

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y --- A	GYOHDA et al. Purification and characterization of the R64 shufflon-specific recombinase. J Bacteriol. May 2000 Vol 182 No 10 Pages 2787-2792. Especially pg 2787 col 1 para 3, pg 2787 col 2 para 3, pg 2788 col 2 para 3, pg 2788 fig 1B,C, pg 2789 col 1 para 4-5, pg 2790 col 1 para 1.	71-75 --- 1-67
Y	US 2012/0213764 A1 (NAUMER et al.) 23 August 2012 (23.08.2012). Especially para [0001], [0017], [0098].	71-75
A	US 2003/0073824 A1 (Sherman et al.) 17 April 2003 (17.04.2003). Especially claims 1, 3, 5; sheet 34 fig 24	1-67
A	PEIKON et al. In vivo generation of DNA sequence diversity for cellular barcoding. July 10 2014. Nucleic Acids Res Vol 42 No 16 Pages e127 1-10. Especially pg 5 fig 2A, B.	1-67
A	Gyohda et al. Structure and function of the shufflon in plasmid R64. Adv Biophys 2004 Vol 38 Pages 183-213. entire article.	1-67

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

* 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	Date of mailing of the international search report
10 November 2016	01DEC 2016
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774