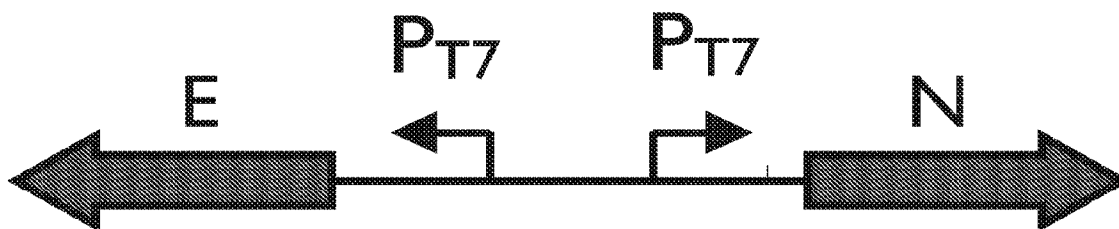




(86) Date de dépôt PCT/PCT Filing Date: 2012/06/14
 (87) Date publication PCT/PCT Publication Date: 2012/12/20
 (45) Date de délivrance/Issue Date: 2023/10/24
 (85) Entrée phase nationale/National Entry: 2013/12/10
 (86) N° demande PCT/PCT Application No.: US 2012/042502
 (87) N° publication PCT/PCT Publication No.: 2012/174271
 (30) Priorité/Priority: 2011/06/16 (US61/497,781)

(51) Cl.Int./Int.Cl. *C12N 15/09* (2006.01),
C12N 15/31 (2006.01), *C12N 15/63* (2006.01)
 (72) Inventeurs/Inventors:
MIRSKY, ETHAN, US;
TEMME, KARSTEN, US;
VOIGT, CHRIS, US;
ZHAO, DEHUA, US
 (73) Propriétaire/Owner:
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA,
US
 (74) Agent: SMART & BIGGAR LP

(54) Titre : FAMILLES MULTIGENIQUES DE SYNTHÈSE
 (54) Title: SYNTHETIC GENE CLUSTERS



(57) Abrégé/Abstract:

Methods, systems and tangible computer readable media are provided for making synthetic gene clusters. The method involves replacing native regulation of genes with synthetic regulation by changing codons with non-native codons having maximal distance from codons of a native coding sequence and organizing the coding sequences into one or more synthetic operons and operably linking one or more heterologous transcriptional regulatory sequences to the operons.

ABSTRACT

Methods, systems and tangible computer readable media are provided for making synthetic gene clusters. The method involves replacing native regulation of genes with synthetic regulation by changing codons with non-native codons having maximal distance from
5 codons of a native coding sequence and organizing the coding sequences into one or more synthetic operons and operably linking one or more heterologous transcriptional regulatory sequences to the operons.

SYNTHETIC GENE CLUSTERS

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

5 [0001] The present application claims priority to US Patent Application No. 61/497,781, filed June 16, 2011.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

10 [0002] This invention was made with United States government support under grant nos. CFF0943385 and EEC0540879 awarded by the National Science Foundation. The United States government has certain rights in the invention.

BACKGROUND OF THE INVENTION

15 [0003] Genetically programming cells require sensors to receive information, circuits to process the inputs, and actuators to link the circuit output to a cellular response (Andrianantoandro E, *et al.*, *Mol Syst Biol* 2 (2006); Chin JW *Curr Opin Struct Biol* 16: 551–556 (2006); Voigt CA *Curr Opin Biotech* 17: 548–557 (2006); Tan C, *Mol Biosyst* 3: 343–353 (2007)). In this paradigm, sensing, signal integration, and actuation are encoded by distinct ‘devices’ comprised of genes and regulatory elements (Knight TK, Sussman GJ
20 *Unconventional Models of Computation* 257–272 (1997); Endy D *Nature* 438: 449–453 (2005)). These devices communicate with one another through changes in gene expression and activity. For example, when a sensor is stimulated, this may lead to the activation of a promoter, which then acts as the input to a circuit.

25

BRIEF SUMMARY OF THE INVENTION

[0004] Embodiments of the present invention provide a polynucleotide comprising a synthetic operon, wherein the operon comprises at least two coding sequences under the

control of a heterologous transcriptional regulatory sequence, wherein each coding sequence is operably linked to a heterologous ribosome binding site (RBS). In some embodiments, the coding sequences are from the same native operon and the heterologous RBSs regulate translation of the coding sequences in a ratio that is substantially similar to the ratio of native translation from the native operon. In some embodiments, the coding sequences are from different native operons and the heterologous RBSs regulate translation of the coding sequences in a ratio that is substantially similar to the ratio of native translation from the native operon. In some embodiments, the coding sequences are from the same native operon and the coding sequences in the operon comprise one or more altered codon compared to the native operon. In some embodiments, codons of one or more coding sequence have been selected for maximal distance from codon usage of a corresponding coding sequence in the native operon.

[0005] In some embodiments, at least two coding sequences encode different proteins encoded by the *Klebsiella pneumoniae* nif gene cluster. In some embodiments, the proteins are selected from the group consisting of nifJ, nifH, nifD, nifK, nifY, nifE, nifN, nifU, nifS, nifV, nifW, nifZ, nifM, nifF, nifB, and nifQ (e.g., wherein the coding sequences are substantially identical to those listed in Figure 18). In some embodiments, the operon comprises coding sequences for *Klebsiella pneumoniae* nifH, nifD, nifK, and nifY. In some embodiments, the operon comprises coding sequences for *Klebsiella pneumoniae* nifE and nifN. In some embodiments, the operon comprises coding sequences for *Klebsiella pneumoniae* nifU, nifS, nifV, nifW, nifZ, and nifM. In some embodiments, the operon comprises coding sequences for *Klebsiella pneumoniae* nifB and nifQ.

[0006] In some embodiments, at least two coding sequences encode different proteins of the *Salmonella Typhimurium* Type III secretion system. In some embodiments, the proteins are selected from the group consisting of PrgH, PrgI, PrgJ, PrgK, OrgA, OrgB, InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS (e.g., wherein the coding sequences are substantially identical to those listed in Figure 24). In some embodiments, the operon comprises coding sequences for *Salmonella Typhimurium* PrgH, PrgI, PrgJ, PrgK, OrgA, and OrgB. In some embodiments, the operon comprises coding sequences for *Salmonella Typhimurium* InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS.

[0007] Embodiments of the present invention also provide for a host cell (optionally isolated) comprising a polynucleotide as described above or elsewhere herein. In some

embodiments, the host cell is a prokaryotic or eukaryotic cell (including but not limited to a mammalian or plant or fungal cell).

[0008] Embodiments of the present invention also provide a system comprising a set of two or more different synthetic operons, the two or more operons each comprising at least two coding sequences under the control of a heterologous transcriptional regulatory sequence, wherein each coding sequence is operably linked to a heterologous ribosome binding site (RBS), wherein the transcriptional regulatory sequence of each operon in the set is controlled by the same transcriptional activator or repressor polypeptide(s).

[0009] In some embodiments, the system further comprises an expression cassette comprising a promoter operably linked to a polynucleotide encoding the transcriptional activator or repressor polypeptide(s). In some embodiments, the promoter of the expression cassette is an inducible promoter. In some embodiments, the polynucleotide in the expression cassette encodes a transcriptional repressor. In some embodiments, the polynucleotide in the expression cassette encodes a transcriptional activator. In some embodiments, the transcriptional activator is an RNA polymerase (RNAP). In some embodiments, the RNAP is T7 RNAP or is substantially similar to T7 RNAP.

[0010] In some embodiments, the transcriptional regulatory sequences of at least two of the operons are different.

[0011] In some embodiments, the coding sequences in the operons are organized such that coding sequences having substantially similar native expression are grouped into the same operon. In some embodiments, the transcriptional regulatory sequence of at least two operons have different promoters that are differentially regulated by T7 RNA polymerase and wherein the different strength of the promoters correspond to the relative strength of native promoters of the coding sequences.

[0012] In some embodiments, the expression cassette and the synthetic operons are expressed in a cell. In some embodiments, the cell is from a different species than the species from which the native operon was isolated. In some embodiments, the cell is from the same species from which the native operon was isolated.

[0013] In some embodiments, the system encodes a nitrogenase. In some embodiments, the system comprises a first operon comprising coding sequences for *Klebsiella pneumoniae* nifH, nifD, nifK, and nifY; a second operon comprising coding sequences for *Klebsiella pneumoniae* nifE and nifN; a third operon comprising coding sequences for *Klebsiella*

pneumonia nifU, nifS, nifV, nifW, nifZ, and nifM; and a fourth operon comprising coding sequences for *Klebsiella pneumoniae* nifB and nifQ. In some embodiments, the first, second, third, and fourth operon comprising a T7 RNA polymerase (RNAP) promoter and the system further comprises an expression cassette comprising a promoter operably linked to a
5 polynucleotide encoding an RNAP substantially identical to T7 RNA polymerase (RNAP).

[0014] In some embodiments, the system encodes a type III secretion system. In some embodiments, the type III secretion system is a *Salmonella Typhimurium* type III secretion system. In some embodiments, the system comprises a first operon comprising coding sequences for *Salmonella Typhimurium* PrgH, PrgI, PrgJ, PrgK, OrgA, and OrgB and a
10 second operon comprising coding sequences for *Salmonella Typhimurium* InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS.

[0015] Embodiments of the present invention also provide a method for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation. In some embodiments, the method comprises providing coding sequences for a set of
15 polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence; organizing the coding sequences into one or more synthetic operon(s); operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling the magnitude of gene expression from the
20 operon(s); and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

[0016] In some embodiments, the polypeptide is heterologous to the cell.

[0017] In some embodiments, the providing comprises obtaining the gene nucleotide
25 sequences and eliminating non-coding sequences.

[0018] In some embodiments, the set of genes is from a gene cluster. In some embodiments, the set of genes are from a prokaryote. In some embodiments, the genes are from a native operon.

[0019] In some embodiments, the at least one regulatory sequence is identified using
30 computation. In some embodiments, the computation comprises searches of coding sequences for ribosome binding sites, terminators, and/or promoters.

[0020] In some embodiments, removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons. In some embodiments, the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence. In some embodiments, the removing
5 comprises selecting non-native codons for optimal expression in a host cell.

[0021] In some embodiments, the method further comprises identifying and removing one or more of transposon insertion sites, sites that promote recombination, sites for cleavage by restriction endonucleases, and sites that are methylated.

[0022] In some embodiments, the organizing comprises grouping coding sequences into
10 operons based on substantially similar native expression level.

[0023] In some embodiments, the organizing comprises ordering coding sequences within operons such that the highest expressing gene (based on native expression) occurs first and the lowest expressing gene (based on native expression) occurs last. In some embodiments, organization is based on native temporal expression, function, ease of manipulation of DNA,
15 and/or experimental design. In some embodiments, magnitude of expression of coding sequences substantially correspond to the ratio of proteins encoded by the coding sequences as measured in the native system. In some embodiments, magnitude of expression of coding sequences is determined by computation. In some embodiments, the computation comprises a numerical optimization algorithm.

[0024] In some embodiments, the numerical optimization algorithm a Nelder-Mead
20 algorithm, a Newton's method, a quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, a ellipsoid method, a Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model.

[0025] In some embodiments, the heterologous transcriptional regulatory sequence(s)
25 comprise a T7 RNAP promoter(s).

[0026] In some embodiments, the heterologous transcriptional regulatory sequence(s) comprise an inducible promoter.

[0027] In some embodiments, the method further comprises operably linking a
heterologous ribosomal binding site (RBS) to one or more coding sequence in the synthetic
30 operon. In some embodiments, different RBSs are operably linked to different coding sequences. In some embodiments, the RBSs regulate translation of the coding sequences in a ratio that is substantially similar to the ratio of native translation from the native operon.

[0028] In some embodiments, the method further comprises operably linking a heterologous transcriptional terminator sequence to one or more coding sequence in the synthetic operon. In some embodiments, the terminators are T7 RNAP terminators. In some embodiments, terminators for different operons are different.

5 [0029] In some embodiments, the method further comprises operably linking a buffer sequences between two functional sequences in an operon wherein the functional sequences are selected from the group consisting of a promoter, ribosome binding site, coding sequence, and terminator. In some embodiments, the buffer sequence is selected from the group consisting of a random sequence, a UP-region of a promoter, an extended 5-UTR sequence,
10 and a RNAase cleavage site.

[0030] In some embodiments, the operons are expressed from a plasmid. In some embodiments, the plasmid has a low copy origin of replication.

[0031] In some embodiments, the polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence is expressed from a control expression
15 cassette, the expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide. In some embodiments, the expression cassette is contained in a control plasmid separate from a plasmid containing the operons. In some embodiments, the control promoter is an inducible promoter.

[0032] In some embodiments, the heterologous polypeptide comprises an RNA polymerase
20 (RNAP). In some embodiments, the RNAP is T7 RNAP. In some embodiments, the expression cassette is an environmental sensor.

[0033] Embodiments of the invention also provide for a method for determining an experimentation point for controlling the magnitude of expression of two or more genes (e.g.,
25 within a synthetic operon). In some embodiments, the method comprises: receiving one or more input data points, wherein the input data points provide information about one or more regulatory elements and a system property; and determining, with a computer, a next data point using a computational method, wherein the next data point provides information about the one or more regulatory elements.

[0034] In some embodiments, the method further comprises using the next data point for
30 further experimentation to optimize expression of the two or more genes. In some embodiments, the regulatory elements include, e.g., ribosomal binding sites and/or transcriptional regulatory elements.

[0035] In some embodiments, the computational method is a numerical analysis technique. In some embodiments, the numerical optimization method is the Nelder-Mead algorithm, the Newton's method, the quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, a ellipsoid method, the Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model. In some embodiments, the numerical optimization method used to determine the next data point for further experimentation requires considering the reflection point, expansion point, or contraction point based on the one or more input data points.

[0036] In some embodiments, the computational method is a design of experiments (DoE) method.

[0037] Embodiments of the invention also provide for a computer program product comprising a tangible computer readable medium storing a plurality of instructions for controlling a processor to perform an operation for determining an experimentation point for controlling the magnitude of expression of two or more genes, the instructions comprising receiving one or more input data points, wherein the input data points provide information about one or more regulatory elements and a system property; and determining, with a computer, a next data point using a computational method, wherein the next data point provides information about the one or more regulatory elements.

[037A] The invention disclosed and claimed herein pertains to a method for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the method comprising: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence, wherein the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence; organizing the coding sequences into one or more synthetic operon(s); operably linking one or more heterologous transcriptional regulatory sequence(s) to the operon(s), thereby controlling the magnitude of gene expression from the operon(s); and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

[037B] The invention disclosed and claimed herein also pertains to a polynucleotide comprising a synthetic operon, wherein the operon comprises at least two coding sequences under the control of a heterologous transcriptional regulatory sequence, wherein each coding sequence is operably linked to a heterologous ribosome binding site (RBS), and wherein codons of one or more coding sequence have been selected for maximal distance from codon usage of the corresponding coding sequence in a native operon thereby removing at least one regulatory sequence within the coding sequence. Also disclosed and claimed is an isolated host cell comprising such a polynucleotide. Also disclosed and claimed is a system comprising a set of two or more different such synthetic operons, wherein the transcriptional regulatory sequence of each operon in the set is controlled by the same transcriptional activator or repressor polypeptide(s).

[037C] The invention disclosed and claimed herein also pertains to a method for determining an experimentation point for controlling the magnitude of expression of two or more genes, the method comprising: receiving one or more input data points, wherein the input data points provide information about one or more regulatory elements and a system property; and causing a processor to perform an operation to determine a next data point according to a plurality of instructions stored in a computer readable medium, wherein the next data point is the experimentation point.

[037D] The invention disclosed and claimed herein also pertains to a method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence; organizing the coding sequences into one or more synthetic operon(s); operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling magnitude of gene expression from the operon(s); and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence, wherein the polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence is expressed

from a control expression cassette, the expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.

5 [037E] An aspect of the disclosure also pertains to a tangible computer readable medium or a computer program product comprising a tangible computer readable medium storing a plurality of instructions for controlling a processor to perform an operation for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the instructions comprising instructions for the steps of a method as disclosed herein.

10 [037F] An aspect of the disclosure also pertains to a computer program product comprising a computer readable memory storing computer executable instructions thereon that when executed by a computer perform the steps of: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence, wherein
15 the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence; organizing the coding sequences into one or more synthetic operon(s); and operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling the magnitude of gene expression from the operon(s); wherein the one or more synthetic operon(s) are configured to be expressed in a cell under the control of a polypeptide that
20 binds directly or indirectly to the heterologous transcriptional regulatory sequence.

[037G] Various embodiments of the claimed invention relate to a method for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the method comprising: causing at least one processor to provide coding sequences for a set of
25 polypeptides encoded by genes collectively associated with a function; causing the at least one processor to change codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the at least one coding sequence, wherein the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence; causing the at least one processor to organize the coding sequences into one or more synthetic
30 operon(s); and causing the at least one processor to operably link one or more heterologous transcriptional regulatory sequence to the one or more synthetic operon(s), thereby controlling magnitude of gene expression from the one or more synthetic operon(s); expressing the one or more

synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the one or more heterologous transcriptional regulatory sequence.

5 [037H] An aspect of the disclosure also pertains to a computer program product comprising a tangible computer readable medium storing computer executable instructions thereon that when executed by a computer perform an operation for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the computer executable instructions comprising instructions for the steps of the method as described herein.

10 [037I] The invention disclosed and claimed herein also pertains to a method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence by removing at least one regulatory
15 sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence; organizing the coding sequences into one or more synthetic operon(s); operably linking one or more heterologous transcriptional regulatory sequence to the operon(s),
20 thereby controlling magnitude of gene expression from the operon(s); and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence; and detecting the magnitude of gene expression by computation, wherein the computation comprises a numerical optimization algorithm, and wherein the numerical optimization algorithm comprises the Nelder-Mead
25 algorithm, the Newton's method, the quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, a ellipsoid method, the Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model.

30 [037J] The invention disclosed and claimed herein also pertains to a method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

changing codon identity within at least one coding sequence by removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence; organizing the coding sequences into one or more synthetic operon(s);
5 operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling magnitude of gene expression from the operon(s); operably linking a heterologous ribosomal binding site (RBS) to one or more coding sequence in the synthetic operon, wherein different RBSs are operably linked to different coding sequences, and wherein the RBSs
10 regulate translation of the coding sequences in a ratio that is similar to a ratio of translation from a native operon, and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

[037K] The invention disclosed and claimed herein also pertains to a method for expressing one
15 or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising: providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function; changing codon identity within at least one coding sequence by removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence
20 comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence; organizing the coding sequences into one or more synthetic operon(s), wherein the synthetic operon comprises two functional sequences selected from the group consisting of a promoter, a ribosome binding site, a coding sequence, and a terminator and the
25 method further comprises operably linking a buffer sequence between two functional sequences, and wherein the buffer sequence is selected from the group consisting of a random sequence, a UP-region of a promoter, an extended 5-UTR sequence, and a RNAase cleavage site; and expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

30 **[037L]** The invention disclosed and claimed herein also pertains to a method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising:

providing a bacterial cell for expressing gene products; providing a gene cluster having a plurality of native bacterial genes having coding sequences; modifying the gene cluster by making at least one modification in at least one location within the gene cluster selected from the group consisting of a coding region and an intergenic region, wherein the gene cluster modification comprises

5 replacing at least one native codon within one of the coding sequences to remove at least one native regulatory sequence using a synonymous codon and wherein the synonymous codon is a maximal distance from a corresponding native codon; operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster; and expressing gene products of the modified gene cluster in the bacterial cell under the

10 control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

[037M] The invention disclosed and claimed herein also pertains to a bacterial nitrogen reduction expression system comprising nucleic acids encoding: at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen

15 fixation within a cell, wherein at least one of the plurality of coding sequences comprises non-native codons in place of a regulatory element, wherein said non-native codons have maximal distance from codons of the native coding sequence; a heterologous promoter region that directs expression of the at least one operon; and a heterologous transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system,

20 wherein the protein binds directly or indirectly to the heterologous promoter region.

[037N] The invention disclosed and claimed herein also pertains to a bacterial nitrogen reduction expression system comprising nucleic acids encoding: at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen

25 fixation within a cell, wherein at least one of the plurality of coding sequences comprises a non-native synonymous codon in place of a native codon, thereby removing a regulatory sequence a non-native synonymous codon in place of a native codon, thereby removing a regulatory sequence; a heterologous promoter region that directs expression of the at least one operon, wherein the heterologous promoter region is from the same species as the genes collectively associated with

30 nitrogen fixation; and a transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system, wherein the protein binds directly or

indirectly to the heterologous promoter region, and wherein the transcriptional controller is not the native transcription controller of the genes collectively associated with nitrogen fixation under native regulation.

5 [037O] The invention disclosed and claimed herein also pertains to a bacterial nitrogen reduction expression system comprising nucleic acids encoding: at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen fixation within a cell, wherein at least one of the plurality of regulatory coding sequences has been synonymously mutated to remove internal regulation, and wherein at least one coding sequence has been modified to reduce a predicted RNA secondary structure; a genetically engineered promoter region that directs expression of the at least one operon; and a transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system, wherein the protein binds directly or indirectly to the heterologous promoter region, and wherein the transcriptional controller does not regulate the genes collectively associated with nitrogen fixation under native regulation.

15 [037P] Various embodiments of the claimed invention also relate to a method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising: providing a bacterial cell for expressing gene products; providing a gene cluster having a plurality of native bacterial genes having coding sequences; modifying the gene cluster by making at least one modification in a coding region or an intergenic region, wherein making the at least one modification in the coding region or the intergenic region comprises replacing at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster wherein the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes; and expressing gene products of the modified gene cluster in the bacterial cell under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

30 [0037Q] Various embodiments of the claimed invention also relate to a method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising: providing a bacterial cell for expressing gene products; providing a gene cluster having a plurality

of native bacterial genes having coding sequences; modifying the gene cluster by making at least one modification in a coding region or an intergenic region, wherein making the at least one modification in the coding region or the intergenic region comprises replacing at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster, wherein the at least one heterologous transcriptional regulatory sequence is from a different species than the plurality of native bacterial genes; and expressing gene products of the modified gene cluster in the bacterial cell under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

[037R] Various embodiments of the claimed invention also relate to a recombinant bacterial cell comprising a modified gene cluster, wherein the modified gene cluster comprises a plurality of native bacterial genes having coding sequences and comprises at least one modification in a coding region or an intergenic region, wherein the at least one modification in the coding region or the intergenic region comprises a replacement of at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; wherein at least one coding sequence within the modified gene cluster is operably linked to at least one heterologous transcriptional regulatory sequence; wherein the at least one heterologous transcriptional regulatory sequence is from a different species than the plurality of native bacterial genes; and wherein the expression of gene products of the modified gene cluster in the bacterial cell is under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

25

DEFINITIONS

[0038] A recitation of "a", "an" or "the" is intended to mean "one or more" unless specifically indicated to the contrary.

[0039] A polynucleotide or polypeptide sequence is "heterologous to" an organism or a second sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if

30

from the same species, a coding sequence which is not naturally associated with the promoter (*e.g.* a T7 RNA polymerase promoter operably linked to a synthetic *nif* operon).

[0040] The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence. In the context of a

5

ribosomal binding site (RBS) and coding sequences, the term refers to the functional linkage of the RBS to the coding sequence wherein the RBS recruits ribosomes for translation of the coding sequence on an RNA.

[0041] A “cognate pair” as used herein refers to a sequence-specific DNA binding polypeptide and a target DNA sequence that is bound by the particular sequence-specific DNA binding polypeptide. For sequence-specific DNA binding polypeptides that bind more than one target nucleic acid, the cognate pair can be formed with the sequence-specific DNA binding polypeptide and any one of the target DNA sequences the polypeptide binds.

[0042] “Orthogonal” transcriptional systems refer to systems (e.g., one, two, three, or more) of transcriptional regulatory elements comprising target DNA sequences regulated by their cognate sequence-specific DNA binding polypeptide such that the sequence-specific DNA binding polypeptides in the system do not have “cross-talk,” i.e., the sequence-specific DNA binding polypeptides do not interfere or regulate transcriptional regulatory elements in the system other than the transcriptional regulatory elements containing the cognate target DNA sequence of the sequence-specific DNA binding polypeptide.

[0043] “Sequence-specific DNA binding polypeptides” refer to polypeptides that bind DNA in a nucleotide sequence specific manner. Exemplary sequence-specific DNA binding polypeptides include, but are not limited to transcription factors (e.g., transcriptional activators), RNA polymerases, and transcriptional repressors.

[0044] A “transcriptional activator” refers to a polypeptide, which when bound to a promoter sequence, activates or increases transcription of an RNA comprising the operably-linked coding sequence. In some embodiments, the transcriptional activator bound to a target sequence in a promoter can assist recruitment of RNA polymerase to the promoter. A “transcriptional repressor” refers to a polypeptide, which when bound to a promoter sequence, blocks or decreases transcription of an RNA comprising the operably-linked coding sequence. In some embodiments, the transcriptional repressor blocks recruitment of the RNA polymerase to the promoter or blocks the RNA polymerase’s movement along the promoter.

[0045] The term “coding sequence” as used herein refers to a nucleotide sequence beginning at the codon for the first amino acid of an encoded protein and ending with the codon for the last amino acid and/or ending in a stop codon.

[0046] The term "host cell" refers to any cell capable of replicating and/or transcribing and/or translating a heterologous gene. Thus, a "host cell" refers to any prokaryotic cell (including but not limited to *E. coli*) or eukaryotic cell (including but not limited to yeast cells, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells), whether located in vitro or in vivo. For example, host cells may be located in a transgenic animal or transgenic plant. prokaryotic cell (including but not limited to *E. coli*) or eukaryotic cells (including but not limited to yeast cells, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells).

[0047] "Transcriptional regulatory elements" refer to any nucleotide sequence that influences transcription initiation and rate, or stability and/or mobility of a transcript product. Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start sites, termination sequences, polyadenylation sequences, introns, etc. Such transcriptional regulatory sequences can be located either 5'-, 3'-, or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription.

[0048] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

[0049] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid

chains of any length, including full-length proteins, wherein the amino acid residues are linked by covalent peptide bonds.

[0050] Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The term "complementary to" is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

[0051] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997), and Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available on the Web through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short wordlength (W) in the query sequence, which either match or satisfy some positive-valued threshold score (T) when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915, (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[0052] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin and Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787, (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between
5 two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0053] "Percentage of sequence identity" is determined by comparing two optimally
10 aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in
15 both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

[0054] The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 25% sequence identity to a designated
20 reference sequence. Alternatively, percent identity can be any integer from 25% to 100%, for example, at least: 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below. One of skill will recognize that the percent identity values above can be appropriately adjusted to determine
25 corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 40%. Percent identity of polypeptides can be any integer from 40% to 100%, for example, at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%,
30 95%, or 99%. In some embodiments, polypeptides that are "substantially similar" share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group

of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Exemplary conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, aspartic acid-glutamic acid, and asparagine-glutamine.

10

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] Fig. 1 depicts a scheme illustrating *nifE* and *nifN* genes under the control of unique T7 promoters.

[0056] Fig. 2 illustrates the quantitative measurement of the capacity of the synthetic operon to complement a *nifEN* knockout strain and recover the ability to fix nitrogen.

15 [0057] Fig. 3 illustrates a library of wild-type and mutant T7 promoters and their strength to control gene expression.

[0058] Fig. 4 illustrates strengths of T7 promoters to control *nifE* and *nifN* genes in selected mutant strains. Fig. 4A depicts strengths of three strains. Fig. 4B depicts the calculated Reflection coordinates.

20 [0059] Fig. 5 illustrates the nitrogen fixation in the Reflection strain and the initial strains.

[0060] Fig. 6 illustrates the method of refactoring nitrogen fixation.

[0061] Fig. 7 illustrates the *nif* gene cluster from *Klebsiella oxytoca*.

[0062] Fig. 8 depicts a scheme of a fluorescent reporter plasmid in which the 150bp surrounding a gene's start codon (from -60 to +90) was fused to the mRFP gene and
25 expressed under the control of the Ptac promoter.

[0063] Fig. 9 illustrates the measured fluorescence by flow cytometry.

[0064] Fig. 10 illustrates the multiple clones used to identify the synthetic ribosome binding site that best matched the native ribosome binding site.

[0065] Fig. 11 illustrates the chimeric operons.

- [0066] Fig. 12 lists the errors in the fully synthetic operons.
- [0067] Fig. 13 shows that each synthetic operon required different levels of IPTG concentration for optimal function. It also shows the performance of individual operons in the T7 Wires system under Ptac promoter control.
- 5 [0068] Fig. 14 shows a table of the control of the synthetic operons in the system.
- [0069] Fig. 15 shows nitrogen fixation from a full synthetic cluster expressed in a complete nif knockout strain.
- [0070] Fig. 16 illustrates the use of either controller #1 or controller #2 to produce the same performance from the full synthetic cluster.
- 10 [0071] Fig. 17 depicts a detailed schematic of the full synthetic cluster.
- [0072] Fig. 18 shows DNA sequences for native genes and synthetic genes, as well as the percent common nucleotide and codon identities between each pair.
- [0073] Fig. 19 shows the names and sequences of parts of the synthetic controller.
- [0074] Fig. 20 lists the names, sequences and strengths of each components of the full
15 cluster.
- [0075] Fig. 21 shows a diagram of the RBS test vector.
- [0076] Fig. 22 depicts schematics of the *inv-spa* and *prg-org* operons and the plasmids used. Fig. 22A shows a schematic of Δ *prg-org* *Salmonella* SL1344 knock-out strain. The *inv-spa* and *prg-org* operons are boxed. Fig. 22B shows a schematic of the *prg-org* operon test vector and reporter plasmid. The control plasmid and reporter plasmid are on the right.
20
- [0077] Fig. 23 shows a western blot of secreted protein expressed from the synthetic *prg-org* operon in Δ *prg-org* knock-out strain. Fig. 23A shows that the Δ *prg-org* knock-out strain does not express the *prg-org* operon. Fig. 23B shows that the synthetic refactored *prg-org* operon in *Salmonella* Δ *prg-org* cells can be controlled by the addition of IPTG.
- 25 [0078] Fig. 24 shows the synthetic RBS and synthetic operon sequences of the T3SS.
- [0079] Fig. 25 shows a block diagram of a computer system.
- [0080] Fig. 26 illustrates the process of refactoring a gene cluster. The wild-type *K. oxytoca* nitrogen fixation gene cluster is shown at top. The genes are colored by function: blue (nitrogenase), green (co-factor biosynthesis, shading corresponds to operons), yellow (e-

transport), and grey (unknown). The thin arrows show the length and orientation of the seven operons and a horizontal bar indicates overlapping genes. The recoded genes are shown as dashed lines. The symbols used to define the refactored cluster and controller are defined in Figs. 29 and 30, respectively

5 [0081] Fig. 27 illustrates the robustness of the nitrogen fixation pathway to changes in the expression of component proteins. (A) The pathway for nitrogenase maturation is shown and proteins are coloured by function (Fig. 26). The metal clusters are synthesized by the biosynthetic pathway (23, 24). Nitrogen fixation catalyzed by the matured nitrogenase is shown with its *in vivo* electron transport chain. (B) The tolerance of nitrogenase activity to changes in the expression of component proteins are shown. Activity is measured via an acetylene reduction assay and the % compared to wild-type *K. oxytoca* is presented. Wild-type operons are expressed from a P_{tac} promoter on a low copy plasmid. The promoter activity is calculated as the output of the P_{tac} promoter at a given concentration of IPTG and compared to a constitutive promoter. The effect of not including NifY (-Y) and NifX (-X) 10 are shown in red. (C) The comparison of the strength of wild-type (black) and synthetic (white) ribosome binding sites (RBSs) is shown. The RBSs were measured through an in-frame transcriptional fusion (-60 to +90) with mRFP. The strength is measured as the geometric average from a distribution of cells measured by flow cytometry. The synthetic RBSs of *nifF* and *nifQ* are not intended to match the wild-type measurement. Error bars represent the standard deviation of at least three experiments performed on different days. 20

[0082] Fig. 28 illustrates converting to T7* RNAP Control. (A) Nitrogenase activity is shown as a function of promoter strength for each refactored operon in respective *K. oxytoca* knockout strains ($\Delta nifHDKTY$, $\Delta nifENX$, $\Delta nifJ$, $\Delta nifBQ$, $\Delta nifF$, and $\Delta nifUSVWZM$). Vertical dashed lines indicate strength of the mutant T7 promoter that controls each operon in the complete refactored gene cluster. (B) A controller plasmid decouples operon expression from the inducible promoter. A T7 RNAP variant (T7*) was designed to reduce toxicity. A set of 4 mutated T7 promoters were used to control the expression of each operon (part numbers and sequences for mutants 1-4 are listed in the *Materials and Methods* section). P_{tac} activity under 1mM IPTG induction is indicated by a dashed horizontal line. (C) Nitrogenase activity is compared for each refactored operon under the control of the P_{tac} promoter at the optimal IPTG concentration (black) and the controller (part D) with 1mM IPTG and expression controlled by different T7 promoters (white). The T7 promoters used are WT for operons HDKY, EN and J; promoter 2 for operons BQ and USVWZM; and promoter 3 for F. 25 30

Error bars represent the standard deviation of at least three experiments performed on different days.

[0083] Fig. 29 shows a comprehensive schematic for the complete refactored gene cluster and controller. Each of the 89 parts is represented according to the SBOL visual standard and the SynBERC Registry part number and part activity are shown. The T7 promoter strengths are measured with red fluorescent protein (mRFP) and reported in REU (*see, Materials and Methods*). Terminator strengths are measured in a reporter plasmid and reported as the fold-reduction in mRFP expression when compared to a reporter without a terminator. The RBS strength is reported in as arbitrary units of expression from the induced P_{tac} promoter (1mM IPTG) and a fusion gene between the first 90 nucleotides of the gene and red fluorescent protein. The nucleotide numbers for the plasmids containing the refactored cluster and controller are shown. The codon identity of each recoded gene as compared to wild-type is shown as a percent.

[0084] Fig. 30 shows the regulation of the complete refactored gene cluster. (A) Nitrogenase activity for the three controllers are shown: IPTG-inducible, aTc-inducible, and IPTG ANDN aTc logic. The gas chromatography trace is shown for each as well as the calculated percent of wild-type activity, ($7.4\% \pm 2.4\%$, $7.2\% \pm 1.7\%$ and $6.6\% \pm 1.7\%$ respectively). Standard deviation is calculated using data from at least two experiments performed on different days. (B) ^{15}N incorporation into cell biomass is shown. Nitrogen fixation from N_2 gas by the refactored gene cluster was traced using $^{15}\text{N}_2$ and measured using isotope ratio mass spectrometry (IRMS). Data are represented as the fraction of cellular nitrogen that is ^{15}N . The standard deviation represents two experiments performed on different days. (C) The effect of ammonia on regulation of nitrogenase expression is shown. Acetylene reduction traces shown with and without addition of 17.5 mM ammonium acetate for wild type cells (left) and cells bearing synthetic *nif* system (right). The synthetic system was induced by Controller #1 using 1mM IPTG and exhibited nitrogenase activity of $1.1\% \pm 0.5\%$ and $6.1\% \pm 0.4\%$ with and without ammonium acetate respectively. (D) T7* RNAP expression of Controller #1 corresponding to Part C is shown. Strains carrying Controller #1 and a RFP reporter plasmid were characterized under 1mM IPTG induction with or without addition of ammonium acetate.

[0085] Fig. 31 shows the *nif* operon deletions used in this study. The solid lines show the region of deleted *nif* operons. The dashed line in NF25 shows the retained *nifLA* operon.

[0086] Fig. 32 depicts a table of construction and verification of all *K. oxytoca* *nif* gene deletion mutants.

[0087] Fig. 33 shows promoter characterization using Relative Expression Units. (A) Conversion of arbitrary units into Relative Expression Units (REU). Promoters were characterized using mRFP1 fluorescent reporter protein in N155 (Measured Promoters). Data were first normalized by the fluorescence of N110 (Internal Standard) and then scaled by the fluorescence of N155(J23100) to account for RBS differences between N155 and N110 (RBS Adjustment). To directly compare our measurements to expression levels of the Kelly et al. standards, we further multiplied by the ratio of N110 fluorescence to the fluorescence of a Kelly standard plasmid expressing mRFP1 (RFP Promoter Standard). A final conversion factor is applied to compare all measurements to the Kelly et al. J23101-EGFP promoter standard based on a strong linear correlation of promoter strength (RPU) between constructs expressing mRFP and EGFP. Solid and dashed boxes were drawn to indicate which plasmids were measured at different facilities. Asterisked and non-asterisked units were measured in different facilities and correspond to the conversion factors directly above. (B) Promoter characterization for P_{lac} promoter (left) and P_{tet} promoter (right). The promoter strengths of P_{lac} promoter and P_{tet} promoter were measured under varied concentrations of inducers (IPTG or aTc). The strengths of T7 promoters (WT and mutants, Fig. 28B) are shown as horizontal dotted lines.

[0088] Fig. 34 illustrates debugging of the refactored operons. (A) The process is shown for the identification of problem sequences within a refactored operon. After design and synthesis, the problematic DNA is crossed with wild-type to create a chimeric library, which is screened. This is done iteratively to reduce the size of the problematic region until the specific errors are identified. (B) The debugging process led to the correction of RBS strengths, the recoded sequence of *nifH*, and numerous nucleotide errors found in the sequenced cluster in the database. Amino acid mutations to correct errors in the synthetic sequence are shown.

[0089] Fig. 35 depicts a table of DNA sequence errors in *nif* cluster sequence X13303.1.

[0090] Fig. 36 shows cell growth supported by nitrogen fixation. The dotted line indicates initial seeding density of OD600 0.5. Wild-type *Klebsiella* grew to an OD600 2.57 ± 0.07 after 36 hours of incubation in depression conditions. Eliminating the full *nif* cluster severely inhibited cell growth (Δnif , OD600 0.76 ± 0.02). Complementing the knockout strain with

the refactored cluster and Controller #1 under 1mM IPTG induction yielded growth of OD600 1.10 ± 0.03 .

[0091] Fig. 37 shows expression of synthetic *nifH* variants. Western blot assay to detect the expression of synthetic *nifH*_{v1} (left) and synthetic *nifH*_{v2} (right). All
5 constructs bore P_{tac}-*nifHDK* with the synthetic gene indicated. Cultures were induced with 50μM IPTG.

[0092] Fig. 38 depicts a table of DNA sequences of synthetic parts.

[0093] Fig. 39 shows maps of key plasmids. SBOL graphical notation is used to describe genetic parts: the BioBrick prefix and suffix are open squares, and terminators are in the
10 shape of a T.

DETAILED DESCRIPTION

I. Introduction

[0094] The present invention relates to gene cluster engineering. It has been discovered
15 how to recombinantly and computationally manipulate and select native gene cluster coding sequences and heterologous regulatory sequences such that the coding sequences are under control of heterologous regulation and produce the functional product of the gene cluster (e.g., a native operon). By eliminating native regulatory elements outside of, and within, coding sequences of gene clusters, and subsequently adding synthetic regulatory systems, the
20 functional products of complex genetic operons and other gene clusters can be controlled and/or moved to heterologous cells, including cells of different species other than the species from which the native genes were derived.

[0095] As demonstrated below, the inventors have re-engineered the *Klebsiella oxytoca* *Nif*
25 gene cluster as well as a *Salmonella* Type III protein secretion system, thereby generating functional products (e.g., nitrogen fixing enzymes and peptide secretion complexes, respectively) under control of a heterologous regulatory system. Once re-engineered, the synthetic gene clusters can be controlled by genetic circuits or other inducible regulatory systems, thereby controlling the products' expression as desired.

II. Generation of Synthetic Gene Clusters

[0096] It is believed that the methods described herein can be used and adapted to re-engineer regulation of essentially any operon or other gene cluster. Generally, the native operons or gene clusters to be engineered will have the same functional product in the native host. For example, in some embodiments, at least a majority of the gene products within the native operon or gene cluster to be re-engineered will each function to produce a specific product or function of the native host. Functional products can include, for example, multi-component enzymes, membrane-associated complexes, including but not limited to complexes that transport biological molecules across membranes, or other biologically active complexes. For example, in some embodiments, the functional products are, e.g., a Type III protein secretion system, a bacterial microcompartment, a gas vesicle, a magnetosome, a cellulosome, an alkane degradation pathway, a nitrogen fixation complex, a polybiphenyl degradation complex, a pathway for biosynthesis of Poly (3-hydroxybutyrate), nonribosomal peptide biosynthesis enzymes, polyketide biosynthesis gene cluster products, a terpenoid biosynthesis pathway, an oligosaccharide biosynthesis pathway, an indolocarbazole biosynthesis pathway, a photosynthetic light harvesting complex, a stressosome, or a quorum sensing cluster. See, Fischbach and Voigt, *Biotechnol. J.*, 5:1277-1296 (2010).

[0097] Native operons or gene clusters used in embodiments of the present invention can be derived (originated) from prokaryotes or eukaryotes.

[0098] As used herein, “native” is intended to refer to the host cell or host genome from which an operon or gene cluster is originally derived (e.g., as the operon is found in nature). Thus, “native expression” of an operon refers to the specific expression levels and patterns of a set of genes in an operon or gene cluster in a native host.

[0099] An operon refers to a unit of DNA comprising multiple separate coding sequences under the control of a single promoter. The separate coding sequences are typically expressed within a single RNA molecule and subsequently translated separately, e.g., with varying translation levels due to the strength of ribosomal binding sites (RBSs) associated with the particular coding sequences. Operons are most typically found in prokaryotic cells.

[0100] [0001] Gene clusters refer to sets of genes having a common function or function product. Genes are typically found within physical proximity to each other within genomic DNA (e.g., within one centiMorgan (cM)). Gene clusters can occur in prokaryotic or eukaryotic cells.

A. Coding Sequences

[0101] Once a native operon or gene cluster has been identified for re-engineering, the coding sequences to be re-engineered can be identified. Generally, it will be desirable to start with only the coding sequences from the native operon or gene cluster, thereby removing native promoters and other non-coding regulatory sequences. Depending on the function of the various gene products of the native operon or gene cluster, in some embodiments all of the coding sequences of a native operon or gene cluster are re-engineered.

[0102] Alternatively, one or more coding sequences can be omitted from the re-engineering process. For example, it may be known that one or more of the gene products in a native operon or gene cluster do not contribute to the function product of the operon or may not be necessary for generation of the operon's or cluster's product. For example, as described in the examples below, in re-engineering the *Nif* operon, the *nifT* gene had no known function and notably it was known that elimination of *nifT* did not significantly affect the ultimate function of the operon, i.e., nitrogen fixation. Thus, *nifT* was not included in the re-engineering process.

[0103] In some embodiments, the operon or gene cluster will include coding sequences for regulatory proteins that regulate expression or activity of one or more of the other products of the operon or gene cluster. In such embodiments, it can be desirable to omit such regulatory proteins from the re-engineering process because synthetic regulation will be employed instead. For example, as described in the examples below, in re-engineering the *nif* operon, *nifL* and *nifA* were known to act as regulatory genes for the *nif* operon and thus were omitted so that synthetic regulation could be instead used.

[0104] Once the set of gene products to be re-engineered has been identified, one can start with the native coding sequence, or the amino acid sequences of the gene products. For example, in some embodiments, the amino acid sequences of the gene products can be used to produce a synthetic coding sequence for expression in the host cell in which the re-engineered products are to be ultimately expressed.

[0105] In some embodiments, the native coding sequences of the set of gene products to be re-engineered are used as a starting point. In this case, in some embodiments, sequences not essential to production of the gene products is eliminated. For example, ribosome binding sites, terminators, or promoters within the coding sequences can be eliminated. In some embodiments, the nucleotide sequences of the coding sequences are analyzed using an

algorithm (i.e., in a computer) to identify ribosome binding sites, terminators, or promoters within the sequence(s).

[0106] Nonessential regulatory sequences within the coding sequences can be reduced or eliminated by altering the codons of the native coding sequence(s). Regulatory sequences
5 comprising codons can be disrupted, for example, by changing the codons to synonymous codons (i.e., encoding the same amino acid) thereby leaving the encoded amino acid sequence intact while changing the coding sequence. One or more codons of one or more coding sequences can be altered.

[0107] In some embodiments, at least 5%, 10%, 15%, 20% or more codons of one or more
10 native coding sequence to be inserted into a synthetic operon are replaced. In some embodiments, at least 5%, 10%, 15%, 20%, 30%, 40%, 50% or more codons of each of the native coding sequences to be inserted into a synthetic operon are replaced.

[0108] In some embodiments, replacement codons can be selected, for example, to be significantly divergent from the native codons. The codon changes can result in codon
15 optimization for the host cell, i.e., the cell in which the polynucleotide is to be expressed for testing and/or for ultimate expression. Methods of codon optimization are known (e.g., Sivaraman et al., *Nucleic Acids Res.* 36:e16 (2008); Mirzahoseini, et al., *Cell Journal* (Yakhteh) 12(4):453 Winter 2011; US Patent No. 6,114,148) and can include reference to commonly used codons for a particular host cell. In some embodiments, one or more codon
20 is randomized, i.e., a native codon is replaced with a random codon encoding the same amino acid. This latter approach can help to remove any cis-acting sequences involved in the native regulation of the polypeptide. In some embodiments, codons are selected to create a DNA sequence that is maximally distant from the native sequence. In some embodiments, an algorithm is used to eliminate transcriptionally functional sequences in a gene encoding the
25 polypeptide. For example, in some embodiments, ribosome binding sites, transcriptional regulatory elements, terminators, or other DNA sequences bound by proteins are removed from the native coding sequence. Notably, the functional sequences removed can be functional in the native species (from which the sequence was originally derived), in the heterologous host cell, or both. In some embodiments, optimizing comprises removal of
30 sequences in the native coding sequence that are functional for heterologous transcriptional activators or repressors to be used to regulate the synthetic operons to be generated.

[0109] Generation of synthetic coding sequences, as well as the remaining portions of the synthetic operon, in many cases will be performed *de novo* from synthetic oligonucleotides.

Thus, in some embodiments, codons are selected to create a DNA sequence that does not generate difficulties for oligonucleotide production or combination. Thus, in some embodiments, codon sequences are avoided that would result in generation of oligonucleotides that form hairpins.

5 [0110] In some embodiments, as noted above, codon alteration will depend on the host cell used. Host cells can be any prokaryotic cell (including but not limited to *E. coli*) or eukaryotic cell (including but not limited to yeast cells, mammalian cells, avian cells, amphibian cells, plant cells, fish cells, and insect cells).

[0111] Nonessential regulatory sequences within native sequences can be identified, in
10 some embodiments, using an algorithm performed by a processor executing instructions encoded on a computer-readable storage medium. For example, in some embodiments, ribosome binding sites are identified using a thermodynamic model that calculates the free energy of the ribosome binding to mRNA. In some embodiments, promoters are identified with an algorithm using a position weighted matrix. In some embodiments, transcriptional
15 terminators are identified by an algorithm that identifies hairpins and/or poly-A tracks within sequences. In some embodiments, an algorithm identifies other transcriptionally functional sequences, including but not limited to transposon insertion sites, sites that promote recombination, sites for cleavage by restriction endonucleases, and/or sequences that are methylated.

20 [0112] In view of the alterations described above, in some embodiments, a coding sequence in a synthetic operon of the invention is less than 90, 85, 80, 75, or 70% identical to the native coding sequence. In some embodiments, the coding sequence encodes a protein sequence that is identical to the native protein or is at least 80, 85, 90 or 95% identical to the native protein. In some embodiments, less than 70%, 60%, or 50% of codons in one, two or
25 more coding sequences in a synthetic operon are identical to the codons in the native coding sequence.

B. Organizing coding sequences into synthetic operons

[0113] Once coding sequences have been selected (e.g., and substantially “cleaned” of
30 native or spurious regulatory sequences), the coding sequences are organized into one or more synthetic operon(s). Organization of the synthetic operon(s) includes insertion of various heterologous transcriptional and translational sequences between, before, and/or after

the coding sequences so that expression of each coding sequence is controlled as desired. Thus, for example, 5' promoter sequences can be selected to drive expression of an operon RNA comprising the coding sequences of the operon. Selection of one or more terminator of appropriate strength will also affect expression levels. Moreover, the order of the coding sequences within a synthetic operon and/or selection of RBSs for the coding sequences allows for control of relative translation rates of each coding sequence, thereby allowing several levels of control for absolute and relative levels of the final protein products.

5 [0114] Because each synthetic operon can have its own promoter, different synthetic operons can be expressed at different strengths. Thus, in some embodiments, coding sequences are organized into different operons based on the relative native expression levels. Said another way, in some embodiments, coding sequences are organized into operons by grouping coding sequences expressed at substantially the same native level in a particular synthetic operon.

10 [0115] Moreover, because coding sequences at the 5' (front) end of an RNA can be expressed at a higher level than coding sequences further 3', in some embodiments, coding sequences are ordered within a synthetic operon such that the highest expressing coding sequence (in the native context) occurs first and the lowest expressing gene occurs last. In some embodiments, organization of genes within operons is based on native temporal expression, function, ease of manipulation of DNA, and/or experimental design.

15 [0116] In designing the transcriptional (e.g., promoters) and translational (e.g., RBSs) controls of the synthetic operons, the ratio of proteins measured in the native system can be considered. Thus, in some embodiments, two or more coding sequences that are expressed in a native context at substantially the same level and/or that are desirably expressed in an approximately 1:1 ratio to achieve functionality (e.g., where two or more members are part of a functional complex in a 1:1 ratio) are placed in proximity to each other within a synthetic operon. "Proximity" will generally mean that coding sequences are adjacent to each other in the synthetic operon.

20 [0117] In some embodiments, relative expression levels of coding sequences within and, in some embodiments, between synthetic operons is determined by testing one or more test operons for desired expression and/or desired functionality and then improving expression based on the initial results. While this method can be performed in a "trial and error" basis, in some embodiments, a numerical optimization method is employed to guide selection of regulatory elements in order to alter gene expression and to improve desired system

properties. Such methods, for example, can be performed by a processor executing instructions encoded on a computer-readable storage medium (discussed further below). Exemplary numerical optimization methods include but are not limited to, a Nelder-Mead algorithm, a Newton's method, a quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, an ellipsoid method, a Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model. In some embodiments, a computational design of experiments (DoE) method is employed to alter gene expression and to improve desired system properties in the synthetic operons.

[0118] Transcriptional regulatory elements, ribosomal binding sites, terminators, and other sequences affecting transcription or translation can be selected from existing collections of such sequences, and/or can be generated by screening of libraries generated by design or by random mutation. Exemplary regulatory sequences include cis-acting nucleotide sequences bound by a sequence-specific DNA binding polypeptide, e.g., a transcriptional activator or a transcriptional repressor. Exemplary transcriptional activators include, but are not limited to, sigma factors, RNA polymerases (RNAPs) and chaperone-assisted activators. In some embodiments, the transcriptional activator/cis-acting sequence cognate pair will be orthogonal to the host cell. Said another way, the regulatory sequence will not be bound by other host cells proteins except for the heterologous transcriptional activator that binds the cis-acting sequence.

20 i. **Sigma factors**

[0119] In some embodiments, the sequence-specific DNA binding polypeptide is a sigma (σ) factor and the regulatory sequence of the synthetic operon comprises the sigma factor's cognate cis-acting nucleotide sequence. Sigma factors recruit RNA polymerase (RNAP) to specific promoter sequences to initiate transcription. The σ 70 family consist of 4 groups: Group 1 are the housekeeping σ s and are essential; groups 2-4 are alternative σ s that direct cellular transcription for specialized needs (Gruber and Gross, *Annu. Rev. Microbiol.*, 57:441-466 (2003)). Group 4 σ s (also known as ECF σ s; extracytoplasmic function) constitute the largest and most diverse group of σ s, and have been classified into 43 subgroups (Staron *et al.*, *Mol Microbiol* 74(3): 557-81 (2009)).

[0120] In some embodiments, the set of sequence-specific DNA-binding polypeptides comprise multiple sigma factors. In some embodiments, the set comprises sigma factors from Group 1, Group 2, Group 3, and/or Group 4 Sigma factors. The ECF subgroup of Group 4 is thought to recognize different promoter sequences, making these σ s particularly

useful for constructing orthogonal σ -promoter systems. However, it will be appreciated that any group of sigma factors can be used according to the methods of the embodiments of the invention to develop cognate pairs.

Table 1

Group Nr ^a	ID ^b	GI ^c	SPECIES ^d	CLASS ^d	PHYLUM ^d
ECF01	>3473	109899616	<i>Pseudoalteromonas atlantica</i> T6c	Gamma ⁺ proteobacteria	Proteobacteria
ECF01	>4085	114562024	<i>Shewanella frigidimarina</i> NCIMB 400	Gamma ⁺ proteobacteria	Proteobacteria
ECF03	>1198	29350055	<i>Bacteroides thetaiotaomicron</i> VPI-5482		Bacteroidetes
ECF03	>1244	34541012	<i>Porphyromonas gingivalis</i> W83		Bacteroidetes
ECF05	>965	28868416	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	Gamma ⁺ proteobacteria	Proteobacteria
ECF05	>1054	67154316	<i>Azotobacter vinelandii</i> AvOP	Gamma ⁺ proteobacteria	Proteobacteria
ECF07	>980	67154823	<i>Azotobacter vinelandii</i> AvOP	Gamma ⁺ proteobacteria	Proteobacteria
ECF07	>1134	15598606	<i>Pseudomonas aeruginosa</i> PAO1	Gamma ⁺ proteobacteria	Proteobacteria
ECF09	>3581	15597622	<i>Pseudomonas aeruginosa</i> PAO1	Gamma ⁺ proteobacteria	Proteobacteria
ECF09	>1009	70730971	<i>Pseudomonas fluorescens</i> Pf-5	Gamma ⁺ proteobacteria	Proteobacteria
ECF11	>3726	28868260	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	Gamma ⁺ proteobacteria	Proteobacteria
ECF11	>987	28899132	<i>Vibrio parahaemolyticus</i> RIMD 2210633	Gamma ⁺ proteobacteria	Proteobacteria
ECF13	>1146	33152898	<i>Haemophilus ducreyi</i> 3500UHP	Gamma ⁺ proteobacteria	Proteobacteria
ECF13	>1025	37524103	<i>Photobacterium luminescens</i> subsp. laumondii TTO1	Gamma ⁺ proteobacteria	Proteobacteria
ECF15	>436	77464848	<i>Rhodobacter sphaeroides</i> 2.4.1	Alphaproteobacteria	Proteobacteria
ECF15	>524	16127705	<i>Caulobacter crescentus</i> CB15	Alphaproteobacteria	Proteobacteria
ECF17	>1691	15607875	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF17	>1458	21221399	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF19	>3197	15607586	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF19	>1315	21219164	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF21	>1280	29350128	<i>Bacteroides thetaiotaomicron</i> VPI-5482		Bacteroidetes
ECF21	>2825	89889680	<i>Flavobacterium bacterium</i> BBFL7		Bacteroidetes

ECF23	>231	15895043	<i>Clostridium acetobutylicum</i> ATCC 824		Firmicutes
ECF23	>1851	30261806	<i>Bacillus anthracis</i> str. Ames		Firmicutes
ECF25	>1645	170078575	<i>Synechococcus</i> sp. PCC 7002		Cyanobacteria
ECF25	>1643	17230772	<i>Nostoc</i> sp. PCC 7120		Cyanobacteria
ECF27	>4265	21222299	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF27	>1331	31795084	<i>Mycobacterium bovis</i> AF2122/97		Actinobacteria
ECF29	>371	13476734	<i>Mesorhizobium loti</i> MAFF303099	Alphaproteobacteria	Proteobacteria
ECF29	>2688	71281387	<i>Colwellia psychrethra</i> 34H	Gammaproteobacteria	Proteobacteria
ECF31	>2963	85713274	<i>Idiomarina baltica</i> OS145	Gammaproteobacteria	Proteobacteria
ECF31	>34	16080921	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		Firmicutes
ECF33	>375	27378153	<i>Bradyrhizobium japonicum</i> USDA 110	Alphaproteobacteria	Proteobacteria
ECF33	>423	39934888	<i>Rhodopseudomonas palustris</i> CGA009	Alphaproteobacteria	Proteobacteria
ECF35	>3582	15598092	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF35	>1119	24375055	<i>Shewanella oneidensis</i> MR-1	Gammaproteobacteria	Proteobacteria
ECF37	>3390	89094252	<i>Oceanospirillum</i> sp. MED92	Gammaproteobacteria	Proteobacteria
ECF37	>2513	83718468	<i>Burkholderia thailandensis</i> E264	Betaproteobacteria	Proteobacteria
ECF39	>1438	21223369	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF39	>2973	84494624	<i>Janibacter</i> sp. HTCC2649		Actinobacteria
ECF41	>491	16127496	<i>Caulobacter crescentus</i> CB15	Alphaproteobacteria	Proteobacteria
ECF41	>1141	77459658	<i>Pseudomonas fluorescens</i> Pf0-1	Gammaproteobacteria	Proteobacteria
ECF43	>4437	21244845	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	Gammaproteobacteria	Proteobacteria
ECF43	>3477	109897287	<i>Pseudoalteromonas atlantica</i> T6c	Gammaproteobacteria	Proteobacteria

[0121] In addition to native sigma factors, chimeric or other variant sigma factors can also be used in the method of the invention. For example, in some embodiments, one or more sigma factor are submitted to mutation to generate library of sigma factor variants and the resulting library can be screen for novel DNA binding activities.

[0122] In some embodiments, chimeric sigma factors formed from portions of two or more sigma factors can be used. Accordingly, embodiments of the invention provide for generating a library of polynucleotides encoding chimeric sigma factors, wherein the chimeric sigma factors comprise a domain from at least two different sigma factors, wherein each of the domains bind to the -10 or -35 region of a regulatory element; and expressing chimeric sigma factors from the library of polynucleotides, thereby generating a library of chimeric sigma factors. For example, in some embodiments, chimeric sigma factors are generated comprising a "Region 2" from a first sigma factor and a "Region 4" from a second sigma factor, thereby generating chimeric sigma factors with novel DNA binding activities. "Region 2" of sigma factors is a conserved domain that recognizes -10 regions of promoters. "Region 4" is a conserved domain of sigma factors that recognizes -35 regions of promoters. It will be appreciated that chimeric sigma factors can be generated from any two native sigma factors that bind different target DNA sequences (e.g., different promoter sequences). It has been found that chimeric sigma factors formed from the ECF2 and ECF11 subgroups have unique DNA binding activities useful for generating orthogonal sets as described herein. Exemplary chimeric sigma factors include, but are not limited to, ECF11 ECF02 (containing amino acids 1-106 from ECF02_2817 and 122-202 from ECF11_3726) and ECF02_ECF11 (containing amino acids 1-121 from ECF11_3726 and 107-191 from ECF02_2817).

[0123] The ECF11_ECF02 amino acid sequence is as follows:

20 1 MRITASLRTFCHLSTPHSDSTTSRLWIDEVTAVARQRDRDSEFMRIYDHFAPRLLRYLTGL
 61 NVPEGQAEELVQEVLLKLWHKAESFDPSKASLGTWLFRIARNLYIDSVRKDRGWVQVQNS
 121 LEOLERLEAISNPENLMLSEELROIIVFRTIESLPEDLRMAITLRELDGLSYEEIAAIMDC
 181 PVGTVRSRIFRAREAIDNKVQPLIRR*

25 [0124] The ECF02_ECF11 amino acid sequence is as follows:

 1 MSEQLTDQVLVERVQKGDQKAFNLLVVRYQHKVASLVSRYVPSGDVDPVVQEAFFIKAYRA
 61 LDSFRGDSAFYTWLYRIAVNTAKNYLVAQGRPPSSDVDAIEAENFEQLERLEAPVDRTL
 121 DYSQRQEQQLNIAIQNLPTDQAKVLRMSYFEALSHREISERLDMPLGTVKSCRLRALFQKL
 181 RSRIEES*

30 ii. **RNA Polymerases**

[0125] In some embodiments, the sequence-specific DNA-binding polypeptide is a polypeptide having DNA binding activity and that is a variant of the T7 RNA polymerase (RNAP) and the RNAP's cognate cis-acting sequence (e.g., a promoter recognized by the RNAP) is operably linked to the synthetic operon to control the operon's expression. The T7
5 RNAP amino acid sequence is as follows:

```

1  mntiniaknd fsdielaaip fntladyge rlareqlale hesyemgear frkmferqlk
61 agevadnaaa kplittllpk miarindwfe evkakrgkrp tafqflqeik peavayitik
121 ttlacltsad nttvqavasa igraiedear fgrirdleak hfkknveeql nkrvghvykk
181 afmqvveadm lskgllggee wsswhkedsi hgvrcieml iestgmvsllh rqnagvvgqd
10 241 setielapey aeaiatraga lagispmfqp cvvppkpwtg itgggywang rrplalvrth
301 skkalmryed vympevykai niaqntawki nkkvlavanv itkwkxhcvpe dipaiereel
361 pmkpedidmn pealtawkra aaavyrkdka rksrrislef mleqankfan hkaiwfpynm
421 dwrgrvyavs mfnpggndmt kgl1tlakgk pigkegyywl kingancagv dkvpfperik
481 fieenhenim acaksplent wvaeqdspfc flafcfeyag vqhhglsync slplafdgsc
15 541 sqiqhfsaml rdevggravn llpsetvqdi ygivakkvne ilgadaingt dncvvtvtde
601 ntgeisekvk lgtkalagqw laygvtrsvt krsvmtlayg skefgfrqqv ledtiqpaid
661 sgkglmftqp nqaagymakl iwesvsvtvv aaveamnwlk saakllaaev kdkktgeilr
721 krcavhwvtp dgfpvwqeyk kpiqtrlnlm flgqfrlqpt intnkdseid ahkqesgiap
781 nfvhsqdqsh lrktvwahe kygiesfali hdsfgtipad aanlfkavre tmvdtyescd
20 841 vladfydqfa dqlhesqldk mpalpakgnl nlrdelesdf afa

```

[0126] The T7 RNAP promoter has also been characterized (see, e.g., Rong *et al.*, *Proc. Natl. Acad. Sci. USA*, 95(2):515-519 (1998)) and is well known.

[0127] Methods have been discovered for generating orthogonal pairs of RNAP variants
25 and target promoter variants. Due to toxicity of expression of native T7 RNAP, a series of mutations and modifications can be designed such that a library of RNAP variants can be expressed and tested for activity in cells without excessive toxicity. Accordingly, embodiments of the invention provide for one or more of the following modifications (and thus, for example, an embodiment of the invention provides for host cells comprising
30 expression cassettes, or nucleic acids comprising expression cassettes, wherein the expression cassette encodes a RNAP variant substantially identical to T7 RNAP, wherein the expression cassette comprises one or more of the following):

Expression of the T7 RNAP variant can be expressed from a low copy plasmid. Expression of the RNAP can be controlled by a separately encoded protein from a separate vector, thereby blocking expression of the RNAP until a second vector is added to the cells promoting RNAP expression;

- 5 Translational control: a GTG start codon; weak ribosomal binding sites, and/or random DNA spacers to insulate RNAP expression can be used;

A molecular tag to promote rapid degradation of the RNAP. For example, an Lon N-terminal tag will result in rapid degradation of the tagged RNAP by the Lon protease system.

- 10 A mutated RNAP active site (e.g., within amino acids 625-655 of T7 RNAP). For example, it has been discovered that a mutation of the position corresponding to amino acid 632 (R632) of T7 RNAP can be mutated to reduce the RNAP's activity. In some embodiments, the RNAP contains a mutation corresponding to R632S.

- [0128] Moreover, a variety of mutant T7 promoters have been discovered that can be used
15 in a genetic circuit. Thus, in some embodiments, the regulatory sequence of a synthetic operon comprises a mutant sequence as set forth in the table below.

Promoter Name	Sequence TAATACGACTCACTANNNNAGA	Strength (2009.10.02 to 2009.10.09)
WT	TAATACGACTCACTATAGGGAGA	5263
Mut1	TAATACGACTCACTACAGGCAGA	365
Mut2	TAATACGACTCACTAGAGAGAGA	366
Mut3	TAATACGACTCACTAATGGGAGA	577
Mut4	TAATACGACTCACTATAGGTAGA	1614
Mut5	TAATACGACTCACTAAAGGGAGA	1018
Mut6	TAATACGACTCACTATTGGGAGA	3216

- [0129]** A number of different stem loop structures that function as terminators for T7 RNAP have also been discovered. *See*, Table directly below. Accordingly, an embodiment of the invention provides for a synthetic operon comprising a promoter functional to a native T7 RNAP or an RNAP substantially identical thereto, wherein the operably linked polynucleotide comprises a terminator selected from the table directly below. Terminators with different sequences can be selected for different transcripts to avoid homologous recombination.

Terminator Name	Sequence TANNNNAACSSWWSSSSSSTCWWW WCGSSSSSWWSSGGTTTTTTGT	Strength (2009.12.16 Assay)
52	TATAAACGGGGGGCTAGGGGTTTTT GT	107
23	TACTCGAACCCCTAGCCCGCTTTATC GGCGGCTAGGGTTTTTTGT	714
72	TAGCAGAACCGCTAACGGGGCGAAG GGTTTTTTGT	1051
48	TACTCGAACCCCTAGCCCGCTTTATC GGCGGCTAGGGTTTTTTGT	1131
1	TACATATCGGGGGGTAGGGTTTTTT GT	1297
2	TACATATCGGGGGGTAGGGTTTTTT GT	1333
WT	TAGCATAACCCCTTGGGGCCTCTAAC GGTCTTGAGGGTTTTTTGT	1395
31	TACCCTAACCCCTTCCCGGTCATCG GGCGGATGGGGTTTTTTGT	1586
58	TAGACCAACCCCTTGGGGCCTCAATCG GGGGGATGGGGTTTTTTGT	1608
25	TACTCTAACCCCATCGCCGTCTTAGG GGTTTTTTGT	1609
17	TACCTCAACCCCTTCCGCCCTCATATC GCGGGCATGCGGTTTTTTGT	1887

- 10 **[0130]** In some embodiments, RNAP variants can be designed comprising an altered specificity loop (corresponding to positions between 745 and 761). Thus in some embodiments, an RNAP is provided that is identical or substantially identical to T7 or T3 RNAP but has a Loop Sequence selected from those in the tables directly below between positions 745 and 761.

A

RNAP Family	Scaffold	RNAP Plasmid	Promoter Plasmid	Loop Sequence	Promoter Sequence
T7	N249	N249	N155	VWQEYKKPIQTRLNLMFLGQFRLQPTINTNKDSEI DAHK	TAATACGACTCACTATA GGGAGA
T3	N115	N377:115	N352	VWQEYKKPIQKRLDMIFLGQFRLQPTINTNKDSEI DAHK	TAATAACCCTCACTATA GGGAGA
K1F	N115	N421:115	N353	VWQEYKKPIQTRLNLMFLGSFNLQPTVNTNKDSEI DAHK	TAATAACTATCACTATA GGGAGA
N4	N77	W78	W74	VWQEYKKPIQTRIDCVILGTHRMALNTINTNKDSEI AHK	TAATAACCACACTATA GGGAGA

B

	T7 promoter	T3 promoter	K1F promoter	N4 promoter
T7 RNAP	2177	24	17	14
T3 RNAP	83	1062	14	14
K1F RNAP	45	26	463	13
N4 RNAP	51	147	46	2616

iii. Activators Requiring Chaperones

[0131] In some embodiments, the set of sequence-specific DNA-binding polypeptides
 5 comprise polypeptides having DNA binding activity and that require a separate chaperone
 protein to bind the sequence-specific DNA-binding polypeptide for the sequence-specific
 DNA-binding polypeptide to be active. Exemplary transcriptional activators requiring a
 chaperone for activity include, but are not limited to activator is substantially similar to InvF
 from *Salmonella Typhimurium*, MxiE from *Shigella flexneri*, and ExsA from *Pseudomonas*
 10 *aeruginosa*. These listed activators require binding of SicA from *Salmonella Typhimurium*,
 IpgC from *Shigella flexneri*, or ExsC from *Pseudomas aeruginosa*, respectively, for
 activation.

Sequence information for the above components are provides as follows:

Name	Type	DNA sequence encoding the named polypeptide	Optional Mutation
<i>sicA</i>	Gene	atggattatacaaaataatgtcagcgaagaacgtgttgcggaatgattgggatgccgttagtgaag gcgccacgctaaaagacgttcatgggatccctcaagatagatggacggtttatagctcatgcta lgagtttataaccaggacgactggatgaagctgagacgttcttctgntcttatgcattatgatt tacaatcccgattacacatgggactggcagatgccaactgaaaaacaatttcagaaagc atgtgacctttagcagtagcgtttacgttacttaaaaatgattatgcccccgtttttaccgggcagt gtcaattattaatgcgtaaggcagcaaaagccagacagtgtttgaactgtcaatgaactactga agatgagctctctcgggcataaagcgttggctatctggagcgcctaaaanccggcgagacagag cagcacagtgaacanganaaggaataa	
<i>sicA*</i>	Mutant <i>sicA</i>	atggattatacaaaataatgtcagcgaagaacgtgttgcggaatgattgggatgccgttagtgaag gcgccacgctaaaagacgttcatgggatccctcaagatagatggacggtttatagctcatgcta tgagtttataaccaggacgactggatgaagctgagacgttcttctgntcttatgcattatgatt ttacaatcccgattacacatgggactggcagatgccaactgaaaaacaatttcagaaagc atgtgacctttagcagtagcgtttacgttacttaaaaatgattatgcccccgtttttaccgggcagt gtcaattattaatgcgtaaggcagcaaaagccagacagtgtttgaactgtcaatgaactactga	The large "t" of the <i>sicA</i> sequence above was mutated to "a" by error- prone PCR. This mutation was made to reduce cross talk between SicA and

		agatgagictctcggggcaaaagcgttggctctatcggagggcgtaaaaacggcggagacagag cagcacaagtgaacaagaaaagaataa	MxiE.
<i>invF</i>	Gene with new start codon	atgctaatacgcaggaaactctaaagaaggagagaagcggaaaatccgagcccggaagca tggttatcacagcgttccgcgcaaaaagctgcata atg tattttctgaagccgacacaatga aaattgctgattcaggaagggcgcgtctgttttgcgagcagccgcttgcaccagatcagg agacctggttttcaccgttaaaaatgaagctacagcaaatctggcattatcagatggcgcag gattatggacacgacatacgtgaatccgataaaagggtttgctgagtcctgagttcgcctatt ggcaagatcgtaaacgctgcgagctactgttttgcagcaaatatcagccttccggccttcaat aaggtaactggcgtgttacgaaaaagcgaaggtactgttggttggctattactcgtcagtcgaa ccagcggcaacacgatgaatgctgggagaagactatggcgtttctataccaatttctcgttt gtgcagcagagcgttggcgaaaaagcgaagagtgaaatcagaactggcgtatggcgaatc gctgctgaatagtgtagaaggccacgagaacatcacccaatagccgttaatcaggttactc gcttccacattttctagtgagatcaaaagctgacggcgttccgcggaaaatataaatatata tcaattggcagcaaatga	The accepted start codon (the large "atg") was determined to be incorrect and a correct upstream start codon was found.
<i>psicA</i>	Promoter	ccacaagaacgagggtacggcattgagccgctaaggcagtagcagatgattcattggcgtttt lgaatgttcaataaccaccgtcgggttataactgcatcagataaacgagtcgtaagtctlaca aagtcggtpacagataacaggagtaagta	
<i>ipgC</i>	Gene	atgtcttaaatatcaccgaaaatgaaagcactctctactcagtaattgatgcaattaactctggcgt acactgaagataataatgcaattcctgatgataatgatgatgacattatcattatgcttatgacttta caacaaaaggaaagataggaagcgtgaaagttcttcaggtttatgatacagactttacaatgt agactacattatgggactcgcagctattatcagataaaaagaacagttccaacaagcagcagact ttatgctgtcgttttctattgaaaaaaatgactatacaccagatcctacatggcaaatgctcagct tcggttgaaagcccccttaaaagctaaagagtgctcgaacgtgaattcaacacagcaatgatga aaaataaaaaataaaagcacaatcactctggagcgaattcaggatatacaaggagtaa	
<i>mxiE</i>	Gene with codon optimization	A lgagtaataataaaggcctgaaccaccgacaacatgttctacatctacagctctggctatgaacc ggtagaacgttgaactggtagaagataaaagcgttaacatcagaaactggcncggcgtggaaa gg ctttttctttt gtcgttaaccagaacatcaaatcagcgataacgtaactaccactacc gcttcaacatcaactctgcccataaactctgctgtttgggattatttagcggcgccttggtaac atctcaccagaaaaatgcaaccattctacacgaaaacgactcgcgtgatacgttaaacgga atctatgctggataaacgatgctgcgcttctatttagtagcagacagaacgtgtcfaatgccctggc aatgatccgtatgaccgaaatgattcatctgttctgtaacctgctgacgattgaaaaaagaaa gaaglycgcatacaaaagcctgaccgaaactatggcgtttctgaagcgtacttctgtagctgtc gcaaaagcgtgggtggcaaaagtgaaagacagcgtgancacgtggcgcctggtaagcctg tggatgttttctgtaatacagaccattacgagcgcggccagcaaatggtatgctctaccag tcacttcagcaatgaaataaacgcgtctggccttagtgcgccggaactgagcaacatcaccttc ctggtgaaagaaaataatgaaaaaatctaa	The wide type gene has "ttttttt" in this enlarged sequence region. One more "t" was added to make "ttttttt" and then the entire gene was codon optimized by GenScript. The additional "t" was added to make this ORF in-frame. In addition, the wide-type gene starts with "g" and this synthetic gene starts with "a."
<i>pipaH9.8</i>	Promoter	gcgaaaaatgacatcaaaaacgccattaacctgatgttctgggaaatataaatgtaggctagggtc aaaaatcgtggcgttgacaaaatggctgcgttacgicattgagcatatccaggactggccggcaa accgggtacgcgactgttgccttgaaaagttagctgacctcagtaaatatcaatagcgttctga cagaccgctaccgttcaaatatgaaatgacgatgttaactaacgaaaaacaagaacaatacgg gcaaacaggccattcacggttaactgaaacagatcgttttttacagccaatttgtttatccttata t A ataaaaaagtgc	
<i>pipaH9.8*</i>	Promoter with mutation	gcgaaaatgacatcaaaaacgccattaacctgatgttctgggaaatataaatgtaggctagggtc aaaaatcgtggcgttgacaaaatggctgcgttacgicattgagcatatccaggactggccggcaa accgggtacgcgactgttgccttgaaaagttagctgacctcagtaaatatcaatagcgttctga cagaccgcttaccgttcaaatatgaaatgacgatgttaactaacgaaaaacaagaacaatacgg gcaaacaggccattcacggttaactgaaacagatcgttttttacagccaatttgtttatccttana ag ataaaaaagtgc	The enlarged "ta" above of <i>pipaH9.8</i> was mutated to "ag" by saturation mutagenesis. This mutation was made to reduce leaky expression of <i>pipaH9.8</i> .
<i>exsC</i>	Gene	atggalltaacagcaaggtcaaccgactgcttccggagttcgcagccggtatcggtttgccttcc ctgtcctcgcagaggaggcagatggcgagcctcctgttcgacgaacagggtggcgtcaccctgt tgcgtcgcggagcgcgagcgtctgttctggaggccgaltggcgggcatcagatgctggg cgaaggatcttccgacgctgccagctcaaccgcaatggcaccgttctgactgcaattcggc ttcgagagctgaccgcaaggttccagttgatgcgagatctcgcagcgaactgaccctcga atgcttcgagggcaccctggccaatctgctgacacgcccaggttctggcagcgcctgctgccc ggcagactgactcggggcgtcgtcgtcggcctgaggggttga	
<i>exsD</i>	Gene	atggagcaggaaagcagataagcagtaactcccagaaagcgggttctgctggcaggcgggtatcc gtgtggcctcgcagcccctcgcggggtcgggtccgggttacgcatcagcagtttctatc gtgagtcggaaatcagtgccgcaactggcgttctgctgagcggatgctgcccggcctggc gctggagcaactgttccgtcgtgagtggttcagcagcgcctggcgcggcctggcgtggc gctcgaagaggtgcggcaatctcctcgcgcggcaggaagcagcagcgttggctcga actgggcgaccgggtcaactcgccttggcagctgagatcagctgggttctgctggcgtt latggctgggtgggaaagcctgctcagccagcgcaccccggctggcgccttctgctgggtgagc lggagaccagtcggcaactcagagcaagtcgcaatctgctccgctggcggcagctgga gcccggagcagcccgaggaaactggccaggttgcgaagtggcagcggcgcaccagga acaggtggccgaaactggccgcaagctggagcggcctggcactggcgaagagcgcctggc cgaactggcagcggggcatggcagcgtctcgcagcggcggcgtggcggctcggagccg atcccgggtcctcgaatgctcttggcaacctctctcggcgtggagcagcagctggcggg cggacggctccaggcctggctgacagcaacgcaactgtgcccaggcagcagatcacttactg gcaagactga	
<i>exsA</i>	Gene	atgcaagggccaaatcttggcgaagcagataacgtctgcatggaaactccaacttccg aatlacagggttaacaaggaagggcgtatatttctcctcagggcgaactgacctccaagga	

		<p>caicgattccatttttctgctggcgcctggcaggttgccttttgcgccgcggaagctatgctgtaa gtaccaagggaaggacagccgaatacttggattccattatctgcccagtttctacaaggcttctgt ccagcgttccggcgcctgtgagtgaaagtcgagcgttgcgacgagccgtgccgggcatc gcgttgcctgccacgecttgcctggcgggttgcgtaagggttgaaggaaatgcttgcagag catccgccgatgctcgcctgctgaagatcaggagttgctgatgcttccggttccagtcgcag gggccgctgctgatgctggcttgcggcaactgagcaaccggcatgctgagcgtctgcaactt catggaagcactacctcaacgagtggaagctgctgacttcccgcgagtcggcatgggg ctgaccacctcaaggagctgttccgagctctatgggtttccgccgcctggatcagcga gcggagaatccttatgccatcagttgctgctcaacagcgacatgagcatgctgacatgccat ggaggcgggttttccagtcagctctatttaccacagagctatgccgccgtttcggctgacgccc ggccgctcggcggcggggaaggacgaatgccggcctaaaaataactga</p>	
<i>pexsD</i>	Promoter	<p>gaaaggacgaatgccgggctaaaaataactgacgtttttgaaagcccggtagcggctgcatgagt agaatcggcccaant</p>	
<i>pexsC</i>	Promoter	<p>gatgtagctttttttaaagaaaagctctcagtgacaaaaagcagatgcatagcccggctgtagca tgcgctgagctt</p>	
<i>rfp</i>	Gene	<p>atggcttccctccgaagacgttacaagagtcacgctttcaaaagttcgtatggaagttccgltaa cggtcacgagttcgaatcgaagggtgaagggtcgtcgtacgaaaggtaacgagaccgt aaactgaaagtaccaaaagggtgctccgctgccgttggacatccctgcccagttccag tacggttccaaagcttacgttaaacaccgggtgacatccggactaccgaaactgctctccg gaaagttcaaaagggaacgtgttatgaactcgaagacgggggtttgtaccgttacccaggact cctccctgcaagacgggtgagttcactcaaaagttaaacctgctgggtaacttcccgtccgacg gtcgggttatgcgaaaaaacatgggttgggaagcttccaccgaacgtatgtaccggaaagac ggtgctctgaaaggtgaaatcaaaatgctgctgaaactgaaagacgggtggtcactacgacgtga agttaaaccacactacatggctaaaaaaccgggtcagctgccgggtgcttacaaccgacatca aactggacatcacctcccacaacgaagactacaccatcgttgaacagtacgaactgctgaaagt cgtcactccaccggtgctgcaacnaacgacgaaactacgcttaa</p>	

C. Controlling Operon Expression

[0132] As noted above, the one or more synthetic operons are controlled by regulatory elements responsive to a sequence-specific DNA binding polypeptide (e.g., a transcriptional
 5 activator). Where more than one operon is used, it can be desirable that each operon be responsive to the same transcriptional activator, albeit with a different regulatory sequence that controls the “strength” of expression of a particular operon. As noted above, in some embodiments, the transcriptional activator is a T7 RNAP or a variant thereof.

[0133] Expression of the sequence-specific DNA binding polypeptide can be controlled on
 10 a separate expression cassette, the expression cassette comprising a promoter operably linked to a polynucleotide encoding the sequence-specific DNA binding polypeptide. In some embodiments, the promoter is inducible, thereby imparting control of expression of the operon based on the inducer. Exemplary inducible promoters (with inducer in parentheses) include, e.g., Ptac (IPTG), Ptrc (IPTG), Pbad (arabinose), Ptet (aTc), Plux (AI-1).

15 Alternatively, in some embodiments, the promoter is constitutive.

[0134] In some embodiments, additional “buffer” nucleotide sequences are inserted between promoters and ribosomal binding sites, between coding sequences and terminators, and/or between coding sequences and a subsequent ribosomal binding site. These sequences act as “buffers” in that they reduce or eliminate regulatory cross-talk between different
 20 coding sequences. In some embodiments, the spacer forms a stem loop, is a native sequence from a metabolic pathway, or is from a 5'-UTR, e.g., obtained from a phage. In some embodiments, the stem loop is a ribozyme. In some embodiments, the ribozyme is RiboJ. In

some embodiments, the buffer sequence is selected from sequences of a given length with nucleotides selected at random. In some embodiments, the buffer sequence is a UP-region of a promoters. UP regions can positively influence promoter strength and are generally centered at position -50 of a promoter (as measured from the start of transcription). *See, e.g.,*

5 Estrem, *et al.*, *PNAS*, 95 (11): 9761-9766 (1988). In some embodiments, the buffer sequence is an extended 5-UTR sequence.

[0135] Exemplary buffer sequences include those listed in the table below:

Sources	Sequences
T5 phage	agttcgaatgagcgcgataccctctacaataatttgttaa
T5 phage	ataaattgataacaaaacctctacaataatttgttaa
T5 phage	ataaattgagagagggtcctctacaataatttgttaa
T5 phage	attaaagaggagaattaacctctacaataatttgttaa
T5 phage	aaacctaatggatcgacctctctacaataatttgttaa
T7 phage	atcggaggggacacggcgcacctctacaataatttgttaa
T7 phage	gctaggtaaacactagcagcctctacaataatttgttaa
T7 phage	atgaaacgacagtgatcgacctctacaataatttgttaa
T7 phage	agggagaccacaacggttcctctacaataatttgttaa
High-transcription escape	attaaaaacctgctaggatcctctacaataatttgttaa
High-transcription escape	ataaaggaaaacggtcaggtcctctacaataatttgttaa
High-transcription escape	ataggttaaaagcctgtcatcctctacaataatttgttaa
Carbon utilization	acaataaaaaatcattacatgttctctacaataatttgttaa
Carbon utilization	agaaagcagcgcgcaaaaatcagctgctctacaataatttgttaa
Carbon utilization	atgagttcattcagacaggcaaatcctctacaataatttgttaa
Carbon utilization	aacttgcagtatttactggatcctctacaataatttgttaa
Carbon utilization	agcccaaaaaaaagtcagtgttgcctctacaataatttgttaa
Carbon utilization	acacagtcactatcttttgaaaaaggtcctctacaataatttgttaa
Anti-escaping sequences	atccggatcctctcccggcctctacaataatttgttaa
	aacaaaaiaaaaagggtcgtcaccctctacaataatttgttaa
T5 phage	agttcgaatgagcgcgatacagttccagattcaggaactataa
T5 phage	ataaattgataacaaaagggtccagattcaggaactataa
T5 phage	ataaattgagagagggttagttccagattcaggaactataa
T5 phage	attaaagaggagaattaacgattccagattcaggaactataa
T5 phage	aaacctaatggatcgacctctctacaataatttgttaa
T7 phage	atcggaggggacacggcgaagtccagattcaggaactataa
T7 phage	gctaggtaaacactagcagcgtccagattcaggaactataa
T7 phage	atgaaacgacagtgatcagttccagattcaggaactataa
T7 phage	agggagaccacaacggttccagattcaggaactataa
High-transcription escape	attaaaaacctgctaggatagttccagattcaggaactataa
High-transcription escape	ataaaggaaaacggtcaggtagttccagattcaggaactataa
High-transcription escape	ataggttaaaagcctgtcagattccagattcaggaactataa
Carbon utilization	acaataaaaaatcattacatgtttagttccagattcaggaactataa
Carbon utilization	agaaagcagcgcgcaaaaatcagctgattccagattcaggaactataa
Carbon utilization	atgagttcattcagacaggcaaatgctccagattcaggaactataa
Carbon utilization	aacttgcagtatttactggatagttccagattcaggaactataa
Carbon utilization	agcccaaaaaaaagtcagtgtttagttccagattcaggaactataa
Carbon utilization	acacagtcactatcttttgaaaaaggtcagattcaggaactataa
Anti-escaping sequences	atccggatcctctcccggatccagattcaggaactataa
	aacaaaaiaaaaagggtcgtcagattcaggaactataa
Stem loops	gatcaccaggggatccccggggaaggat
Stem loops	gatcaccaggggatccccggggaaggat
Stem loops	gatcaccaggggatccccggggaaggat
Stem loops	atgatctggttataaaggatcgggtcattta
Stem loops	gttctccaggggtggatccccctggtggaaatcg
Stem loops	agcagggtaaaagtctatgcaccaa
Stem loops	acgtcgaactatctcagggatattgttgcggtac
Stem loops	acgtcgaactatctcagggatattgttgcggtac
Stem loops	acgtcgaactatctcagggatattgttgcggtac
Stem loops	gactgtcaccggatgtcttccggctgatgactcgtgaggaacag
Stem loops (Ribozyme)	gatcaccaggggatccccggggaaggatcctctacaataatttgttaa

Stem loops	Gatcgccaccggcagctgccggggcgatcaaggatcctacaataatttgtttaa
Stem loops	gatcatcggtagagtaataatgagcagatccccggggaaggatcctacaaataatttgtttaa
Stem loops	attgacitggttattaaaggtaacgggicattitaccctacaataatttgtttaa
Stem loops	Gitccacggggtgggatgagccctcgtggggaatggccctacaataataatttgtttaa
Stem loops	agcatgaggtaaaatgcatgacccaacctcacaataatttgtttaa
Stem loops	Acgtcgactatctcggatgagataattgtgacggtaacctcacaataatttgtttaa
Stem loops	Acgtcgactatctcggatgagataattgtgacggtaacctcacaataatttgtttaa
Stem loops	acgtcgactatctcggatgagataattgtgacggtaacctcacaataatttgtttaa
Stem loops (Ribozyme)	gactgtcaccggatgtgcttccggctgatgagtcctgagggacgaacagccctcacaataatttgtttaa

[0136] The synthetic operons and/or the expression cassette for expressing the sequence-specific DNA binding polypeptide can be carried on one or more plasmids, e.g., in a cell. In some embodiments, the operon and the expression cassette are on different plasmids. In some embodiments, the expression cassette plasmid and/or operon plasmid(s) are low copy plasmids. Low copy plasmids can include, for example, an origin of replication selected from PSC101, PSC101*, F-plasmid, R6K, or IncW.

III. Synthetic Operons

[0137] Embodiments of the present invention also provide for synthetic operons, for example as generated by the methods described herein.

IV. Systems of Synthetic Operons

[0138] Embodiments of the invention also provide for systems comprising synthetic operons and one or more controlling expression cassettes, wherein the expression cassette encodes a sequence-specific DNA binding polypeptide controlling expression of the synthetic operon(s). In some embodiments, the controlling expression cassette(s) are genetic circuits. For example, the expression cassettes can be designed to act as logic gates, pulse generators, oscillators, switches, or memory devices. In some embodiments, the controlling expression cassette are linked to a promoter such that the expression cassette functions as an environmental sensor. In some embodiments, the environmental sensor is an oxygen, temperature, touch, osmotic stress, membrane stress, or redox sensor.

[0139] As explained above, in some embodiments, the expression cassette encodes T7 RNAP or a functional variant thereof. In some embodiments, the T7 RNAP is the output of the genetic circuit(s).

[0140] The operons and expression cassettes can be expressed in a cell. Thus in some embodiments, a cell contains the systems of the invention. Any type of host cell can comprise the system.

V. *Computation*

5 [0141] In some aspects, the invention utilizes a computer program product that determines experimental values for controlling the magnitude of expression of two or more genes. This may be used for example to optimize a system property (e.g. nitrogen fixation levels). In one embodiment, the program code receives one or more input data points, wherein the input data points provide information about one or more regulatory elements and a system property. It
10 then uses a computational method to determine a next data point. In one aspect, the computational method may be a design of experiments (DoE) method.

[0142] In some embodiments, the program code-generated next data point can then be used for further experimentation, e.g., to see if the suggested next data point results in optimized expression level for two or more genes, leading to an improvement in a desired system
15 property. In one aspect, the generation of next data points is repeated until a desired system property level is obtained. In another aspect, the next data points are iteratively generated until the magnitude of expression of two or more genes reaches a desired level.

[0143] In some embodiments, the computer program code may use a computational method that employ numerical analysis or optimization algorithms. In some aspects, the numerical
20 optimization methods may use the is the Nelder-Mead algorithm, the Newton's method, the quasi-Newton method, the conjugate gradient method, an interior point method, a gradient descent, a subgradient method, an ellipsoid method, the Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model..

[0144] In one specific embodiment, the computer program to generate the next data point
25 for experimentation uses the Nelder-Mead algorithm. The computer-implemented method will receive one or more input data points and calculate the reflection point, expansion point or contraction point to computationally determine the next data point to experiment with, based on the input data points.

[0145] In one implementation of the Nelder-Mead algorithm, the program code will take
30 the received input data points as the simplex vertices of an n-dimensional space, having n+1 simplex vertices. Then the objective function will be evaluated for each vertex of the simplex, and the algorithm uses this information to propose a sequence of new coordinates

for evaluation. New coordinates will be determined by the computer code according to the following algorithmic logic:

1. Order the simplex vertices: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
 - 5 2. Calculate x_0 , the center of gravity of all points except x_{n+1} .
 3. Calculate a Reflection coordinate: $x_r = x_0 + \alpha(x_0 - x_{n+1})$
 4. Calculate an Expansion coordinate: $x_e = x_0 + r(x_0 - x_{n+1})$
 5. Calculate a Contraction coordinate: $x_c = x_{n+1} + \rho(x_0 - x_{n+1})$
 6. Calculate Reduction coordinates: $x_i = x_1 + \sigma(x_i - x_1)$ for all $i \in \{2, \dots, n+1\}$
- 10 [0146] The objective function is evaluated at these points and used to determine a new simplex according to the following criteria:
1. If the Reflection, Expansion or Contraction coordinates are better than the worst simplex point, x_{n+1} , define a new simplex by replacing the worst simplex point with the best of the three (Reflection, Expansion or Contraction).
 - 15 2. Otherwise, define a new simplex by combining the best simplex point with the Reduction coordinates.

[0147] In one embodiment, a computer program product is provided comprising a tangible computer readable medium storing a plurality of instructions for controlling a processor to perform an operation for determining an experimentation point for controlling the magnitude
 20 of expression of two or more genes, the instructions comprising receiving one or more input data points, wherein the input data points provide information about one or more regulatory elements and a system property; and determining, with a computer, a next data point using a computational method, wherein the next data point provides information about the one or more regulatory elements.

25 [0148] Fig. 25 shows a block diagram of an example computer system 600 usable with system and methods according to embodiments of the present invention. The computer system 600 can be used to run the program code for various method claims according to embodiments of the present invention.

30 [0149] Any of the computer systems mentioned herein may utilize any suitable number of subsystems. Examples of such subsystems are shown in Fig. 25 in computer apparatus 600. In some embodiments, a computer system includes a single computer apparatus, where the subsystems can be the components of the computer apparatus. In other embodiments, a computer system can include multiple computer apparatuses, each being a subsystem, with
 35 internal components.

[0150] The subsystems shown in Fig. 25 are interconnected via a system bus 675. Additional subsystems such as a printer 674, keyboard 678, fixed disk 679, monitor 676, which is coupled to display adapter 682, and others are shown. Peripherals and input/output (I/O) devices, which couple to I/O controller 671, can be connected to the computer system by any number of means
5 known in the art, such as serial port 677. For example, serial port 677 or external interface 681 can be used to connect computer system 600 to a wide area network such as the Internet, a mouse input device, or a scanner. The interconnection via system bus 675 allows the central processor 673 to communicate with each subsystem and to control the execution of instructions from system memory 672 or the fixed disk 679, as well as the exchange of information between subsystems.
10 The system memory 672 and/or the fixed disk 679 may embody a computer readable medium. Any of the values mentioned herein can be output from one component to another component and can be output to the user.

[0151] A computer system can include a plurality of the same components or subsystems, e.g., connected together by external interface 681 or by an internal interface. In some embodiments,
15 computer systems, subsystem, or apparatuses can communicate over a network. In such instances, one computer can be considered a client and another computer a server, where each can be part of a same computer system. A client and a server can each include multiple systems, subsystems, or components.

[0152] It should be understood that any of the embodiments of the present invention can be
20 implemented in the form of control logic using hardware and/or using computer software in a modular or integrated manner. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will know and appreciate other ways and/or methods to implement embodiments of the present invention using hardware and a combination of hardware and software.

[0153] Any of the software components or functions described in this application may be
25 implemented as software code to be executed by a processor using any suitable computer language such as, for example, Java™, C++ or Perl™ using, for example, conventional or object-oriented techniques. The software code may be stored as a series of instructions or commands on a computer readable medium for storage and/or transmission, suitable media include random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard-drive or a floppy
30 disk, or an optical medium such as a compact disk (CD) or DVD (digital versatile disk), flash memory, and the like. The computer readable medium may be any combination of such storage or transmission devices.

[0154] Such programs may also be encoded and transmitted using carrier signals adapted for transmission via wired, optical, and/or wireless networks conforming to a variety of protocols, including the Internet. As such, a computer readable medium according to an embodiment of the present invention may be created using a data signal encoded with such programs. Computer readable media encoded with the program code may be packaged with a compatible device or provided separately from other devices (e.g., via Internet download). Any such computer readable medium may reside on or within a single computer program product (e.g. a hard drive, a CD, or an entire computer system), and may be present on or within different computer program products within a system or network. A computer system may include a monitor, printer, or other suitable display for providing any of the results mentioned herein to a user.

[0155] Any of the methods described herein may be totally or partially performed with a computer system including a processor, which can be configured to perform the steps. Thus, embodiments can be directed to computer systems configured to perform the steps of any of the methods described herein, potentially with different components performing a respective steps or a respective group of steps. Although presented as numbered steps, steps of methods herein can be performed at a same time or in a different order. Additionally, portions of these steps may be used with portions of other steps from other methods. Also, all or portions of a step may be optional. Additionally, any of the steps of any of the methods can be performed with modules, circuits, or other means for performing these steps.

[0156] The specific details of particular embodiments may be combined in any suitable manner or varied from those shown and described herein without departing from the spirit and scope of embodiments of the invention.

[0157] The above description of exemplary embodiments of the invention has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form described, and many modifications and variations are possible in light of the teaching above. The embodiments were chosen and described in order to best explain the principles of the invention and its practical applications to thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated.

EXAMPLES

[0158] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Use of the Nelder-Mead method to optimize efficiency of operon discovery

5 [0159] This examples illustrates how to recombinant and computationally manipulate and select native gene cluster coding sequences and heterologous regulatory sequences. We have termed this process “refactoring”, which comprises optimization of multiple genes, regulation of the gene cluster, and establishment of the genetic context for the biological circuit. Refactoring complex gene clusters and engineering metabolic pathways requires numerous
10 iterations between design, construction and evaluation in order to improve a desired system property, e.g. higher product titers, lower toxicity, or improved nitrogen fixation.

[0160] One common way to affect these properties is to modify gene expression levels within the system, even if the direct relationship between gene expression and the system property is unknown. Making quantitative changes to gene expression can be achieved
15 through the use of regulatory elements, e.g. promoters and ribosome binding sites, that exhibit rationally predictable behavior.

[0161] It is possible to utilize numerical optimization methods to guide selection of regulatory elements in order to alter gene expression and to improve desired system properties. One relevant algorithm is the Nelder-Mead method, a nonlinear optimization
20 algorithm that minimizes an objective function in multidimensional space. We use the Nelder-Mead method to optimize a system property where each dimension in algorithmic space corresponds to expression of a gene in the engineered system. Points in this space represent a particular combination of expression levels for the genes in the system. As a result, each point may be considered a uniquely engineered strain. The algorithm is used to
25 suggest new coordinates in space that improve the system property. New strains can be engineered by modifying regulatory elements to attain the suggested levels of gene expression. After evaluating the performance of the new strains, the algorithm can be used to predict subsequent modifications. This process iterates until the system property has been improved a desired amount.

30 [0162] The Nelder-Mead method relies on the concept of a simplex, which is an object in N dimensional space having N+1 vertices. The objective function is evaluated at each vertex of the simplex, and the algorithm uses this information to propose a sequence of new

coordinates for evaluation. New coordinates are proposed according to the following process:

1. Order the simplex vertices: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
2. Calculate x_o , the center of gravity of all points except x_{n+1} .
3. Calculate a Reflection coordinate: $x_r = x_o + \alpha(x_o - x_{n+1})$
- 5 4. Calculate an Expansion coordinate: $x_e = x_o + r(x_o - x_{n+1})$
5. Calculate a Contraction coordinate: $x_c = x_{n+1} + \rho(x_o - x_{n+1})$
6. Calculate Reduction coordinates: $x_i = x_1 + \sigma(x_i - x_1)$ for all $i \in \{2, \dots, n+1\}$

[0163] The objective function is evaluated at these points and used to determine a new simplex according to the following criteria:

- 10 1. If the Reflection, Expansion or Contraction coordinates are better than the worst simplex point, x_{n+1} , define a new simplex by replacing the worst simplex point with the best of the three (Reflection, Expansion or Contraction).
2. Otherwise, define a new simplex by combining the best simplex point with the Reduction coordinates.

15 [0164] These steps constitute an iteration of the algorithm. The newly defined simplex becomes the seed for generating new coordinates during the next iteration of the algorithm. Iterations typically continue until one of the coordinates in the simplex crosses a desired threshold for objective function evaluation. We have optimized the performance of a nitrogen fixation operon by varying the selection of promoters that control expression of individual genes. We initially refactored the *nifEN* operon so
 20 that each gene was expressed under the control of a unique T7 promoter (Fig. 1). To assess the impact of refactoring the *nifEN* operon, we quantitatively measured the capacity of the synthetic operon to complement a *nifEN* knockout strain and recover the ability to fix nitrogen (Fig. 2). Our refactored system showed limited ability to fix nitrogen (20% of wild-type activity).

[0165] We subsequently applied the Nelder-Mead method to optimize *nifE* and *nifN* gene expression
 25 with the goal of improving nitrogen fixation rates. Our algorithmic space consisted of two dimensions, *nifE* and *nifN* expression. Our coordinate system was scaled to the strength of the promoters controlling these genes. To enable varied levels of gene expression, we generated and characterized a library of mutant T7 promoters (Fig. 3). Our library covers three order of magnitude of gene expression (This is the same library that is described in U.S. Patent Publication No. 2013-0005590. Here, it is characterized
 30 for behavior in *Klebsiella oxytoca*). We then randomly selected mutants from the library of T7 promoters to generate two additional strains with rationally altered levels of *nifE* and *nifN*

expression. The strength of T7 promoters used in these three strains defined our initial simplex. We evaluated nitrogen fixation for each strain in the simplex (strain 1: 20%, strain 2: 9%, strain 3: 12%) and used the algorithm to calculate Reflection coordinates (Fig. 4). To construct the strain that matched the Reflection coordinates, we chose promoters from our library nearest to the coordinates in strength. We evaluated nitrogen fixation in this Reflection strain and found that it significantly outperformed (52%) our initial strains (Fig. 5).

[0166] Our improved strain had surprising results and surpassed expectations, and performed sufficiently for downstream applications. To reach higher levels of gene expression, stronger promoters can be engineered and used in the methods of the invention. Alternatively, complimentary changes to multiple regulatory elements, *e.g.*, the promoter and ribosome binding site for a given gene, can be used to achieve desired expression levels. This involves describing the strengths of each type of element in common units of expression. This example demonstrates that new strains can be engineered by modifying regulatory elements to attain the desired levels of gene expression. The example also illustrates the use of numerical optimization methods, such as, but not limited to the Nelder-Mead method, to guide selection of regulatory elements in order to alter gene expression and to improve desired system properties.

Example 2: Refactoring Nitrogen Fixation

[0167] This example demonstrates the method of refactoring the nitrogen fixation gene cluster. The method includes steps that comprise: 1) removing host regulation and implement synthetic, orthogonal regulation; 2) tracking the contribution of each regulatory part to gene cluster function; 3) promoting modularity and integration with synthetic circuits; and 4) creating a platform amenable to rational optimization. In certain embodiments, the method of refactoring nitrogen fixation comprises reducing cluster to characteristic genes and assembling synthetic cluster.

[0168] The *nif* gene cluster from *Klebsiella oxytoca* has been one of the primary models for study of the nitrogenase enzyme (Fig. 7; *see*, Rubio and Ludden, Maturation of Nitrogenase: a Biochemical Puzzle, *J. Bacteriology*, 2005). It is a concise gene cluster, encompassing 20 genes in 7 operons within 25kb of DNA. The nitrogenase enzyme is composed of two major units, Component I and Component II, that interact to facilitate the reduction of multiply bonded gases like N₂. Within the enzyme complex, multiple Fe-S clusters are responsible for active site chemistry and electron transfer to the active site. The majority of the genes in the

gene cluster are involved in Fe-S cluster biosynthesis, chaperoning and insertion into the final enzyme complex.

[0169] Nearly every *nif* gene produces a protein with a function known to be essential to nitrogenase assembly or function (*see*, Simon, Homer and Roberts, Perturbation of *nifT* expression in *Klebsiella pneumoniae* has limited effect on nitrogen fixation, *J. Bacteriology*, 1996 and Gosink, Franklin and Roberts, The product of the *Klebsiella pneumoniae nifX* gene is a negative regulator of the nitrogen fixation (*nif*) regulon, *J Bacteriology*, 1990). Two genes, *nifL* and *nifA*, encode the master regulatory proteins. The *nifT* gene has no known function, and eliminating it has little effect on nitrogen fixation. Additionally, while elimination of *nifX* has minor effect on nitrogen fixation, its overexpression detrimentally reduces enzyme activity. For these reasons, we chose to eliminate *nifL*, *nifA*, *nifT* and *nifX* from our refactored gene cluster.

[0170] We designed synthetic genes by codon randomizing the DNA encoding each amino acid sequence. Protein coding sequences were based on the sequence deposited in the NCBI database (X13303; *see*, Arnold *et al.*, Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *JMB*, 1988). Codon selection was performed by DNA2.0 using an internal algorithm and two guiding criteria. We specified that our genes express reasonably well in both *E. coli* and *Klebsiella*. Also, we specified that our codon usage be as divergent as possible from the codon usage in the native gene. While designing synthetic genes, we scanned each proposed sequence for a list of undesired features and rejected any in which a feature was found. The feature list includes restriction enzyme recognition sites, transposon recognition sites, repetitive sequences, sigma 54 and sigma 70 promoters, cryptic ribosome binding sites, and rho independent terminators. Fig. 18 shows DNA sequences for native genes and synthetic genes, as well as the percent common nucleotide and codon identities between each pair.

[0171] Synthetic ribosome binding sites were chosen to match the strength of each corresponding native ribosome binding site. To characterize the strength of a given native ribosome binding site, we constructed a fluorescent reporter plasmid in which the 150bp surrounding a gene's start codon (from -60 to +90) were fused to the mRFP gene (Fig. 8). The chimera was expressed under control of the Ptac promoter, and fluorescence was measured via flow cytometry (Fig. 9). To generate synthetic ribosome binding sites, we constructed a library of reporter plasmids using 150bp (-60 to +90) of a synthetic expression cassette. Briefly, a synthetic expression cassette consisted of a random DNA spacer, a degenerate sequence encoding an RBS library, and the coding sequence for each synthetic

gene. We screened multiple clones to identify the synthetic ribosome binding site that best matched the native ribosome binding site (Fig. 10).

[0172] We constructed synthetic operons that consisted of the same genes as the native operons. This strategy enabled us to knock out a native operon from *Klebsiella* and
5 complement the deletion using the synthetic counterpart.

[0173] Each synthetic operon consisted of a Ptac promoter followed by synthetic gene expression cassettes (random DNA spacer, synthetic rbs, synthetic coding sequence) and a transcription terminator. The random DNA spacer serves to insulate the expression of each synthetic coding sequence from preceding cassettes. Each synthetic operon was scanned to
10 remove unintended regulatory sequences (similar to the process used during synthetic gene design and synthesis).

[0174] In two cases, we encountered synthetic operons that showed no functional complementation in the corresponding knockout strain (nifHDKTY and nifUSVWZM). To debug the synthetic operons, we broke the operon into constituent gene expression cassettes.
15 We then constructed chimeric operons, wherein some cassettes had synthetic components and other cassettes were native genes and their ribosome binding sites (Fig. 11). This strategy enabled us to test each chimeric operon for functional complementation and quickly identify the problematic synthetic expression cassettes. With further analysis of problematic expression cassettes, we were able to diagnose and correct errors in the fully synthetic
20 operons. Fig. 12 illustrates a list of errors in the two operons.

[0175] Each synthetic operon was initially designed to be controlled by a Ptac inducible promoter. By titrating IPTG concentration, we could precisely specify promoter strength and corresponding synthetic operon expression. This enabled us to vary expression level to identify optimal operon function. We found that each synthetic operon required different
25 levels of IPTG concentration for optimal function (Fig. 13).

[0176] We utilized the T7 Wires system to decouple the Ptac promoter from each synthetic operon. By inserting the wire between the promoter and transcriptional unit, we achieved two significant milestones. First, we gained the ability to modulate the transcriptional signal through the use of various mutant T7 promoters. This allowed us to shift optimal operon
30 function to a single inducer concentration by selecting corresponding mutant T7 promoters. Second, we modularized control of the synthetic operon (Fig. 14). That is, any genetic circuit can control the synthetic operon provided that it can produce the necessary T7 RNAP concentration to drive each wire.

[0177] We adopted a hierarchical approach to assembling individual operons into a fully synthetic cluster. First, we assembled three operons into half clusters (*nifJ-nifHDKY-nifEN* and *nifUSVWZM-nifF-nifBQ*) and demonstrated the ability of each synthetic half cluster to complement function in a corresponding knockout strain. Next, we combined the two half clusters into a full synthetic cluster and demonstrated nitrogen fixation in a complete *nif* knockout strain (Fig. 15).

[0178] We have shown that the use of T7 Wires produces a modular synthetic gene cluster. We have demonstrated that the use of either controller #1 or controller #2 produces the same functional performance from the synthetic cluster (Fig. 16). In controller #1, T7 RNAP is under control of the Ptac promoter. In controller #2, T7 RNAP is under control of the Ptet promoter.

[0179] Fig. 17 shows a schematic of the full biological cluster, with each part detailed. Fig. 19 shows the parts list of the synthetic controllers. Fig. 20 shows names, sequences and strengths of each component of the full cluster.

[0180] We have further demonstrated that complex genetic circuits can be used to produce functional performance of the synthetic gene cluster. We constructed a genetic circuit encoding the logic “A and not B” and used this circuit to control T7 RNAP. In this circuit, the “A and not B” logic corresponds to the presence or absence of the inducers, IPTG and aTc, such that the cell computes “IPTG and not aTc.” The circuit was constructed by modifying controller #1 to include the *cl* repressor binding sites OR1 and OR2 in the Ptac promoter to produce controller #3. Additionally, plasmid pNOR1020 (*see, e.g.,* Tamsir and Voigt *Nature* 469:212–215 (2011)) encodes the repressor *cI* under control of the Ptet promoter. When pNOR1020 and controller #3 are co-transformed, they produce the logic circuit “IPTG and not aTc.”

Ptac controller #1 promoter sequence:
 tattctgaaatgagctgttgacaattaatcatcggctcgtataatgtgtggaattgtgagcggataacaatt

Controller #3 promoter sequence:
 tattaacaccgtgcgtgttgacagctatacctctggcggttataatgctagcggattgtgagcggataacaatt

Input	Expected Logic Output	Synthetic Nitrogen Fixation Performance (%WT)
No inducer	0	<0.5%
1 mM IPTG	1	9%
50 ng/ml aTc	0	<0.5%
1 mM IPTG and 50 ng/ml aTc	0	<0.5%

[0181] In this experiment, we also included controller #1 as a performance reference. Under inducing conditions (1mM IPTG), controller #1 exhibits 12% of WT fixation.

Example 3: Refactoring the Bacterial type III secretion system (T3SS)

5 [0182] This example illustrates the use of the method described herein to completely refactor the Bacterial type III secretion system (T3SS). This example also illustrates that the refactored synthetic operons of T3SS are controllable and function independently of all native control and regulation.

[0183] Bacterial type III secretion systems (T3SS) are valuable because, unlike
10 conventionally used Sec and Tat pathways, they translocate polypeptides through both inner and outer membranes. This enables the delivery of protein directly to culture media, which can be one of the critical requirements in engineered bacterial technology. For example, toxic proteins can be removed from the cytoplasm without being allowed in the periplasm and functional enzymes (*e.g.*, cellulases) which need to work outside the cell, can be
15 delivered directly into the media.

[0184] However, the difficulty with utilizing T3SS in engineered bacterial systems is twofold. T3SS generally exist in pathogenic bacteria which utilize these mechanisms for invasion of host cells. Thus, T3SS are very tightly regulated in the cell and are difficult to control independently. Because of this, we chose to use methods of the present invention to
20 completely refactor T3SS and test the function of the refactored operons in knockout cells.

[0185] The term “refactoring” refers to a process that involves optimization of multiple genes, regulation of a gene cluster, and establishment of the genetic context for a biological circuit. Refactoring complex gene clusters and engineering biological pathways requires numerous iterations between design, construction and evaluation in order to improve a
25 desired system property. Briefly, refactoring includes breaking down a biological system into its component parts and rebuilding it synthetically. It also involves removing all native control and regulation of the biological system in order to replace it with a mechanism that provides independent control.

[0186] This example illustrates a method of recoding 18 genes of the bacterial type III
30 secretion systems. The term “recoding” refers to a method of removing or replacing sequence of a gene in order to reduce or eliminate any native regulation elements, while also preserving the protein sequence encoded by the gene. The genes of the type III secretion

system were recoded using an algorithm provided by DNA2.0 (Menlo Park, CA) in which individual codons of each gene are re-selected such that the gene encodes the same protein, but with maximum dissimilarity with the native sequence.

- 5 [0187] The 18 genes are arranged in two bacterial operons. Each gene is a recoded version of a native gene from *Salmonella Typhimurium*. Each gene is coupled to a synthetic ribosome binding site (RBS) sequence that sets an appropriate expression level for each individual gene. Details of the synthetic RBS selection are described below. The operons can be induced with any desired promoter. In this example, simple inducible promoters are used. The recoded T3SS operons can be attached to any genetic control circuit as needed.
- 10 [0188] To select a synthetic RBS sequence that best matches the native expression level of each of the 18 genes of the bacterial type III secretion systems, we measured the expression of each gene in the natural system. We cloned the 36-base region upstream on the start codon, along with the 36-bases of coding region fused to an RFP (Red Fluorescent Protein). This was cloned into a plasmid with a constitutive promoter.
- 15 [0189] This construct was transformed into *Salmonella Typhimurium* SL1344 and grown overnight at 37 °C in PI-1 inducing media (LB with 17g/L NaCl). The culture was subcultured into fresh inducing media to an OD₂₆₀ of 0.025, grown for 2 hours at 37 °C until cells reached log-phase. Fluorescence was measured on a cytometer. The geometric mean of RFP fluorescence across at least 10,000 cells was used as the measure of protein expression.
- 20 [0190] To find ribosomal binding sequences to test, we utilized the Ribosome Binding Site Calculator, identified known RBS sequences from the Registry of Standard Biological Parts and generated a series of randomized sequences. The randomized sequences comprise the following formats: CTTGGGCACGCGTCCATTAANNAGGANNAATTAAGC;
TGGGCACGCGTCCATTAANNAGGANNAATTATTAGC;
25 TACTTGGGCACGCGTCCATTAANNAGGANNAATAGC;
CTTGGGCACGCGTCCATTAANAAGGAGNAATTAAGC;
CTTGGGCACGCGTCCATTANTAAGGAGGNATTAAGC.
- [0191] All RBS sequences were cloned into the RBS test vector (Fig. 21) along with the first 36 bases of the synthetic gene they were generated to drive. We followed the same experimental procedure
30 used to measure the expression of each gene in the natural system. Of the randomized RBS, 12 - 48 colonies of each randomized sequence was tested. The

synthetic construct that best matched the native expression level was selected and sequenced. This sequence was then used in the construction of the refactored operons.

[0192] Two operons were assembled. The first, “prg-org” contains 6 genes, and the second “inv-spa” contains 13 genes. These genes are allocated to each operon in the same manner as
5 in the wild-type system. However, the order of genes in each operon is arranged on the basis of measured expression level from strongest to weakest. Operons were assembled by placing the selected synthetic RBS in front of its corresponding synthetic gene sequence. Restriction enzyme binding sites were added between genes or pairs of genes in order to facilitate future manipulation. The entire sequence was synthesized by DNA2.0. The synthetic operon was
10 cloned into a low-copy test vector and placed under the control of an inducible promoter (e.g., pTac or pBad – IPTG or Arabinose induction). A reporter plasmid was created containing a native *Salmonella* secretable effector protein which was fused to a FLAG epitope tag for identification. This reporter was placed under a strong constitutive promoter. Fig. 22B shows a schematic of the prg-org operon test vector and a reporter plasmid.

[0193] We also generated two operon knockout (prg-org and inv-spa) *Salmonella* SL1344 cell lines using the method described in Datsenko, Wanner, *Proc. Natl. Acad. U.S.A.*, 2000. Fig. 22A shows a schematic of Δ prg-org *Salmonella* SL1344 knock-out strain. The inv-spa and prg-org operons are boxed. Fig. 23A shows that the Δ prg-org knock-out strain does not express the prg-org operon.

[0194] The test plasmid (or the control plasmid) and the reporter plasmid were transformed into the appropriate knockout strain. The strains were grown from colony overnight in low-salt media (LB with 5g/L NaCl) at 37 °C. The cultures were subcultured to an OD₂₆₀ of 0.025 in fresh low-salt media and grown for 2 hours. The cultures were diluted 1:10 into high-salt, inducing media (LB with 17g/L of NaCl) in 50mL un baffled flasks and grown for 6-8 hours.
25 1 mL of each culture was spun down at 3000xg for 5 minutes, then the supernatant filtered through a 0.2 μ M filter. This culture was then run on an SDS-PAGE gel and a western blot performed with an anti-FLAG antibody.

[0195] Fig. 23B shows that the synthetic refactored prg-org operon in *Salmonella* Δ prg-org cells can be controlled by the addition of IPTG. The level of expression is comparable to that
30 generated from the natural PprgH promoter.

Example 4: Refactoring Nitrogen Fixation Gene Cluster from *Klebsiella oxytoca*

[0196] Bacterial genes associated with a single trait are often grouped in a contiguous unit of the genome known as a gene cluster. It is difficult to genetically manipulate many gene

clusters due to complex, redundant, and integrated host regulation. We have developed a systematic approach to completely specify the genetics of a gene cluster by rebuilding it from the bottom-up using only synthetic, well-characterized parts. This process removes all native regulation, including that which is undiscovered. First, all non-coding DNA, regulatory proteins, and non-essential genes are removed. The codons of essential genes are changed to create a DNA sequence as divergent as possible from the wild-type gene. Recoded genes are computationally scanned to eliminate internal regulation. They are organized into operons and placed under the control of synthetic parts (promoters, ribosome binding sites, and terminators) that are functionally separated by insulator parts. Finally, a controller consisting of genetic sensors and circuits regulates the conditions and dynamics of gene expression. We applied this approach to an agriculturally relevant gene cluster from *Klebsiella oxytoca* encoding the nitrogen fixation pathway for converting atmospheric N₂ to ammonia. The native gene cluster consists of 20 genes in 7 operons and is encoded in 23.5kb of DNA. We constructed a refactored gene cluster that shares little DNA sequence identity with wild-type and for which the function of every genetic part is defined. This work demonstrates the potential for synthetic biology tools to rewrite the genetics encoding complex biological functions to facilitate access, engineering, and transferability.

INTRODUCTION

[0197] Many functions of interest for biotechnology are encoded in gene clusters, including metabolic pathways, nanomachines, nutrient scavenging mechanisms, and energy generators (1). Clusters typically contain internal regulation that is embedded in the global regulatory network of the organism. Promoters and 5'-UTRs are complex and integrate many regulatory inputs (2, 3). Regulation is highly redundant; for example, containing embedded feedforward and feedback loops (4). Regulation can also be internal to genes, including promoters, pause sites, and small RNAs (5, 6). Further, genes often physically overlap and regions of DNA can have multiple functions (7). The redundancy and extent of this regulation makes it difficult to manipulate a gene cluster to break its control by native environmental stimuli, optimize its function, or transfer it between organisms. As a consequence, many clusters are cryptic, meaning that laboratory conditions cannot be identified in which they are active (8).

[0198] Gene clusters have been controlled from the top-down by manipulating the native regulation or adding synthetic regulation in an otherwise wild-type background (9). For example, either knocking out a repressor or overexpressing an activator has turned on clusters encoding biosynthetic pathways (10-14). When the cluster is a single operon, it has been

shown that a promoter can be inserted upstream to induce expression (15). The entire echinomycin biosynthetic cluster was transferred into *E. coli* by placing each native gene under the control of a synthetic promoter (16).

[0199] In engineering, one approach to reduce the complexity of a system is to “refactor” it, a term borrowed from software development where the code underlying a program is rewritten to achieve some goal (e.g., stability) without changing functionality (17). This term was first applied to genetics to describe the top-down simplification of a phage genome by redesigning known genetic elements to be individually changeable by standard restriction digest (18). Here, we use it to refer to a comprehensive bottom-up process to systematically eliminate the native regulation of a gene cluster and replace it with synthetic genetic parts and circuits (Fig. 26). The end product is a version of the gene cluster whose DNA sequence has been rewritten, but it encodes the same function. The design process occurs on the computer, and then the resulting DNA sequence is constructed using DNA synthesis (19). The first step of the process is to remove all non-coding DNA, and regulatory genes. Next, each essential gene is recoded by selecting codons that produce a DNA sequence that is as distant as possible from the wild-type sequence. The intent is to introduce mutations throughout the gene to eliminate internal regulation (including that which is undiscovered), such as operators, promoters, mRNA secondary structure, pause sites, methylation sites, and codon regulation. Recoded sequences are further scanned by computational methods to identify putative functional sequences, which are then removed. The recoded genes are organized into artificial operons and the expression levels are controlled by synthetic ribosome binding sites (RBSs), and insulator sequences physically separate the genes. The end result is a refactored gene cluster whose native regulation has been removed and has been organized into a set of discrete, well-characterized genetic parts.

[0200] Once the native regulation has been removed, synthetic regulation can be added back to control the dynamics and conditions under which the cluster is expressed. Constructing such regulation has been a major thrust of synthetic biology and involves the design of genetic sensors and circuits and understanding how to connect them to form programs (20). In our design, we genetically separate the sensing/circuitry from the refactored pathway by carrying them on different low copy plasmids (Fig. 26). The plasmid containing the sensors and circuits is referred to as the “controller” and the output of the circuits lead to the expression of an engineered T7 polymerase (T7*). The refactored cluster is under the control of T7 promoters. One advantage of this organization is that T7

polymerase is orthogonal to native transcription and the T7 promoters are tightly off in the absence of the controller. In addition, changing the regulation is simplified to swapping the controller for one that contains different sensors and circuits, so long as the dynamic range of T7* is fixed.

5 [0201] As a demonstration, we have applied this process to refactor the gene cluster encoding nitrogen fixation in *Klebsiella oxytoca* (21). Nitrogen fixation is the conversion of atmospheric N₂ to ammonia (NH₃), so that it can enter metabolism (22). Industrial nitrogen fixation through the Haber-Bosch process is used to produce fertilizer. Many
10 microorganisms fix nitrogen and the necessary genes typically occur together in a gene cluster, including the nitrogenase subunits, the metallocluster biosynthetic enzymes and chaperones, e- transport, and regulators (Fig. 27A) (23, 24). The gene cluster from *K. oxytoca* has been a model system for studying nitrogen fixation and consists of 20 genes encoded in 23.5kb of DNA (Fig. 26, top) (25). The biosynthesis of nitrogenase is tightly regulated by a two-layer transcriptional cascade in response to fixed nitrogen, oxygen, and
15 temperature (26). The complete cluster has been transferred to *E. coli*, thus demonstrating that it has all of the genes necessary for nitrogen fixation (27). The encoding of this function is complex, many of the genes overlap, the operons are oriented in opposite directions, and there are many putative hidden regulatory elements, including internal promoters and hairpins (25). The purpose of refactoring is to reorganize the cluster, simplify its regulation, and
20 assign a concrete function to each genetic part.

RESULTS

Tolerance of the Native Gene Cluster to Changes in Expression

[0202] Prior to refactoring a cluster, a robustness analysis is performed to determine the tolerances of a gene or set of genes to changes in expression level (Fig.27B). This informs
25 the grouping of genes into operons and the selection of synthetic parts to obtain desired expression levels. In the wild-type background, genes are knocked out and complemented under inducible control. The tolerance is obtained by measuring nitrogenase activity as a function of the activity of the inducible promoter.

[0203] Nitrogenase function is notably sensitive to expression changes and each tolerance
30 has a clear optimum (Fig. 27B). The chaperone NifY is required to achieve full activity and broadens the tolerance to changes in expression level. NifT did not have an effect on activity, as observed previously (28), and it is frequently absent from homologous clusters (29).

The genes controlling electron transport (*nifJ* and *nifF*) need to be expressed at low levels, and activity falls rapidly as expression increases. The optima for genes participating in the metal cluster biosynthetic pathways vary. The *nifUSVWZM* operon, which encodes proteins for early Fe-S cluster formation and proteins for component maturation, needs to be
5 expressed at low levels, whereas *nifBQ*, encoding proteins for FeMo-co core synthesis and molybdenum integration, need to be expressed 10-fold higher. NifEN is tolerant to varied expression levels. However, activity is lost with the inclusion of *nifX*, which has been characterized as a negative regulator (30). The native cluster also includes the regulatory proteins NifL and NifA, which integrate environmental signals (26). The genes *nifT*, *nifX*,
10 and *nifLA* are not included in the refactored cluster.

The Complete Refactored Gene Cluster

[0204] The nitrogenase activities of the refactored operons were measured as a function of the IPTG-inducible P_{lac} promoter (Fig. 28A). Each operon has a different optimum. To combine the operons, the P_{lac} promoters were replaced with T7 promoters that have a strength
15 close to the measured optimum (Fig. 28B and *Materials and Methods* section). The nitrogenase genes (*nifHDK*) are highly expressed in *Klebsiella* under fixing conditions (up to 10% of cell protein) (31), so the strongest promoter was used to control this operon (T7.WT, 0.38 REU) (32). A long operon was built to include the *nifEN* and *nifJ* genes, where the lower expression required for *nifJ* was achieved through transcriptional attenuation. The *nifF*
20 gene was encoded separately under the control of a medium strength promoter (T7.3, 0.045 REU). Finally, the *nifUSVWZM* and *nifBQ* operons were controlled by weak promoters (T7.2, 0.019 REU). Each of the individual refactored operons under the control of a T7 promoter was able to recover the activity observed from the P_{lac} promoter and corresponding optimal IPTG concentration (Fig. 28C).

25 [0205] Transitioning the control to T7* and T7 promoters facilitates the assembly of the complete cluster from refactored operons. We first assembled half-clusters using Gibson Assembly (33) and verified their function in strains with the corresponding genes deleted. The first half-cluster consisted of the *nifHDKYENJ* operon. The second half-cluster was assembled from the *nifBQ*, *nifF*, and *nifUSVWZM* operons. The half-clusters were able to
30 recover $18\% \pm 0.7\%$ and $26\% \pm 8.4\%$ of wild-type activity, respectively. The full synthetic cluster was assembled from both half-clusters (Fig. 29), and its activity measured in a strain where the full cluster is deleted. The synthetic gene cluster recovers nitrogenase activity at

7.4% \pm 2.4% of the wild-type (Fig. 30A). Strains carrying the synthetic gene cluster utilized ambient N₂ as a nitrogen source, growing 3.5-fold slower than the wild-type strain (Fig. 37) and incorporating ¹⁵N-labelled nitrogen into 24% \pm 1.4% of their cellular nitrogen content, as measured by isotope ratio mass spectrometry (IRMS) (Fig. 30B).

5 [0206] The complete refactored cluster consists of 89 genetic parts, including a controller, and the function of each part is defined and characterized. Therefore, the genetics of the refactored system are complete and defined by the schematic in Fig. 29. However, the process of simplification and modularization reduces activity (18). This is an expected outcome of refactoring a highly evolved system.

10 **Swapping Controllers to Change Regulation**

[0207] The separation of the controller and the refactored cluster simplifies changing the regulation of the system. This can be achieved by transforming a different controller plasmid, as long as the dynamic range of the T7* RNAP expression is preserved. To demonstrate this, we constructed two additional controllers (Fig. 30A). Controller #2
15 changes the chemical that induces the system by placing the expression of T7* RNAP under the control of the aTc-inducible P_{tet} promoter. When induced, Controller #2 produces nitrogenase activity identical to Controller #1 (7.2% \pm 1.7%). The controller can also serve as a platform to encode genetic circuits to control regulatory dynamics or to integrate multiple sensors. To this end, Controller #3 contains two inducible systems (IPTG and aTc)
20 and an ANDN gate (34, 35). In the presence of IPTG and the absence of aTc, nitrogen fixation is 6.6% \pm 1.7% of wild-type activity. These controllers represent the simplicity by which the regulation of the refactored cluster can be changed.

[0208] In addition to making it possible to add new regulation, the process of refactoring eliminates the native regulation of the cluster. This is demonstrated through the decoupling
25 of nitrogenase activity from the environmental signals that normally regulate its activity. For example, ammonia is a negative regulator that limits overproduction of fixed nitrogen (26). In the presence of 17.5 mM ammonia, no nitrogenase activity is observed for the wild-type cluster (Fig. 30C). In contrast, the refactored gene cluster maintains activity in the presence of ammonia (1.1% \pm 0.5%). Interestingly, this 7-fold reduction of activity is not due to
30 residual regulation present in the system. Rather, it occurs because the addition of ammonia to the media reduces the output of the controller by 4.5-fold (Fig. 30C). In theory, this could

be fixed by increasing the expression level of T7* RNAP, but it speaks to the need to create genetic circuits that are robust to environmental context.

DISCUSSION

[0209] The objective of refactoring is to facilitate the forward engineering of multi-gene systems encoded by complex genetics. Native gene clusters are the product of evolutionary processes; thus, they exhibit high redundancy, efficiency of information coding, and layers of regulation that rely on different biochemical mechanisms (36-38). These characteristics inhibit the quantitative alteration of function by part substitution, because the effect can become embedded in a web of interactions. Here, modularizing the cluster, physically separating and insulating the parts, and simplifying its regulation have guided the selection and analysis of part substitutions. The information gleaned from screening the permutations in a refactored system can be cleanly fed back into the design cycle.

[0210] The refactored cluster can also serve as a platform for addressing questions in basic biology. First, it allows for the impact of regulatory interactions to be quantified in isolation. For example, in the natural system, one feedback loop could be embedded in many other regulatory loops. Systematically removing such regulation provides a clean reference system (potentially less active and robust than wild-type) from which improvements can be quantified as a result of adding back regulation. It also serves as a basis for comparison of radically different regulatory programs or organizational principles; for example, to determine the importance of temporal control of gene expression (4, 39) or the need for genes to be encoded with a particular operon structure (40, 41). Second, the process of reconstruction and debugging is a discovery mechanism that is likely to reveal novel genetics and regulatory modes. In this work, the improvement from 0% to 7% revealed only minor changes: misannotations in genes and improper expression levels. However, the debugging process itself is blind to the mechanism – it simply identifies problematic regions of DNA.

[0211] One of the immediate applications of refactoring is in the access of gene clusters from genomic sequence information. This could be necessary either because the cluster is silent, meaning that it that cannot be activated in the laboratory, or because the desired cluster is from a metagenomic sample or information database and the physical DNA is unavailable (42). There are have been many elegant methods to activate a gene cluster, including the placement of inducible promoters upstream of the natural operons and the division of the cluster into individual cistrons, which can then be reassembled (43). With advances in DNA synthesis technology, it is possible to construct entire gene clusters with complete control over the identity of every nucleotide in the design. This capability eliminates the reliance on

the natural physical DNA for construction and enables the simultaneous redesign of components in the complete system. Fully harnessing this technology will require the marriage of computational methods to select parts and scan designs, characterized part libraries, and methods to reduce their context dependence.

5 MATERIAL AND METHODS

Strains and Media

- [0212] *E. coli* strain S17-1 was used for construction and propagation of all plasmids used in *Klebsiella oxytoca* knockout mutant construction. *K. oxytoca* strain M5a1 (Paul Ludden, UC Berkeley) and mutants derived from M5a1 were used for nitrogen fixation experiments.
- 10 Luria-Bertani (LB)-Lennox was used for strain propagation. All assays were carried out in minimal medium containing (per liter) 25g of Na_2HPO_4 , 3g of KH_2PO_4 , 0.25g of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1g of NaCl , 0.1g of $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 2.9mg of FeCl_3 , 0.25mg of $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$, and 20g of sucrose. Growth media is defined as minimal media supplemented with 6ml (per liter) of 22% NH_4Ac . Derepression media is defined as minimal
- 15 media supplemented with 1.5ml (per liter) of 10% serine. The antibiotics used were $34.4\mu\text{g ml}^{-1}$ Cm, $100\mu\text{g ml}^{-1}$ Spec, $50\mu\text{g ml}^{-1}$ Kan, and/or $100\mu\text{g ml}^{-1}$ Amp.

Codon Randomization

- [0213] Initial gene sequences were proposed by DNA2.0 to maximize Hamming distance from the native sequence while seeking an optimal balance between *K. oxytoca* codon usage
- 20 and *E. coli* codon preferences experimentally determined by the company (44). Rare codons (<5% occurrence in *K. oxytoca*) were avoided, and mRNA structure in the translation initiation region was suppressed. Known sequence motifs, including restriction sites, transposon recognition sites, Shine-Dalgarno sequences and transcriptional terminators, were removed by the DNA2.0 algorithm.

25 Elimination of Undesired Regulation

- [0214] Each synthetic operon was scanned prior to DNA synthesis to identify and remove undesired regulation. Multiple types of regulation were identified using publicly available software. The RBS Calculator was used (Reverse Engineering, 16S RNA: ACCTCCTTA) to identify ribosome binding sites throughout the proposed DNA sequence of the operon (45).
- 30 The Prokaryotic Promoter Prediction server was used to identify putative $\sigma 70$ promoter sites (e-value cutoff of 5, sigma.hmm database) (46). The PromScan algorithm was used to identify putative $\sigma 54$ promoter sites using default options (47). TransTermHP software was

used with default parameters to identify terminator sequences in both the forward and reverse directions (48). RBSs greater than 50 AU and all identified promoters and terminators were considered significant.

Nitrogenase Activity Assay

5 [0215] *In vivo* nitrogenase activity is determined by acetylene reduction as previously described (49). For *K. oxytoca* whole-cell nitrogenase activity assay, cells harboring the appropriate plasmids were incubated in 5ml of growth media (supplemented with antibiotics, 30°C, 250r.p.m.) in 50ml conical tubes for 14 hours. The cultures were diluted into 2ml derepression media (supplemented with antibiotics and inducers) to a final OD₆₀₀ of 0.5 in
10 14ml bottles, and bottles were sealed with rubber stoppers (Sigma Z564702). Headspace in the bottles was repeatedly evacuated and flushed with N₂ past a copper catalyst trap using a vacuum manifold. After incubating the cultures for 5.5 hours at 30°C, 250r.p.m, headspace was replaced by 1 atmosphere Ar. Acetylene was generated from CaC₂ using a Burreis bottle, and 1ml was injected into each bottle to start the reaction. Cultures were incubated for 1 hour
15 at 30°C, 250 r.p.m before the assay was stopped by injection of 300µl of 4M NaOH solution into each bottle. To quantify ethylene production, 50µl of culture headspace was withdrawn through the rubber stopper with a gas tight syringe and manually injected into a HP 5890 gas chromatograph. Nitrogenase activity is reported as a percentage of wild-type activity. Briefly, ethylene production by strains was quantified by integrating area under the peak
20 using HP Chemstation software and dividing ethylene production of experimental strains by the ethylene production of a wild type control included in each assay.

N₂-dependent Growth and ¹⁵N₂ Incorporation Assay

[0216] Nitrogen fixation by synthetic *nif* cluster in *K. oxytoca* is further demonstrated by N₂-dependent growth and ¹⁵N₂ incorporation. Cells are diluted as described in the acetylene
25 reduction assay. The headspace of the bottles is replaced by normal N₂ gas or by stable isotope nitrogen, ¹⁵N₂ (¹⁵N atom 99.9%, Icon Isotopes, Cat#: IN 5501). After incubating the cultures for 36 hours at 30°C, 250r.p.m, N₂-dependent growth of the cells is determined by measuring optical density at 600 nm (OD₆₀₀). To do the ¹⁵N₂ incorporation assay, the ¹⁵N-enriched cells with corresponding control cultures under normal nitrogen gas are collected by
30 centrifugation, the cell pellets are dried in a laboratory oven at 100°C for 12 hours. The dried pellets are analysis for ¹⁵N/¹⁴N ratio at the Center for Stable Isotope Biogeochemistry at the University of California, Berkeley using the Finnigan MAT Delta plus Isotope Ratio Mass Spectrometer.

***K. oxytoca* Knockout Strains**

[0217] All *K. oxytoca* mutants are constructed from M5al by allele exchange using suicide plasmid pDS132 carrying the corresponding *nif* gene deletion (pDS132 was graciously provided by the Paul Ludden lab at UC Berkeley as a gift from Dr. Dominique Schneider at
5 Université Joseph Fourier) (49). We made a slight modification to a previously published protocol (50). Here, a kanamycin resistance cassette was cloned into the suicide plasmid upstream of the left homologous exchange fragment. These operon deletions in *nif* gene cluster span the promoter and the complete amino acid coding sequences except when specifically designated. All mutants were verified by DNA sequencing of the PCR product
10 of the corresponding gene region to confirm physical DNA deletion and by whole-cell acetylene reduction assay to confirm the lack of nitrogenase activity.

Promoter Characterization

[0218] As described in this example, the output of promoters is reported as relative expression units (REU). This is simply a linear factor that is multiplied by the arbitrary units
15 measured by the flow cytometer. The objective of normalizing to REU is to standardize measurements between labs and projects. The linear factor is 1.66×10^{-5} and the division by this number back converts to the raw arbitrary units. This number was calculated to be a proxy to the RPU (relative promoter units) reported by Kelly and co-workers (51). Our original standardized measurements were made prior to the Kelly paper and involved a
20 different reference promoter, fluorescent protein (mRFP), RBS, and plasmid backbone. Because of these differences, one cannot calculate RPU as defined by Kelly, et al. Instead, a series of plasmids was made (Fig. 33A) to estimate the relative expression of reporter protein from experimental constructs compared the standard construct in Kelly, et al. Conversion factors between constructs were measured and multiplied to obtain the linear factor above.
25 We renamed the unit to REU (relative expression units) because it is intended to be a simple normalization of fluorescent units (akin to a fluorescent bead) and not a direct measurement of the activity of a promoter (*e.g.*, the polymerase flux).

[0219] Cells were grown as in the Acetylene Reduction Assay with two modifications. The initial flush of headspace with N₂ was not performed, and the assay was halted after the 5.5
30 hour incubation. To halt the assay, 10 μ l of cells were transferred from each bottle to a 96-well plate containing phosphate buffered saline supplemented with 2mg ml⁻¹ kanamycin. Fluorescence data was collected using a BD Biosciences LSRII flow cytometer. Data were gated by forward and side scatter, and each data set consisted of at least 10,000 cells. FlowJo

was used to calculate the geometric means of the fluorescence distributions. The autofluorescence value of *K. oxytoca* cells harboring no plasmid was subtracted from these values to give the values reported in this study. The strengths of T7 promoter mutants were characterized by swapping them in place of the P_{tac} promoter in plasmid N149 (SBa_000516),
5 co-transforming with Controller #1 (plasmid N249), and measuring fluorescence via flow cytometry under 1mM IPTG induction.

[0220] To replace the P_{tac} promoter by a T7 promoter in each synthetic operon, we followed a simple process. First, we identified the IPTG concentration corresponding to the maximal functional activity of each synthetic operon. Second, we translated this IPTG
10 concentration into REU based on characterization of the P_{tac} promoter (Fig. 33B, left). Third, we selected the T7 mutant promoter with the closest strength in REU. For the synthetic *nifF* operon, we observed broad, robust fixation under the P_{tac} promoter. We found that T7 mut 3 produced inducible functional activity with a maximum at 1mM IPTG induction of the T7 RNAP. For the synthetic *nifJ* operon, our method suggests that we use a weak T7 mutant
15 promoter. However, we found that a WT T7 promoter produced inducible activity with a maximum at 1mM IPTG. We attribute this deviation to a change in RBS strength due to contextual differences between P_{tac} and the T7 promoter.

Debugging Synthetic Operons

[0221] Some of the initial designs for refactored operons showed little or no activity. When
20 this occurs, it is challenging to identify the problem because so many genetic changes have been made simultaneously to the extent that there is almost no DNA identity with the wild-type sequence. To rapidly identify the problem, a debugging method was developed that can be generalized when refactoring different functions (Fig. 34A). Chimeric operons are created by replacing a wild-type region of DNA with its synthetic counterpart. The function of each
25 chimera in this library is assessed to identify which region of synthetic DNA caused a loss of activity. New chimeras are then be constructed with increasingly fine-resolution changes between synthetic and wild-type DNA. This approach “zooms in” on the problematic region of DNA, which can then be fixed. The most common problem is due to errors in the reference DNA sequence (Genbank, X13303.1) (52). Refactored genes were designed using
30 only the amino acid sequence information from the database; thus, they were sensitive to sequencing errors leading to missense mutations that reduced or eliminated activity. Indeed, 18 such mutations were identified and confirmed by carefully resequencing the wild-type cluster (Fig. 35). Fifteen of the 18 mutations occurred in refactored operons that required

debugging and were corrected (Fig. 34B). This demonstrates the challenge of reconstituting biological functions using only database information and DNA synthesis (55).

[0222] Modifying synthetic RBS strength was also important to debugging. The function of the synthetic *nifUSVWZM* operon was significantly improved by changing RBSs to match a 1:1 ratio of NifU:NifS. The initial selection of RBSs led to an observed 10:1 ratio in their respective RBS strengths. After debugging, *nifU* and *nifS* RBS strength was better balanced (1.25:1) and this improved activity. For one RBS, the measurement method proved to be inaccurate. We found the measured strength of the wild-type *nifQ* RBS was extremely low (Fig. 27C), and the synthetic *nifBQ* operon showed low activity when the synthetic *nifQ* RBS was matched to the measured strength. In contrast, the robustness analysis showed a requirement for high expression level of the *nifBQ* operon (Fig. 27B). Thus, a strong synthetic RBS near the strength of the *nifB* RBS was used and significantly improved *nifBQ* operon activity. In one case, our initial recoded *nifH* gene did not express well using either wild-type or synthetic regulation (Fig. 37). We designed a new synthetic gene, requiring that it diverge in DNA sequence from both the native and first synthetic DNA sequences and found that the new synthetic gene both expressed well and recovered activity.

Growth by Nitrogen Fixation

[0223] Cells capable of nitrogen fixation should exhibit measurable growth on media that lacks nitrogen by utilizing atmospheric N_2 as a source of nitrogen. Conversely, cells incapable of nitrogen fixation should not grow on nitrogen-free media.

[0224] In parallel to the $^{15}N_2$ incorporation assay, we monitored strain growth under nitrogen-limited media conditions and 100% $^{15}N_2$ atmosphere (Methods, N_2 -dependent Growth Assay). Cells were grown on derepression media as used in the Nitrogenase Activity Assay. Depression media is not strictly nitrogen-free, containing 1.43 mM serine in order to promote ribosomal RNA production and hasten nitrogenase biosynthesis (54).

[0225] Strains containing Controller #1 and the refactored gene cluster grew nearly 30% as much as wild-type strains. In contrast, minimal growth was observed in Δnif strains, consistent with the limited nitrogen available from serine and cell lysis products (55). Fig. 37 illustrates cell growth supported by nitrogen fixation.

30 Western Blot Assay for Synthetic *nifH* Expression

[0226] The first synthetic *nifHDK* did not exhibit nitrogenase activity under induction ranging from 0 to 1mM IPTG, and the *nifH* gene (synthetic *nifH_{v1}*) was identified as a

problematic part using the debugging protocol shown in Fig. 34. However there was no mutation found. Western blots were further used to confirm problematic synthetic *nifH* expression.

5 [0227] A western blot for NifH protein in Fig. 37 (left) showed that wild type *nifH* expressed well with either synthetic *nifD* or *nifK* (construct N10, N12, N14), whereas synthetic *nifH_{v1}* was not expressed regardless of the context of *nifDK* (construct N1 and N19). A second synthetic *nifH* (synthetic *nifH_{v2}*) was used to replace synthetic *nifH_{v1}*. The western blot in Fig. 37 (right) showed the synthetic *nifH_{v2}* (construct N38) expressed well.

10 [0228] Samples for western blots were prepared by boiling collected *K. oxytoca* cells in SDS-PAGE loading buffer and run on 12% SDS-Polyacrylamide gels (Lonza Biosciences). Proteins on the gels were transferred to PVDF membranes (BioRad Cat#: 162-0177) using Trans-Blot SD Semi-Dry Transfer Cell (BioRad Cat#: #170-3940). Blocking the membrane and Antibody binding were performed using SNAP i.d. Protein Detection System (Millipore Cat#WBA VDBA). The membranes were blocked by TBST-1% BSA (TBS-Tween20™). The anti-NifH and anti-NifDK antibodies (kindly provided by Paul Ludden Lab at UC-Berkeley) were used as the primary
15 antibodies. The anti-NifH antibody was a universal anti-NifH made against a mixture of purified NifH proteins from *Azotobacter vinelandii*, *Clostridium pasteurianum*, *Rhodospirillum rubrum*, and *K. oxytoca*. The anti-NifDK antibody was made against purified NifDK protein from *Azotobacter vinelandii*. The anti-NifH and anti-NifDK antibodies were used at 1:500 and 1:2000 respectively. The secondary antibody (Goat anti-Rabbit IgG-HRP, Sigam Cat#: A0545) was used at 1:10,000.
20 Development was done using an enhanced chemiluminescent substrate for HRP (Pierce Cat#: 32209) and captured on film (Kodak: Cat#:178-8207).

Construction of Plasmids and Parts

25 [0229] Plasmids were designed *in silico*. Synthetic parts (promoters, RBS, terminators and insulators) were combined with the initial synthetic gene sequences proposed by DNA2.0 in ApE (A Plasmid Editor) and GeneDesigner (56) to create synthetic operons. Synthetic operons were computationally scanned to eliminate unintended regulation (Methods, “Elimination of Undesired Regulation”), and parts containing such regulation were replaced. This reiterative process continued until the synthetic operons included only designed regulation.

30 [0230] Physical DNA was constructed using standard manipulation techniques. Assembly methods followed published protocols and included BioBrick (57), Megawhop (58), Phusion

Site-Directed Mutagenesis or Gibson Assembly methods (59). We found that Gibson Assembly was the most efficient DNA assembly method, except when making small (<10bp) changes in plasmids under 10kb in size. We noted assembly failures were infrequent, more common in assemblies above 15kbp, and linked to the presence of homology within ~500bp of part termini. In these cases, we observed annealing of unexpected parts to create non-intended junctions.

[0231] Plasmid pIncW (pSa, SpR) was generated from pEXT21 (pSa, SpR) by deletion of *osa*, *nuc1*, the Tn21 integrase gene, and ORF18 (60). Plasmid pSB4C5 (pSC101, CmR) was obtained from the Registry of Standard Biological parts and serves as the base vector for wild-type complementation, RBS characterization, and synthetic operons (57). Plasmid N58 (pSC101, CmR) was generated by inserting the P_{tac} cassette (SynBERC Registry, SBa_000561) between the BioBrick prefix and BioBrick suffix of pSB4C5. Plasmid N292 (SBa_000566) was generated by inserting a terminator characterization cassette between the BioBrick prefix and BioBrick suffix of pSB4C5. The cassette consists of the PT7 promoter, RBS (SBa_000498), GFP, the wild-type T7 terminator, RBS D103 (SBa_000563) from Salis *et. al.* (13), and mRFP (SBa_000484). Plasmid N149 (SBa_000516) was constructed by inserting the P_{tac} promoter cassette (SBa_000563), RBS D103 (SBa_000563) from Salis *et. al.* (13), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5. Plasmid N505 (SBa_000517) was constructed by inserting the P_{tet} promoter cassette (SBa_000562), RBS D103 (SBa_000563), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5. Plasmid N110 (SBa_000564) was constructed by inserting a constitutive promoter (SBa_000565), a strong RBS (SBa_000475), and mRFP (SBa_000484) between the BioBrick prefix and BioBrick suffix of pSB4C5. Plasmid N573 (SBa_000559) was constructed by inserting the AmpR resistance marker in pNOR1020 (14).

[0232] It has been shown that the multicopy expression of some nitrogen fixation genes can eliminate nitrogenase maturation and function (*i.e.*, multicopy inhibition) (63, 64). An additional uncertainty is that the replacement of the native promoter with an inducible promoter could disrupt their function. To examine these effects, we constructed plasmids to complement the activities of the knockout strains (Fig. 31) and tested their activity under inducible control. These plasmids are also the basis for the experiments to quantify the robustness to changes in expression (Fig. 27).

[0233] Complementation plasmids were constructed by inserting the DNA encoding each wild-type operon between the P_{tac} promoter and BioBrick suffix of plasmid N58 (pSC101,

CmR). One exception was plasmid Nifl8 which was constructed by cloning the nifHDKTY operon into the multi-cloning site of pEXT21 (60). Wild-type operon sequences were defined by published transcription initiation sites (65).

[0234] Wild-type RBS characterization vectors were constructed by inserting the region from -60bp to +90bp for each native gene and mRFP (SBa_000484) between the P_{tac} cassette (SBa_000561) and the BioBrick suffix of plasmid N58 (pSC101, CmR). The native gene sequence from +1bp to +90bp formed an in-frame fusion with mRFP. In cases where the gene transcript does not extend to -60bp, a shorter cassette was cloned into N58. RBS strength was characterized using the Promoter Characterization Assay described herein.

10 [0235] Synthetic RBSs of sufficient length to capture the full ribosome footprint (~35bp) were generated with the RBS Calculator (61). The strength of each was measured using a synthetic RBS characterization vector. These vectors were constructed similar to the wild-type RBS characterization vectors using -60bp to +90bp of the designed synthetic gene. This region includes part of a buffer sequence, the synthetic RBS, and the region from +1bp to
15 +90bp of the synthetic gene. If the synthetic and wild-type RBSs differed by more than 3-fold in expression, new RBS sequences were generated and screened. Insulator parts consisting of ~50bp of random DNA precede each synthetic RBS (66).

[0236] Synthetic operons were cloned into the pSB4C5 (pSC101, CmR) backbone between the BioBrick prefix and BioBrick suffix.

20 Synthetic Part Generation

[0237] T7 * RNA Polymerase: The T7 RNA polymerase was modified to be non-toxic to both *Klebsiella* and *E. coli* at high expression levels. The RNAP was expressed from a low-copy origin (pSa) under control of a weak RBS (SBa_000507,

TATCCAAACCAGTAGCTCAATTGGAGTCGTCTAT) and N-terminal degradation tag
25 (SBa_000509,

TTGTTTATCAAGCCTGCGGATCTCCGCGAAATTGTGACTTTTCCGCTATTTAGCGA
TCTTGTTCAAGTGTGGCTTTCCTTCACCGGCAGCAGATTACGTTGAACAGCGCATC
GATCTGGGTGGC). The start codon was changed from ATG to GTG, and the active site contained a mutation (R632S).

30 [0238] T7 promoters: T7 promoters were generated from a random library. The T7 promoter seed sequence was TAATACGACTCACTANNNNNAGA. For the sequences of individual promoters, see Fig. 38.

[0239] T7 terminators: T7 terminators were generated from a random library and inserted into the terminator characterization vector N292 (SBa_000566). The T7 terminator seed sequence was TANNNAACSSWWSSNSSSSTCWWCGSSSSSSWWSSGTTT.

Terminator plasmids were co-transformed with plasmid N249 and characterized (Methods, Fluorescence Characterization) under 1mM IPTG induction of T7* RNAP. RFP expression was measured for each terminator, and data are reported as the fold reduction in measured fluorescence when compared to a derivative of N292 carrying no terminator. For the sequences of individual terminators, see Fig. 38

[0240] Ribosome binding sites: The RBS Calculator was used to generate an RBS that matched the measured strength of the wild-type RBS. In three cases, synthetic RBSs were selected from existing parts (SBa_000475 for *nifJ* and *nifQ*, and SBa_000469 for *nifH*). In cases where the strength of the initial synthetic RBS differed from the WT RBS by more than 3-fold (*nifV*, *nifZ*, and *nifM*), a library of synthetic RBS was constructed by replacing the 15bp upstream of the start codon with NNNAGGAGGNNNNNN. We screened mutants in each library to identify synthetic RBSs within three fold of the WT RBS strength. Ribosome binding site strength is reported in arbitrary fluorescence units measured using the fluorescence characterization assay.

[0241] Insulator sequences (spacer sequences): Insulator sequences were generated using the Random DNA Generator using a random GC content of 50% (66).

[0242] ANDN Logic: We constructed a genetic circuit encoding the logic A ANDN B and used this circuit to control T7* RNAP in Controller #3. In this circuit, the A ANDN B logic corresponds to the presence or absence of the inducers, IPTG and aTc, such that the cell computes IPTG ANDN aTc. The circuit was constructed by modifying the Ptac promoter in Controller #1 (SBa_000520) to include the *cl* repressor binding sites OR1 and OR2 to produce plasmid N639 (SBa_000560). Additionally, plasmid pNOR1020 encodes the repressor *cl* under control of the Ptet promoter (62). We modified pNOR1020 by changing the resistance marker to confer ampicillin resistance to produce N573 (SBa_000559). When N639 and N573 are co-transformed, they produce the logic circuit IPTG ANDN aTc.

Ptac (SBa_000512) sequence:

tattctgaaatgagctgttgacaattaatcatcggctcgtataatgtgtggaattgtgagcggataacaatt

Ptac plus OR1 and OR2 (SBa_000506) sequence:

tattaacaccgtgcgtgttgacagctatacctctggcggttataatgctagcgggaattgtgagcggataacaatt

[0243] Fig. 39 illustrates maps for key plasmids.

[0244] The *nif* gene cluster in *K. oxytoca* Ma5L was re-sequenced from PCR fragments. The re-sequenced DNA sequence was compared to the reference sequence from Genbank, X13303.1 (52). Sequence differences are listed in Fig. 35. The nucleotide locations are
 5 numbered relative to X13303.1. Amino acid mutations to correct errors in the X13303.1 record are shown (Impact).

References

1. Fischbach M, Voigt, C.A. (2010) Prokaryotic gene clusters: A rich toolbox for synthetic biology. *Biotechnol. J.* 5:1277-1296.
- 10 2. Ishihama A (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev* 34(5):628-645.
3. Mandal M & Breaker RR (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 5(6):451-463.
4. Temme K, *et al.* (2008) Induction and relaxation dynamics of the regulatory network
 15 controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *J Mol Biol* 377(1):47-61.
5. Georg J & Hess WR (2011) cis-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* 75(2):286-300.
6. Guell M, Yus E, Lluch-Senar M, & Serrano L (2011) Bacterial transcriptomics: what
 20 is beyond the RNA hori-zome? *Nature reviews. Microbiology* 9(9):658-669.
7. Johnson ZI & Chisholm SW (2004) Properties of overlapping genes are conserved across microbial genomes. *Genome Res* 14(11):2268-2272.
8. Zazopoulos E, *et al.* (2003) A genomics-guided approach for discovering and
 expressing cryptic metabolic pathways. *Nat Biotechnol* 21(2):187-190.
- 25 9. Medema MH, Breitling R, Bovenberg R, & Takano E (2011) Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nature reviews. Microbiology* 9(2):131-137.
10. Gottelt M, Koi, S., Gomez-Escribano, J.P., Bibb, M., Takano, E. (2010) Deletion of a
 regulatory gene within the *cpk* gene cluster reveals novel antibacterial activity in
 30 *Streptomyces coelicolor* A3(2). *Microbiology* 156:2343-2353.
11. Lombo F, Brana, A.F., Mendez, C., Salas, J.A. (1999) The mithramycin gene cluster of *Streptomyces argillaceus* contains a positive regulatory gene and two repeated DNA sequences that are located at both ends of the cluster. *J. Bacteriol.* 181:642-647.
12. Medema MH, Bretiling, R., Takano, E. (2011) Synthetic biology in *Streptomyces*
 35 bacteria. *Methods Enzymol* 497:485-502.
13. Pickens LB, Tang, Y., Chooi, Y-H. (2011) Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.* 2:1-26.
14. Smanski MJ, Peterson, R.M., Rajski, S.R., Shen, B. (2009) Engineered *Streptomyces platensis* strains that overproduce antibiotics platensimycin and platencin. *Antimicrob. Agents Chemother.* 53:1299-12304.
- 40 15. Biggins JB, Liu, X., Feng, Z., Brady, S.F. (2011) Metabolites from the induced expression of cryptic single operons found in the genome of *Burkholderia pseudomallei*. *JACS* 133:1638-1641.
16. Watanabe K, Hotta, K., Praseuth, A.P., Koketsu, K., Migita, A., Boddy, C.N., Wang, C.C.C., Oguri, H., Oikawa, H. (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nature Chemical Biology*:1-6.
- 45

17. Fowler M & Beck K (1999) *Refactoring : improving the design of existing code* (Addison-Wesley, Reading, MA) pp xx1, 431 p.
18. Chan LY, Kosuri S, & Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* 1:2005 0018.
- 5 19. Czar MJ, Anderson JC, Bader JS, & Peccoud J (2009) Gene synthesis demystified. *Trends Biotechnol* 27(2):63-72.
20. Purnick PE & Weiss R (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* 10(6):410-422.
21. Stacey GS, Burris RH, & Evans HJ (1992) *Biological nitrogen fixation* (Chapman & Hall, New York) pp xii, 943 p.
- 10 22. Burris RH (1991) Nitrogenases. *J Biol Chem* 266(15):9339-9342.
23. Hu Y, Fay AW, Lee CC, Yoshizawa J, & Ribbe MW (2008) Assembly of nitrogenase MoFe protein. *Biochemistry* 47(13):3973-3981.
24. Rubio LM & Ludden PW (2005) Maturation of nitrogenase: a biochemical puzzle. *J Bacteriol* 187(2):405-414.
- 15 25. Arnold W, Rump A, Klipp W, Priefer UB, & Puhler A (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *J Mol Biol* 203(3):715-738.
26. Dixon R & Kahn D (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* 2(8):621-631.
- 20 27. Dixon RA & Postgate JR (1972) Genetic transfer of nitrogen fixation from *Klebsiella pneumoniae* to *Escherichia coli*. *Nature* 237(5350):102-103.
28. Simon HM, Homer MJ, & Roberts GP (1996) Perturbation of *nifT* expression in *Klebsiella pneumoniae* has limited effect on nitrogen fixation. *J Bacteriol* 178(10):2975-2977.
- 25 29. Fani R, Gallo R, & Lio P (2000) Molecular evolution of nitrogen fixation: the evolutionary history of the *nifD*, *nifK*, *nifE*, and *nifN* genes. *J Mol Evol* 51(1):1-11.
30. Gosink MM, Franklin NM, & Roberts GP (1990) The product of the *Klebsiella pneumoniae nifX* gene is a negative regulator of the nitrogen fixation (*nif*) regulon. *J Bacteriol* 172(3):1441-1447.
- 30 31. Orme-Johnson WH (1985) Molecular basis of biological nitrogen fixation. *Annu Rev Biophys Biophys Chem* 14:419-459.
32. Kelly JR, *et al.* (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 3:4.
- 35 33. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343-345.
34. Yokobayashi Y, Weiss R, & Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A* 99(26):16587-16591.
- 35 35. Tamsir A, Tabor JJ, & Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469(7329):212-215.
- 40 36. Alon U (2007) *An introduction to systems biology : design principles of biological circuits* (Chapman & Hall/CRC, Boca Raton, FL) pp xvi, 301 p., 304 p. of plates.
37. Kitano H (2002) Systems biology: a brief overview. *Science* 295(5560):1662-1664.
38. Palsson B (2006) *Systems biology : properties of reconstructed networks* (Cambridge University Press, Cambridge ; New York) pp xii, 322 p.
- 45 39. Kalir S, *et al.* (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* 292(5524):2080-2083.
40. Zaslaver A, Mayo A, Ronen M, & Alon U (2006) Optimal gene partition into operons correlates with gene functional order. *Phys Biol* 3(3):183-189.
- 50 41. Kovacs K, Hurst LD, & Papp B (2009) Stochasticity in protein levels drives colinearity of gene order in metabolic operons of *Escherichia coli*. *PLoS Biol* 7(5):e1000115.

42. Wenzel SC & Muller R (2005) Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr Opin Biotechnol* 16(6):594-606.
43. Welch M, *et al.* (2009) Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* 4(9):e7002.
44. Salis HM, Mirsky EA, & Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27(10):946-950.
45. Zomer AL (2011) PPP: Perform Promoter Prediction.
46. Studholme D (2011) PromScan.
47. Kingsford CL, Ayanbule K, & Salzberg SL (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* 8(2):R22.
48. Stewart WD, Fitzgerald GP, & Burris RH (1967) In situ studies on nitrogen fixation with the acetylene reduction technique. *Science* 158(3800):536.
49. Philippe N, Alcaraz JP, Coursange E, Geiselmann J, & Schneider D (2004) Improvement of pCVD442, a suicide plasmid for gene allele exchange in bacteria. *Plasmid* 51(3):246-255.
50. Zhao D, Curatti L, & Rubio LM (2007) Evidence for nifU and nifS participation in the biosynthesis of the iron-molybdenum cofactor of nitrogenase. *J Biol Chem* 282(51):37016-37025.
51. Kelly JR, *et al.* (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard. *J Biol Eng* 3:4.
52. Arnold W, Rump A, Klipp W, Priefer UB, & Puhler A (1988) Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of *Klebsiella pneumoniae*. *J Mol Biol* 203(3):715-738.
53. Bayer TS, *et al.* (2009) Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *J Am Chem Soc* 131(18):6508-6515.
54. Jacob GS, Schaefer J, Garbow JR, & Stejskal EO (1987) Solid-state NMR studies of *Klebsiella pneumoniae* grown under nitrogen-fixing conditions. *J Biol Chem* 262(1):254-259.
55. Mason CA & Hamer G (1987) Cryptic Growth in *Klebsiella-Pneumoniae*. *Appl Microbiol Biot* 25(6):577-584.
56. Villalobos A, Ness JE, Gustafsson C, Minshull J, & Govindarajan S (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* 7:285.
57. Shetty RP, Endy D, & Knight TF, Jr. (2008) Engineering BioBrick vectors from BioBrick parts. *J Biol Eng* 2:5.
58. Miyazaki K (2003) Creating random mutagenesis libraries by megaprimer PCR of whole plasmid (MEGA WHOP). *Methods Mol Biol* 231:23-28.
59. Gibson DG, *et al.* (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6(5):343-345.
60. Dykxhoorn DM, St Pierre R, & Linn T (1996) A set of compatible tac promoter expression vectors. *Gene* 177(1-2):133-136.
61. Salis HM, Mirsky EA, & Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27(10):946-950.
62. Tamsir A, Tabor JJ, & Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469(7329):212-215.
63. Buck M & Cannon W (1987) Frameshifts close to the *Klebsiella pneumoniae* nifH promoter prevent multicopy inhibition by hybrid nifH plasmids. *Mol Gen Genet* 207(2-3):492-498.

64. Riedel GE, Brown SE, & Ausubel FM (1983) Nitrogen fixation by *Klebsiella pneumoniae* is inhibited by certain multicopy hybrid nif plasmids. *J Bacteriol* 153(1):45-56.
- 5 65. Beynon J, Cannon M, Buchanan-Wollaston V, & Cannon F (1983) The nif promoters of *Klebsiella pneumoniae* have a characteristic primary structure. *Cell* 34(2):665-671.
66. Maduro M (2011) Random DNA Generator.

[0245] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

10

What is Claimed is:

1. A method for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the method comprising
providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;
changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence, wherein the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence;
organizing the coding sequences into one or more synthetic operon(s);
operably linking one or more heterologous transcriptional regulatory sequence(s) to the operon(s), thereby controlling the magnitude of gene expression from the operon(s); and
expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.
2. The method of claim 1, wherein the polypeptide is heterologous to the cell.
3. The method of claim 1 or 2, wherein the providing comprises obtaining nucleotide sequences of the genes and eliminating non-coding sequences.
4. The method of claim 1, 2 or 3, wherein the set of genes is from a gene cluster.
5. The method of any one of claims 1 to 4, wherein the set of genes are from a prokaryote.
6. The method of any one of claims 1 to 5, wherein the genes are from a native operon.
7. The method of any one of claims 1 to 6, wherein the at least one regulatory sequence is identified using computation.
8. The method of claim 7, wherein the computation comprises searches of coding sequences for ribosome binding sites, terminators, promoters, or a combination thereof.
9. The method of any one of claims 1 to 8, wherein the removing the at least one regulatory sequence comprises selecting non-native codons for optimal expression in a host cell.

10. The method of any one of claims 1 to 9, further comprising identifying and removing one or more of: transposon insertion sites, sites that promote recombination, sites for cleavage by restriction endonucleases, and sites that are methylated.

11. The method of any one of claims 1 to 10, wherein the organizing comprises grouping coding sequences into operons based on similar native expression level.

12. The method of any one of claims 1 to 10, wherein the organizing comprises ordering coding sequences within operons such that the highest expressing gene based on native expression occurs first and the lowest expressing gene based on native expression occurs last.

13. The method of any one of claims 1 to 12, wherein magnitude of expression of coding sequences-corresponds to the ratio of proteins encoded by the coding sequences as measured in the native system.

14. The method of any one of claims 1 to 13, wherein magnitude of expression of coding sequences is determined by computation.

15. The method of claim 14, wherein the computation comprises a numerical optimization algorithm.

16. The method of claim 15, wherein the numerical optimization algorithm comprises the Nelder-Mead algorithm, the Newton's method, the quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, an ellipsoid method, the Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model.

17. The method of any one of claims 1 to 16, wherein the heterologous transcriptional regulatory sequence(s) comprise a T7 RNA polymerase promoter(s).

18. The method of any one of claims 1 to 16, wherein heterologous transcriptional regulatory sequence(s) comprise an inducible promoter.

19. The method of any one of claims 1 to 18, further comprising operably linking a heterologous ribosomal binding site (RBS) to each of one or more coding sequences in the synthetic operon.

20. The method of claim 19, wherein different RBSs are operably linked to different coding sequences.

21. The method of claim 19 or 20, wherein the RBSs regulate translation of the coding sequences to which they are linked in a ratio that is similar to the ratio of native translation from the native operon.
22. The method of any one of claims 1 to 21, further comprising operably linking one or more heterologous transcriptional terminators to one or more coding sequences in the synthetic operon.
23. The method of claim 22, wherein the terminators are T7 RNA polymerase terminators.
24. The method of claim 22, wherein terminators for different synthetic operons are different.
25. The method of any one of claims 1 to 24, further comprising operably linking one or more buffer sequences between two functional sequences in an operon wherein the functional sequences are selected from the group consisting of a promoter, ribosome binding site, coding sequence, and terminator.
26. The method of claim 25, wherein the one or more buffer sequences are selected from the group consisting of a random sequence, a UP-region of a promoter, an extended 5-UTR sequence, and a RNAase cleavage site.
27. The method of any one of claims 1 to 26, wherein the operons are expressed from a plasmid.
28. The method of claim 27, wherein the plasmid has a low copy origin of replication.
29. The method of any one of claims 1 to 28, wherein the polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence is expressed from a control expression cassette, the expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.
30. The method of claim 29, wherein the expression cassette is contained in a control plasmid separate from any plasmid containing the synthetic operons.
31. The method of claim 29 or 30, wherein the control promoter is an inducible promoter.
32. The method of claim 29, 30 or 31, wherein the heterologous polypeptide comprises an RNA polymerase (RNAP).

33. The method of claim 32, wherein the RNAP is T7 RNAP.

34. The method of any one of claims 29 to 33, wherein the expression cassette is an environmental sensor.

35. A polynucleotide comprising a synthetic operon, wherein the operon comprises at least two coding sequences under the control of a heterologous transcriptional regulatory sequence, wherein each coding sequence is operably linked to a heterologous ribosome binding site (RBS), and wherein codons of one or more coding sequence have been selected for maximal distance from codon usage of the corresponding coding sequence in a native operon thereby removing at least one regulatory sequence within the coding sequence.

36. The polynucleotide of claim 35, wherein the coding sequences are from the same or different native operons and the heterologous RBSs regulate translation of the coding sequences to which they are linked in a ratio that is similar to the ratio of native translation from the native operon.

37. The polynucleotide of claim 35 or 36, wherein the coding sequences are from the same or different native operons and the coding sequences in the operon comprise one or more altered codon compared to the native operon.

38. The polynucleotide of claim 35, 36 or 37, wherein at least two coding sequences encode different proteins encoded by the *Klebsiella pneumoniae* nif gene cluster.

39. The polynucleotide of claim 38, wherein the proteins are selected from the group consisting of nifJ, nifH, nifD, nifK, nifY, nifE, nifN, nifU, nifS, nifV, nifW, nifZ, nifM, nifF, nifB, and nifQ.

40. The polynucleotide of claim 38, wherein the operon comprises coding sequences for *Klebsiella pneumoniae* nifH, nifD, nifK, and nifY.

41. The polynucleotide of claim 38, wherein the operon comprises coding sequences for *Klebsiella pneumoniae* nifE and nifN.

42. The polynucleotide of claim 38, wherein the operon comprises coding sequences for *Klebsiella pneumoniae* nifU, nifS, nifV, nifW, nifZ, and nifM.

43. The polynucleotide of claim 38, wherein the operon comprises coding sequences for *Klebsiella pneumoniae* nifB and nifQ.

44. The polynucleotide of claim 35, 36 or 37, wherein at least two coding sequences encode different proteins of the *Salmonella Typhimurium* Type III secretion system.

45. The polynucleotide of claim 44, wherein the proteins are selected from the group consisting of PrgH, PrgI, PrgJ, PrgK, OrgA, OrgB, InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS.
46. The polynucleotide of claim 44, wherein the operon comprises coding sequences for Salmonella Typhimurium PrgH, PrgI, PrgJ, PrgK, OrgA, and OrgB.
47. The polynucleotide of claim 44, wherein the operon comprises coding sequences for Salmonella Typhimurium InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS.
48. An isolated host cell comprising the polynucleotide of any one of claims 35 to 47.
49. A system comprising a set of two or more different synthetic operons as recited in any one of claims 35 to 47, wherein the transcriptional regulatory sequence of each operon in the set is controlled by a transcriptional activator or repressor polypeptide, and the transcriptional regulatory sequence of each operon in the set is controlled by the same transcriptional activator or repressor polypeptide(s).
50. The system of claim 49, further comprising an expression cassette comprising a promoter operably linked to a polynucleotide encoding the transcriptional activator or repressor polypeptide(s).
51. The system of claim 50, wherein the promoter of the expression cassette is an inducible promoter.
52. The system of claim 49, 50 or 51, wherein the polynucleotide in the expression cassette encodes said transcriptional repressor polypeptide(s).
53. The system of claim 49, 50 or 51, wherein the polynucleotide in the expression cassette encodes said transcriptional activator polypeptide(s).
54. The system of claim 53, wherein the transcriptional activator is an RNA polymerase (RNAP).
55. The system of claim 54, wherein the RNAP is T7 RNAP.
56. The system of any one of claims 49 to 55, wherein the transcriptional regulatory sequences of at least two of the operons are different.
57. The system of claim 56, wherein the transcriptional regulatory sequence of the at least two operons have different promoters that are differentially regulated by T7 RNA

polymerase and wherein the different strength of the promoters correspond to the relative strength of native promoters of the coding sequences.

58. The system of any one of claims 49 to 57, wherein the coding sequences in the operons are organized such that coding sequences having similar native expression are grouped into the same operon.

59. The system of any one of claims 49 to 58, wherein the system is expressed in a cell.

60. The system of claim 59, wherein the cell is from a different species than the species from which said native operon is native.

61. The system of claim 59, wherein the cell is from the same species from which said native operon is native.

62. The system of any one of claims 49 to 61, wherein the system encodes a nitrogenase.

63. The system of claim 62, the system comprising:
a first operon comprising coding sequences for *Klebsiella pneumoniae* nifH, nifD, nifK, and nifY;
a second operon comprising coding sequences for *Klebsiella pneumoniae* nifE and nifN;
a third operon comprising coding sequences for *Klebsiella pneumoniae* nifU, nifS, nifV, nifW, nifZ, and nifM; and
a fourth operon comprising coding sequences for *Klebsiella pneumoniae* nifB and nifQ.

64. The system of claim 63, wherein the first, second, third, and fourth operons comprise a T7 RNA polymerase (RNAP) promoter and the system further comprises an expression cassette comprising a promoter operably linked to a polynucleotide encoding the RNAP.

65. The system of any one of claims 49 to 61, wherein the system encodes a type III secretion system.

66. The system of claim 65, wherein the type III secretion system is a *Salmonella Typhimurium* type III secretion system.

67. The system of claim 66, the system comprising:
a first operon comprising coding sequences for *Salmonella Typhimurium* PrgH, PrgI, PrgJ, PrgK, OrgA, and OrgB; and

a second operon comprising coding sequences for Salmonella Typhimurium InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR, and SpaS.

68. A method for replacing native regulation of a set of genes collectively associated with a function with synthetic regulation, the method comprising:

causing at least one processor to provide coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

causing the at least one processor to change codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the at least one coding sequence, wherein the removing comprises selecting non-native codons having maximal distance from codons of the native coding sequence;

causing the at least one processor to organize the coding sequences into one or more synthetic operon(s); and

causing the at least one processor to operably link one or more heterologous transcriptional regulatory sequence to the one or more synthetic operon(s), thereby controlling magnitude of gene expression from the one or more synthetic operon(s);

expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the one or more heterologous transcriptional regulatory sequence.

69. A method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising:

providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

changing codon identity within at least one coding sequence, thereby removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence;

organizing the coding sequences into one or more synthetic operon(s);

operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling magnitude of gene expression from the operon(s); and

expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence,

wherein the polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence is expressed from a control expression cassette, the expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.

70. The method of claim 69, wherein the polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence is heterologous to the cell.

71. The method of claim 69 or 70, wherein the providing comprises obtaining gene nucleotide sequences of the genes and eliminating non-coding sequences.

72. The method of claim 69, 70, or 71, wherein the genes are from a gene cluster.

73. The method of any one of claims 69 to 72, wherein the genes are from a prokaryote.

74. The method of any one of claims 69 to 73, wherein the genes are from a native operon.

75. The method of any one of claims 69 to 74, wherein the at least one regulatory sequence is identified using computation.

76. The method of claim 69, wherein the computation comprises searches of coding sequences for ribosome binding sites, terminators, promoters, or a combination thereof.

77. The method of any one of claims 69 to 76, further comprising identifying and removing one or more of: transposon insertion sites, sites that promote recombination, sites for cleavage by restriction endonucleases, and sites that are methylated.

78. The method of any one of claims 69 to 77, wherein the organizing comprises grouping coding sequences into operons based on similar native expression level.

79. The method of any one of claims 69 to 77, wherein the organizing comprises ordering coding sequences within operons such that the highest expressing gene, based on native expression, occurs first and the lowest expressing gene, based on native expression, occurs last.

80. The method of any one of claims 69 to 79, wherein the heterologous transcriptional regulatory sequence comprise a T7 RNA polymerase promoter.

81. The method of any one of claims 69 to 79, wherein heterologous transcriptional regulatory sequence(s) comprise an inducible promoter.
82. The method of any one of claims 69 to 81, further comprising operably linking one or more heterologous transcriptional terminator sequences to one or more coding sequences in the synthetic operon.
83. The method of claim 82, wherein the one or more heterologous transcriptional terminator sequences comprise a T7 RNA polymerase terminator.
84. The method of claim 82 or 83, wherein heterologous transcriptional terminator sequence for different synthetic operons have different sequences.
85. The method of any one of claims 69 to 84, wherein the operon(s) are expressed from a plasmid.
86. The method of claim 85, wherein the plasmid has a low copy origin of replication.
87. The method of any one of claims 69 to 86, wherein the expression cassette is contained in a control plasmid separate from any plasmid containing the synthetic operon(s).
88. The method of any one of claims 69 to 87, wherein the control promoter is an inducible promoter.
89. The method of any one of claims 69 to 88, wherein the heterologous polypeptide comprises an RNA polymerase (RNAP).
90. The method of claim 89, wherein the RNAP is T7 RNAP.
91. The method of any one of claims 69 to 90, wherein the expression cassette is an environmental sensor.
92. The method of any one of claims 69 to 91, wherein relative magnitude of expression of the coding sequences in the synthetic operon(s) correspond to relative protein levels in a native system containing the native codons.
93. A method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising:
providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

changing codon identity within at least one coding sequence by removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence;

organizing the coding sequences into one or more synthetic operon(s);

operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling magnitude of gene expression from the operon(s); and

expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence; and

detecting the magnitude of gene expression by computation, wherein the computation comprises a numerical optimization algorithm, and wherein the numerical optimization algorithm comprises the Nelder-Mead algorithm, the Newton's method, the quasi-Newton method, a conjugate gradient method, an interior point method, a gradient descent, a subgradient method, a ellipsoid method, the Frank-Wolfe method, an interpolation method and pattern search methods, or an ant colony model.

94. A method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising:

providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

changing codon identity within at least one coding sequence by removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence;

organizing the coding sequences into one or more synthetic operon(s);

operably linking one or more heterologous transcriptional regulatory sequence to the operon(s), thereby controlling magnitude of gene expression from the operon(s);

operably linking a heterologous ribosomal binding site (RBS) to one or more coding sequence in the synthetic operon, wherein different RBSs are operably linked to different coding sequences, and wherein the RBSs regulate translation of the coding sequences in a ratio that is similar to a ratio of translation from a native operon, and

expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

95. A method for expressing one or more synthetic operons collectively associated with a function in a cell by replacing native regulation of a set of genes with synthetic regulation, the method comprising:

providing coding sequences for a set of polypeptides encoded by genes collectively associated with a function;

changing codon identity within at least one coding sequence by removing at least one regulatory sequence within the coding sequence, wherein removing the at least one regulatory sequence comprises replacement of native codons in the coding sequence with non-native synonymous codons and comprises selecting non-native codons having maximal distance from the native codons of the coding sequence;

organizing the coding sequences into one or more synthetic operon(s), wherein the synthetic operon comprises two functional sequences selected from the group consisting of a promoter, a ribosome binding site, a coding sequence, and a terminator and the method further comprises operably linking a buffer sequence between two functional sequences, and wherein the buffer sequence is selected from the group consisting of a random sequence, a UP-region of a promoter, an extended 5-UTR sequence, and a RNAase cleavage site; and

expressing the one or more synthetic operon(s) in a cell under the control of a polypeptide that binds directly or indirectly to the heterologous transcriptional regulatory sequence.

96. A method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising:

providing a bacterial cell for expressing gene products;

providing a gene cluster having a plurality of native bacterial genes having coding sequences;

modifying the gene cluster by making at least one modification in at least one location within the gene cluster selected from the group consisting of a coding region and an intergenic region, wherein the gene cluster modification comprises replacing at least one native codon within one of the coding sequences to remove at least one native regulatory sequence using a synonymous codon and wherein the synonymous codon is a maximal distance from a corresponding native codon;

operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster; and

expressing gene products of the modified gene cluster in the bacterial cell under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

97. The method of claim 96, wherein at least two coding sequences of the plurality of native bacterial genes have at least one native codon replaced with a synonymous codon.

98. The method of claim 96 or 97, wherein at least one native regulatory sequence of the plurality of native regulatory sequences is identified using computation.

99. The method of claim 98, wherein the computation comprises searches of coding sequences for ribosome binding sites, terminators, promoters, or a combination thereof.

100. The method of any one of claims 96 to 99, wherein the heterologous transcriptional regulatory sequence is from the same species from which the plurality of native bacterial genes are native.

101. The method of any one of claims 96 to 99, wherein the heterologous transcriptional regulatory sequence is from a different species from which the plurality of native bacterial genes are native.

102. The method of any one of claims 96 to 101, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is expressed from a control expression cassette, the control expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.

103. The method of any one of claims 96 to 102, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is heterologous to the cell.

104. The method of any one of claims 96 to 102, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is from the same species from which the plurality of native bacterial genes are native.

105. The method of any one of claims 96 to 102, further comprising: detecting magnitude of the expressing by computation.

106. The method of claim 105, wherein the computation comprises a numerical optimization algorithm.

107. The method of any one of claims 96 to 106, wherein the gene cluster modification comprises replacing at least one intergenic region to remove at least one native regulatory sequence selected from the group consisting of a ribosome binding site, a terminator, and a promoter.

108. The method of claim 107, wherein at least one native regulatory sequence of the plurality of native regulatory sequences is identified using computation.

109. The method of any one of claims 96 to 106, wherein the gene cluster modification comprises:

altering at least one intergenic region of the plurality of intergenic regions within the gene cluster that contains the plurality of native regulatory sequences.

110. A bacterial nitrogen reduction expression system comprising nucleic acids encoding:

at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen fixation within a cell, wherein at least one of the plurality of coding sequences comprises non-native codons in place of a regulatory element, wherein said non-native codons have maximal distance from codons of the native coding sequence;

a heterologous promoter region that directs expression of the at least one operon; and

a heterologous transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system, wherein the protein binds directly or indirectly to the heterologous promoter region.

111. The bacterial nitrogen reduction expression system of claim 110, wherein the genes collectively associated with nitrogen fixation within a cell are selected from the group

consisting of: *nifJ*, *nifH*, *nifD*, *nifK*, *nifY*, *nifE*, *nifN*, *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, *nifM*, *nifB*, and *nifQ*.

112. The bacterial nitrogen reduction expression system of claim 110 or 111, wherein the heterologous promoter region is not the native promoter of a gene associated with nitrogen fixation.

113. A bacterial nitrogen reduction expression system comprising nucleic acids encoding:

at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen fixation within a cell, wherein at least one of the plurality of coding sequences comprises a non-native synonymous codon in place of a native codon, thereby removing a regulatory sequence;

a heterologous promoter region that directs expression of the at least one operon, wherein the heterologous promoter region is from the same species as the genes collectively associated with nitrogen fixation; and

a transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system, wherein the protein binds directly or indirectly to the heterologous promoter region, and wherein the transcriptional controller is not the native transcription controller of the genes collectively associated with nitrogen fixation under native regulation.

114. The bacterial nitrogen reduction expression system of claim 113, wherein the genes collectively associated with nitrogen fixation within a cell are selected from the group consisting of: *nifH*, *nifD*, *nifK*, *nifY*, *nifE*, *nifN*, *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, *nifM*, *nifB*, and *nifQ*.

115. The bacterial nitrogen reduction expression system of claim 113 or 114, wherein the coding sequences have been modified to reduce a predicted RNA secondary structure.

116. The bacterial nitrogen reduction expression system of any one of claims 110 to 115, wherein the heterologous promoter of the at least one operon causes a coding sequence of the operon to be expressed at an expression level which causes maximal nitrogenase activity.

117. The bacterial nitrogen reduction expression system of any one of claims 110 to 115, wherein the heterologous promoter of the operon causes two or more coding sequences to be expressed at an expression level which causes maximal nitrogenase activity.

118. The bacterial nitrogen reduction expression system of any one of claims 110 to 115, wherein the heterologous promoter causes each coding sequence of the operon to be expressed at an expression level which causes maximal nitrogenase activity.

119. The bacterial nitrogen reduction expression system of claim 116, 117, or 118, wherein an expression level of the coding sequence which causes maximal nitrogenase activity is determined by obtaining a wildtype bacteria with nitrogen reduction activity, deleting a native occurrence of said coding sequence, and providing a heterologous occurrence of said coding sequence under the control of an inducible promoter.

120. A bacterial nitrogen reduction expression system comprising nucleic acids encoding:

at least one operon comprising a plurality of coding sequences for a set of polypeptides encoded by genes collectively associated with nitrogen fixation within a cell, wherein at least one of the plurality of regulatory coding sequences has been synonymously mutated to remove internal regulation, and wherein at least one coding sequence has been modified to reduce a predicted RNA secondary structure;

a genetically engineered promoter region that directs expression of the at least one operon; and

a transcriptional controller coding sequence that encodes a protein that directs expression of the at least one operon of the expression system, wherein the protein binds directly or indirectly to the heterologous promoter region, and wherein the transcriptional controller does not regulate the genes collectively associated with nitrogen fixation under native regulation.

121. The bacterial nitrogen reduction expression system of claim 120, wherein the genes collectively associated with nitrogen fixation within a cell are selected from the group consisting of: *nifH*, *nifD*, *nifK*, *nifY*, *nifE*, *nifN*, *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, *nifM*, *nifB*, and *nifQ*.

122. The bacterial nitrogen reduction expression system of claim 120 or 121, wherein the heterologous promoter of the at least one operon causes a coding sequence of the operon to be expressed at an expression level which causes maximal nitrogenase activity.

123. The bacterial nitrogen reduction expression system of claim 120 or 121, wherein the heterologous promoter of the operon causes two or more coding sequences to be expressed at an expression level which causes maximal nitrogenase activity.

124. The bacterial nitrogen reduction expression system of claim 120 or 121, wherein the heterologous promoter causes each coding sequence of the operon to be-expressed at an expression level which causes maximal-nitrogenase activity.

125. The bacterial nitrogen reduction expression system of claim 122, 123, or 124, wherein an expression level of the coding sequence which causes maximal nitrogenase activity is determined by obtaining a wildtype bacteria with nitrogen reduction activity, deleting a native occurrence of said coding sequence, and providing a heterologous occurrence of said coding sequence under the control of an inducible promoter.

126. A method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising: providing a bacterial cell for expressing gene products; providing a gene cluster having a plurality of native bacterial genes having coding sequences; modifying the gene cluster by making at least one modification in a coding region or an intergenic region, wherein making the at least one modification in the coding region or the intergenic region comprises replacing at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster wherein the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes; and expressing gene products of the modified gene cluster in the bacterial cell under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

127. The method of claim 126, wherein at least two coding sequences of the plurality of native bacterial genes have at least one native codon replaced with a synonymous codon.

128. The method of claim 126 or 127, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is expressed from a control expression cassette, the control expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.

129. The method of claim 126, 127, or 128, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is heterologous to the cell.

130. The method of any one of claims 126 to 129, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes.

131. The method of any one of claims 126 to 130, further comprising: detecting the magnitude of gene expression of the expressed gene products by computation.

132. The method of claim 131, wherein the computation comprises a numerical optimization algorithm.

133. The method of any one of claims 126 to 132, wherein modifying the gene cluster comprises replacing at least one intergenic region to remove at least one native regulatory sequence selected from the group consisting of a ribosome binding site, a terminator, and a promoter.

134. The method of any one of claims 126 to 132, wherein modifying the gene cluster comprises: altering at least one intergenic region within the gene cluster to modify a native regulatory sequence.

135. The method of claim 134, further comprising: identifying the native regulatory sequence using computation.

136. The method of claim 134 or 135, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes.

137. The method of claim 134, 135, or 136, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is expressed from a control expression cassette, the control expression cassette comprising a control promoter operably linked to a polynucleotide sequence encoding the polypeptide.

138. The method of any one of claims 134 to 137, further comprising: detecting the magnitude of gene expression of the expressed gene products by computation.

139. A method of altering regulation of a plurality of native bacterial genes associated with a function in a cell, comprising: providing a bacterial cell for expressing gene products; providing a gene cluster having a plurality of native bacterial genes having coding

sequences; modifying the gene cluster by making at least one modification in a coding region or an intergenic region, wherein making the at least one modification in the coding region or the intergenic region comprises replacing at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; operably linking at least one heterologous transcriptional regulatory sequence to at least one coding sequence within the modified gene cluster, wherein the at least one heterologous transcriptional regulatory sequence is from a different species than the plurality of native bacterial genes; and expressing gene products of the modified gene cluster in the bacterial cell under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

140. The method of claim 139, wherein the bacterial cell is from a nitrogen fixing bacterial species.

141. The method of claim 139 or 140, wherein the gene cluster comprises genes collectively associated with nitrogen fixation.

142. The method of claim 139, 140, or 141, wherein the gene cluster comprises *nif* genes.

143. The method of any one of claims 139 to 142, wherein modifying the gene cluster comprises making at least one modification in a native regulatory sequence.

144. The method of claim 143, further comprising identifying the native regulatory sequence using a computational algorithm.

145. The method of any one of claims 139 to 144, wherein modifying the gene cluster comprises making at least one modification in a promoter.

146. The method of any one of claims 139 to 145, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is heterologous to the bacterial cell.

147. The method of any one of claims 139 to 146, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes.

148. The method of any one of claims 139 to 147, further comprising detecting the expressed gene products.

149. The method of any one of claims 139 to 148, wherein the gene cluster is heterologous to the bacterial cell.

150. The method of any one of claims 139 to 149, further comprising expressing the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence from an expression cassette.

151. A recombinant bacterial cell comprising a modified gene cluster, wherein the modified gene cluster comprises a plurality of native bacterial genes having coding sequences and comprises at least one modification in a coding region or an intergenic region, wherein the at least one modification in the coding region or the intergenic region comprises a replacement of at least one native codon within one of the coding sequences to modify at least one native regulatory sequence using a synonymous codon, wherein the synonymous codon is a maximal distance from a corresponding native codon; wherein at least one coding sequence within the modified gene cluster is operably linked to at least one heterologous transcriptional regulatory sequence; wherein the at least one heterologous transcriptional regulatory sequence is from a different species than the plurality of native bacterial genes; and wherein the expression of gene products of the modified gene cluster in the bacterial cell is under the control of a polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence.

152. The recombinant bacterial cell of claim 151, wherein the bacterial cell is from a nitrogen fixing bacterial species.

153. The recombinant bacterial cell of claim 151 or 152, wherein the modified gene cluster comprises genes collectively associated with nitrogen fixation.

154. The recombinant bacterial cell of claim 151, 152, or 153, wherein the modified gene cluster comprises *nif* genes.

155. The recombinant bacterial cell of any one of claims 151 to 154, comprising at least one modification in a native regulatory sequence.

156. The recombinant bacterial cell of any one of claims 151 to 155, comprising at least one modification in a promoter.

157. The recombinant bacterial cell of any one of claims 151 to 156, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is heterologous to the cell.

158. The recombinant bacterial cell of any one of claims 151 to 157, wherein the polypeptide that binds directly or indirectly to the at least one heterologous transcriptional regulatory sequence is from the same species as the plurality of native bacterial genes.

159. The recombinant bacterial cell of any one of claims 151 to 158, wherein the modified gene cluster is heterologous to the bacterial cell.

160. The recombinant bacterial cell of any one of claims 151 to 159, further comprising an expression cassette comprising a polynucleotide sequence encoding the polypeptide.

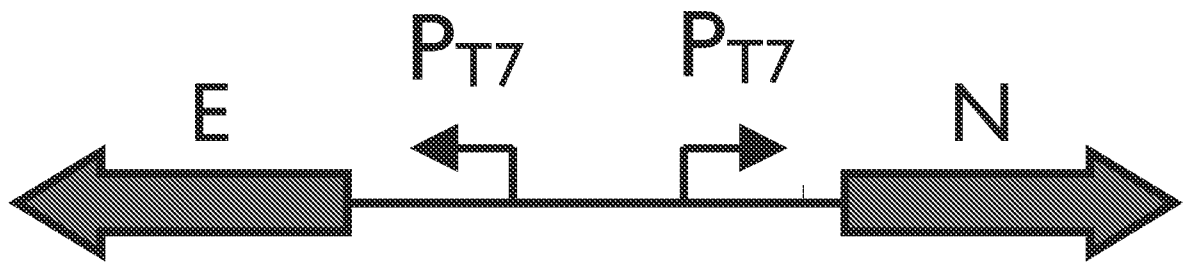


FIG.1

2009.12.10 - Kpn T7 promoter lib at 1 mM IPTG

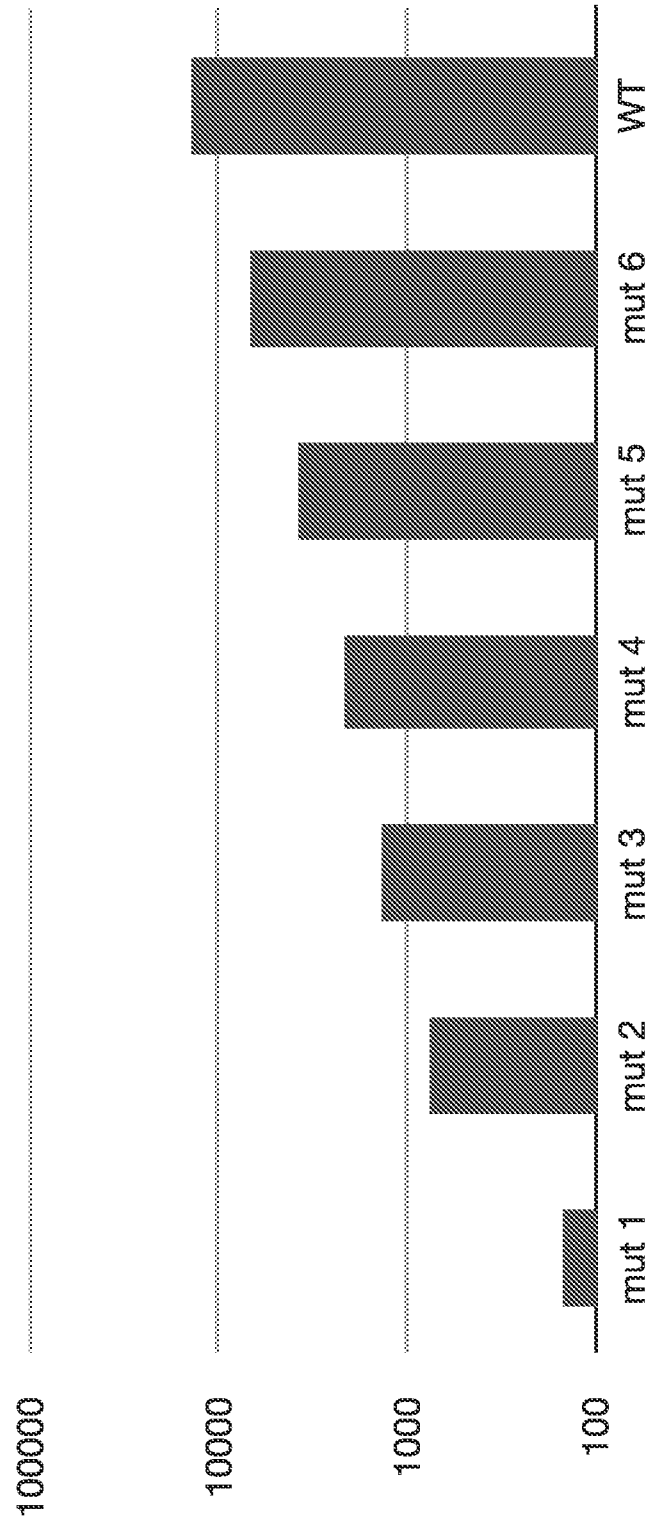
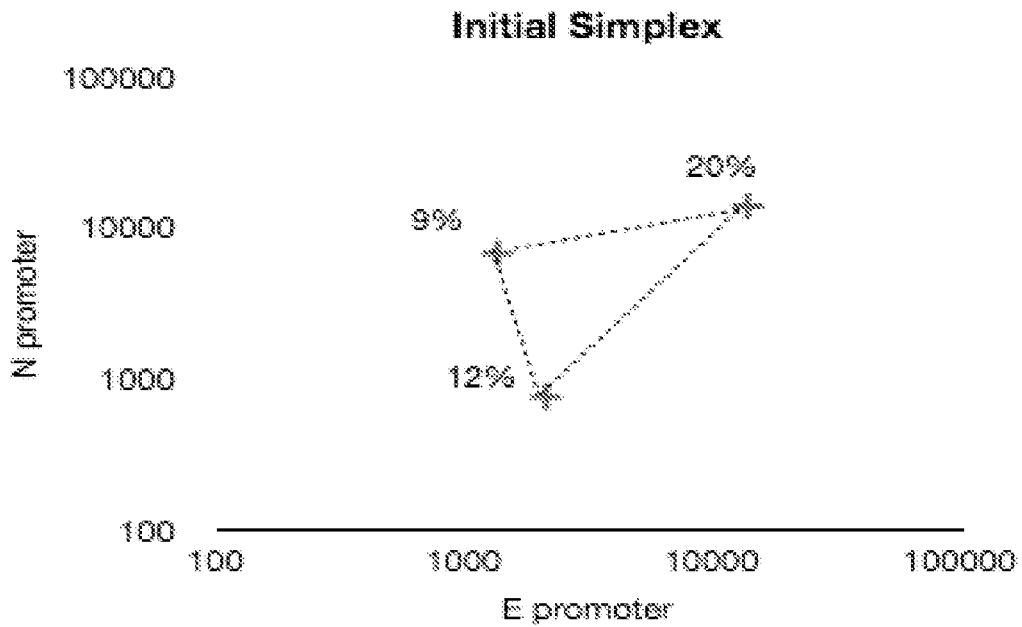


FIG. 2

Promoter Name	Sequence TAATACGACTCACTANNNNNAGA	Strength
WT	TAATACGACTCACTATAGGGAGA	13780
Mut1	TAATACGACTCACTACAGGCAGA	147
Mut2	TAATACGACTCACTAGAGAGAGA	752
Mut3	TAATACGACTCACTAATGGGAGA	1346
Mut4	TAATACGACTCACTATAGGTAGA	2127
Mut5	TAATACGACTCACTAAAGGGAGA	3738
Mut6	TAATACGACTCACTATTGGGAGA	6710

FIG. 3

A



B

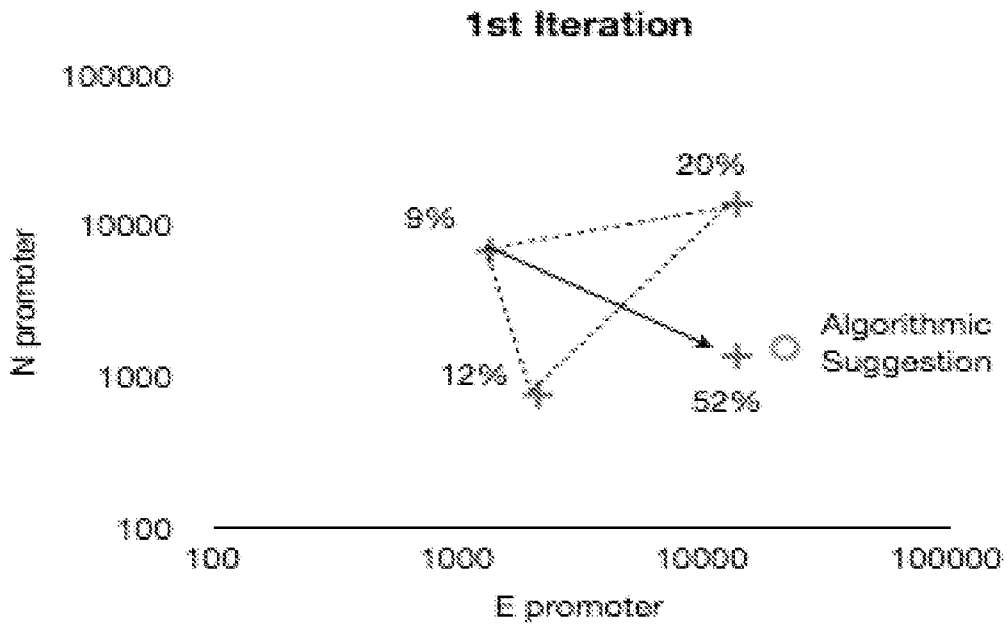


FIG. 4

Strain Name	E Promoter	N Promoter	Fixation
#1 (N259)	WT	WT	20%
#2 (N438)	Mut 4	Mut 5	9%
#3 (N307)	Mut 4	Mut 2	12%
Reflection (N319)	WT	Mut 3	52%

FIG. 5

Refactoring Nitrogen Fixation

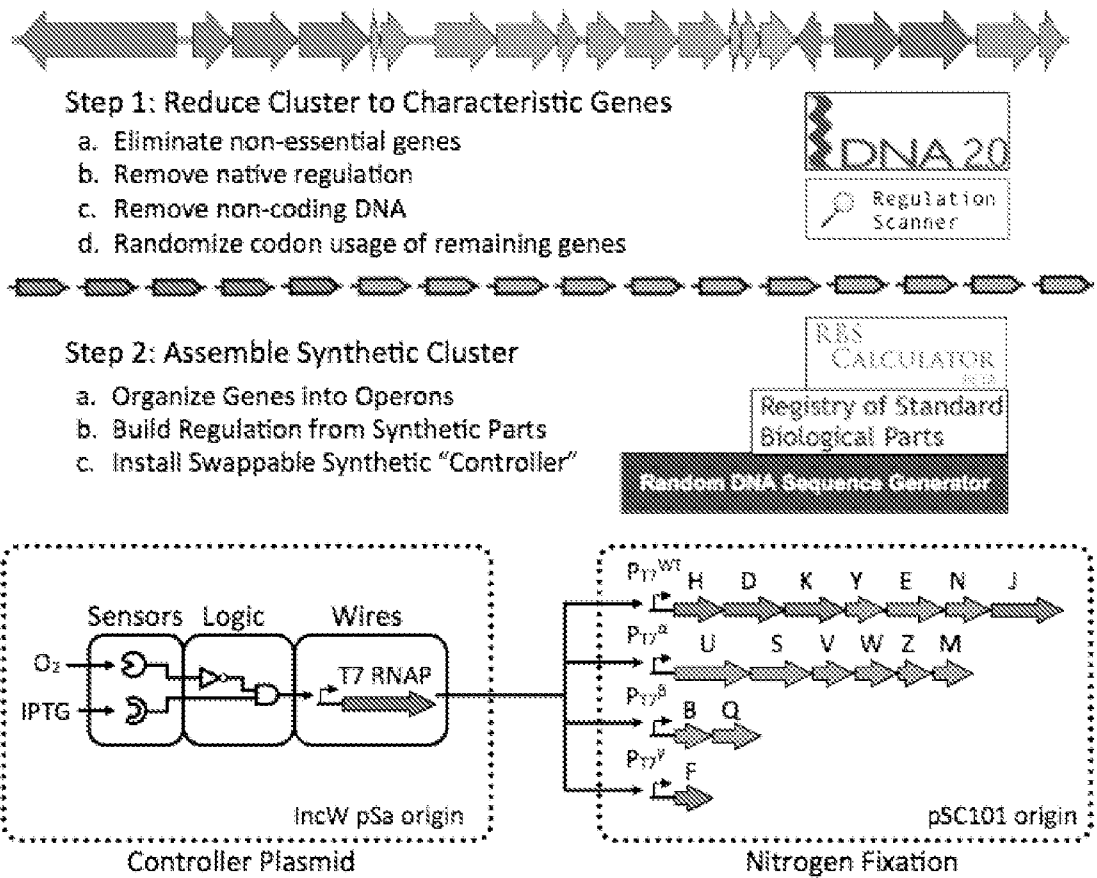


FIG. 6

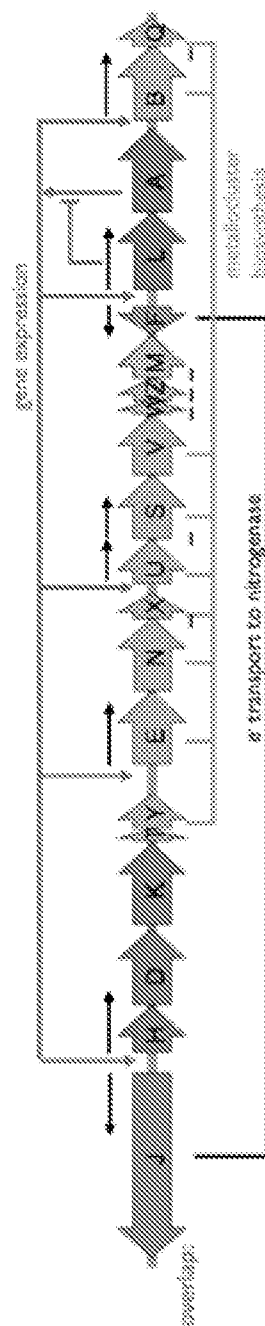
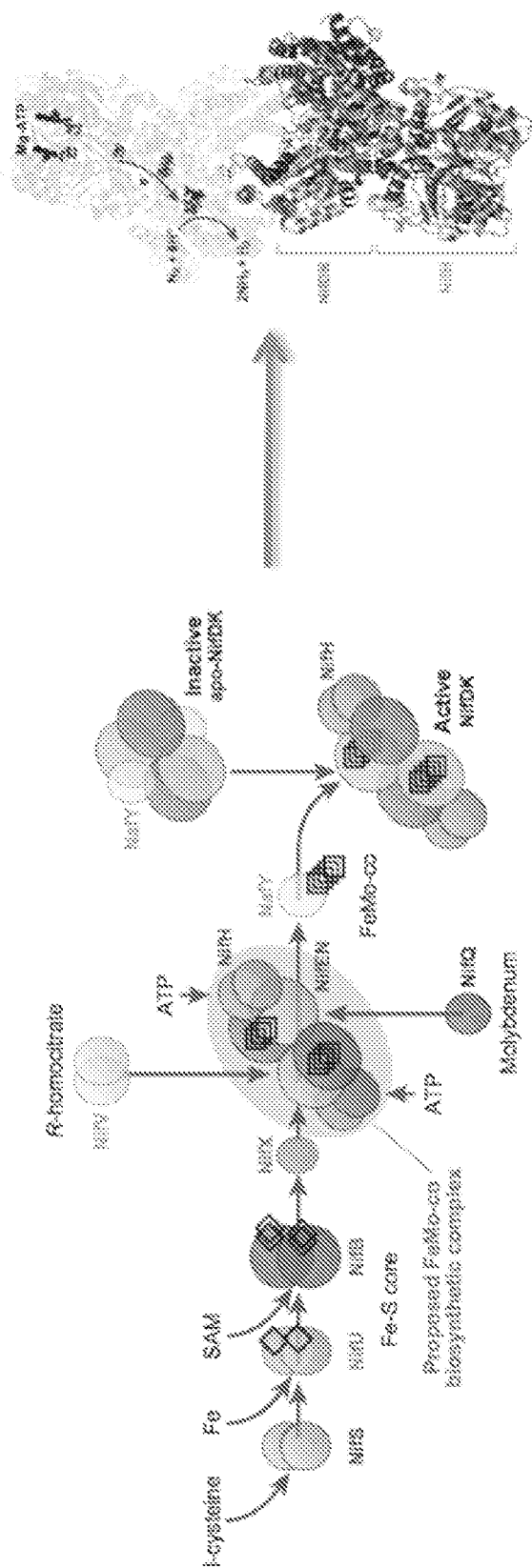


FIG. 7

Ptac

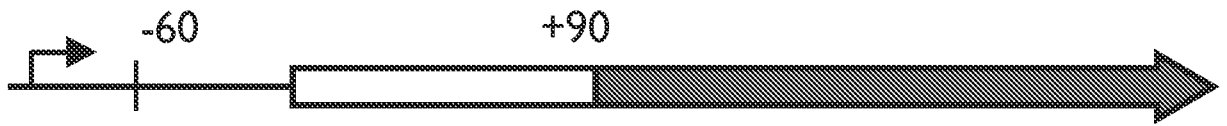


FIG. 8

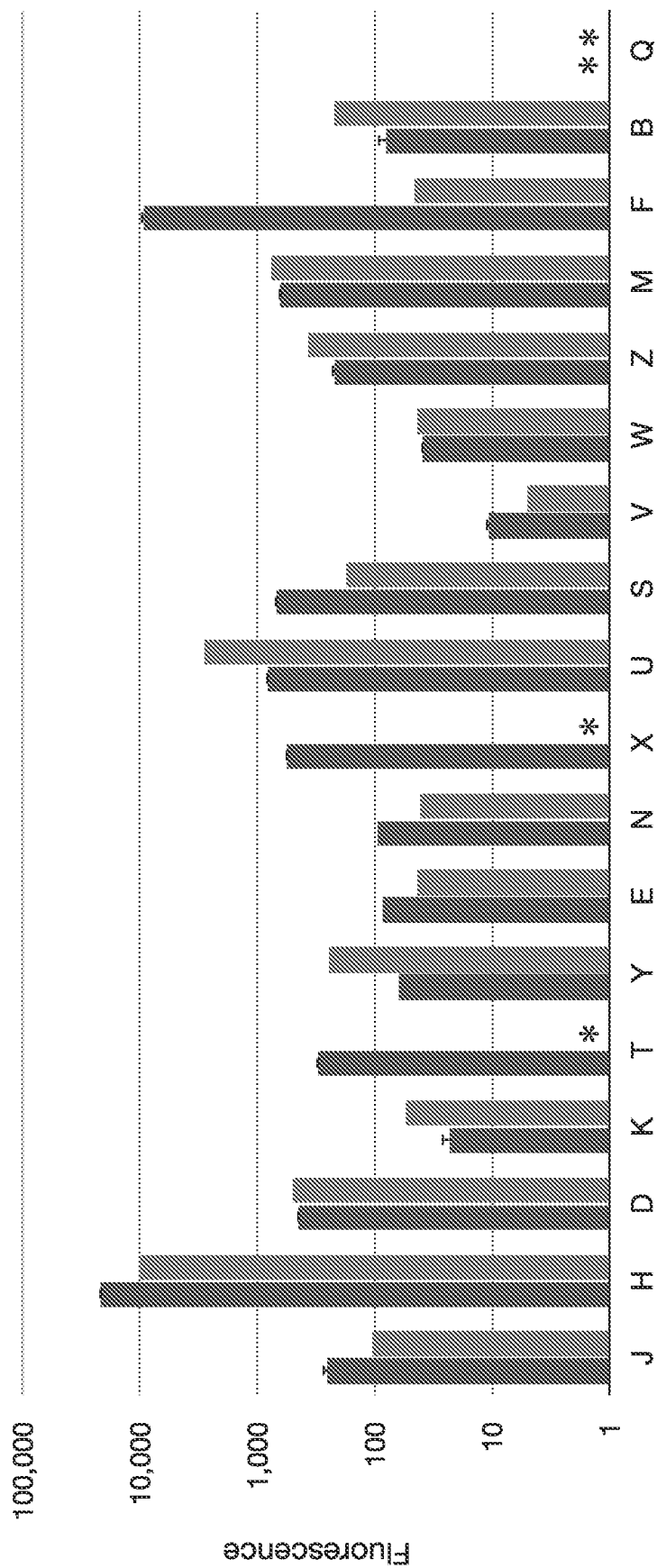


FIG. 9

Gene	Native RBS Strength	Synthetic RBS Strength
J	254	104
H	21641	9903
D	446	496
K	23	54
T	303	Eliminated
Y	62	243
E	85	43
N	95	41
X	558	Eliminated
U	814	2832
S	686	174
V	11	5
W	39	43
Z	218	366
M	637	754
F	9283	46
B	80	221
Q	Not Detectable	Not Detectable

FIG. 10

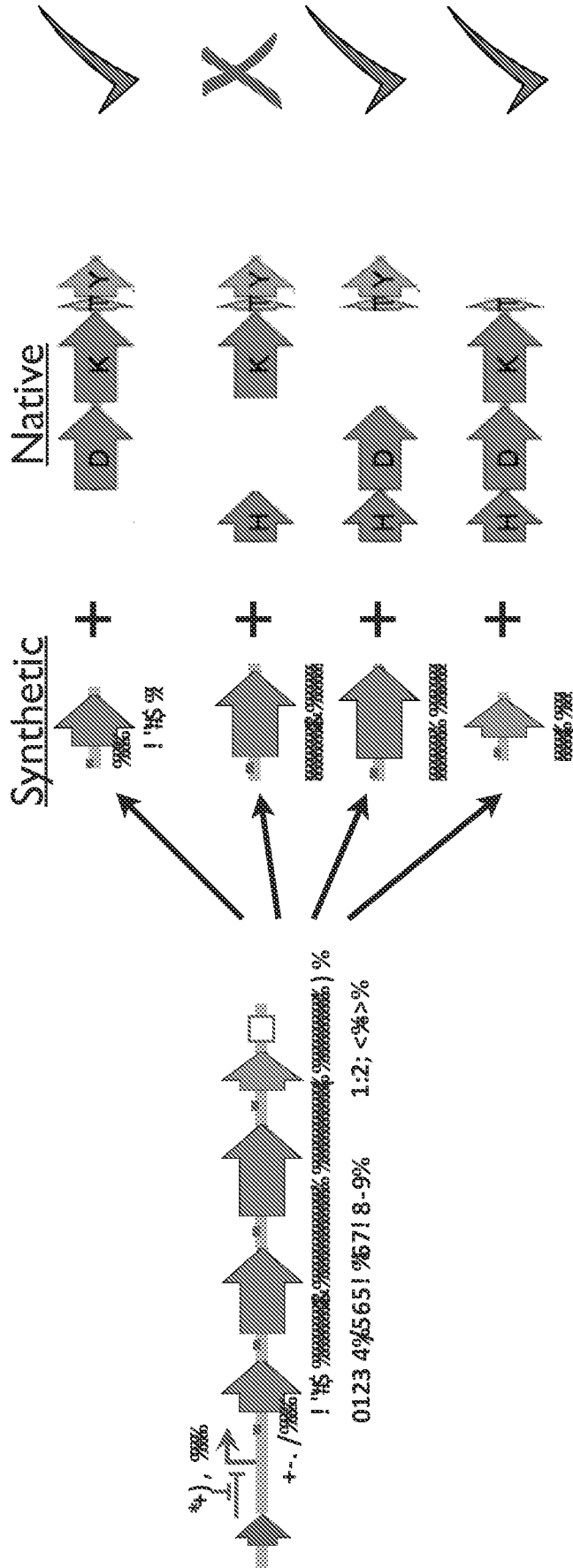


FIG. 11

Operon	Error Location	Error
HDKY	H	Transposon insertion
USVWZM	S	Sequencing errors in Arnold NCBI entry (L110M, Q113E, S124G, R290A)
USVWZM	V	Sequencing errors in Arnold NCBI entry (L257S, D295H)
USVWZM	M	Cloning mutation
USVWZM	Ptac	Cloning error yielded constitutive promoter

FIG. 12

Performance of Individual Operons

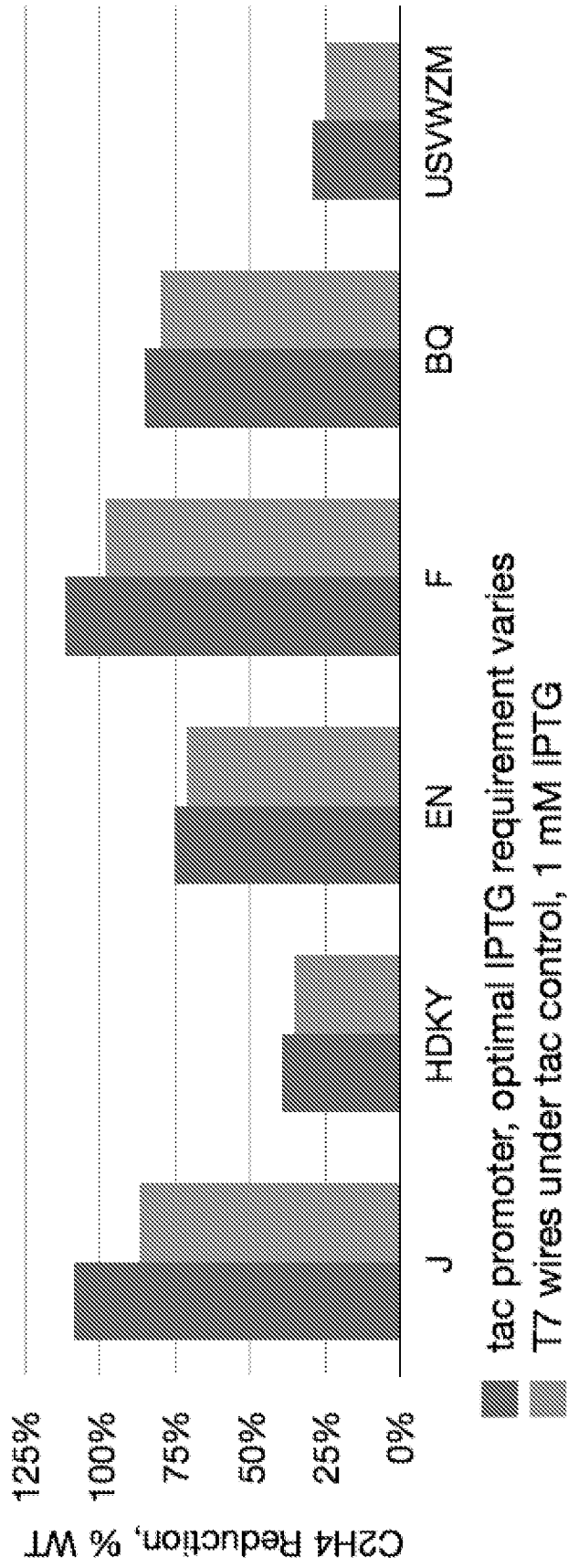


FIG. 13

Operon	Function with Ptac promoter	T7 promoter	Function with T7 promoter
J	109%	WT	87%
HDKY	39%	WT	35%
EN	75%	WT	71%
USVWZM	29%	Mut 2	25%
F	112%	Mut 3	98%
BQ	85%	Mut 2	80%

FIG. 14

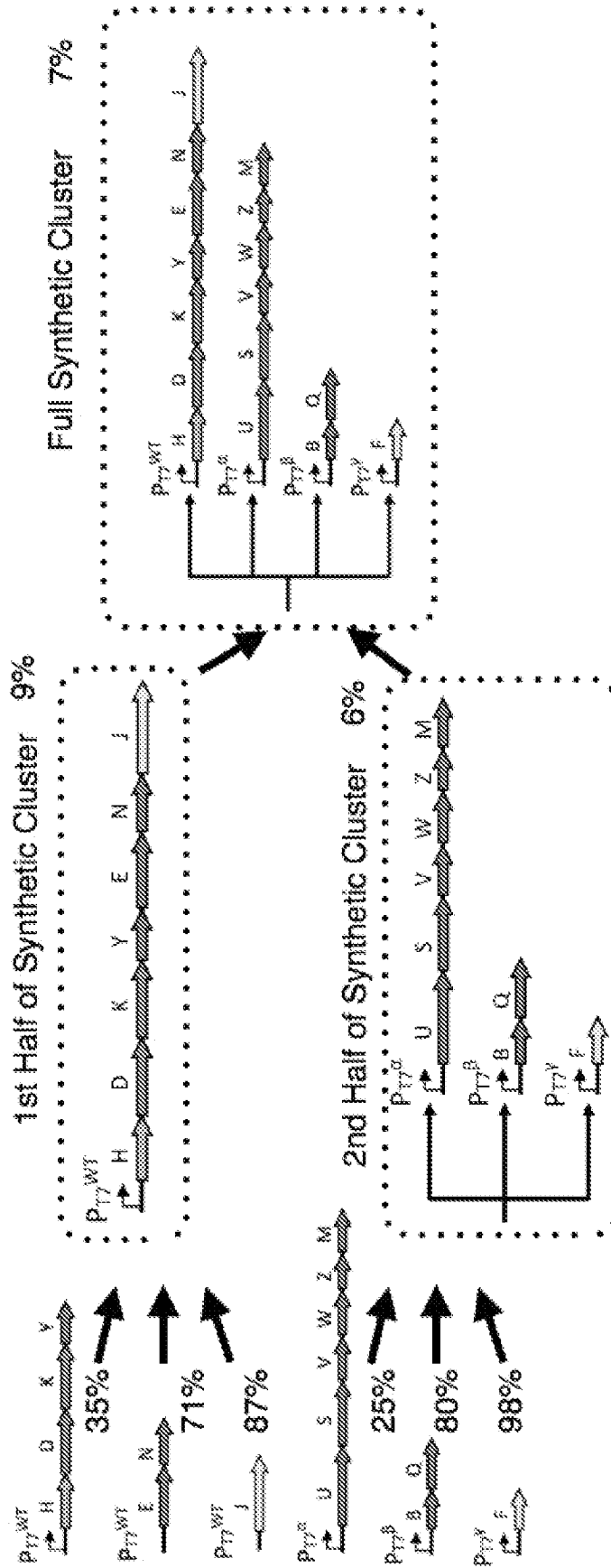


FIG. 15

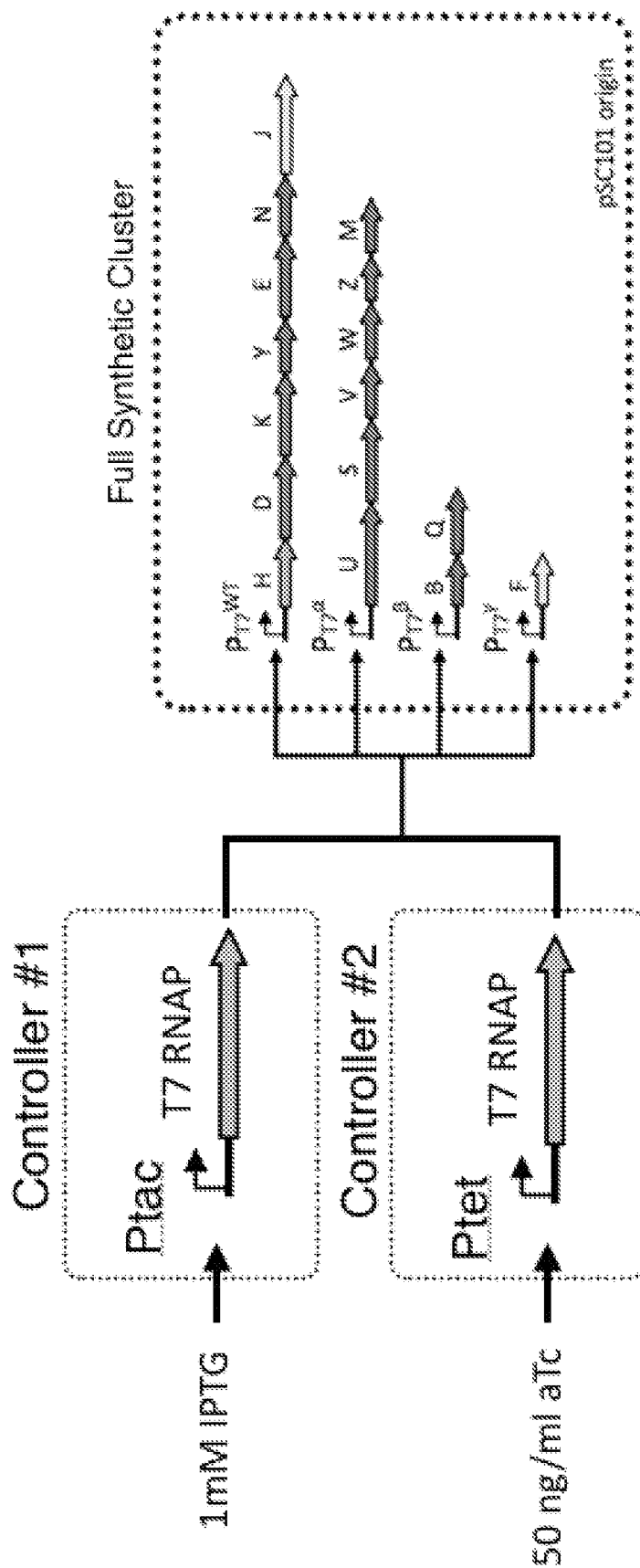


FIG. 16

Fig. 18

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
J	WT Gene J is SEQ ID NO:1 Synthetic Gene J is SEQ ID NO:2	atgaaaacaatggatggcaacgccgcggcgcc tggatctcttalgcccttaccgaggtcgcggcgattt accccataaaccctccacgcgatggcgga acgctcagcagtgggcgccgaggggaaaaag aaccttttggccagccggctgcctaatggagat gcagtcggaggccggcgcggcagggcggtcc acggcgcgctgcagggccggggcgctcaccacc acctatacggctccaggggctgctgctgatgat ccccaacatgtacaaaatcggcggtgaactgctg ccggggcgtcttccagctcagcggccggcgctgg cgaccaatfcgctgaalatitllggcgatcaccagg atgtgatggcggctccggcagaccggctgcggat gctggcgggagaacaacgtgcagcaggigatgga ttgtcggcggtggcgaictggcgcgatlaagg gacgcacccgttllglaaacttctcagcggtttgg caactcgcacgaaatcaaaaaatcgagggtgtgg aalatgagcagctggcgacgctgctggaccggc ccgcgcgcgacagcttccgcgtaacgcgctgca tcgggatcaccggctcaccgggaacggcgag aacccggatatacttccaggagcgggagggcg ggcaaccgttlltatacggcgctgcccgaactgtc gaaagctalatgacgcaaaatcagcgcctcaccg gccgggaglatcacttcttaactataccggcgcg gggatgcccgaacgggtgallatcgcaatgggat cggctcgcgataaccgltcaggaagtgggtgatac gctgaacgcggcgggagagaaggctggcgctgct cagcgtgcatctgttcccttcttctgctgcccac ttcttcccagctgcccgaaccgtgcagcggat tgccgtcttgaccgtaccaaaagagccggcgct caggctgaaccgctgtgctggtatgtaaaaacg ccttctaccaccacgagatgcgccgtaaatcgtc ggcggccgctatgcctggcgggaaggatgtc tgcccaacgataatcggccgtcttggataacctca ataaacctgctccgatggacgggtllaccctcgg gattgtcgatgatgaaccttacttctgtaaccgg gccagcagacgctggcggctcgcgatgacggc atcaccgcctgtaaatitggggcatgggctcgg cggcaccgctggggccaalaaaagcgcgatcaa aattatcggcgataaaaacggcctctacggcgag gectactttctatgactegaaaaatccggcggc attaccggttcaatitacgcttggcgaccggcgg atcaatcgcctacctgatccatcgggcccatttt atctctgttcgcagcagctctacgttgagcgtac gatctgctggacggatlaaacggcgggcggaact tittactcaactgcagctggtcagatcggagctg gagcagcatctgcggctggcttlaaacgctatct ggcgcgggaaaaatccttittataccctgaacg ccgtggatatacggcgagctcgggctggcg ggcgcttlaatatgctgatgcagcggcgcttctta agctggcggcgattatcagccgcagaccggg cggattaccicaagcaggggttgaaaaaagcta cggcagcaaaaggggcggtgallgagatgaa	ATGAAAACATATGGACGGTAA CGCTGCGGCTGCATGGATTA GCTACGCCTTTACCGAAGTGG CTGCGATCTACCCGATTACGC CGAGCACCCCGATGGCGGAA AATGTGGACGAATGGGCTGC GCAGGGCAAGAAGAACCTCT TCGGCCAGCCGGTGC GCCTG ATGGAGATGCAGTCCGAAGC GGGTGCAGCAGGTGCTGTGC ATGGCGCCTTGCAAGCTGGC GCACTGACGACCACCTACAC CGCGTCGCAGGGCCTGTGCT GATGATCCCAAACATGTACA AAATCGCGGGTGAACCTGCTG CCGGGTGTCTTTCATGTTTCG GCACGCGCACTGGCCACCAA TAGCCTCAACATCTTTGGCGA TCATCAGGATGTAATGCCGG TGCGCCAAACGGGCTGCGCG ATGTTGGCCGAGAATAACGT CCAGCAAGTTATGGATTTGTC CGCGGTAGCCCACTTGCCAG CGATCAAAGGTTCGATTCCGT TCGTGAACTTCTTCGATGGCT TTCGCACCAGCCACGAAATC CAGAAGATCGAGGTTCTGGA ATATGAACAGCTGGCCACCTT GTTGGATCGTCCGGCCCTGGA CAGCTTCCGCCGTAACGCCCT TCACCCGGACCACCCGGTCAT CCGTGGCACCCGCCAGAACC CGGACATCTACTTCCAGGAA CGTGAGGCCGGTAACCGTTTC TATCAGGCGCTCCCGGATATT GTGGAATCTTACATGACCCA GATTTCTGCCCTGACTGGTCC CGAGTATCACCTGTTTAACTA CACTGGTGCTGCGGATGCGG AGCGCGTGATCATCGCGATG GGCTCTGTCTGTGACACCGTC CAAGAGGTGGTTGACACGCT GAATGCAGCGGGTGAGAAAG TTGGTCTGTCTCCGTTTCATC TTTTCCGCCCGTTTTTCGTTAG CGCACTTCTTCCGCCAACTGC CGAAAACCTGTACAGCGTATC GCAGTATTGGACCGTACGAA AGAGCCAGGTGCTCAAGCAG AGCCGCTGTGCCTCGATGTGA AGAATGCCTTTTACCACCATG ACGATGCCCCGTTGATTGTGG	80.8	48

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
		ccagcgggogcagcagcctgggcatggcctcgcg gcatcagggtgacgattccggcgcactggggcagc gctggatgaaccccgggcgaagcagcagccat gatgccggattcaccggcagatattctgcagccgat gaaccgcagtgccggcaccagcagccgggtgag cgcgttcgctggatggaggacgggaccttccct cgggcaccgcccgctgggagaaacgcgggac gcgctggaagtggcggctggcagccggagggc tgcacgcagtgtaaccagtgcccttattctgcc gcatcggcgatccgccggcgctgctcaacgg cgaagagcagcagccggcgggtgggctgct gagcaaacccggcagggagcgaaggagatc actaccatctggctatctgcgcctggattgtccg gctggcgaactgcgctgatactgctccgcggc cggcaaggcgttaaaaagcagctctcctgafagc agcgtcagatggcggcggctgggactatgctg ggcgctgacgccgaagagcaatcgttctgaag acgacggctcaaggcagccagttgaaaccccg ctgctggagtttccggcgcctggcgggatgctg gtgaaacgccttatgccggcctgataaccagctg ttggcgaccggatgctgctgctaacgcccagc gctgctctctatctggggagccagcggcggctg attccctataaccaccaaccacggccatggcc cggcatggcgaacgcctgttggagataatgc ggagttggcctggcagatgctggcggccag gcggctgctcaacaatggcagatgatgaccg ccgctggcctaccggcagcagcgaacttag cgaagcaatggcagctggcggcgaagcaggga tgaaggcggggcaccggcggcggcgggacc ggctcagcgaacggctggcggcggaaaaagag ggctgcccgtgttggagcagctgtggcagaacc ggcactatttggctgctgctggcagtgatttgg cggcggcggcggcctacgatalcggcttggc ggctcagcagctgctggcggcggggaagac gtaataatctgggtgtgacaccaggttactcca afaccggcggcagcctgctaaatgacccgggt ggccggcctcggaaagtggcggcagggcga acgacggcggaaaaagatcggcagatggc gatgagctacggcaatgtgacgtggcccaggct gcgatggcggctgataaagatcagaccctacgg gcgacggcggggcgaagcctggcgggacc gtcgctgggtgattgctacggcggctgcttaacc acgggctgaaagccggatgctgctgagccagc gcgaggcgaacggcgggtgaggcgggatact ggcaactgtggcctatcaccagcggggaag cggaaaggtaagacggcgttattgctgatccgaa gagccggaggagagctccggcacttttctgg cgaagtgcctacgctcctgctgacaaaaagac gccgcatctggcggatgctgctttagccgaacc gaggaggacggcggggccggcttggccagat cggcggctggcggcggcggaggatgag	GTGGTCGCTATGCCTTGGGCG GTAAGGACGTGTTGCCGAAC GATATTGCGGCCGTGTTTGAT AACCTGAACAAACCGCTGCC GATGGACGGCTTCACGCTGG GTATCGTGGACGATGTTACCT TCACCTCTCTCCCGCCAGCGC AGCAGACCCTGGCGGTTTCTC ACGACGGCATCACGGCATGT AAGTTTTGGGGCATGGGCTCC GACGGCACGGTTGGTGGCAA CAAGTCCGCGATCAAGATTA TCGGCGACAAAACGCCACTG TATGCGCAAGCGTACTTTTCC TACGACTCGAAGAAGAGCGG TGGTATTACCGTCAGCCATCT GCGTTTTGGTGATCGCCCCGAT CAACTCCCCGTATTTGATCCA TCGCGCGGATTTTCATCTCGTG CAGCCAGCAAAGCTATGTTG AACGCTACGATCTGCTGGAT GGCTTAAACCGGGTGGCAC CTTTCTGCTGAACTGCTCCTG GAGCGATGCCGAACTGGAGC AACATCTGCCGGTTCGGTTCA AACGTTATCTGGCACGCGAG AATATCCACTTCTACACTCTC AACGCTGTGGACATCGCCCCG TGAGCTTGGTTTGGGTGGCCG TTTCAACATGCTGATGCAGGC TGCTTCTTCAAACCTGGCCGC GATCATTGACCCGCAGACTG CTGCGGACTATCTGAAGCAG GCTGTTGAGAAAAGCTATGG CAGCAAAGGTGCGGCGGTCA TCGAGATGAACCAGCGTGCC ATCGAGCTTGGCATGCCAG CCTGCACCAGGTGACGATCC CGGCACATTGGGCCACCCCTG GATGAGCCAGCGGCGCAGGC GTCCGCGATGATGCCGGACTT TATCCGCGACATCCTGCAACC GATGAACCGTCAGTGC GGCG ACCAGCTTCCGGTGTGGCTT TTGTGCGCATGGAAGATGGC ACCTTCCCGTCCGGCACGGCC GCATGGGAGAAAACGTGGCAT CGCCCTTGAGGTGCCAGTCTG GCAGCCGGAAGGCTGCACGC AGTGCAACCAGTGC GCCTTC ATTTGTCCGCACGCCGCGATT CGTCCGGCGTTGTTGAATGGC GAAGAGCATGATGCTGCCCC GGTTGGCCTGCTGAGCAAAC		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			CGGCACAAGGCGCTAAAGAA TATCACTATCATCTGGCGATT AGCCCGCTGGACTGCTCCGG CTGTGGCAACTGCGTTGACAT TTGTCCAGCTCGTGGCAAAGC GTTGAAGATGCAGTCTCTGG ATAGCCAACGCCAGATGGCT CCGGTGTGGGATTATGCGCTG GCGCTGACCCCGAAGTCTAA CCCGTTTCGTAAAACCACCGT CAAAGGCTCGCAGTTCGAAA CCCCCTGCTGGAGTTTAGCG GTGCGTGCCTGGTTGTGGCG AAACGCCGTATGCGCGCCTC ATTACCCAGCTGTTTGCGGAC CGCATGCTGATTGCCAATGCC ACCGGCTGTTCCAGCATCTGG GGCGCATCTGCGCCGAGCAT CCCGTATAACCACCAATCATCG TGGTCATGGTCCGGCCTGGGC GAATAGCCTGTTTGAGGACA ATGCCGAATTTGGTTTAGGTA TGATGCTGGGCGGTCAAGCT GTGCGTCAACAGATCGCGGA CGATATGACGGCTGCGTTAG CGCTCCCGGTTTCCGATGAGC TGAGCGACCGATGCGCCAG TGGTTGGCGAAACAGGACGA GGGTGAAGGCACGCGTGAGC GTGCGGACCGTCTGAGCGAG CGCTTAGCCGCGGAGAAAGA GGGCGTTCGCTGTTAGAGC AGCTGTGGCAAATCGTGAT TACTTTGTGCGTTCGAGCCAG TGGATTTTCGGCGGTGACGGC TGGGCCATGATATTGGCTTC GGTGGCCTGGACCACGTCCTC GCCAGCGGTGAGGATGTGAA CATTCTGGTATTTGACACCGA AGTCTACTCGAACACCGGCG GTCAAAGCAGCAAATCGACC CCGGTCGCGGCCATCGCCAA GTTCGCGGCTCAGGGCAAGC GCACCCGCAAGAAAGACCTG GGTATGATGGCGATGAGCTA CGGCAACGTCTATGTAGCCC AGGTGGCGATGGGTGCGGAT AAAGATCAAACCTCTGCGCGC CATTGCGGAAGCTGAAGCGT GGCCAGGCCCGTCGCTGGTG ATTGCGTATGCGGCCTGCATC AATCATGGCCTGAAGGCCGG TATGCGTTGCAGCCAACGTG AGGCGAAGCGCGCTGTTGAG		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			GCGGGCTACTGGCACCTGTG GCGTTATCACCCGCAGCGCG AAGCGGAAGGCAAGACGCCG TTTATGTTAGATAGCGAAGA ACCGGAAGAGTCGTTCCGTG ACTTCTGTTGGGTGAGGTGC GCTACGCATCCCTGCACAAG ACCACCCCGCACCTCGCCGAT GCCCTTTTCAGCCGTACCGAA GAAGATGCGCGTGCGCGCTT TGCGCAATACCGTCGCCTGGC TGGCGAAGAGTAA		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
H	WT Gene H is SEQ ID NO:3 Synthetic Gene H is SEQ ID NO:4	atgaccatgogtcaatgcegtatttacggtaaagg cggatcgggtaaatccaccaccaacgcagaacctc gtcgccgcgctggcggagatgggtaagaaagtg atgalegtoggtgcgatcgaaggcggactcca ccgctctgattctgcacgcaaaagcacagaacac caftatggagatggccgggaagtggctcggtc gaggacctcgaactcgaagacgtgctgcaaattg gtacggcgatgtgcgctgcgcggaatccggcg gcccggagccaggcgtcggtcgcgcgggacgc ggcgtgatacggcgatcaactttctgaagaaga aggcgccctacgaggacgatctcgattcgtgtteta tgacgtgctcggcgacgtggctcggcggtctc gccatgccgatccgcgaaacaagcccaggag atclacatcgtctgctcggcgaaatgatggcgat gtacgcggccaacaatatctcaagggatcgta aafacgccaaatccggcaaggtgcgctcggcg gcctgafctgtaactcagtcagaccgaecgtgaa gacgaactgatattgcccctggcggaagactcg gtaccagatgatccatttgtgccccgcgacaac atcgtgcagcgcggagatccgccgatgacg gtatcagtagtacgaccccgctgtaaacaggcca acgaataccgacccctggcgagagatcgteaa caacaccatgaaagtgggtgcgacgccctgacc atggatgagctggaatcgtctgatggagtcgg catcatggaagagggaagacaccagcatattggc aaaaccgcccggaagaaaacggcgctga	ATGACCATGCGTCAGTGCGC GATTTATGGCAAAGGTGGTA TTGGCAAAGCACGACGACC CAGAACTTGGTGGCGGCCCT GGCCGAGATGGGTAAAAAGG TTATGATTGTGGGTTGCGACC CGAAGGCCGACAGCACGCGC CTGATTCTGCACGCGAAAAGC ACAAAACACGATTATGGAGA TGGCTGCCGAGGTTGGTAGC GTGGAGGATCTGGAGCTGGA GGACGTTCTGCAAATTGGTTA CGGTGATGTTTCGTTGCGCAGA GAGCGGTGGTCCGGAACCCAG GTGTCGGCTGTGCGGGTCTGTG GTGTGATTACCGCTATCAATT TCCTGGAAGAAGAGGGTGCG TACGAAGATGATCTGGATTTT GTTTTCTACGATGTGCTGGGT GATGTCGTGTGCGGTGGTTTT GCAATGCCGATTCGCGAGAA TAAGGCACAAGAAATTTACA TTGTCTGTAGCGGCGAGATG ATGGCAATGTACGCTGCTAA CAACATCAGCAAGGGTATTG TTAAATACGCAAAAAGCGGT AAGGTTTCGCTTGGGTGGTTTTG ATTTGCAACAGCCGTCAGAC CGACCGTGAGGACGAACTGA TCATCGCCCTGGCTGAGAAA CTGGGCACCCAAATGATCCA CTTCGTGCCACGCGATAATAT TGTTCAACGTGCAGAAATCC GCCGTATGACCGTCATTGAGT ATGACCCGGCATGCAAGCAA GCCAACGAGTACCGCACCTT GGCACAGAAAATCGTGAACA ACACCATGAAGGTTGTTCCG ACGCCGTGTACGATGGACGA GCTGGAGAGCCTGCTGATGG AGTTCGGCATTATGGAGGAG GAGGACACCAGCATTATCGG TAAGACCGCAGCGGAGGAGA ATGCGGCATAA	79	41.8

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
D	WT Gene D is SEQ ID NO:5 Synthetic Gene D is SEQ ID NO:6	atgatgaccaacgcaacggcgcaacgtaatctggcgtgatccaggaagtcctggagggttcccgga aaccgcgcgaaaagagcgcagaaagcacatgat ggtcagc gatccgaaaatgaagagcgtcggc aagtgcaattatcttaaccgcaaatcacaacccggc gtaatgaccgtacggcgtcgcctacgcgggtc caaaggggtggtatttggcccgattaaggatagg cccatatttcgacggaccggctggctcggccca gtaftcccgcgccgaacgacgcaactactacacc ggagtcagcggcgtgatagctcggcaocgtga acttcacctctgatfllcaggagcgcgacalcgtct cggcggcgataaaaagctcagcaagctgaltgaa gagatggagttctgttccgcaccaaagggat caccaticagtcggaatgccgggtgggctgac ggtgatgatcagcgcggcggccaacgccagc agcaaggcgcctggalaaccgggtgacccgggtac gctgcgaaggcttctcggcgtctcgcagctctg gggcaccatctgccaacgacgtgggtgcgcgact ggafctcgaacaatcgcgaaggacagccgttga aaccaccccctacgatgtggcgalcctcggcgact acaacatcggcggcgacgcctggccctcgcgca tctctctggaagagatgggctacgggtagtcgc gcagtggtccggcgacggcagcgtgggtggagat ggagaatcccccattcgtcaagctgaaccctggctc actgctaccgttcgatgaactatcgcgccafat tggaggagaaaatcagatctcgtggatggagta caactctcggccgacaaaatcgcggaatcgc ctgcgcaaaatcgcgaccagctcgcgatacca tctgcgcgaacgcgcaagcgggtgatcgcgggt atgaggggcagatgctggcgatctcgcgaataa tgcgccgcgcctggaggggcgtagggtgctgctc latatcggaggccctcggccgcgccacgtattgg cgcctatgaggatctcgggatggagatcgcgc gccggctacgagtttgcataacgatgaltacga ccgcaacctgcgggatcgaagaggggcacgct gctgctgatgacgccagcagctacgagctggaa gcgttcgtcaaggcgcctgaagcccgaacctatcg gctccggcatcaaggaaaaatatactccagaaa atggggcgtccgctcggccagatgcactcgtggg actatctcggccglatccagggctacgatggttc gccatttcggccgcatatggatagaccctgaa caaccggcgtggaacgaactgaccgctccgtg gctgaagctcgcgtga	atgatgactaatgctactggcgcaacgtaacctgg cactgattcaagaagtactggaggttcccgga aaccgcgcgcaagagcgcgtaaacacatg atggttctgaccgaaaatgaaatctgtgggta aatgcatcactcttaatcgaatctcagccggg tgcatactgctcgtggctgctgctacgcaggtt ctaaagggtcgtattcggcccgatcaaaagatf ggcgcatatctctatggcccggcaggctgtgg ccagtactctcggcggcaacgtcgtactacta cacggcgtctctcggctgactcttccggcacg ctgaactcaccctcgtactccaggaaactgaca tctgttctgggtggcgataaaaagctcgtc aaaaact gatcgaagaatggaactgctgttccgcctgact aaaggcattactatcaaaagcgaatctcgggtg ggtctgatcgggtgatgacatcagcgcggctgca aacgcactctcaaaagccctggataagccgggtg atcccgggtctgtcgcgagggctcgcggcgtt ctacgtctctgggtcactacatcgaacaacgatgt gtgcgtgactggattcgaacaaccgtgaaggt cagccttttgaaccaccccctatgactgtggat taltggcgactataacatcggcggcgacgcctg ggcatcccgcactcctcgtggaggagatgggtct gcgtgtgtcgcacagtggtcggcgatggcac cctgtgtgaaatggaaaacaccccgttltlaaa ctgaaccctggctactgctacgcctccatgaact acattgcccgctacatggagaaaaatcaga tcccctggatggaatacaactctcggctcgcact aaaaatcgcgaatccctgcgtaaaatcgcggat cagttgatgataccatcgcgcgaacgctgaag cagtaattgcgcgctacgaaaggccagatggca gcaalcatctcgaagtacgctcgcgcctggaa ggtcgtlaaagctcctcgtacatgggtgtctgc gtcacgctcaltgctgctggctcagggagacc tggcatggagatcctgcagcgggtlacgaat ttgcacacaacgacgactatgatctacgctgc cagacctgaagaaggctacgctgctgtttgacg acgccagctctatgaactggaaagcctctgtaaa agcgtcgaaccagacctgacggctcggca tcaaggaaaaatacatttccagaaaatggcggt gccgttccgcaagatgcactctgggactactc cggctcgtacacggctacgacgggttctcgtatc ttcgctcgtgacatggatgacccfgaataacc cagcgtggaatgaactgaccgacccgtggctg aaatctgcaaaa	<p align="center">82.6</p>	<p align="center">51.2</p>

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
K	WT Gene K is SEQ ID NO:7 Synthetic Gene K is SEQ ID NO:8	atgagcctaaacgattgataaaaitaatagctgttat ccgctatfcgaacaggatgaataaccaggagctgtt ccgcaataaagcggcagctggaagagggcgcacg atgcgcagcgcgtgcaggaggctttgctggac caccaccgccgagtgaagcgtgaatfcca cgcgaggcgtgaccgttgaccggcgaaagcc tggcagccgctggcgccgtgtttgctcgtggg atttgccaacacctgcccgtatgtgcacggctca ggggtgctgtgacctacttgcacctatftaacg ccatttcaaagagccgatcgcctgctgctccgact cgatgaccgaagacggcggtcttccggcgca acaacaataigaaccgggctgcagaaccag cgcgctgtacaaccgggagatcattgggtgctc accacctgcatggcggaagtatcggcgatgacc tgcaggcgtttatcccaacgctaaaaaagatggc ttcgctgacagcagcctcgcggcggcggcggc atacgccaagctttatcggcagccaccgaccgg ctgggataacatggttgaaggcttcgccaacact cactgcccactaccagggcgagccgggcaait gccgaagctcaatcgtgtgaccggcttgaacct atctcggcaacttcgctgtatfaagcggatgatg gaacagatggcggtgcccgtgcaccctgctcgg atcgcctgggaagctcgcacaccccgcggacgg tcaclatcggatgtatccggcgccaccacgcagc aggagatgaaagagggcccctgacgccatcgata cgcctgctcgcagccgtggcagctgctgaagag caaaaaagtgtgtcaggagatgtggaaccagcc cgccaccgaggctgcctatccgctggggctggc cgcaccagatgaactgctgatgaccgtcagccag cttagcggcaagccgallgcccagccctcacc ttgagcggcggcggcgtggtgacatgatgctga ctcccacacctggctgcaccgcaagaagtggc ctgtacggcgatccggacttctgtatgggctcac ccgcttctgctggagctgggctgcccagcaacg gtgatcctgagccataacgccaacaacgctggc aaaaagcgtatgaacaaatgctcgtatgctcgc gtacgggcgcgatacgaagtgttataactgctg atttgtggcacttccgtcgtgatgttaccctca gccggactttatgacggcaactctacggcaagt ttatccagcgcgataccctggcgaagggtlaaagc ctftgaagtgcgcctatccgctcggcttccgct gttcagccaccatctgaccggcagacaacc tggggtatgaaggggcgtatgaacatgtgacga cgtctgtgaacgccgtcgtggaactggatag cgataccagccagctgggcaaacgattacagc ttcgatctcgtcgttaa	atgtctcaaacatcagataaaatcaactctgttac ccgctgttcgagcaggacgaatafcaggaaactg tccgtaacaaacgctcagctggaagaagcgcac gaogcacagcgcgtgcaggaaagtgtcgcacg gaccaccaccgggaatacgaagctctgaact ccgtcgcgaagccctgacgggtgatccggcga aagcgtgccagccctctgggtgcccgtctgtgca gccctgggtttgccaacacctgcccgtatgtcca cggttccaggcgtgctgtagcctacttccgctacc taiftcaaccgccactftaaagaaccaatcgcgt gcgtgtccgacagcatgacggaggacggcggc agtttccggtggttaacaacaacatgaacctgggc ctgcaaaatgcttccgcactgtacaaccggaa atcaltcgcagctctaccacctgcatggcagag gttattgggtgatgctgcagcatttatggcaac gcaaaagaaagacgggttctgtgacagctctatc cgggtccgcagcctcatalaccctcctcctcgg ttctcagctaacgtgtgggacaacatgttcgaa ggcttcgcaaaaactttaccgcagactatcaag gccaacgggttaacgtccgaagctgaacctg gtgaccggcttgaacctacctgggcaacttcc gtgtcttgaagcgcctatgtagcagatggcg gttccgtgttctctgctgtctgaccctgtagggt ctggacactccagcggagggcactatcgcac gtattctgggtggcaacctcagcaggaaatgaa agaggcccagacgcgattgacacctgtctgc tgcacaacctggcagctgctgaaaagcaagaaa gttgttcaggaaatgtggaaccagccggcaacg gaagtgtcaatcccgctgggtctggcagctact gacgaactgctgatgaccgtgtcccaactgagc ggcaaaaccaatcgcggatgctctgaccctgga acgcccgtcgcctgggtgacatgatgctggaca gccacacgtggctgcatggcaagaaatftggc tgtacgggtgaccggacttctglaalgggctga cccgttctcgttggaaactgggctgcccagcga ctgtatctgtctcaaacgctaaacaacgttgg cagaaggccatgaacaaatgctggatgcgag cccatagggcgtgatagcgaagtgttcaaa ctgagacctgtggcatttccgctctctgatgttac gcgtcagccggatctcagatcggtaactcttac ggcaaaftcaltccagcgtgacactctggcctaaa ggcaaaagcgttgaagtgcgctgattctctgg gcttccgctgttcgaccgtcaccacctgcaccg ccagaccacctggggttacgaaggcgcgcatga acatcgtactactctgtgtaaacgcagctctgga aaagctggacagcagatctccagctgggca aaaccgactatcttctgatctggctcgttaa	82.1	51.2

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
Y	WT Gene Y is SEQ ID NO:9 Synthetic Gene Y is SEQ ID NO:10	atgtccgacaacgataccctattctggcgtatgctg gcgctgttcaagtctctgccggacctacagccggc gcaaatcgtgactggctggcgcaggagagcgg cgagacgctgacgccagagcgtctggcgacct gaccagccgcagctggccgccagcttccctcc gcgacggcggatgatgccccgcctgctggtcgc gggtgatggcgagcctgcaggcgcgctgcccg ccatttacgcacgttgcctgccagcgcacg ccgacgctgctggcggcatttgcctccaggatgg gctggtgattaacggccatttcggccaggagacga ctgttttatctacgcttccgatgaacaaggcggct ggtgtacgatctgcgcctatccctccgcccc caccagcaggaggccaacgaagtgcgcgcccg gcttatggagactgtcagctgctgtttgcccagga gataggcgggcccgcgcgcgcggctgatcc gccatcgcacccgatgaaagcgcagcccg ggacgacgattcaggcacagtgcgaggcgalca atacgtctctggcggcgttgcgcgctggct ggcgaagcggcctaacagggataaccctctggaa gaacgcgtttttaa	ATGTCTGACAATGATACCCTG TTTTGGCGCATGCTGGCGCTG TTTCAGTCGCTGCCGATTG CAGCCGGCTCAAATCGTCGA TTGGCTGGCGCAGGAATCCG GCGAAACCCTGACGCCGGAG CGCCTTGCCACCCTGACCCAA CCGCAACTCGCGGCGTCGTTT CCATCCGCGACGGCAGTGAT GAGCCCGGCTCGCTGGAGCC GCGTTATGGCTTCTCTGCAAG GCGCCCTCCAGCCCACTTGC GCATCGTACGTCCGGCGCAG CGTACCCCGCAACTGCTCGCC GCGTTTTGCAGCCAAGACGG CCTTGTTATCAATGGTCATTT CGGCCAGGGTCGTCTGTTCTT CATTTACGCCTTTGACGAGCA GGGCGGCTGGCTGTATGACTT GCGCCGCTATCCGAGCGCAC CGCACCAGCAGGAAGCGAAT GAGGTGCGTGCTCGTCTGATT GAAGATTGCCAGCTGCTGTTT TGCCAGGAGATTGGCGGTCC GGCAGCAGCGCTCCGATCC GCCACCGCATCCATCCGATG AAGGCGCAGCCGGTACTAC GATTCAGGCGCAGTGTGAAG CTATCAACACCCTGCTGGCCG GTCGCCTGCCGCGTGGCTCG CCAAACGTTTGAACCGTGAT AACCCGCTGGAAGAGCGTGT GTTTTAA	79.8	45.2

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
E	<p>WT Gene E is SEQ ID NO:11</p> <p>Synthetic Gene E is SEQ ID NO:12</p>	<p>atgaagggaatgaaattctggcgtgctggtgatga accggcctgtgaacacaaccataaacaataatcc ggctgcagcgcgcccacccggcgccaccgcg cgcgggctgcgcgttcgacggcgcgagataac cctgctgcccacgcccggcggcggcgcgcatctggc cacggcccacggcgtgcggcgggaagctcatgg gataaccggcgagcggcagctccggccccacc cttaateggctcgggtcaccaccgatctcaacga acaggacgtgattatgggcccggcggaacggcg actgfttcacggcgtgcccacatctgaccggccta tcaatcggcgggcggtcttatacaaacacctgct accggccatggagggggatgaccgggaagcgggt atgccaggccgcgagaccgccaccggcgtaacc ggttatgctatggacggcggcgggtttctacggca gtaaaaatctcggtaaccggcggcgggcgagc tcattggcacaacgggtcaccggcggcgagcgc cgccccctggcgggagagcagctcttggcccgg gagcagcgtcacgataltggcctgattggcgaatt caatattgcccggcgagttctggcatalcagccgct gctcagcagaactggggatccgcgtgctcggcag cctctcgggtgatggcggcttgcggagatccag accatgcaaccggcgagcccaatalgctggtct gctcggggcgttaaltaacgtgccagagccct ggagcagcgtacggcagcggcgtggtggaagg cagctttacgggaccgcgccacctctgagccc tggccagctggcggcgctgctggggcagcagc acctcggcagcgcaccgaagcgtgatggcgg ggaggaaacaggcggcggaacggcgctacagc cgtggcgcaacagctgcccggcggcgaagcgt ctgctctataccggcgggtgaaatctggtcgggt ggtaicggcgctcaggaattggcctgaccgtg gtggcaaccggcagcggtaaatccaccgaagag gataaacagcggatccgcgagctgatggcgaa gaggcggfaatgctgggaagaggcgaacggcgg cacgctgctggatggtctatcgtatcaggccg acctgatgattggcggcgagcgaatatgtacacc gectataaaagcaggctgcccgttctcgtatcaat caggagcgcgaacagccttgcgtggctatcagg ggatcgtcaccctgcccggcagctgtgtcagac catcaacagccccatctggcggcaaacattctc cgccccgtggcgctaa</p>	<p>ATGAAGGGTAACGAGATTCT TGCTCTGCTGGACGAACCGG CCTGTGAACACAACCATAAA CAGAAATCCGGCTGTAGCGC CCCAAAGCCGGGTGCGACGG CGGCTGGCTGCGCTTTCGATG GTGCGCAGATCACCTGCTCC CGATTGCGGACGTTGCCACCC TCGTGCATGGCCCAATCGGTT GCGCAGGTAGCTCTTGGGAC AACCGTGGCAGCGCCTCCAG CGGTCCGACCCTGAATCGTTT GGGCTTTACCACTGACTTGAA TGAACAAGATGTGATCATGG GTCGCGCGGAGCGTCGCTG TTCCACGCTGTGCGCCATATF GTCACCCGTTACCACCCAGCG GCAGTATTCATCTACAATAACG TGCGTGCCGGCTATGGAAGG CGATGACCTGGAGGCCGTGT GTCAGGCAGCCCAGACTGCG ACCGGCGTCCCGGTAATCGC AATTGATGCGGCTGGCTTCTA CGGTTTGAAGAACCTGGGCA ACCGTCCGGCAGGCGATGTC ATGGTTAAACGCGTCATTGGC CAACGTGAGCCAGCGCCGTG GCCGGAGAGCACCTGTTTG CCCCGGAGCAACGTCATGAC ATTGGCTTGATCGGTGAGTTC AACATTGCGGGCGAGTTTTG GCACATTCAGCCGCTGCTTGA TGAGCTGGGTATCCGCGTTTT GGGTTTCGCTCAGCGGCGATG GTCGTTTTGCGGAGATTCAA CCATGCACCGTGCCCAGGCG AACATGCTGGTGTGCACCG TGCTCTGATCAATGTTGCGCG TGCTCTGGAACAGCGCTATG GCACCCCGTGGTTTTGAAGGCT CGTTCTATGGTATCCGCGCGA CCAGCGACGCCCTGCGCCAG TTAGCGGGCGCTGCTGGGCGA TGACGACCTCCGTCAGCGCA CCGAGGCGCTGATCGCGCGT GAAGAACAGGCGGCTGAGCT GGCCCTGCAACCGTGGCGTG AACAGCTGCGTGGCCGCAAG GCCCTGCTCTACACGGGTGGT GTCAAAGCTGGTCTGTGGT GTCCGCGCTTACAGGATCTGGG TATGACCGTGGTTGCCACGG GCACGCGTAAGAGCACGGAA GAGGATAAACAGCGCATCCG</p>	<p>80.8</p>	<p>47.6</p>

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			CGAATTGATGGGCGAAGAGG CCGTGATGCTTGAAGAAGGC AACGCACGTACCTTATTGGAT GTAGTTTATCGCTATCAAGCA GACCTGATGATTGCCGGTGG CCGCAACATGTATACCGCCTA CAAAGCGCGCTTGCCGTTCT GGACATCAACCAGGAACGCG AGCACGCGTTTGCGGGCTAC CAAGGCATCGTGACCTTAGC GCGCCAGCTGTGCCAAACGA TTAACAGCCCAGATCTGGCCGC AGACTCATTCCCAGCGCACCGT GGCGCTAA		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
N	WT Gene N is SEQ ID NO:13 Synthetic Gene N is SEQ ID NO:14	atggcagacattttccgaccgataagccgctggc ggtcagcccccataaaaaccggccagccgctcgg cgcaatcctcgcagcctcgggatogaacacagc atccctcgtggtccacggcgcgcaggggtgcagc gccttcgccaaagtctttttatcaacatttccacga ccgggtccctcgcagtcgacggcgatggaccacc acgtcgcagattatgggcgcggacggcaatattt taccgcccctggataccctctgcagcgcacaacac cgcaggetatcgtactgctcagcacegggctgct ggaggcccagggcagcgatatttcccgctgggt cggcagtttcgcgaagagtatcccggcatalaagg ggggtggcgalattgacgggtaaacgcggatittt atggctccatggagaacggctcagcgcgggtgta gagagcgtcattgagcagtggtgcccggcggcg ccggcggcggctcagcgcacacgcggggtcaat ctgctggctcagccatctctgctcggcgatatac gagtggctgcgccgatgcgtcgaagccttggctc gcagccgataatcctcgcggaccctggcgcaatcg atggacggccacctggcgcagggcgatttctcgc cgcctgaccagggcgggacgcggctgcggcag atagagcagatgggcaaacctgtgcagcttcg ceattggcgtctccctcctcgcggcctcctcgc tggccccgcctcgcggcggcgggtatcgcct gccgcacctgatgacctcgaacgctgcgacgc ctttatcacaacggcgaatctccggacgcgc cgttccgagtggtggaacgccagcggcggcca gctacaggatgcgatgacgactgccafatgtggc tccaggggcagcgcgatggcgalagcggcggaa ggcgatttgcctggcggcgtggtgtgatttcgcaa cagccaggggatgcagcccggccccttgggtg cccctacgggtcaccagcctgcggcagctgcc ggtggaaacgggtgggtgcccggggatctggagga tctgcaaacctcgtgtgcggcgcateccgcggac ctgctgggtggcgaactcgcacgcccgcgacctg gcggagcagtttgcgctgccgctggtgcgcgcg gggttccgctctllgacaagctcggcgaaitccgc cgggtgcgacaggggtatagcgggatgcgcgat acgctgtttgagctggcaaacctgatacgcgagc gtcaccaccacctcggccactaccgatcgcgct gcgccaagaacccgaatcgtcactctccacagga ggcgcttatgccgcatgaa	ATGGCAGACATTTTCCGCACT GATAAGCCGTTGGCTGTGTCTG CCGATCAAGACCCGGCCAGCC GCTGGGTGCGATCCTGGCGTC CCTGGGTATCGAGCACTCGAT TCCGCTGGTACATGGCGCGC AGGGCTGTTCCGGCTTTTGCCA AGTTTTTCTTTATCCAGCACT TCCACGATCCGGTCCCCTGTC AAAGCACGGCAATGGACCCG ACCAGCACCATCATGGGCGC TGATGGTAACATCTTCACCGC GCTGGACACTCTCTGCCAACG CAATAACCCGCAAGCAATTG TGCTGCTGAGCACCCGGCTCT CCGAGGCGCAGGGCAGCGAC ATTTCCCGTGTAGTGCGTCAG TTCCGTGAAGAATATCCGCGT CATAAAGGCGTGGCGATTCT GACTGTTAACACCCCGGACTT TTACGGTAGCATGGAGAACG GCTTTTCCGCTGTCTGGAGT CTGTGATTGAACAGTGGGTTT CGCCAGCCCCACGTCCGCGC CAGCGCAATCGTCGCGTCAA TCTTTTGGTGAGCCATCTCTG TAGCCCAGGCGATATTGAGT GGCTGCGCCGTTGCGTCGAG GCCTTCGGTCTGCAACCGATC ATTCTGCCGGATCTGGCTCAG AGCATGGACGGCCACCTTGC TCAGGGTGACTTTTCGCCGCT GACGCAGGGCGGCACGCGCT TGCGCCAAATCGAGCAGATG GGCCAGAGCCTTTGCTCTTTT GCGATTGGCGTCAGCCTGCA CCGTGCGAGCAGCCTGTGCG CTCCGCTTGTGCTGGCCGAAG TCATCGCCTTGCCGCACCTCA TGACCTTGGAACGCTGCGAC GCCTTTATCCATCAGTTGGCG AAAATCAGCGGTTCGCGCCGT TCCGGAGTGGCTGGAACGCC AGCGCGGTCAGCTGCAAGAC GCCATGATCGATTGCCACATG TGGCTGCAAGGCCAGCGCAT GCGGATTGCCGCCGAAGGCG ACCTGCTGGCAGCGTGGTGC GATTTCCGGAACCTCTCAAGGT ATGCAGCCGGGTCCACTGGTT GCTCCGACGGGTCTATCCGAG CCTGCGTCAGTTGCCGGTGG GCGCGTGGTGGCGGGTGATC TGGAGGATCTTCAGACCTCT	79.7	47

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			TATGCGCACATCCGGCCGACT TACTGGTGGCGAACTCCCAC GCCCGTGATTTAGCAGAGCA ATTCGCCCTGCCGCTGGTGCG CGCAGGCTTCCCGCTGTTTGA CAAACCTGGGCGAATTCGTC GTGTTTCGCCAGGGTTATAGCG GTATGCGTGATAACCCTGTTCG AGTTGGCGAACCTGATCCGT GAACGCCATCATCATCTGGCT CATTATCGCAGCCCGCTGCGC CAGAACCCAGAATCCTCGTT GTCTACGGGTGGCGCGTACG CAGCGGATTAA		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
U	WT Gene U is SEQ ID NO:15 Synthetic Gene U is SEQ ID NO:16	atgtggaattactccgagaaagtgaagaccatttt ttaaacccecgcaatgcgcgctggggacaacg ccaacgcggtaggcgcgctcgggtcgttaagctg cggcgcacgccctgcgctgatgctgcgcgctcga cccgcaaaagcgaatcattgaggaggcgggcttc cagaccttcggctcggcagcgcaccatcgectct ctcccgctgacggagctgattatcgccatacc ctcgccgaagccgggcagataaccaatcagcag attgccgattatctcgacggactcggccgggaga aaatgcactgctcgggtgatggccaggaggccct gcgcggcggccatcgccaacttcggggcgaag ccttgaagaggagcagcagcaggcgaagctgat ctgcaaatgctcggcgtcgtgatgaaggccatattc gccgcgcgggtacagaacaacgggctgaccacce ttccgagggtgatcaactacaccaaaagcggcgg cggctgcacctcttccacgaaaaaatcgagctg ggcctggcggagatctcggccagcagccgcag acgacgccagccgtggccagcggcaagatcc gcactggcagagcgtcgtcgtataccatcgagaa ctgcggccgcatattcaggccgacggcggcgat atggcgctactcagcgtaccaaccaccagggtga cgtcagcctctccggcagctgtagcggctgat gatgaccgalatgacctggcctggcgcagcga aaactgatggaacgtaccggctgltatggaagt ggtggcggcctga	ATGTGGAACTACAGCGAGAA AGTCAAGGACCATTTCTTCAA TCCGCGCAACGCGCGTGTGT GGATAACGCAAATGCGGTGG GCGACGTCGGCAGCTTATCTT GTGGCGATGCTCTCCGCTTGA TGCTGCGCGTGGACCCGCAG AGCGAAATCATCGAAGAAGC GGGCTTTCAGACCTTCGGCTG CGGCAGCGCGATTGCGTCGT CCAGCGCACTGACGGAGCTG ATCATCGGTACACCCTGGCG GAAGCGGGTCAGATCACCAA CCAGCAGATCGCCGACTACT GGACGGCTTACCGCCGAAAA AGATGCACTGCTCTGTAATGG GCCAGGAAGCTCTTCGTGCG GCCATTGCTAACTTTCGCGGT GAATCGCTGGAAGAGGAGCA TGACGAGGGTAAGCTGATCT GCAAGTGCTTCGGCGTCGAT GAAGGCCATATTCGCCGTGCT GTCCAGAACAACGGTCTTAC GACTCTGGCCGAGGTGATCA ATTACACCAAGGCAGGTGGC GGTTGTACCAGCTGCCATGA GAAAATCGAGCTGGCCCTGG CCGAGATTCTCGCCAACAG CCGCAAACCACCCCGGCAGT TGCGTCCGGTAAAGATCCGC ACTGGCAGAGCGTTCGTGGAT ACCATCGCTGAACTGCGTCCA CATATCCAAGCGGACGGTGG TGACATGGCGCTGTTGTCCGT GACGAACCACCAAGTGACTG TTTCGCTGTCCGGCAGCTGTT CTGGCTGCATGATGACCGAC ATGACCCTGGCGTGGCTGCA ACAGAAATTGATGGAGCGTA CCGGCTGCTATATGGAAGTTG TTGCCGCCTAA	82.3	53.5

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
S	WT Gene S is SEQ ID NO:17 Synthetic Gene S is SEQ ID NO:18	atgaaacagggttatctcgataacaacgccaccac ccgtctggaccggatggctctggaagcgatgatg cccttttgaccgattttacggcaaccctctgctga tacacgattttggcattccggcccaggcgctctg gaacgcgcgcatcagcaggctgcggcgctgctg ggcgcggagfatccagcagagatcattttacctc ctgcgccaccgaagccaccgccaccgccatcgc ctcggcgatcgccctgctgctgagcgtcgcgaa atcaccaccagcgtgctgaacatccggcgaccg tggcgccctcgcgagcacatggagcgcgagggct accggatcaccgcatcgcggtagatggcgaggg ggcgctggacatggcgagttccgcggcgct cagcccgcgcgtcgcgtggcagcgtgatggg gcgaaataacgaaaccgggggtgtttcccgatcg gcgaaatggcggagctggccatgaacaaggg ggcgtgttcaactcgcagtcgggtcaggtggcgg gaaaataccgacgccgtgggcccagaccggcat cgatatgctcctcctgcgcgcataagttccacg ggcctaaaaggcgtaggctgtttatctcggcg gggaacgcgcttcccccgtcctgcggcggt caccaggagtagcggcggcgagccgggacaga aaatatctcgggaatcgtcggcatggcgccg tgcgagctggcgaatcctcgcgggaatgac gcatacggccaattcgcgaacaggctggagcat cgctcctggccagcgtccctcggcatgggga tggcgccggccagccggcggtgccggcacg gtgaatcggcctttagtttattgaaggtgaagcc attctgctcgtgtaaacaggccgggaltcggcgc ctcagcggcagcgcctgcacctcaggctcgcctg gaacctcccacgtgatggggcgatgaatattcc cttacaccgcgcccacggccaccatccgctttct ctctcgcgtacaccgggagaaaagagatcgalt acgtcgtcggcacgctcggcgccgattatcgaccg gctcgcggcgctgctgcctactggcagaacgg caagccgcgcccggcgagcccgattacgccc ggtttacggctaa	ATGAAACAAGTGTACCTGGA CAACAACGCGACCACCCGCC TGGACCCGATGGTTCTGGAA GCGATGATGCCGTTTCTCACG GATTTCTATGGCAATCCGTCC AGCATCCATGACTTCGGC ATC CCGGCACAAAGCGGCGCTGGA ACGTGCGCACCAGCAAGCTG CGGCACTGCTGGGCGCAGAG TACCCGCTGTAATCATTTC ACGAGCTGTGCGACCGAGGC CACTGCAACCGCCATTGCGTC GGCCATCGCGTTATTGCCGA ACGCCGCGAAATCATCACCT CGGTAGTGGAGCACCCGGCT ACGCTGGCGGCGTGGCAGCA CATGGAACGCGAAGGCTATC GCATCCATCGCATTGCGGTGG ATGGCGAAGGTGCGCTGGAC ATGGCCCAGTTCGCTGCAGC GCTCTcgCCGCGTGTGCGGTTG GTGAGCGTGATGTGGGCCAA CAACGAAACCGGCGTGCTGT TCCCGATTGGCGAAATGGCC GAGCTTGCCACGAGCAGGG CGCTCTGTTCCACTGCGATGC CGFTCAGGTGCTTGGCAAAT CCCAATTGCTGTTGGCCAGAC GCGCATCGACATGCTGTCTTG CTCCGCGCACAAAGTTTCATGG TCCGAAGGGTGTGGTTGCTT GTACTTACGTGCTGGCACGCG CTTTCGTTCCGCTGCTTCGCGG TGGCCATCAAGAATATGGTC GCCGTGCCGGCACTGAGAAT ATCTGTGGCATCGTCGGCATG GGCGCTGCGTGGCAACTGGC GAACATCCATCTGCCGGGTAT GACCCATATTGGCCAGTTACG CAATCGCCTGGAGCACCGTCT GCTCGCCAGCGTGCCGTCCGT GATGGTTATGGGCGGTGGTC AGCCGGCTGTACCGGGTACT GTCAACCTGGCGTTCGAGTTT ATCGAAGGTGAAGCGATCCT GCTCTTGCTGAACCAGGCTGG CATTGCCGCAAGCTCCGGCTC CGCGTGTACCTCTGGCAGCTT GGAGCCGAGCCATGTGATGC GCGCCATGAACATTCCATAC ACCGCGGCTCACGGCACCAT TCGTTTTAGCCTGAGCCGTTA TACGCGCGAGAAAGAGATCG ACTACGTGCTTGGCACCCTCC	79.1	43.6

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			CGCCAATCATTGATCGTCTGC GTGCCTTGTCCTCCCGTATTGGC AGAATGGTAAGCCGCGTCCG GCAGATGCAGTCTTTACCCCG GTTTACGGTTAA		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
V	<p>WT Gene V is SEQ ID NO:19</p> <p>Synthetic Gene V is SEQ ID NO:20</p>	<p>atggaacgcgtgctgattaacgataccacctgcg cgacggcgagcagagccccggcgctgcctttcg caccagcgaaaaggctgccattgccgagggcgtt tacggcgcaggaataacggcgatggaggctggc accccggcgatgggcgacgaggagatcgcggc gatccagctggtgcgtcggcagctgccgacgc gacctgatgacctggtgtcggatgaacgcgctg gagatccggcagagcggcgatcgggcatcgact gggtggatactcgattccggctcggataagctg cggcagfacaaactgcgcgagccgctggcggtg ctgctggagcggctggcgatgftatccatcttgcg cataccctcggcctgaaggtatgcatcggctgcg aggacgctcgcggccagcggccagacctg cgcgclatcgccgaggtcgcgcagcaatgcgcc ggcggccgctcgcgctatgcgatacggcggcc tgcctgacctttaccaccggcgcaaatctcg ggcctgcgcgacgtctggctcggcgaaatcga tgcatagccataaacgatctgggtatggcgaccg caatacgtggcgcgtaagcgcggggccac cagcgtgaatacagcggctcctcggctcggcgag cggcgggcaacgcggcgtggaaaccgtcgc gctggccttgaacgctgcttggcgtggagacc ggcgtgcallttcggcgtgcccgcgtcctgtca gagggtcgcggaaagccgcgcagcggccatcg accgcagcagcgcgtggctggcgagctgggtgt tacctatgagtcaggctgacacgtggcggcgtg ctcggcgcagcgcgagctaccagtcacatcgcc ccttccctgatggcgcagctaccggctgggtgt gggcaaacacccggcgctcaggcggtaacgg cgtttttgaccagatgggctatcaactcaacgcc cgcagattaaccagctgctgcccgcacatcgccg cttcggcgagaactggaagcgcagccgaaaga ttacgagctgggtggctatctacgacgagctgtgcg gtgaatccgctcgcggggcaggggggtaa</p>	<p>ATGGAGCGCGTCTTGATCAA CGATACTACCCTGCGTGATGG CGAACAATCTCCGGGCGTAG CGTTTCGTACCTCCGAGAAAG TTGCCATCGCGGAGGCACTGT ACGCTGCGGGTATCACCGCG ATGGAAGTCGGCACTCCGGC GATGGGTGATGAAGAGATCG CCCCGATTACGCTGGTGCCTC GTCAACTGCCGGACGCGACG CTTATGACCTGGTGCCGTATG AACGCTCTGAAATCCGTCA GAGCGCGGATCTGGGTATTG ACTGGGTGGATATCTCGATCC CAGCATCCGACAAGCTGCGT CAGTACAAGCTGCGTGAGCC GCTGGCCGTGCTGCTGGAGC GCCTTGCGATGTTTATCCATC TGGCCACACGTTAGGCCCTCA AAGTATGTATTGGTTGCGAG GATGCGAGCCGTGCGTCTGG TCAGACCCTGCGCGCCATTGC CGAGGTGGCCAGCAATGCG CGGCTGCGCGCTTGCGTTACG CTGACACCGTGGGCCTGCTG GACCCGTTACCACCGCAGC CCAGATCAGCGCCCTGCGTG ACGTTTGGTCCGGCGAGATC GAGATGCATGCTACAATGA TCTGGGCATGGCTACCGCGA ACACGCTGGCGGCAGTTTCG GCTGGCGCCACGTGGTGAA CACTACCGTCTCGGTCTGGG TGAACGTGCAGGCAACGCAG CCCTGGAACCCTTGCCTG GGCTTGAACGCTGCTGGG CGTGGAAACCGGCTCCATTT CAGCGCGCTCCAGCGAGCT GTCAGCGCGTCCGCGAGGCT GCACAGCGCGCAATCGACCC GCAACAGCCGCTGGTGGGTG AATTGGTTTTTACCCACGAGT CTGGTGTTCACGTGCGGCGC TGCTGCGCCACAGCGAATCCT ATCAATCTATTGCCCAAGCC TCATGGGCGTAGCTACCGTC TGGTGTCTCGGCAAGCATTCG GGTCGTCAGGCTGTCAACGG TGTTTTCGACCAGATGGGTTA CCACCTGAATGCGGCGCAGA TCAATCAGTTGCTGCCGGCCA TTCGCCGCTTCGCCGAGAATT GGAAACGCTCTCCGAAAGAC TACGAACTGGTTGCGATCTAT</p>	80.8	48.3

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			GACGAATTGTGCGGTGAATC CGCCCTTCGTGCTCGCGGCTA A		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
W	WT Gene W is SEQ ID NO:21 Synthetic Gene W is SEQ ID NO:22	atggagtggittatcaaattccggcgctggacga actcgcctcgcgcaatctttttcagttttcgcct cccctatcagcccagctgcttggccgctgcagc ctgccggctgctggcaacgtttcatcgcaaacctcg cgcggaggctcccgctgcaaaaccggctcgagga taacgaccgcgcgccctggctgctggcgcgaag actgctcgcggagagctatcagcaacagttcag gagagcggaaatga	ATGGAGTGGTTTTACCAGATT CCGGGTGTAGACGAATTGCG CAGCGCTGAATCCTTCTTTCA GTTCTTCGCGGTTCCATACCA GCCGGAAGTGTGGCCCGCT GCTCGCTTCCGGTGTAGCGA CGTTCACCGTAAACTGCGTG CGGAGGTCCCCTGCAAAAC CGTCTGGAGGACAATGATCG TGCGCCGTGGCTCTTGGCGCG CCGCCTCCTGGCCGAATCTTA TCAGCAGCAATTCAGGAGA GCGGCACCTAA	77.9	43

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
Z	WT Gene Z is SEQ ID NO:23 Synthetic Gene Z is SEQ ID NO:24	atgagaccgaaatcacctttagegaagaggtcgcg cgtcgtacgcgcgattcgtaacgacggcaccggtg gcgggcttcgcgcccgccgcgcgtcgtggtcagg cgcggcagcaccggctttgtgcgcgactggggc gttttttgaagatcagattatctaccagatccactt tcggaaaccgatcggatcctcggctgcgcgcgag caggagctgatcccatcaccagccggtggctgg ccggaatttgcaatacaggatagcgtgacctg ccagatggcgctcgcggtcaacggcgatgtggtc gtgagcggccggcagcggggacgcgtgagget accgatcggggagagctcggcgacagctacacc gtcgaacttagcggcgcgtggtcagggtccgggt gcaggccatcgccttatagaggaaagagaaga atga	ATGCGCCCGAAATTCACCTTC TCTGAAGAGGTCCGCGTAGTT CGCGCGATTTCGTAATGATGG CACCGTGGCGGGTTTTGCGCC AGGTGCGCTGCTGGTTCGTCCG CGGTTCCGACGGGCTTTGTGCG TGA CTGGGGTGTGTTCTCTGCA AGACCAGATCATCTATCAAA TCCACTTTCGGAAACCGACC GCATTATCGGCTGTCGCGAGC AGGAGTTAATCCCGATTACCC AGCCGTGGTTGGCTGGTAAC CTCCAGTATCGTGACAGCGTC ACGTGCCAAATGGCACTGGC TGTCAACGGTGACGTGGTTGT GAGCGCCGGTCAACGTGGCC GTGTGGAGGCCACTGATCGT GGCGAACTTGGCGATTCTAC ACCGTGGACTTCAGCGGCCG TTGGTTCGCGTTCGCGTCCA GGCCATCGCGCTGATTGAAG AGCGCGAAGAATAA	80.1	48.3

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
M	WT Gene M is SEQ ID NO:25 Synthetic Gene M is SEQ ID NO:26	atgaaccctgggcaacgtttgcccgccagcggc tggcgcgcagccgctggaatcgcgacccggcgg ccctggatcggccgacacgcggctttgaaca ggcctggcaacgccagtgccatatggagcagac gatcgtcgcgcgggtccctgaaggcgataatccg gcgcggttgctggagaatacctgctccctcctgc catctggctcgcaggggggattitgcgcgccc gagcgcgctgccaicgctgcgccaacacgcggg ctggaactcgccttcgccaatcgcggccagg cgccgcagccggaatctcaccggtaacggcatg gtatctgcgccaccagacgcagttatgcggccg gaacagcgtcigaccgccatctactgctgacggg cgataacgaccgcgaagccgtgaccagcggat cctcggcctglatcggcaaatcaacgcctcggg gacgccttcgcgcgctggcccagcgcattccc actgcccagcgcgctgggaagagggtcgtttagg ctggallagcggctggcctgctctatcgcagctcg agaccgcgctgtttcactggcggaaaacgcgct aagccttccatcgcagcgaactgggctgcat ctttatggtgcgaagegattcgccegcgcggcc catggagccgcagcagggcctggagagcgcgc gagattatcttggcagcagaccagcagcgcca tcagcggcagtggtggaaacagatgattcgcg agccgggactgtcgggtag	ATGAATCCGTGGCAGCGCTTT GCCCCGTCAACGCCTTGCTCGC AGCCGCTGGAACCGTGATCC GGCTGCTCTCGACCCAGCCG ATACCCCAGCGTTCGAGCAG GCGTGGCAGCGTCAATGCCA TATGGAACAAACCATCGTAG CGCGTGTCCCGGAAGGCGAT ATTCCGGCTGCCTTACTGGAA AACATCGCGGCCAGCCTGGC GATCTGGCTGGACGAGGGTG ACTTCGCTCCGCCGAGCGC GCTGCGATTGTGCGTCATCAT GCACGTCTGGAGCTGGCGTTT GCCGACATTGCCCGCCAGGC ACCGCAACCGGATCTGAGCA CGGTTCAAGCGTGGTATCTGC GTCACCAGACTCAATTCATGC GTCCGGAGCAGCGTCTGACC CGTCACCTGCTCCTGACGGTC GATAATGATCGCGAGGCGGT GCATCAACGCATCCTTGGCCT GTATCGTCAGATCAACCGCA GCCGTGACGCCTTCGCCCCAC TGGCACAGCGCCACTCTCATT GCCCCGTCCGCTTGAAGAA GGCCGTCTGGGCTGGATCTCC CGTGGTCTGCTGTACCCGCAG CTCGAAACCGCGTTGTTTAGC CTGGCGGAAAACGCACTGTC GCTGCCGATTGCGTCGGAATT GGGTTGGCACCTGTTATGGTG CGAGGCCATTCGTCCGGCAG CCCCGATGGAGCCGCAACAG GCCCTTGAATCTGCGCGCGAC TACTTGTGGCAGCAGAGCCA GCAGCGCCACCAGCGTCAAT GGCTGGAGCAGATGATTTCC CGCCAACCGGGCCTGTGTGG TTAA	79.7	45.3

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
F	WT Gene F is SEQ ID NO:27 Synthetic Gene F is SEQ ID NO:28	atggcgaaacattggatattttttggcaccgataacc ggtaaaaccgcgaaaatcgcaaaatgatccata agcaactgggcgagctggcgatgccccgggtca acattaaccgcacgacgctggacgacttatggcc taccgggtgctgctgctgggcacgcccagctcg gcgacggccagctgcccgggctggaggccgga tgcgaaagcgagtcattggagcgaattatcagcg gctcgcagcagccagcctgaangggaaaacog tggcgtgttcggcctcggcgatcagagaggcta tccggacaacttcgacgcccggatgcccgcctg tfcgacgccctgagcgcgcggcgccgagatg attggcagctggccaaatgagggtatgaattcag cgcgtcctcggcgtcggagggcaccgcttgtt ggcgtggtgctggatcaggataaccagttcgacc agaccgaagcgcgctggcgagctggcttgagg agatfaaacgcaccgtgctgtag	ATGGCGAACATCGGCATCTTC TTTGGTACGGATACCGGCAA AACCCGCAAGATTGCGAAGA TGATTCACAAACAGCTGGGC GAGCTGGCCGATGCCCCGGT TAACATCAATCGTACCACTTT GGATGACTTTATGGCTTACCC AGTCCTGTTGCTCGGCACGCC GACGCTTGGTGATGGTCAACT GCCGGGCTTAGAGGCGGGCT GCGAGAGCGAAAGCTGGTCT GAGTTTATCTCCGGTCTGGAT GACGCTTCCCTGAAGGGCAA AACCGTGGCGCTGTTTGGCCT GGCGACCCAGCGTGGTTACC CGGACAACCTTCGTGTCGGGT ATGCGTCCGCTGTTTCGACCGG CTGAGCGCCCGTGGCGCCCA GATGATTGGTAGCTGGCCGA ACGAAGGTTATGAGTTTAGC GCATCGTCCGCGCTGGAAGG CGACCGCTTCGTGCGCTTGGT GCTGGATCAAGACAATCAGT TCGACCAGACCGAAGCGCGC CTGGCGTCTTGGCTTGAAGAG ATCAAACGCACCGTTCTGTAA	81.5	51.4

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
B	WT Gene B is SEQ ID NO:29 Synthetic Gene B is SEQ ID NO:30	atgacttcctgctcctctttctggcggcaaaagcct gcccggcggcgatgacagcgcattgacgcgcg ttgtggccgataaagctgcgcgcaccctgctac tctcgcctatgggcatcaccgttctgcgcggatgca tctgcccgtcgcgcgcctgcaattgacagtgca actactgtaafcgcaaatcgattgcagcaacgagt ccgcgccggggtatcgtcaacgctgctgacgcc tgaacaggcggctcgtgaaagtgcgacaggtcgcg caggcgatcccgcagctttcgggtgggcatcgc ccgggcccggcgatcgcctcgcgaatatogccc gcaccttgcaccctggagctgacccgcgaaca gctgcccggacctgaaatgatcctgctgaccaacg gactggtgctgctgacgcgggtggaccgctgct ggatgctggcgtgacacgctacggcaccatla acaccctcgcgcggagattgcgcgcgaatata cgctggctatggctggacggcgaacgctacag cgggcgcgaagcgggagagatcctgattgcccg tcagcttgaggcgtacgcagcgtgaccgcaaaa ggcgtgctgggtgaaaataaattcggctgctgaccc cggatcaacgatagcggcatggccggcgtgag ccgcgcctgcccggcgcgcgcggttataccat aataatgatccctgctgacccagcgcggagcacg gcaaggcttggcctcaacggccagccggagc cggacgcgcgagacgctcgcgcaccgcgcgc cggctgcggcgaagtgatccgcgagatgaccac tgcaccagctgcgcgcgacgccattgggatgc tcggcgaagaccgcgcgcagcagcttaeccagct tcggcgcagagagctcctccggcctggctgccc gatcctcaccagcgcgcgcagctgacgcgcag cattgagaccgcgcgcgaatcgaagccgatgac gcccctgctggctgcgcgtggcgtcaagccgcggg gacgtcattgattgctacttggctacgcgcaccgg ttctacattfacagcctcctcggccgcgcgtatggg ctggctcaacgagcgttllacgcccataatigtca ggggcgcgatgactgcgagccgcaggataacg cagcccgttggcggcgtcctcgaactgctggc ggacgttaaaagccgtatctgctgctgctatcggcc atacgcctggcgaacagctggaacaggaaggca tgaacctgctgctgacggcgcgtggcgcggcgt ctccgaagctgctgcccgcgtggctggcaacagcgt cgggggagctggcctgcccgttggccgataaag ggggtcgcctga	ATGACCTCTTGTTCGTCGTTTT TCTGGCGGTAAAGCGTGCCG TCCGGCCGATGACTCCGCGCT GACTCCGCTGGTGGCCGACA AGGCAGCTGCGCACCCCGTGC TATAGCCGCCACGGCCATCA CCGCTTCGCGCGTATGCACCT GCCAGTGCCTCCGGCCTGCA ACTTACAATGCAACTACTGCA ACCGCAAGTTCGATTGCAGC AATGAAAGCCGTCCGGGCGT GTCCTTACCCTGCTGACGCC GGAACAGGCTGTGGTGAAGG TGCGCCAGGTCGCCCAAGCT ATCCCGCAGCTGctgGTGGTCG GTATTGCTGGTCCGGGCGATC CGCTTGCGAATATCGCCCCGA CCTTCCGTACCTTGGAGCTTA TTCGCGAACAGTTGCCGGAC CTGAAACTGTGCCTGAGCAC CAACGGCTTGGTGTCTGCCAG ATGCCGTTGATCGTCTGTCTCG ATGTGGGCGTGGATCACGTT ACCGTCACCATTAACACCCTG GACGCAGAAATCGCAGCGCA AATCTACGCGTGGTTGTGGCT GGATGGCGAACGCTACTCCG GTCGCGAAGCCGGCGAAATT CTCATTGCCCCGCCAGCTGGAA GGCGTACGTCGCCTGACCCGC GAAAGGTGTGCTCGTCAAGA TCAACAGCGTATTGATTCCGG GCATCAATGACAGCGGCATG GCGGGTGTAGCCGTGCGCT GCCGCGCTCTGGTGCCTTCAT CCACAACATCATGCCACTGAT TCCGCGTCCGGAGCATGGCA CTGTTTTCGGTCTGAACGGCC AGCCGGAACCGGACGCGGAA ACCTTGGCGGCGACGCGCTC CCGCTGCGGCGAGGTTATGC CACAAATGACCCACTGCCAC CAGTGCCGTGCCGACGCGAT TGGCATGCTTGGTGAGGATC GCTCGCAACAGTTTACGCAAT TACCGGCTCCGGAGTCCCTCC CGGCCTGGCTGCCGATCCTGC ATCAGCGTGCTCAGTTGCATG CGAGCATCGCCACGCGCGGT GAGAGCGAAGCCGATGACGC CTGCCTGGTGGCCGTTGCGTC GAGCCGTGGCGATGTAATTG ACTGCCATTTCCGGCCATGCCG ACCGTTTCTATATCTATAGCC	80.2	45.4

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
			TGTCTGCGGCTGGTATGGTTC TGGTTAACGAACGTTTCACCC CGAAATACTGCCAGGGTCGC GATGACTGCGAGCCGCAGGA CAATGCCGCACGCTTTGCTGC CATCCTTGAGTTGCTGGCGGA CGTCAAAGCGGTGTTTTGTGT GCGTATCGGCCATACCCCGTG GCAACAGCTGGAGCAGGAAG GCATCGAACCGTGCGTGGAT GGCGCCTGGCGTCCGGTATCC GAGGTCCTGCCGGCATGGTG GCAGCAGCGCCGTGGTAGCT GGCCGGCTGCATTGCCGCAC AAAGGCGTTGCGTAA		

Gene	SEQ ID NO:	WT Sequence	Synthetic Sequence	% Nucleotide Identity	% Codon Identity
Q	WT Gene Q is SEQ ID NO:31 Synthetic Gene Q is SEQ ID NO:32	atgccgccctcgactgggtgcccgcctatggct gctgtaccacgcggggaaaggcagcttccgctg cgcatggggcctagcccgcgcgatggcaggcg ctcggcggcgccctgggcgaggtggaaacgcc gctcgaggcgagacgctcaccgctcggcgcct gatggcggagctcaacgccaccgcgaagagg agcggcagcagctgggcgcctggctggcgggt ggatgcagcaggatgccggccgatggcgcag attatgccgagggttcgctggcgttfaaccatctt ggcaggatcttggtctgcatcgccgcggaatt gcgctcctgatgagcactgcttccacagctgg tggatgaacgaacaatatcgctggaaaaa gttctttatcgtcagcgtgttctgcaacagggg gaagtatctccgctcgccaagctgcgacgagtg ctgggaacgcagegctgtttgagtag	ATGCCGCCATTGGACTGGTTG CGTCGTTTGTGGTFACTCTAT CACGCCGGCAAAGGCAGCTT TCCGCTTCGTATGGGCTTGTG GCCGCGTGACTGGCAAGCTC TGCGCCGTGCCTGGGCGAG GTGAAACGCCGCTGGATGG CGAAACCCTGACCCGTCGCC GTCTGATGGCGGAGCTGAAT GCGACCCGCGAAGAAGAACG CCAGCAGCTGGGTGCCTGGC TGGCCGGTTGGATGCAACAG GATGCCGGTCCGATGCCGCA GATTATCGCAGAGGTGAGCC TGGCGTTCAACCATCTCTGGC AGGACCTTGGCCTCGCGAGC CGCGCTGAACTGCGTCTGCTG ATGTCTGACTGCTTCCCGCAG CTGGTTGTTATGAACGAGCAC AACATGCGCTGGAAGAAATT CTTTTACCGCCAGCGTTGCCT GCTGCAACAGGGCGAAGTCA TCTGTGCGAGCCCGTCTTGCG ATGAATGCTGGGAACGTTCT GCGTGCTTTGAGTAA	81.3	54.2

Fig. 19. Ptac-T7 Synthetic Controller.

Order #8-19: backbone is pincW (KT 96)

SEQ ID NO:	Order	Name	Sequence
SEQ ID NO:33	1	lacI	caattegegctaactlacattaatfgegtfgegfcactgcccgcttccagtcgggaaacctgtcgtgccaggctgcatfaatgaatcggccaacgcgcggggagaggcgggttgcgtattgggcgccagggtggttttc tttccaccagtgagacggccaacagctgattgcccttcaccgccctggccclgagagaggtgacagcaagc ggtcaccgctgggttggcccagcaggcgaataatcctgtttagtggtggtgacggcgggatafaacatg agctgtcttcggatcgtcgtatccactaccagatatacgcaccaaacgcgcagcccggaactcggtaat ggcggcgaftgcggccagcgcacatgafctgttggcaaccagcaicgcagtgggaaacgatgccctcat tcagcaattgcaftgggttggtaaaacgggacatggcaccagcgccttccggttcgctatcggciga atttgattgcgagtgagataattatgccagccagccagacgcagacgcgcggagacagaactaattggg cccgctaacagcgcgattgctggtagcccaatgcgaccagatgctccacgccagtcgcgtaccgtct tcattgggagaaaataactgtgtggtggtgtcggtcagagacatcaagaataaacgccggaacattag tgcagggcagctccacagcaatggcactcgggtcaccagcggatagttatgacgccaccatgacgc gttgcgcgagaagattgtgcaccgcccttacaggctcgcagccgctgcttaccatcgacaccac cacgctggcaccagtgatcggcgcgagattlaatcggcgcgacaattggcagcggcgcgtgcaggg ccagactggagggtggcaacgcccaatcagcaacgactgttggccggcagttgtgtgcccacgcgggtg ggaaatgtaattcagctccgccatcggcgttccacttttcccgcttttcgcagaaacgtggcctggcctg gttcaccacgcgggaaacgggtctgataagagacaccggcactctgcggacatcgtataacgttactgg tttcac
SEQ ID NO:34	2	lacI promoter	atcaccacctgaattgactctctccggcgctatcagccataaccgcgaaagggtttgcaccattcgat gggtgcaacgtaaatgcatgcggctcgccttcgcgcgcgaattggccgccatgcccggcgataatggcc tgctctcgcgaaacgttgggtggcgggaccagtgacgaaggcttgagcgaaggcgtgcaagattcc gaataccgcaagcgacaggccgatcctcgcgctccagcgaaggcggctcgcggaaaatgacc cagagcgtcggcgaccctgtctacaggtgcatgataaagaagacagtcataagtgccggcagcat agtcattgccccgcgccaccggaggagctgactgggtgaaggctcgaaggcctcggcggcagct latcgactgcacgggtgcaccaatgcttctggcgtcaggcagccatcggagctgtggtatggctgtgca ggctgtaaatcactgcataaattcgtgtcctcaaggcgcactccggtctggataatgtttttgcgccgac atcatacgggtcggcaaatatcgaatgagctg
SEQ ID NO:35	3	tac promoter	ttgacaattaatcctcggctcgtataatgtgtggaattgtgagcggataacaatt
SEQ ID NO:36	4	insulator 1	TGCAGTTTTATTCTCTCGCCAGCACTGTAATAGGCACTAA
SEQ ID NO:37	5	T7 RNAP rbs	TATCCAAACCAGTAGCTCAATTggagtCGTCTAT
SEQ ID NO:38	6	umuD degradation tag	Gtgtgtttatcaagcctgcggatcctccgcaaatgtgactttccgctatitagcgtcttgttcagtggtg ctttcttcaccggcagcagattfacgtgaaacagcgcacatgctg

SEQ ID NO:	Order	Name	Sequence
SEQ ID NO:39	7	T7 RNAP scaffold	ggtggcAACACGATTAACATCGCTAAGAACGACTTCTCTGACATC GAACTGGCTGCTATCCCGTTCAACACTCTGGCTGACCATTACG GTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTC TTACGAGATGGGTGAAGCACGCTCCGCAAGATGTTTGAGCGT CAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAG CCTCTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCA ACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCC CGACAGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGT AGCGTACATCACCATTAAGACCACTCTGGCTTGCCTAACCAGT GCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTC GGGCCATTGAGGACGAGGCTCGCTTCGGTCGTATCCGTGACCT TGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAACAACCTAA CAAGCGCGTAGGGCACGTCTACAAGAAAGCATTATGCAAGT TGTCGAGGCTGACATGCTCTAAGGGTCTACTCGGTGGCGAG GCGTGGTCTTCGTGGCATAAAGGAAGACTCTATTCATGTAGGAG TACGCTGCATCGAGATGCTCATTGAGTCAACCGGAATGGTTAG CTTACACCGCCAAAATGCTGGCGTAGTAGGTC AAGACTCTGAG ACTATCGAACTCGCACCTGAATACGCTGAGGCTATCGCAACCC GTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTG CGTAGTTCCTCCTAAGCCGTGGACTGGCATTACTGGTGGTGGC TATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTGCGTACTC ACAGTAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGC CTGAGGTGTACAAAGCGATTAACATTGCGCAAAACACCGCAT GGAAAATCAACAAGAAAGTCTAGCGGTGCGCAACGTAATCA CCAAGTGAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGA GCGTGAAGAACTCCCGATGAAACCGGAAGACATCGACATGAA TCCTGAGGCTCTCACC GCGTGGAAACGTGCTGCCGCTGCTGTG TACCGCAAGGACAAGGCTCGCAAGTCTCGCGTATCAGCACTTG AGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCAATAAGGC CATCTGGTTCCTTACAACATGGACTGGCGCGTTCGTGTTTAC GCTGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAG GACTGCTTACGCTGGCGAAAGGTAACCAATCGGTAAGGAAG GTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCGGGTGT CGACAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAA AACCACGAGAACATCATGGCTTGCCTAAGTCTCCACTGGAGA AACTTGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGC GTTCTGCTTTGAGTACGCTGGGGTACAGCACCAACGGCCTGAGC TATAACTGCTCCCTCCGCTGGCGTTTGACGGGTCTTGCTCTGG CATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGT CGCGCGGTTAACTTGCTTCTAGTGAAACCGTTCAGGACATCT ACGGGATTGTTGCTAAGAAAGTCAACGAGATTCTACAAGCAG ACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCG ATGAGAACACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCA CTAAGGCACTGGCTGGTCAATGGCTGGCTTACGGTGTTACTCG CAGTGTGACTAAGAGTTCAGTCATGACGCTGGCTTACGGGTCC AAAGAGTTCGGCTTCCGTC AACAAGTGCTGGAAGATAACCATT AGCCAGCTATTGATTCCGCAAGGGTCTGATGTTCACTCAGCC GAATCAGGCTGCTGGATACATGGCTAAGCTGATTTGGGAATCT GTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACTGG CTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATA AGAAGACTGGAGAGATTCTTCGCAAGCGTTCGCTGTGCATTG GGTAACCTCTGATGATTTCCCTGTGTGGCAGGAATAACAAGAAG CCTATTCAGACCGCTTGAACCTGATGTTCCCTCGGTGAGTTC GCTTACAGCCTACCATTAACACCAACAAGATAGCGAGATTG ATGCACACAAAACAGGAGTCTGGTATCGCTCCTAACTTTGTACA CAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGGC ACACGAGAAGTACGGAATCGAATCTTTTGCCTGATTACAGAC TCCTTCGGTACGATTCCGGCTGACGCTGCGAACCTGTTCAAAG CAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGT ACTGGCTGATTTCTACGACCAGTTCGCTGACCAGTTGCACGAG

SEQ ID NO:	Order	Name	Sequence
SEQ ID NO:40	8	Biobrick suffix (BBa_G00001)	Tactagtagcggccgctgcag
SEQ ID NO:41	9	Translational stop (BBa_B0042)	ttagttagitag
SEQ ID NO:42	10	Terminator (BBa_B0054)	attagcagaaagtcaaaagcctccgaccggaggcttttgactaaaacttccttgggggtatcattggg
SEQ ID NO:43	11	VR primer sequence (BBa_G00102)	gctcactcaaaggcggtaat
SEQ ID NO:44	12	Terminator (BBa_B0062)	cagataaaaaaatccttagcttccgctaaggatgatttct

SEQ ID NO:	Order	Name	Sequence
SEQ ID NO:45	13	pSa origin	gctagagctgtcagaccagagctccaaccaacgttttatctafaccgtacagggtatttaattctatfita atctgctagaaatgaggcatgtttaaaccgaatctgacgtttccctgcaaatgccaaaatactafgctat ctccgggttctgctgacggccaagaccgggaaaaccaaaatacgggttctgctgaatacgcgaacgc caaaggcttctgctgctacagatcgaggagaatgcccagcagagcaaaaggactggcgaacgc caagctcggagcaatcgcgaaaaggccgagcgtggcgacgtgctactgacgccggagattacgcgc attgcccgttccgctcgcctggccctggaaatctcaaagcggcgagcgcgcggcctaategtccat gtgaccaaacagaagatcatctggacggcagcctacaagcgacatctggcaaccgtgcttggctt ggctgcacagatcgagcggcatttcaatgagcagctaccaccgagcgcctacaagctgccagagag gcggcaagacgctcggcgaccacaaggcagcaaatcgagcgccttgaagctggacagcctgatt gatgaagtacaggcatactgtaaccttggcttggcgaagctcgcgagcaggttggtagcgtcag cctcacaccttggcctgttcaatcaaacgggaacatcaaacccacaacaactagaccaccatcac catgcccggggagggaacaacatgcttaagaacaacaagccccggccatctatcaacgagatcat caagacgagcctcgcctcgaatggaggatgcccgcgaagctggcttagctggctacatggcccgtt gcccttggcaagcgacctgcccacaccgaccccgaagaccagctacttggagcgcaccaatggcctc gtcaccttggatcatgggcaagccgagcctcggcctgcccctacggctctatgcccggcaccttgg ctggatgaccgagggcgtgcaacgaaagaccctgcttgaaccttggccggctcgaatgggaa ttctacaaggctcggatgcacaccgatggccgctacacggccacccttcgcaatcagggcgaacg cctgtttcatcatgatttgccttggcggcgagcaaggcaatgacttggcattgagaacgctgctattg caagcgcgttcttcttctggaatcccagcggccagaagatcggggcctatgggatagcaccctcac cctcacaggcgttcttctgaggaagtaccggctcaccggctctacccaatcgactacactgcatgctt tggcggcagctcctgcttgcgatggacatttacacgtggctgacctatcgcgtgttctgttggggcca gggccggccccttctgcaaatcccttgggtcgccttgaagcgcaatctggctctatctatggcagccg cgcacgcaactgcccgaactggacgataaggcccggagcgggcaagcggcagcactcgcaca gcttcaatacaactcaaaaagcgcctacgcgaagctgttgaattgtctatcccggaggcaagcactgcat cgaagatgacggcgaatgcttgcctcaataccacacgccttcaatgacaccggcgcaccggcaag ggccgctgcacggcccctcggactgacaggccaacgctacgcttggcttggtaagccttccca tccaacagcccggcgtcggagcgggctttttatcccgggaagccttggatagaggggtgattatccag gaaaccgctaatgcccgcgaagccttgatcaggggcttccggccgcctcaaaaatctccacgt gaaatcgttaatcagggctgctgaaatcgttaatcggagctgtaaatcgttaataaggctcgtgaa atcgttaatacaaaaggcagctgagaacgctaaatgccccttccagatcaaacgttgcgaacaccctc gctcggcaagtagttacagcaagtagtattcaatagctttcaatataatataatataatatttgg cggccttggcttgggaacalggctacggcaccggcctcggcctgggacaaccgcaagcggcttgc ccaccgtcggcggccagccttggccacaaccggcggccgcaacagatcgtttataaattttttt tgaaaaagaaaaagcccgaaggcggcaacctcgggcttctggattccgatcaacgcaggagctg ttggaaagtagctgttccagaattataggcgcagagacaccagattcaagatgctctgttaattgtt gtagtattgtagtatacaaacatactacagtacagagggcccgaagaatggcaatcactaaacaagac atttggcggcagccgacgaactggacggcgaaggcctcggcccacttggccggcgttggcgaaga aactcggaaagcggtagcttcaaacatttccgaltgcaatggctgaaatgaaaaaccgcaagaccgc accctgcccctacagaccatttggcggctgacgcaacgagcacttggcagcttggcaatgctctat gggtatcggcctggcgcacgcaaacggcgglltgacgaagatcggaaacagatcgaggccgaca aagcggccatcagccagcagcttggcgaagcaatcgaactagccgacaccttcccccgaaaacga ccagctcccggaacgagtgaaatcagctcgaacctatggaaacggcagcgcgacaagctggccgacca gcttggcgaagtgaaagcgcagcggcgaagaactaaaccgctgcaatgaaaagctcaccacaag cgataacgaggctatcgaggcccgcaaacaggcgaaggaggccatcgagcgcggccgcaagcttgc aggtcaggtggaagcccacaagagcaggtcggcaatctcacagccgttggaaaacaggaggcaaa caatgaaaagcggccttggcggccttggcgggctggcggcggcgtgctgctcclaatctcaggttggccg ctggggccgacttccggggtgaagctgctcgaatccttggcggctgacacatcagacgttgggtgaaaccg tcagaccatccgctgtagattggcgaatgacgcaccgggaaagcggccaaagccttggcctcccggtg ctggccaacggctcggcacttggcgttggcttggcggcggcggcggcggcggcggcggcggcggcggc atggccgcaacttggggtgttggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc atcaatgcaatcaggttcaagaaggcctggcctggccttaccgttattaccggcaaaccaaccgagcgc cagatgtaggtagtgaaaaaggggccggcccaacggcctggccttggctagaccggcaatgctca ggaagccttggaaatggcctcggcctcgaaaaatgcccaactgacaccgggcaagcggccttggctg acggcggcagggcagctcgaattaccggcgggacg

SEQ ID NO:	Order	Name	Sequence
SEQ ID NO:46	14	spectinomycin resistance	<p> cccctcgtcccgacacttcagatcgccatagcgcacagcgcctcggagcgggtggtaacggcgcagtg gcgggtttcatggcttgatgactggtttttggggtacagctatgcctcgggcatccaagcagcaagcg cgttacgccgtgggtcgtatggtgatgtatggagcagcaacgatgttacgcagcagggcagtcgccct aaaacaaaagttaaacatcatgagggaagcgggtgacgccgaagtatcgactcaactacagaggtagt ggcgtcagcagcggcctcgaaccgacgttgctggccglacattgtacggctcggcagtggtatggc ggcctgaagccacacagtgaatgatttgcctgggtacgggtgacggtaaggctgatgaaacaacggc cgagctttagatcaacgacctttggaaactcggctccctcggagagagcagattcctcggctgtag aagtcaccattgtgtgcacgacgacatcctcgtggcgttatccagctaacgcgcaactgcaattgga gaatggcagcgcgaatgacattctgcaggatctcgcagccagccagcagcagcattgatctggctactt gctgacaaaagcaagagaacatagcgttgccttggtaggtccagcggcggagggaactcttgatcgg ttcctgaacagatctattgaggcgttaaatgaaaccttaacgctatggaactcggcggcggcactggc tggcgtatgagcgaatgtatgcttactgttcccgcttgggtacagcgcagtaaccggcaaatcgc ggcgaaggatgtcgtcgcgactgggcaatggagcggcctgccggccagtaicagcccgctacttg aagctagacagccttatctggacaagaagaagatcgttggcctcggcgcagatcagttggaagaat ttgtccactacgtgaaaggcagatcaccaggtagtcggcaataatgctaacaaatcgtcaagccg acggcctcgcggcggcgttaactcaagcgttagatgcaactaagcacaataatgctcacagcacaac tctcaggctcaagtctgcttttatttttaagcgtgcataataagccctacacaatgggtacc </p>
SEQ ID NO:47	15	Terminator (BBa_B0053)	tcggcaaaaaaacgggcaagggtgaccaccctgcccctttttctttaaaccgaaaagacttccggt
SEQ ID NO:48	16	VF2 primer sequence (BBa_G000101)	tgccacctgacgtctaaagaa
SEQ ID NO:49	17	Terminator (BBa_B0055)	aaggaatattcagcaatttcccgtgccgaagaaaggccaccctggaaggtagcagtgagtgatgatt gctacgtaa
SEQ ID NO:50	18	Translational stop (BBa_B0042)	ttagttagttag
SEQ ID NO:51	19	Biobrick prefix (Bba_G00000)	gaattcggggccgctctagag

Fig. 20. Parts List of Full Synthetic Cluster

Order # 57-67 have pSB4C5 backbone.

	Order	Name	Strength (AFU)	Sequence
SEQ ID NO:52	1	T7 promoter WT	9000	TAATACGACTCACTATAGGGAGA
SEQ ID NO:53	2	nifH rbs	5500	ACAATAAACTAACATAAGGAGGATAAAT
SEQ ID NO:54	3	nifH		ATGACCATGCGTCAGTGC GCGATTTATGGCAAAGGTGGTATTG GCAAAAAGCACGACGACCCAGA AACTTGGTGGCGGCCCTGGCCG AGATGGGTAAAAAGGTTATGATTGTGGGTTGCGACCCGAAGG CCGACAGCACGCGCCTGATTCTGCACGCGAAAAGCACAAAACA CGATTATGGAGATGGCTGCCGAGGTTGGTAGCGTGGAGGATC TGGAGCTGGAGGACGTTCTGCAAATTTGGTTACGGTGATGTTCCG TTGCGCAGAGAGCGGTGGTCCGGAACCAGGTGTCCGGCTGTGC GGGTTCGTGGTGTGATTACCGCTATCAATTTCTGGAAGAAGAG GGTTCGTACGAAGATGATCTGGATTTCTGTTTCTACGATGTGC TGGGTGATGTCGTGTGCGGTGGTTTTGCAATGCCGATTCGCGA GAATAAGGCACAAGAAAATTTACATTGTCTGTAGCGGCGAGAT GATGGCAATGTACGCTGCTAACAACATCAGCAAGGGTATTGT TAAATACGCAAAAAGCGGTAAGGTTTCGCTTGGGTGGTTGATT TGCAACAGCCGTCAGACCGACCGTGAGGACGAACTGATCATC GCCCTGGCTGAGAACTGGGCACCCAAATGATCCACTTCGTG CCACGCGATAATATTGTTCAACGTGCAGAAATCCGCCGTATGA CCGTCATTGAGTATGACCCGGCATGCAAGCAAGCGAACGAGT ACCGCACCTTGGCACAGAAAATCGTGAACAACACCATGAAGG TTGTTCCGACGCCGTGTACGATGGACGAGCTGGAGAGCCTGCT GATGGAGTTCCGCATTATGGAGGAGGAGGACACCAGCATTAT CGGTAAGACCCGACGGAGGAGAATGCCGCATAA
SEQ ID NO:55	4	insulator 1		GCGTGCGTACACCTTAATCACC GCTTCATGCTAAGGTCCCTGGC TGCATGC
SEQ ID NO:56	5	nifD rbs	500	AAAAATTCACATCCCTATCTA GCGGAGGAGCCGG
SEQ ID NO:57	6	nifD		atgatgacaaatgctactggcgaaacgtaacctggcacigattcaagaaglaclggaagtggtcccgaa accggcgcgcaaaagagcgccgtaaacacatgatgggttctgaccggaaaatgaaatctgtgggtaaatg catcatctcaatcgcaaatctcagccgggtgcatgactgttctgtggcctgtgctgacgcagggttcaaa gtgtgctatcggccgacaaagatgagggcatalctctcagccggcaggcgtgtggccagtaact ctcggcggaacgtcgtactactacacggcggttctggcgttgacttctggcagcgtgaaactc ctctgactccaggaaacgtgacatcgttctgggtggcgataaaaagctgtccaaactgacgaagaaatg gaaactgtgttcccgctgactaaaggcattactatccaaagcgaatgctcgggtgggtctgacgggtatg acatcagcggcgtcgaaacgcattctcaaaagccctggataagccgggtgacccgggtcgtgag ggcttccggcggcgttctcagctctgggtcaccatcgaacgatggtgtgctgactggatctgaa caaccgtgaaggtcagcccttgaaacaccccttatgacgttgcgattatggcgactafaacatcggc ggcgacgcctgggcaccccgatctgctggaggagatgggtctcgtgtgtgctgcacagtggtctgg cgatggcaccctgggtgaaatggaaaacaccccggttgaactgaacctgggtcactcaccgctcc atgaaactaactgcccgtcacaatggaagaaaacatcagalcccttggatggaatacaactctcggctc cgactaaaatcgcaaatccctgcgtaaaatcgccgatcagttgatgataaccatcggcgaacgctg aagcagtaattgcgcgtacgaaggccagatggcagcaatcattgcaagtaccgctccggcctggaa ggctgtaaaagtctgctgacatgggtggctctgctccacgtcatgtatgggtgctcagaggacctg ggcatggagatcatcgcagcgggttacgaattgcacacaacgacgactatgctgacgctgccaga cctgaaagaaggtacgctgctgttggacgacgccagctcttatgaactggaagccctgctgaaagcgt gaaaccagacctgacggctccggcatcaaggaaaaatacatctcagaaaatggcgctgcccctcc gccagatgcactcctgggactactccggctcgtaccacggctacgacgggttgcctatctcgtcgtgta

				catggatagaccctgaataaccagcgtggaatgaactgaccgcaccgtggcctgaaatctgcafaa
SEQ ID NO:58	7	insulator 2		CAAACACCCCATGTCGATACTGAACGAATCGACGCACACTCC CTTCCTTG
SEQ ID NO:59	8	nifK rbs	50	CAATCTCATACTCTCAAAAATTAGGCGAGGTAAC
SEQ ID NO:60	9	nifK		atgtctcaactatcgataaaatcaactctgttaccgctgtctgagcaggacaatatcaggaaactgtt ccglaacaaaacgtcagctggaagaagcgcacgacgcacagcgcgtgcaaggaaagtgttcgcatggac caaccaccggaatacgaagctctgaacttccgtcgcgaagccctgacggctgagccgcaaaagc tgccagcctctgggtcgggtctgtgcagcctgggtttgccaacaccctgcccgtatgtccacggctccc agggctgcgtagcctacttccgtacctattcaaccgcccacttlaaagaaccaatcgcgtgctgtccga cagcatgacggaggacgcggcagtttccgtggttaacaacaactgaacctgggcccgtcaaaatgctt ccgcactgtacaaaaccggaatcaccgcagctgtaccacctgcatggcagagggttattggtgatgat gcaagcalttattgccaacgcaaaagaaagacgggttctgttgacagctctatccggctccgcacgcat accggctctctcaggttctcagtaactgggtgggacaacatgttcgaaggcttgcaaaaactttac cgcagactatcaaggccaaccgggtaaacggcgaagcigaacctgggtgaccggcttgaaacctacc tgggcaacttctgtctgaagcgcagatgatggagcagatggcgggtccgtgttctctgtctgtgacc gtctgagggtctggacactccagcggacggccactatcgalgtattctgtggcaccactcagcagga aatgaaagaggcccagacgcgaltgacaccctgctgctgcaaccgtggcagctgctgaaagcaag aaagttgttcaggaaatgtggaaccagcggcaaccggaagttgcaalcccgtgggtctgagcagctac tgacgaactgctgatgaccgtgtccaaactgagcggcaaaccaatcgcggatgctctgaccctggaac gggtcgcctgggtgacatgatgctggacagccacacgtggctgcatggcaaaatltggcctgtac gggtgaccggacttcgtaatgggectgacccgttctctgctggaaactgggctgagagccgactgtatc ctgtctcaaacgctaacaaactgtggcagaaggccaatgaaacaaatgctggatgagagcccaacgg ccgtgatagcgaagtgttcaactcagcaccgtggcatttccgtctctgatgttacggctcagccgg attcatgatcggtaactctacggcaaatcaccagcgtgacactctggcacaaggcaaaagcgttga agtgccgtgatctgtgggttccgtgttcgaccgtcaccacctgcaaccgagaccacctgggg ttacgaaggcgcgatgaacatcgtactctgtgtaaacgcagtaactggaagcagtaactggaagcgtgacagcagata ctcccagctgggcaaacgactattcttctgatctgggtctgtaa
SEQ ID NO:61	10	insulator 3		CCTGATTGTATCCGCATCTGATGCTACCGTGGTTGAGTTA
SEQ ID NO:62	11	nifY rbs	240	CCATACTCACTCCCGGAGGTACTTCT
SEQ ID NO:63	12	nifY		ATGTCTGACAATGATACCCTGTTTTGGCGCATGCTGGCGCTGT TTCAGTTCGCTGCCGGATTTGCAGCCGGCTCAAATCGTCGATTG GCTGGCGCAGGAATCCGGCGAAACCCTGACGCCGGAGCGCT TGCCACCCTGACCCAACCGCAACTCGCGCGTTCGTTCCCATCC GCGACGGCAGTGATGAGCCCGGCTCGCTGGAGCCGCGTTATG GCTTCTCTGCAAGGCGCCCTCCCAGCCCCTTGCGCATCGTAC GTCCGGCGCAGCGTACCCCGCAACTGCTCGCCGCGTTTTGCAG CCAAGACGGCCTTGTATCAATGGTCAATTCGGCCAGGGTTCGT CTGTTCTTCAATFACGCCCTTTGACGAGCAGGGCGGCTGGCTGT ATGACTTGCGCCGCTATCCGAGCGCACCCGACCAGCAGGAAG CGAATGAGGTGCGTGTCTGCTGATTGAAGATTGCCAGCTGCT GTTCTGCCAGGAGATTGGCGGTCCGGCAGCAGCGCGTCCGAT CCGCCACCGCATCCATCCGATGAAGGCGCAGCCGGGTACTAC GATTCAGGCGCAGTGTGAAGCTATCAACACCCTGCTGGCCGG TCGCCTGCCCGCTGGCTCGCCAAACGTTTGAACCGTGATAAC CCGCTGGAAGAGCGTGTGTTTTAA
SEQ ID NO:64	13	insulator 4		CATTTTTGCCTTGCACAGACCTCCTACTTAGATTGCCAC
SEQ ID NO:65	14	nifE rbs	40	ACTATTCAATACATCACTGGAGGTTATTACAA

SEQ ID NO:66	15	nifE		<p>ATGAAGGGTAACGAGATTCTTGCTCTGCTGGACGAACCGGCC TGTGAACACAACCATAAACAGAAATCCGGCTGTAGCGCCCCA AAGCCGGGTGCGACGGCGGCTGGCTGCGCTTTCGATGGTGCG CAGATCACCTGCTCCCGATTGCGGACGTTGCCACCTCGTGC ATGGCCCAATCGGTTGCGCAGGTAGCTCTTGGGACAACCGTG GCAGCGCCTCCAGCGGTCCGACCCTGAATCGTTTGGGCTTTAC CACTGACTTGAATGAACAAGATGTGATCATGGGTGCGGGCGA GCGTCGCCTGTTCCACGCTGTGCGCCATATTGTCACCCGTTAC CACCCAGCGGCAGTATTCATCTACAATACGTGCGTGCCGGCTA TGGAAGGCGATGACCTGGAGGCCGTGTGTACGGCAGCCAGA CTGCGACCGGCGTCCCGTAATCGCAATTGATGCGGCTGGCTT CTACGGTTCGAAGAACCTGGGCAACCGTCCGGCAGGCGATGT CATGGTTAAACGCGTCATTGGCCAACGTGAGCCAGCGCCGTG GCCGGAGAGCACCTGTTGCCCGGAGCAACGTGATGACAT TGGCTTGATCGGTGAGTTCAACATTGCGGGCGAGTTTGGCAC ATTCAGCCGCTGCTTGATGAGCTGGGTATCCGCGTTTTGGGTT CGCTCAGCGGGCGATGGTCGTTTCGCCGAGATTCAAACCATGCA CCGTGCCCAGGCGAACATGCTGGTGTGCAGCCGTGCTCTGATC AATGTTGCGCGTGCTCTGGAACAGCGCTATGGCACCCCGTGGT TTGAAGGCTCGTTCTATGGTATCCGCGCGACCAGCGACGCCCT GCGCCAGTTAGCGGCGCTGCTGGGCGATGACGACCTCCGTCA GCGCACCGAGGCGCTGATCGCGCGTGAAGAACAGGCGGCTGA GCTGGCCCTGCAACCGTGGCGTGAACAGCTGCGTGGCCGCAA GGCCCTGCTCTACACGGGTGGTGTCAAAAGCTGGTCTGTGTG TCCGCGCTTCAGGATCTGGGTATGACCGTGGTTGCCACGGGCA CGCGTAAGAGCACGGAAGAGGATAAACAGCGCATCCGCGAAT TGATGGGCGAAGAGGCCGTGATGCTTGAAGAAGGCAACGCAC GTACCTTATGGATGTAGTTTATCGCTATCAAGCAGACCTGAT GATTGCCGGTGGCCGCAACATGTATAACGCGCTACAAGCGCG CTTGCCGTTCTGGACATCAACCAGGAACGCGAGCACGCGTTT GCGGGCTACCAAGGCATCGTGACCTTAGCGCGCCAGCTGTGC CAAACGATTAACAGCCCGATCTGGCCGACACTCATTCCCGC GCACCGTGGCGCTAA</p>
SEQ ID NO:67	16	insulator 5		TGTCACGCTAGGAGGCAATTCATAAGAATGCACACTGCA
SEQ ID NO:68	17	nifN rbs	40	CCTAAACCTACCACACCTGGAAGAAGTAATT
SEQ ID NO:69	18	nifN		<p>ATGGCAGACATTTTCCGCACTGATAAGCCGTTGGCTGTGTGCG CGATCAAGACCGGCCAGCCGCTGGGTGCGATCCTGGCGTCCC TGGGTATCGAGCACTCGATTCCGCTGGTACATGGCGCGCAGG GCTGTTCCGCTTTTGCCAAGGTTTTCTTTATCCAGCACTTCCAC GATCCGGTCCCGCTGCAAAGCACGGCAATGGACCCGACCAGC ACCATCATGGGCGCTGATGGTAACATCTTACCAGCGCTGGACA CTCTCTGCCAACGCAATAAACCAGCAAGCAATTGTGCTGCTGAG CACCAGCCTCTCCGAGGCGCAGGGCAGCGACATTTCCCGTGT AGTGCGTCAGTTCCGTGAAGAATATCCGCGTCATAAAGGCGT GGCGATTCTGACTGTTAACACCCCGGACTTTTACGGTAGCATG GAGAACGGCTTTTCCGCTGTCTGAGTCTGTGATTGAACAGT GGGTTCCGCCAGCCCCACGTCCGGCGCAGCGCAATCGTCCGG TCAATCTTTTGGTGAGCCATCTCTGTAGCCAGGCGATATTGA GTGGCTGCGCCGTTGCGTGCAGGGCCTTCGGTCTGCAACCGATC ATTCTGCCGGATCTGGCTCAGAGCATGGACGGCCACCTTGCTC AGGGTGACTTTTCCGCGCTGACGCGAGGGCGGCACGCCGTTGC GCCAAATCGAGCAGATGGGCCAGAGCCTTTGCTCTTTTGGGAT TGGCGTCAGCCTGCACCGTGCAGGAGCAGCCTGCTGGCTCCGCGT TGTCGTGGCGAAGTCATCGCCTTGCCGCACCTCATGACCTTGG AACGCTGCGACGCCTTTATCCATCAGTTGGCGAAAATCAGCG</p>

				GTCGCGCCGTTCCGGAGTGGCTGGAACGCCAGCGCGGTCAGC TGCAAGACGCCATGATCGATTGCCACATGTGGCTGCAAGGCC AGCGCATGGCGATTGCCGCCGAAGGCCGACCTGCTGGCAGCGT GGTGCATTTCGCGAACTCTCAAGGTATGCAGCCGGGTCCACT GGTTGCTCCGACGGGTATCCGAGCCTGCGTCAGTTGCCGGTG GAGCGCGTGGTGCCGGGTGATCTGGAGGATCTTCAGACCCCTCT TATGCGCACATCCGGCCGACTTACTGGTGGCGAACTCCACGC CCGTGATTTAGCAGAGCAATTCGCCCTGCCGCTGGTGCGCGCA GGCTTCCCGCTGTTTGACAAACTGGGCGAATTCGTCGTGTTT GCCAGGGTTATAGCGGTATGCGTGATAACCCTGTTTCGAGTTGGC GAACCTGATCCGTGAACGCCATCATCATCTGGCTCATTATCGC AGCCCGCTGCGCCAGAACCAGAATCCTCGTTGTCTACGGGTG GCGCGTACGCAGCGGATTAA
SEQ ID NO:70	19	nifJ rbs	100	ctagagattaaagaggagaaaitaagc
SEQ ID NO:71	20	nifJ		ATGAAAACATATGGACGGTAACGCTGCGGCTGCATGGATTAGC TACGCCTTTACCGAAGTGGCTGCGATCTACCCGATTACGCCGA GCACCCCGATGGCGGAAAATGTGGACGAATGGGCTGCGCAGG GCAAGAAGAACCCTTTCGGCCAGCCGGTGCGCCTGATGGAGA TGCAGTCGGAAGCGGGTGCAGCAGGTGCTGTGCATGGCGCCT TGCAAGCTGGCGCACTGACGACCACCTACACCGCGTCGCAGG GCCTGTTGCTGATGATOCCAAACATGTACAAAATCGCGGGTG AACTGCTGCCGGGTGTCTTTTCATGTTTCGGCACGCGCACTGGC CACCAATAGCCTCAACATCTTTGGCGATCATCAGGATGTAATG GCGGTGCGCCAAACGGGCTGCGCGATGTTGGCCGAGAATAAC GTCCAGCAAGTTATGGATTTGTCCGCGGTAGCCCACTTGGCAG CGATCAAAGGTGCGATTCCGTTTCGTGAACCTTCTTCGATGGCTT TCGCACCAGCCACGAAATCCAGAAGATCGAGGTTCTGGAATA TGAACAGCTGGCCACCTTGTGGATCGTCCGCCCTGGACAGC TTCCGCGGTAACGCCCTTACCCCGACCAACCCGGTCAATCCGTG GCACCCGCCAGAACCAGGACATCTACTCCAGGAACGTTGAGG CCGGTAACCGTTTCTATCAGGCGCTCCCGGATATTGTGGAATC TTACATGACCCAGATTTCTGCCCTGACTGGTCGCGAGTATCAC CTGTTTAACTACACTGGTGCTGCGGATGCGGAGCGCGTGATCA TCGCGATGGGCTCTGTCTGTGACACCGTCCAAGAGGTGGTTGA CACGCTGAATGCAGCGGGTGAGAAAAGTTGGTCTGCTCTCCGTT CATCTTTTCCGCCCGTTTTTCGTTAGCGCACTTCTTCGCCCACT GCCGAAAACGTACAGCGTATCGCAGTATTGGACCGTACGAA AGAGCCAGGTGCTCAAGCAGAGCCGCTGTGCCTCGATGTGAA GAATGCCTTTTACCACCATGACGATGCCCGTTGATTGTGGGT GGTCGCTATGCCTTGGGCGGTAAGGACGTGTTGCCGAACGAT ATTGCGGCCGTGTTTGATAACCTGAACAAACCGCTGCCGATGG ACGGCTTACGCTGGGTATCGTGGACGATGTTACCTTCACCTC TCTCCCGCCAGCGCAGCAGACCCTGGCGGTTTCTCACGACGGC ATCACGGCATGTAAGTTTTGGGGCATGGGCTCCGACGGCACG GTTGGTGCGAACAAGTCCGCGATCAAGATTATCGGCGACAAA ACGCCACTGTATGCGCAAGCGTACTTTTTCTACGACTCGAAGA AGAGCGGTGGTATTACCGTCAGCCATCTGCGTTTTGGTGATCG CCCGATCAACTCCCGTATTTGATCCATCGCGCGGATTTATC TCGTGCAGCCAGCAAAGCTATGTTGAACGCTACGATCTGCTGG ATGGCCTTAAACCGGGTGGCACCTTTCTGCTGAACGCTGCTCCTG GAGCGATGCCGAACCTGGAGCAACATCTGCCGGTCCGTTTTCAA ACGTTATCTGGCAGCGGAGAATATCCACTTCTACACTCTAAC GCTGTGGACATCGCCCGTGAGCTTGGTTGGGTGGCCGTTTTCA ACATGCTGATGCAGGCTGCCTTCTTCAAACCTGGCCCGATCAT TGACCCGCAGACTGCTGCGGACTATCTGAAGCAGGCTGTTGA GAAAAGCTATGGCAGCAAAGGTGCGGCGTTCATCGAGATGAA CCAGCGTGCCATCGAGCTTGGCATGGCCAGCCTGCACCAGGT

				GACGATCCCGGCACATTGGGCCACCCCTGGATGAGCCAGCGGC GCAGGCGTCCCGGATGATGCCGGACTTTATCCCGGACATCCTG CAACCGATGAACCGTCAGTGCGGGGACCAGCTTCCGGTGTGG GCTTTTGTTCGGCATGGAAGATGGCACCTTCCCGTCCGGCACGG CCGCATGGGAGAAACGTGGCATCGCCCTTGAGGTGCCAGTCT GGCAGCCGGAAGGCTGCACGCAGTGCAACCAGTGCGCCTTCA TTTGTCCGCACGCCGCGATTTCGTCCGGCGTTGTTGAATGGCGA AGAGCATGATGCTGCCCGGTTGGCCTGCTGAGCAAACCGGC ACAAGGCGCTAAAGAATATCACTATCATCTGGCGATTAGCCC GCTGGACTGCTCCGGCTGTGGCAACTGCGTTGACATTTGTCCA GCTCGTGGCAAAGCGTTGAAGATGCAGTCTCTGGATAGCCAA CGCCAGATGGCTCCGGTGTGGGATTATGCGCTGGCGCTGACCC CGAAGTCTAACCCGTTTCGTAAAACCACCGTCAAAGGCTCGC AGTTGAAACCCCGCTGCTGGAGTTTAGCCGTGCGTGCGCTGG TTGTGGCGAAACGCCGTATGCGCGCCTCATTACCCAGCTGTTT GGCGACCGCATGCTGATTGCCAATGCCACCGGCTGTTCCAGCA TCTGGGGCGCATCTGCGCCGAGCATCCCGTATACCACCAATCA TCGTGGTCATGGTCCGGCCTGGGGCAATAGCCTGTTTGAGGAC AATGCCGAATTTGGTTTAGGTATGATGCTGGGCGGTCAAGCTG TGCGTCAACAGATCGCGGACGATATGACGGCTGCGTTAGCGC TCCCGGTTTCCGATGAGCTGAGCGACGCGATGCGCCAGTGGTT GGCGAAACAGGACGAGGGTGAAGGCACGCGTGAGCGTGCGG ACCGTCTGAGCGAGCGCTTAGCCGCGGAGAAAGAGGGCGTTC CGCTGTTAGAGCAGCTGTGGCAAAATCGTGATTACTTTGTGCG TCGCAGCCAGTGGATTTTCGGCGGTGACGGCTGGGCCTATGAT ATTGGCTTCGGTGGCTGGACCACGTCTCGCCAGCGGTGAGG ATGTGAACATTCTGGTATTTGACACCGAAGTCTACTCGAACAC CGGCGGTCAAAGCAGCAAATCGACCCCGGTGCGCGCCATCGC CAAGTTCGCGGCTCAGGGCAAGCGCACCCGCAAGAAAGACCT GGGTATGATGGCGATGAGCTACGGCAACGTCTATGTAGCCCA GGTGGCGATGGGTGCGGATAAAGATCAAACCTCTGCGCGCCAT TGCGGAAGCTGAAGCGTGGCCAGGCCCGTTCGCTGGTGATTGC GTATGCGGCCTGCATCAATCATGGCCTGAAGGCCGGTATGCGT TGCAGCCAACGTGAGGCGAAGCGCGCTGTTGAGGCGGGCTAC TGGCACCTGTGGCGTTATCACCCGCAGCGCGAAGCGGAAGGC AAGACGCCGTTTATGTTAGATAGCGAAGAACCGGAAGAGTCG TTCCGTGACTTCTGTTGGGTGAGGTGCGCTACGCATCCCTGC ACAAGACCACCCCGCACCTCGCCGATGCCCTTTTCAGCCGTAC CGAAGAAGATGCGCGTGGCGCGCTTTGCGCAATACCGTCCGCT GGCTGGCGAAGAGTAATAA
SEQ ID NO:72	21	T7 terminator 25	1600	TACTCTAACCCCATCGGCCGTCTTAGGGGTTTTTTGT
SEQ ID NO:73	22	insulator 6		CCGTGGTTGAGTCAGCGTCGAGCACGCGGC
SEQ ID NO:74	23	T7 promoter mut 2	1500	TAATACGACTCACTAGAGAGAGA
SEQ ID NO:75	24	insulator 7		CGCGACTTCCAGAGAAGAAGACTACTGACTTGAGCGTTCC
SEQ ID NO:76	25	nifB rbs	220	CTCTCTGTAATACATCAAATCAATCATAGGAGGGCTAAA
SEQ ID NO:77	26	nifB		ATGACCTCTTGTTTCGTGTTTTCTGGCGGTAAAGCGTGCCGTC CGGCCGATGACTCCGCGCTGACTCCGCTGGTGGCCGACAAGG CAGCTGCGCACCCGTGCTATAGCCGCCACGGCCATACCCGCTT CGCGCGTATGCACCTGCCAGTCGCTCCGGCCTGCAACTTACAA TGCAACTACTGCAACCGCAAGTTCGATTGCAGCAATGAAAGC

				CGTCCGGGCGTGTCCCTCTACCCTGCTGACGCCGGAACAGGCTG TGGTGAAGGTGCGCCAGGTCGCCCAAGCTATCCCGCAGCTGteg GTGGTCCGGTATTGCTGGTCCGGGCGATCCGCTTGCGAATATCG CCCGCACCTTCCGTACCTTGGAGCTTATTTCGGAACAGTTGCC GGACCTGAAACTGTGCCTGAGCACCAACGGCTTGGTGCTGCC AGATGCCGTTGATCGTCTGCTCGATGTGGGCGTGGATCACGTT ACCGTCACCATTAACACCCTGGACGCAGAAATCGCAGCGCAA ATCTACGCGTGGTTGTGGCTGGATGGCGAACGCTACTCCGGTC GCGAAGCCGGCGAAATTCTCATTGCCCGCCAGCTGGAAGGCG TACGTCGCCTGACCGCGAAAGGTGTGCTCGTCAAGATCAACA GCGTATTGATTCCGGGCATCAATGACAGCGGCATGGCGGGTG TTAGCCGTGCGCTGCGCGCGTCTGGTGCCTTCATCCACAACAT CATGCCACTGATTGCGCGTCCGGAGCATGGCACTGTTTTCGGT CTGAACGGCCAGCCGGAACCGGACGCGGAAACCCTGGCGGGC ACGCGCTCCCGCTGCGGGCAGGTTATGCCACAAATGACCCAC TGCCACCAGTGCCGTGCCGACGCGATTGGCATGCTTGGTGAG GATCGCTCGCAACAGTTTACGCAATTACCGGCTCCGGAGTCCC TCCCGGCTTGGCTGCCGATCCTGCATCAGCGTGTCTCAGTTGCA TGCGAGCATCGCCACGCGCGGTGAGAGCGAAGCCGATGACGC CTGCCTGGTGGCCGTTGCGTTCGAGCCGTGGCGATGTAATTGAC TGCCATTTCCGGCCATGCCGACCGTTTCTATATCTATAGCCTGT TGCGGCTGGTATGGTTCTGGTTAACGAACGTTTCACCCCGAAA TACTGCCAGGGTCGCGATGACTGCGAGCCGACGCAATGCC GCACGCTTTGCTGCCATCCTTGAGTTGCTGGCGGACGTCAAAG CGGTGTTTTGTGTGCGTATCGGCCATAACCCGTGGCAACAGCT GGAGCAGGAAGGCATCGAACCGTGCCTGGATGGCGCCTGGCG TCCGGTATCCGAGGTCCTGCCGGCATGGTGGCAGCAGCGCCG TGGTAGCTGGCCGGCTGCATTGCCGCACAAAGGCGTTGCGTA A
SEQ ID NO:78	27	insulator 8		ACTACGAGATTTGAGGTAACCAAAATAAGCACGTAAGTGGC
SEQ ID NO:79	28	nifQ rbs	5	TGCGTTTAGCAGTTCCCGTAGGAATATTTCTT
SEQ ID NO:80	29	nifQ		ATGCCGCCATTGGACTGGTTGCGTCTGTTTGTGGTTACTCTATC ACGCCGGCAAAGGCAGCTTTCGCTTCGTATGGGCTTGTCCGC GCGTGACTGGCAAGCTCTGCGCCGTCCGCTGGGCGAGGTGGA AACGCCGCTGGATGGCGAAACCCTGACCCGTCCGCTCTGAT GGCGGAGCTGAATGCGACCCGCGAAGAAGAAGCCAGCAGCT GGGTGCCCTGGCTGGCCGTTGGATGCAACAGGATGCCGGTCC GATGGCGCAGATTATCGCAGAGGTGAGCCTGGCGTTCAACCA TCTCTGGCAGGACCTTGGCCTCGCGAGCCGCGCTGAACTGCGT CTGCTGATGTCTGACTGCTTCCCGCAGCTGGTTGTTATGAACG AGCACAAACATGCGCTGGAAGAAATCTTTTACCGCCAGCGTTG CCTGCTGCAACAGGGCGAAGTCATCTGTTCGAGCCCGTCTTGC GATGAATGCTGGGAACGTTCTGCGTGTCTTGAGTAA
SEQ ID NO:81	30	T7 terminator 1	1300	TACATATCGGGGGGGTAGGGGTTTTTTGT
SEQ ID NO:82	31	insulator 9		GTCTGTAGCACGTGCATC
SEQ ID NO:83	32	T7 promoter mut 3	2750	TAATACGACTCACTAATGGGAGA
SEQ ID NO:84	33	nifF rbs	45	GACAAGAGTCTCAATTATAAGGAGGCTTTACTAC

SEQ ID NO:85	34	nifF		ATGGCGAACATCGGCATCTTCTTTGGTACGGATAACCGGCAAA ACCCGCAAGATTGCGAAGATGATTCACAAACAGCTGGGCGAG CTGGCCGATGCCCCGGTTAACATCAATCGTACCACTTTGGATG ACTTTATGGCTTACCCAGTCCTGTTGCTCGGCACGCCGACGCT TGGTGATGGTCAACTGCCGGGCTTAGAGGCGGGCTGCGAGAG CGAAAGCTGGTCTGAGTTTATCTCCGGTCTGGATGACGCTTCC CTGAAGGGCAAAACCGTGGCGCTGTTTGGCTGGGCGACCAG CGTGGTTACCCGGACAACCTTCGTGTCGGGTATGCGTCCGCTGT TCGACGCGCTGAGCGCCCGTGGCGCCAGATGATTGGTAGCT GGCCGAACGAAGGTTATGAGTTTAGCGCATCGTCCGCGCTGG AAGGCGACCGCTTCGTCCGCTTGGTGTGGATCAAGACAATC AGTTCGACCAGACCGAAGCGCGCCTGGCGTCTTGGCTTGAAG AGATCAAACGCACCGTTCTGTAATAA
SEQ ID NO:86	35	T7 terminator 1	1300	TACATATCGGGGGGGTAGGGGTTTTTTGT
SEQ ID NO:87	36	insulator 10		GGTCATTACAACGGTTAT
SEQ ID NO:88	37	T7 promoter mut 2	1500	TAATACGACTCACTAGAGAGAGA
SEQ ID NO:89	38	insulator 11		AACATAGCGTTCCATGAGGGCTAGAATTACCTACCGGCCT
SEQ ID NO:90	39	nifU rbs	2800	CAGATACTGACAAATAAACCAGCGAAGGAGGTTTCTTA
SEQ ID NO:91	40	nifU		ATGTGGAACACTACAGCGAGAAAAGTCAAGGACCATTTCTTCAAT CCGCGCAACGCGCGTGTGTGGATAACGCAAATGCGGTGGGC GACGTCGGCAGCTTATCTTGTGGCGATGCTCTCCGCTTGATGC TGCGCGTGGACCCGCGAGCGGAAATCATCGAAGAAGCGGGCT TTCAGACCTTCGGCTGCGGCAGCGGATTGCGTCCAGCGC ACTGACGGAGCTGATCATCGGTACACCCCTGGCGGAAGCGGG TCAGATCACCAACCAGCAGATCGCCGACTATCTGGACGGCTT ACCGCCGAAAAGATGCACTGCTCTGTAATGGGCCAGGAAGC TCTTCGTGCGGCCATTGCTAACTTTCGCGGTGAATCGCTGGAA GAGGAGCATGACGAGGGTAAGCTGATCTGCAAGTGCTTCGGC GTCCGATGAAGGCCATATTCGCCGTGTGTCAGAACAAACGGT CTTACGACTCTGGCCGAGGTGATCAATTACACCAAGGCAGGT GGCGGTGTACCAGCTGCCATGAGAAAATCGAGCTGGCCCTG GCCGAGATTCTCGCCCAACAGCCGCAAACCACCCCGGCAGTT GCGTCCGGTAAAGATCCGCACTGGCAGAGCGTCCGTTGGATAAC ATCGCTGAACTGCGTCCACATATCCAAGCGGACGGTGGTGAC ATGGCGCTGTTGTCCGTGACGAACCACCAAGTGACTGTTTCGC TGTCGGGCAGCTGTTCTGGCTGCATGATGACCGACATGACCC GGCGTGGCTGCAACAGAAATTGATGGAGCGTACCGGCTGCTA TATGGAAGTTGTTGCCGCCTAA
SEQ ID NO:92	41	insulator 12		CATTGTAATAGCCACCAAAGAGTGATGATAGTCATGGGT
SEQ ID NO:93	42	nifS rbs	175	GATACCCGTAGACCATTCTGAAATCGAAGGAGGTTTTCC
SEQ ID NO:94	43	nifS		ATGAAACAAGTGTACCTGGACAACAACGCGACCACCCGCCTG GACCCGATGGTTCTGGAAGCGATGATGCCGTTTCTCACGGATT TCTATGGCAATCCGTCCAGCATCCATGACTTCGGCATCCCGGC ACAAGCGGCGCTGGAACGTGCGCACCAGCAAGCTGCGGCACT GCTGGGCGCAGAGTACCCGTCTGAAATCATTTCACGAGCTGT GCGACCGAGGCCACTGCAACCGCCATTGCGTCCGGCATCGCG

				<p>TTATTGCCGGAACGCCGCGAAATCATCACCTCGGTAGTGGAG CACCCGGCTACGCTGGCGGCGTGCAGACACCTGGAACGCCAA GGCTATCGCATCCATCGCATTGCGGTGGATAGCGAAGGTGCG CTGGACATGGCCAGTTCCGTGCAGCGCTC_{cg}CCGCGTGTGCG GTTGGTGAGCGTGATGTGGGCCAACAAACGAAACCGGCGTGCT GTTCCCGATTGGCGAAATGGCCGAGCTTGCCACGAGCAGGG CGCTCTGTTCCACTGCGATGCCGTTCAAGTCGTTGGCAAATC CCAATTGCTGTTGGCCAGACGCGCATCGACATGCTGTCTTGCT CCGCGCACAAAGTTTCATGGTCCGAAGGGTGTGGTTGCTTGTA CTTACGTCGTGGCACGCGCTTTCGTCCGCTGCTTCGCGGTGGC CATCAAGAATATGGTCGCCGTGCCGGCACTGAGAATATCTGT GGCATCGTCGGCATGGGCGCTGCGTGCGAACTGGCGAACATC CATCTGCCGGGTATGACCCATATTGGCCAGTTACGCAATCGCC TGGAGCACCGTCTGCTCGCCAGCGTGCCGTCCGTGATGGTTAT GGGCGGTGGTCAGCCG_{cg}TGTACCGGGTACTGTCAACCTGGCG TTCGAGTTTATCGAAGGTGAAGCGATCCTGCTCTTGCTGAACC AGGCTGGCATTGCCGCAAGCTCCGGTCCGCGTGTACCTCTGG CAGCTTGGAGCCGAGCCATGTGATGCGCGCCATGAACATTCC ATACACCGCGGCTCACGGCACCATTGTTTTAGCCTGAGCCGT TATACGCGCGAGAAAGAGATCGACTACGTCGTTGCGACCCTC CCGCCAATCATTGATCGTCTGCGTGCCTTGTCGCCGTATTGGC AGAATGGTAAGCCGCGTCCGGCAGATGCAGTCTTTACCCCGG TTTACGGTTAA</p>
SEQ ID NO:95	44	insulator 13		GAGTTACTGGCCCTGATTTCTCCGCTTCTAATACCGCACA
SEQ ID NO:96	45	nifV rbs	5	GCGACTAGGAGCCTAACTCGCCACAAGGAAACAT
SEQ ID NO:97	46	nifV		<p>ATGGAGCGCGTCTTGATCAACGATACTACCCTGCGTGATGGCG AACAAATCTCCGGGCGTAGCGTTTCGTACCTCCGAGAAAGTTGC CATCGCGGAGGCACTGTACGCTGCGGGTATCACCGCGATGGA AGTCGGCACTCCGGCGATGGGTGATGAAGAGATCGCCCGCAT TCAGCTGGTGCCTCGTCAACTGCCGGACGCGACGCTTATGACC TGGTGCCGTATGAACGCTCTGGAATCCGTCAGAGCGCGGAT CTGGGTATTGACTGGGTGGATATCTCGATCCCAGCATCCGACA AGCTGCGTCAGTACAAGCTGCGTGAGCCGCTGGCCGTGCTGCT GGAGCGCCTTGCATGTTTATCCATCTGGCCACACGTTAGGC CTCAAAGTATGTATTGGTTGCGAGGATGCGAGCCGTGCGTCTG GTCAGACCCTGCGCGCCATTGCCGAGGTGGCCCAGCAATGCG CGGCTGCGCGCTTTCGTTACGCTGACACCGTGGGCCTGCTGGA CCCGTTACACCCGAGCCAGATCAGCGCCCTGCGTGACGTT TGGTCGGGCGAGATCGAGATGCATGCTCAAAATGATCTGGGC ATGGCTACCGGAACACGCTGGCGGCAGTTTCGGCTGGCGCC ACGTCGGTGAACACTACCGTCCCTCGGTCTGGGTGAACGTGCA GGCAACGCAAGCCCTGGAAACCGGTTGCGCTGGCCCTGGAACGC TGCCTGGGCGTGGAACCCGGCGTCCATTTAGCGCGCTCCCAG CGAGCTGTCAGCGCGTCCGCGGAGGCTGCACAGCGCGCAATCG ACCCGCAACAGCCGCTGGTGGGTGAATTGGTTTTACCCACGA GTCTGGTGTTCACGTTGCGGCGCTGCTGCGCCACAGCGAATCC TATCAATCTATTGCCCCAAGCCTCATGGGCCGTAGCTACCGTC TGGTGCTCGGCAAGCAATTCGGGTGCTCAGGCTGTCAACGGTGT TTTCGACCAGATGGGTACCACCTGAATGCGGCGCAGATCAAT CAGTTGCTGCCGGCCATTCCCGCTTCGCCGAGAATTGGAAAC GCTCTCCGAAAGACTACGAACTGGTTGCGATCTATGACGAATT GTGCGGTGAATCCGCCCTTCGTGCTCGCGGCTAA</p>
SEQ ID NO:98	47	insulator 14		GACTCAACACGCTAGGGACGTGAAGTCGATTCTTCGATG

SEQ ID NO:99	48	nifW rbs	40	CAGAAGGCGAGA ACTAGATTTAAGGGCCATTATAG
SEQ ID NO:100	49	nifW		ATGGAGTGGTTTTACCAGATTCCGGGTGTAGACGAATTGCGCAGCGCTGAATCCTTCTTFCAGTTCTTCGCGGTTCCATAACCAGCCGGAAGTACTGGGCGCTGCTCGCTTCCGGTGTAGCGACGTTCCACCGTAAACTGCGTGCGGAGGTCCCGCTGCAAAACCGTCTGGAGGACAATGATCGTGCGCCGTGGCTCTTGGCGCGCCGCTCTGGCCGAATCTTATCAGCAGCAATTTACAGGAGAGCGGCACTAA
SEQ ID NO:101	50	insulator 15		TCGAGAAACAAGGCAGTTCCGGGCTGAAAGTAGCGCCGGG
SEQ ID NO:102	51	nifZ rbs	365	ACAAGTCCCGTATTATAACCGCCTAGGAGGTGTTGG
SEQ ID NO:103	52	nifZ		ATGCGCCCGAAATTCACCTTCTCTGAAGAGGTCCGCGTAGTTCGCGGATTTCGTAATGATGGCACCGTGGCGGGTTTGGCGCCAGGTGCGCTGCTGGTTCGTCGCGGTTCCGACGGGCTTTGTGCGTGA CTGGGGTGTGTTCTTGCAAGACCAGATCATCTATCAAATCCACTTCCGGAAACCGACCGCATTATCGGCTGTGCGGAGCAGGAGTTAATCCCGATTACCCAGCCGTGGTTGGCTGGTAACCTCCAGTATCGTGACAGCGTCACGTGCCAAATGGCACTGGCTGTCAACG GTGACGTGGTTGTGAGCGCCGGTCAACGTGGCCGTGTGGAGGCCACTGATCGTGGCGAACTTGGCGATTCTACACCGTGGACTTCAGCGGCCGTTGGTTCGCGGTTCCGGTCCAGGCCATCGCGCTGATTGAAGAGCGCGAAGAATAA
SEQ ID NO:104	53	insulator 16		ACGCCACGCGTAGTGAGACATAACAGTTCGTTGGGTTTAC
SEQ ID NO:105	54	nifM rbs	750	TCAGAGACTGAAGTTATTACCCAGGAGGTCTATA
SEQ ID NO:106	55	nifM		ATGAATCCGTGGCAGCGCTTTGCCCGTCAACGCCTTGCTCGCAGCCGCTGGAACCGTGATCCGGCTGCTCTCGACCCAGCCGATACCCAGCGTTCGAGCAGGCGTGGCAGCGTCAATGCCATATGGAACAAACCATCGTAGCGCGTGTCCCGGAAGGCGATATTCCGGCTGCCTTACTGGAAAACATCGCGGCCAGCCTGGCGATCTGGCTGACGAGGGTGACTTCGCTCCGCGGAGCGCGCTGCGATTGTGCGTCATCATGCACGCTCTGGAGCTGGCGTTTGCCGACATTGCCGCCAGGCACCCGCAACCGGATCTGAGCAGCGTTCAAGCGTGGTATCTGCGTCACCAGACTCAATTCATGCGTCCGGAGCAGCGCTGACCCGTCACCTGCTCCTGACGGTCGATAATGATCGCGAGCGGTGCATCAACGCATCCTTGGCCTGTATCGTCAGATCAACCGGAGCCGTGACGCTTCGCCCCACTGGCACAGCGCCACTCTCATTGCCCGTCCGCCTTGGAAGAAGGCCGTCTGGGCTGGATCTCCCGTGGTCTGCTGTACCCGCAGCTCGAAACCGCGTTGTTTAGCCTGGCGAAAACGCACTGTCGCTGCCGATTGCGTCCGGAATTGGGTGGCACCTGTTATGGTGGCAGGCGATTTCGTCGGCAGCCCCGATGGAGCCGCAACAGGCCCTTGAATCTGCGCGGACTACTTGTGGCAGCAGGCCAGCAGCGCCACCAGCGTCAATGGCTGGAGCAGATGATTTCCCGCAACCGGGCCTGTGTGGTTAA
SEQ ID NO:107	56	T7 terminator WT	1400	TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT
SEQ ID NO:108	57	Biobrick suffix (BBa_G00001)		Tactagtagcgccgctgcag
SEQ ID	58	Translational		ttagttagttag

NO:109		stop (BBa_B0042)	
SEQ ID NO:110	59	Terminator (BBa_B0054)	attagcagaaagtcaaaagcctccgaccggaggcttttgac laaacctcccttggggttatcatiggg
SEQ ID NO:111	60	VR primer sequence (BBa_G00102)	gctcactcaaagggcggtaat
SEQ ID NO:112	61	Terminator (BBa_B0062)	cagataaaaaaaaaatccttagcttgcgtaaggatgattct
SEQ ID NO:113	62	pSC101 origin (BBa_I50042)	ctgtcagaccaagttacgagctcgttggactcctgttgatagatccaglaatgacctcagaactccatc tggattgttcagaacgcctgggtgccgcggggcgtttttatgggtgagaatccaagcactaggggacagt aagacgggtaagcctgttgatgataccgctgcttactgggtgcallagccagctgaaatgacctgtcac gggataatccgaagtggtcagactggaaaatcagagggcaggaaactgctgaacagcaaaaagtcag atagccacacatagcagaccgccafaaaaagccctgagaagccctgacgggctttctgtattatg ggtagttccttgcataatccataaaaagggcctgtgtagtgccttaccgccactgaccagagccgt gagcgcagcgaactgaatgtcacgaaaagacagcgaactcaggctgctgtagtggcggagacaaaag gaatattcagcgaattgcccgagcttgcgaggggtgclactaagcctttagggttttaaggctgtttttag aggagcaaacagcgtttgcgacatectttgtaatactgcggaaactgactaaaagtagtgagtatacaca gggctgggatacttctttatctttttatctttttatctttttatctttataaaatataaccactgaataaaca aaacacacaaaaggctcagcggaaattacagaggggtctagcagaattacaagtttccagcaaaaggctc gcagaattacagatacccacaactcaaaaggaaggaatgtaattatcattgacagccatctcaatt ggtagtattgataaaatcacctagaccaattgagatgtatgtcgaattagttgtttcaaaagcaaatgaact agcagatgctgctatgacttaacggagcatgaaaccaagcgaattatgctgtgtggcactactcaacc ccagatgaaaaccctacaaggaaaggaacggagggatcgtcactataaccaatagcctcagatga tgaacatcagtagggaaaatgcttatgggtgataagcclaaagcaaccagagagctgactgacgagaactg tggaaaacagggaatccttgggttaaaggctttgagatttccagtggaacaaactatgccaagttctcaagc gaaaaatagaattagttitagtgaagagataatgccctatctttccagttaaaaaaticataaaatataat ctggacaatgttaagcttttgaaaaacaaatactctatgaggattatgagtggttattaaaagaactaacac aaaaagaactcacaaggcacaatataagagatgagccttgaatgaattaaagttatgtaalgctgaaata actaccatgagttttaaaggcttaaccaatgggttttgaaccaataagtaaaagattaaacacttacagc aatafgaaattgggtggfalaagcagggcccgactgatacgttgaatttccaagttgaactagatag acaatggatctcgtaacggaaacttgagaacaaccagataaaaatgaatgggtgacaaaataccaacaa ccattacatcagattcctaccctaacggactaagaaaacactacacagatgctttaaactgcaaaaat cagctcaccagttttagggcaaaattttgagtgacatgcaaaagtaagcatgactcaatgggtcgttca tggctcacgcaaaaacaacgaaccacactagagaacatactggcctaaatacggaaaggatctgaggctc ttatggctctgtatctatcagtgaaagcatcaagactaacaacaaaagtagaacaactgttaccgcttag atatcaaaagggaactgtccatagcagagatgaaaacgggtgtaaaaaagatagatacagagcctt ttacgagtttgggtgactttaaagctgttaccatgaacagatcgacaattgtaac
SEQ ID NO:114	63	chloramphenic ol resistance (BBa_P1004)	gttgatcgggcacgtaagagggtccaacttccaccataatgaataagatcactaccgggcgatttttgg agttatcgagatttccaggagcctaaggaaactaaaatggagaaaaaafaccgggataaccaccgtg atatafcccaatggcctcgtaaaagaactttgaggcatttcagtcagttgctcaatgtaacctafaaccaga ccgttcagctggatattacggcctttttaaagaccglaaagaaaaataagcacaagtttaccggccttta ttcacattctgcccgcctgtagaacgctcaccggagttcgtatggccatgaaagacgggtgagctgggt gatctgggatagtggtacccttggtaaccggtttccatgagcaaacigaaacgtttcctcctggagt gaataccacgacgatttcggcagtttccacatatactcgaagatgtggcgtgttacgggtgaaaacct ggcctatttccaaagggtttatgagaatatttttggctcagccaatccctgggtgagtttaccagtt ttgattaaacgtggcacaatggacaacttctgcccccggttccagatgggcaaatattatacggcaag gcaacaagggtgctgagcgcctggcagcaggttcatcatgccgtttgtgaggttccatgtcggcc gcatgcttaattgaattacaacagctactgtgtagtggcagggcggggcgtataaa
SEQ ID NO:115	64	Terminator (BBa_B0053)	tccggcaaaaaacgggcaagggtgaccaccctgcccttttctttaaaccgaaaagattacttccg tt
SEQ ID NO:116	65	VF2 primer sequence (BBa_G00101)	tgccacctgacgtctaagaa

SEQ ID NO:117	66	Terminator (BBa_B0055)	aaggaatattcagcaattgcccgtgccgaagaaaggcccaccggtgaaggtagccagtgagttgat tgctacgtaa
SEQ ID NO:118	67	Translational stop (BBa_B0042)	ttagttagttag

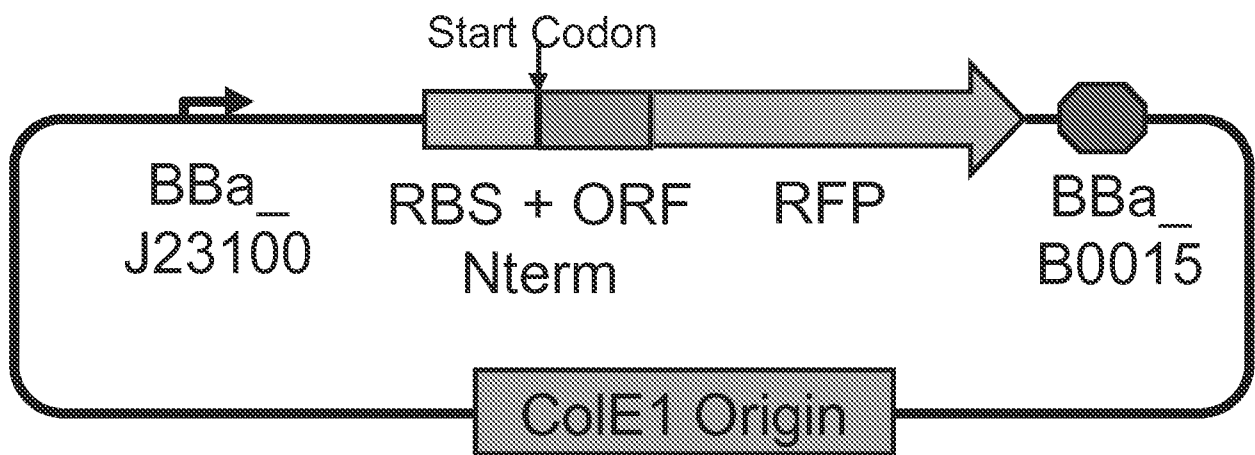
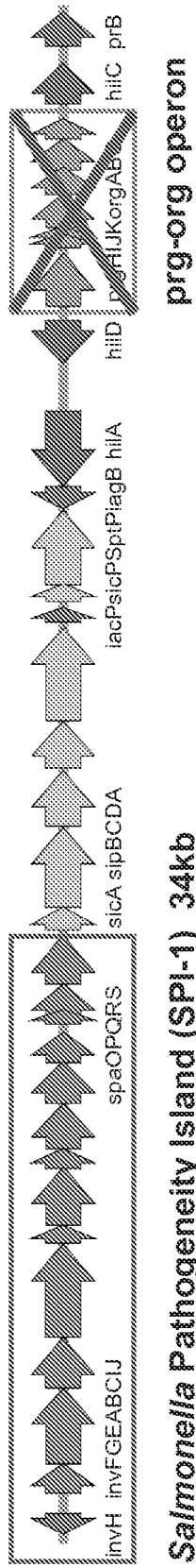


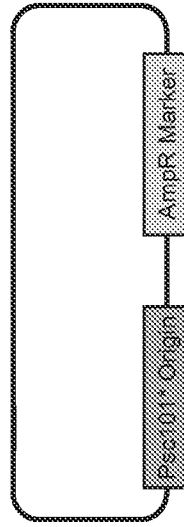
FIG. 21

A

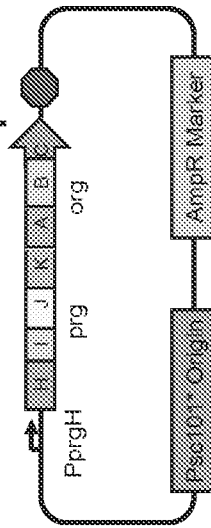


B

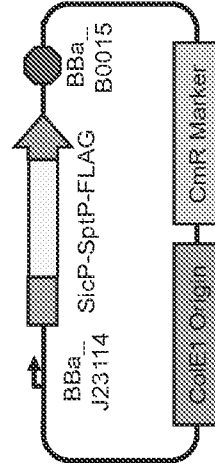
Control Setup



Test Setup



synthetic prg-org operon



reporter plasmid

FIG. 22



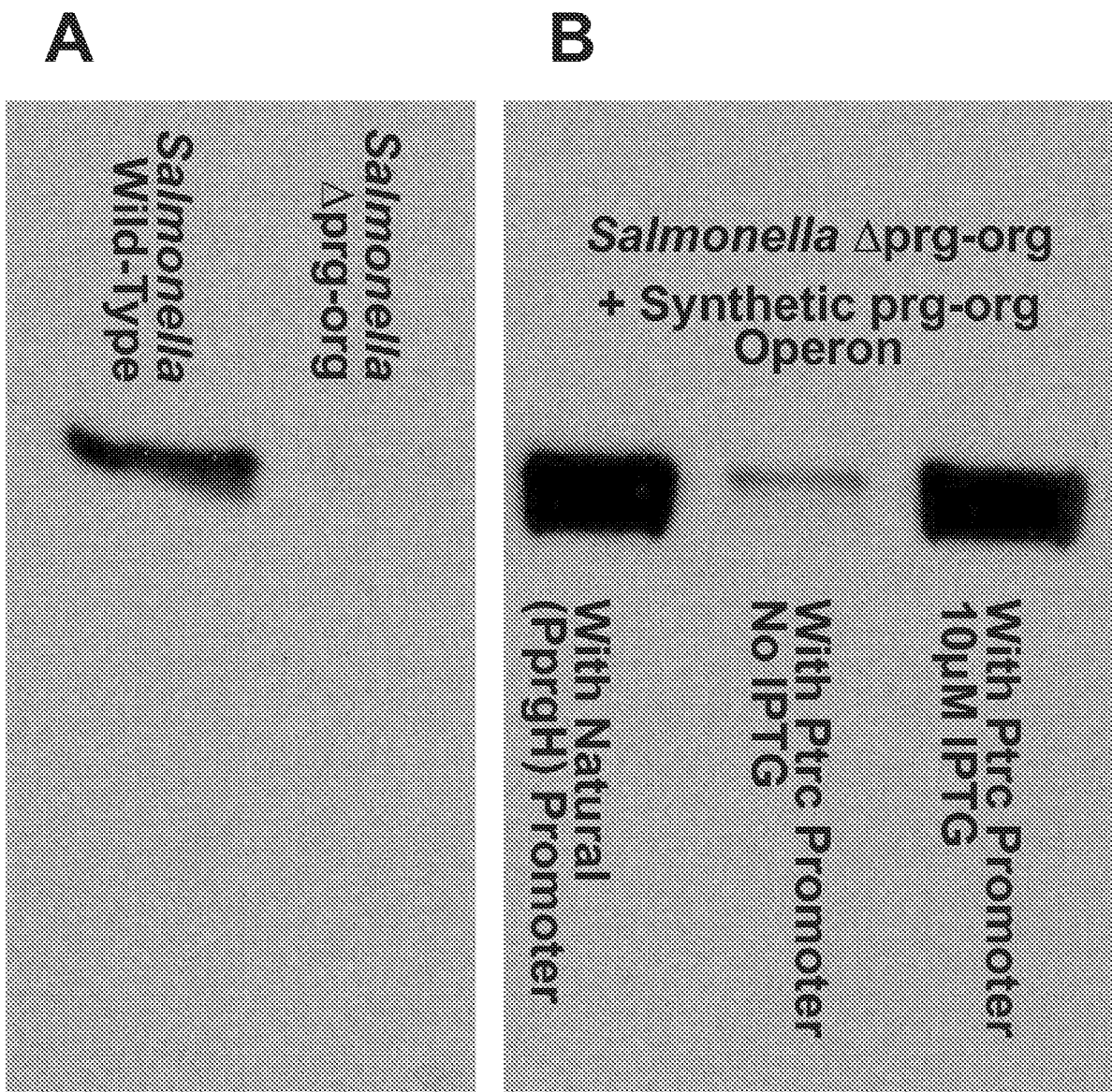


FIG. 23

Fig. 24.

SEQ ID NO:	Name	Sequence
SEQ ID NO:119	SynPrgH_RBS	ATTAAACAGGATAATAAGC
SEQ ID NO:120	SynPrgH	ATGGAAACTAGCAAGGAGAAAAACGATTACGTCCCCTGGTCCGTATATCGTT CGTCTGTTGAACTCGTCCCTGAACGGTTGTGAATTTCCGCTGCTGACTGGTC GTACGCTGTTTCGTTCGTGGGTGTCAGAGCGATGCTCTGACCGCGTCTGGTCAGCT GCCGGACATTCCCTGCCGACTCCTTCTTCATCCCGCTGGATCATGGCGGTGTT AATTTGAGATTCAAGTGGACACTGACGCGACGGAAATCATCCTGCACGAA CTGAAAGAGGGCAACTCCGAGAGCCGCTCCGTGCAACTGAACACCCCGATC CAAGTTGGTGAACCTGCTGATTTTGATTTCGTCCGGAGAGCGAGCCGTGGGTG CCGGAACAGCCGGAGAAGTTGGAAACTTCTGCGAAAAAGAACGAACCCGCG CTTTAAAAACGGCATCGTTCGCGGCACTGGCCGGTTCCTTTATCCTGGGTATC GGCACGGTTGGCACCCCTGTGGATTCTGAACTCGCCGCAACGTCAAGCAGCC GAATTGGACAGCCTGTTGGGTGTCAGGAGAAGGAGCGTTTTTCAGGTGCTGCCG GGTCCGACAAAGATGCTGTATGTTCGCGCGCAAAACGAACGCGACACCCTG TGGGCACGTCAAGTCTGGCACGCGCGGATTACGATAAAAACGCACGCGTT ATTAATGAAAATGAGGAGAATAAACGTATCAGCATCTGGCTGGACACGTAT TATCCACAACCTGGCATAATTACCGTATCCATTTTGATGAACCACGTAAGCCGG TGTTTTGGCTGTCCCCTCAACGCAACACGATGAGCAAGAAAGAGCTGGAGG TGCTGTCCCAGAAATGCGTGCCTGATGCCGTACGCCGACAGCGTCAATAT TACTCTGATGGATGACGTGACCGCAGCAGGCCAAGCCGAGGCAGGTCTGAA ACAACAGGCGCTGCCATACAGCCGCCGTAACCACAAAGGTGGTGTACGTT CGTTATTCAGGGCGCCTTGGACGACGTTGAGATTCTGCGTGCAGCCAGTTT GTCGACTCCTATTATCGTACCTGGGGTGGTTCGTTACGTTCAATTCGCAATTG AATTGAAAGACGATTGGCTGAAAGGCCGCTCGTTCCAATACGGTGCAGGAA GCTACATTAAGATGAGCCCAGGTCATTGGTATTTTCCGTCTCCTCTGTAATA G
SEQ ID NO:121	SynPrgI_RBS	ATTAAAGAGGAGAAATTAAGC
SEQ ID NO:122	SynPrgJ_RBS	ATTAAGGAGGATAAATTAAGC
SEQ ID NO:123	SynPrgJ	ATGTCTATCGCGACTATCGTGCCTGAAAATGCCGTTATTGGTCAAGCGGTGA ATATTCGCAGCATGGAACGGACATCGTCAGCTTGGACGACCGTTTGGCTGC AAGCATTTTCGGGCAGCGCCATCGCTACCGCCGTCGATAAGCAGACCATTA CCAATCGCATTGAAGACCCTAATCTGGTTACCGATCCGAAGGAGCTGGCGA TTAGCCAGGAAATGATTTCCGACTACAATCTGTACGTCAGCATGGTTAGCAC CCTGACCGGTAAGGGCGTTGGCGCTGTTGAGACTTTGCTGCGTTCCTGATAG
SEQ ID NO:124	SynPrgK_RBS	ATTAAACAGGATAATAAGC
SEQ ID NO:125	SynPrgK	ATGATCCGCCGTTACCTGTATACTTCTTGCTGGTTATGACTTTGGCCGGCTG TAAAGATAAGGATCTGCTGAAAGGCTTGGACCAAGAGCAAGCGAATGAGG TCATTGCGGTTCTGCAAAATGCACAACATTGAGGCTAACAAGATTGATAGCG GCAAGCTGGGTTACAGCATTACCGTTCGCGGAACCGGATTTACCCGCGGCGAG TGTATTGGATTAAGACCTACCAGTTGCCGCCTCGCCCCGCTGTGCAAAATCGC CCAAATGTTTCCGGCAGACAGCTGGTTAGCTCTCCGCTGCGGAGAAAGC ACGTCTGTACTCGGCGATTGAACAGCGCCTGGAGCAGTCGCTGCAAAACGAT GGAAGGTGTTCTGTGCGCCCGTGTCCACATCAGTATGATATTGATGCGGGC GAGAACGGTTCGTCCGCCTAAGCCGGTGCACCTGTCGGCTTTGGCGGTGTAT GAACGCGGTTCCCTCTGGCCATCAGATTTCCGATATTAAGCGCTTCCTGA AAAACAGCTTCGCGGATGTTGACTATGATAACATCAGCGTGGTTCTGTCCGA

SEQ ID NO:	Name	Sequence
		GCGTAGCGACGCACAGTTGCAGGCCCGGGCACGCCGTTCAAGCGCAATAG CTTCGCTACCTCCTGGATTGTGCTGATTATCCTGTTGTCTGTTATGAGCGCGG GTTTCGGTGTCTGGTACTACAAAAATCACTATGCGCGTAATAAGAAAGGCA TTACTGCCGATGACAAGGCAAAGTCCAGCAACGAGTAATAA
SEQ ID NO:126	SynOrgA_RBS	GAAAGAAGGGACAGACTAG
SEQ ID NO:127	SynOrgA	ATGATTTCGTCGTAACCGTCAAATGAACCGTCAACCACTGCCAATTATCTGGC AACGCATTATTTTCGACCCACTGTCTATATTCACCCACAACGTCTGCAAAT CGCGCCGGAGATGATCGTGGCTCCGGCAGCGCGTCCGGCAGCGAATGAGCT GATTTTGGCGGCGTGGCGTTTGAAGAACGGTGAGAAGGAGTGCATTACAGAA TAGCCTGACGCAGCTGTGGTTGCGTCAATGGCGCCGTCTGCCGAGGTTGCT TACCTGCTGGGTTGCCACAAGTTGCGTGTGACCTGGCCCGTCAAGGTGCTT TATTGGGCCCTGCCGACTGGGCGCAGGCATTCTTGCCGATGCACCAGGGTA CGTCCTTGTCTGGTTTGAATAAGGCGCCGAACCACCGTTTCTGCTGTCCGT TGGTTACGCCCAACTGAACGCGCTGAATGAGTTCCTGCCGGAGAGCTTGGC CCAACGCTTCTCTGCTGTTTCCACCGTTCATCGAGGAGGCACTGAAACAA GATGCAGTGGAGATGAGCATCCTGCTGCTGGCCCTGCAATAACGCGCAAAAG TATCCTAACACCGTCCCGGCGTTTGCCTGCTAATAA
SEQ ID NO:128	SynOrgB_RBS	GAAAGAAGGGACAGACTAG
SEQ ID NO:129	SynOrgB	ATGTTGAAAAATATCCCGATTCCATCCCCTCTGTCTCCGGTTGAAGGTATCC TGATTAACGCAAGACGTTGGAGCGTACTTCTCGATTGAGCGCCTGGAAC AACAGGCGCATCAGCGTGCAAAGCGCATTTTTCGCTGAGGCAGAAGAAGAA GCCAAGACCCTGCGCATGTATGCGTACCAGGAGGGCTACGAGCAGGGCATG ATTGACGCACTGCAACAGGTGGCCGCCTATTTGACCGACAACCAGACGATG GCTTGAAAATGGATGGAGAAAATTCAAATCTATGCGCGTGAGTTGTTACAGC GCGGCTGTGATCACCCGGAACGTTGTTGACGGTGTGACGAGTGGCTG CGTGATTTGATAAGCCGGAAGGTCAGCTGTTTTTGACCCTGCCGGTGAACG CAAAGAAAGATCATCAGAACTGATGGTGTGCTGATGGAAAATTGGCCGG GCACCTTCAATCTGAAGTATCATCAGGAGCAACGTTTTATCATGTCCTGTGG CGATCAGATTGCCGAGTTTTCCCGGAACAATTTGTTGAAACGGCGGTTGGT GTTATCAAGCACCATCTGGATGAGCTGCCTCAGGACTGTGCGACCATTTCGG ACAATGCGATTAACGCGCTGATTGATGAATGGAAGACGAAAACCCAAGCTG AAGTTATTCGCTGATAA
SEQ ID NO:130	SynInvA_RBS	CTTGGGCACGCGTCCATTAAACAGGAGTAATTAAGC
SEQ ID NO:131	SynInvA	ATGCTGCTGTCCCTGCTGAATAGCGCGCGTCTGCGTCCTGAGCTGCTGATTC TGGTTCTGATGGTTATGATCATCAGCATGTTTCGTTATCCCGTTGCCGACCTAT TTGGTTGACTTCTTGATCGCTTTGAACATTGTCTGGCAATTCTGGTGTTCAT GGGCTCCTTCTACATCGACCCGATTCTGAGCTTACGACCTTTCCGGCGGTT CTGCTGATCAGACTCTGTTCCGTTTGGCACTGAGCATCAGCACCAGCCGCC TGATCCTGATTGAAGCAGATGCGGGTGAATCATCGCGACCTTTGGTCAAGT TGTGATCGGTGACAGCCTGGCGGTTGGTTTCGTCGTATTCTCCATCGTGACG GTGGTGCAGTTTATCGTTATTACCAAGGGCAGCGAACGTTGGCGGAGGTC GCCGCTCGCTTACGCTGGACGGCATGCCGGGTAAACAGATGTCTATTGAT GCAGACCTGAAAGCCGGCATTATTGATGCTGATGCAGCGCGGAGCGCCGC AGCGTCTGGAGCGTGAAAGCCAACCTGTACGGTTCCTTCGACGGTGCATG AAGTTCATTAAGGTGATGCGATTGCGGGCATCATTATCATCTTCGTTAACT TCATTGGCGGTATCAGCGTCGGTATGACCCGTCATGGTATGGATCTGAGCAG CGCCCTGAGCACCTACACCATGCTGACGATTGGTGTGGTCTGGTTGCCAA ATCCGGCGTTGCTGATCGCGATTTCTGCGGGCTTCATCGTTACCCGCGTCA

SEQ ID NO:	Name	Sequence
		<p>ACGGTGATAGCGATAACATGGGTCGTAACATTATGACCCAGCTGCTGAATA ATCCGTTTGTCCGTGGTTGTAACGGCGATTTTGACCATCAGCATGGGCACGCT GCCGGGCTTCCGTTGCCGGTTTTTCGTTATTCTGTCTGTTGTGCTGTCCGCTCC TGTTTTACTTTAAGTTCCGCGAGGGCGAAACGTAGCGCTGCGAAAACCAAAAA CGAGCAAGGGCGAGCAACCCTTGTCCATCGAGGAGAAGGAAGGTAGCAGC CTGGGCTGATTGGCGACCTGGATAAAGTTAGCACGGAAAACCGTCCCCTG ATTTTGTGCTGGTGCCGAAATCGCGTCGTGAGGATCTGGAGAAAGCGCAGCTG GCGGAACGTCTGCGCAGCCAATTCTTTATTGATTATGGTGTGCGTCTGCCAG AAGTACTGCTGCGTGACGGTGAGGGTCTGGATGACAACCTATCGTCTCTGCT GATTAATGAGATTGCGGTTGAACAGTTTACTGTCTATTTTGACCTGATGCGT GTGGTTAACTACAGCGACGAGGTGGTGAGCTTTGGCATCAACCCGACCATT CACCAGCAAGGTTCCAGCCAGTACTTTTGGGTGACCCATGAGGAAGGCGAA AAGCTGCGCGAGCTGGGCTACGTCTGCGTAATGCACTGGACGAACTGTAC CACTGTCTGGCGGTGACGCTGGCACGCAATGTGAACGAGTATTTCCGTATCC AAGAAACGAAACACATGCTGGACCAACTGGAAGCAAAGTTTCTGACCTGC TGAAGGAGGTTTTGCGCCACGCCACCGTGCAGCGCATTTCCGGAAGTGCTGC AACGTCTGCTGTCCGAACGCGTGAGCGTCCGTAACATGAAGCTGATCATGG AAGCCCTGGCACTGTGGGCTCCGCGTGAGAAAGATGTGATCAATCTGGTGG AGCACATCCGTGGTGGATGGCGCGTTATATCTGCCACAAGTTCCGAAATG GTGGTGAACCTGCGTGCCGTTATGGTTTTCCGCCGAAGTTGAGGATGTCATTCC TAAAGGCATTTCGTCAAATTTCTGGCTCCACCTTTTTGAGCTTGGACCCGGAG GCTTCGGCAAATCTGATGGACCTGATCACGCTGAAGCTGGACGACCTGTTG ATTGCGCATAAGGACCTGGTCTGTTGACCAGCGTTGACGTGCGTCCGTTTTA TCAAGAAAATGATTGAAGGTCGTTTTCCGATCTGGAGGTCTGTCTCTCCG TGAGATTGCAGATAGCAAAAGCGTGAATGTCATCAAAACCATCTGA</p>
SEQ ID NO:132	SynInvC_RBS	CTTGGGCACGCGTCCATTAAGAGGACCAATTAAGC
SEQ ID NO:133	SynInvC	<p>ATGAAAACCCACGTCTGCTGCAATACCTGGCCTACCCGCAGAAAATCACT GGCCCTATCATTGAAGCAGAAGTGCCTGATGTTGCAATTGGTGAATTGTGCG AGATCCGTCGCGGCTGGCACCAGAAGCAGGTTGTGGCCCGTGCAGCAAGTGG TTGGTTTTGAGCGCGAACGTACCGTCTGAGCCTGATCGGCAATGCCAAG GCCTGAGCCGTGATGTGGTCTTGTACCCGACCGCCGTGCTCTGAGCGCGTG GGTTGGTTACAGCGTTCTGGGCGCAGTACTGGACCCGACGGGTAAAATCGT TGAACGTTTCACCCCGGAAGTCGCACCGATTTCCGAGGAGCGCGTTATCGA CGTGGCACCGCCGAGCTACGCAAGCCGTGTCCGGTGTGCGCGAACCGCTGAT CACGGGTGTCCGCGCAATTGATGGTCTGCTGACGTGTGGTGTGGGCCAGCG TATGGGTATTTTCGCAAGCGCGGGTTGTGGTAAGACCATGTTGATGCACATG CTGATCGAGCAAACCGAAGCGGATGTCTTTGTTATTGGCCTGATTGGCGAGC GTGGTCTGTAAGTTACCGAATTTGTAGACATGTTGCGTGCATCTCATAAGAA AGAAAAATGTGTGCTGGTTTTTGCCACGTCCGACTTCCCAAGCGTTGACCGC TGCAACGCTGCCAGCTGGCAACGACGCTCGCCGAGTATTTCCGCGACCAG GGTAAACGTGTTGTCTGTTTATCGACAGCATGACCCGTTATGCACGCGCGT TGCGTGATGTCGCGCTGGCGAGCGGCGAGCGTCCGGCTCGCCGTGGTTATC CGGCGTCTGTGTTTCGACAATCTGCCGCGTTTTGCTGGAGCGTCCGGGTGCGAC GAGCGAGGGTAGCAATACCGCCTTCTATAACCGTCTGCTGGAGTCGGAAGA AGAAGCGGACCCGATGGCGGACGAGATCCGTTCTATTCTGGACGGTCACCT GTACCTGTCCCGCAAACCTGGCGGGTCAAGGTCATTACCCGGCTATCGATGTG CTGAAGAGCGTGAGCCGTGTGTTTGGTCAAGTGACCACCCGACTCACCGC GAGCAAGCGAGCGCGGTCGTAAGCTGATGACCCGTCTGGAAGAGCTGCAA CTGTTCAATTGACCTGGCGAGTATCGTCCGGGCGAGAACATTGACAATGAT CGTGCGATGCAAATGCGCGATAGCCTGAAGGCGTGGCTGTGTGACCGCTGT GCGCAATACAGCAGCTTCGATGATACGCTGTCCGGCATGAACGCCTTTGCG GATCAGAACTGA</p>

SEQ ID NO:	Name	Sequence
SEQ ID NO:134	SynInvE_RBS	TACTTGGGCACGCGTCCATTAATTAGGATCAATAGC
SEQ ID NO:135	SynInvE	ATGATTCCGGGCAGCACCTCCGGTATTTCCFTTAGCCGTATCCTGAGCCGTC AGACCTCCCACCAGGATGCAACCCAGCATAACCGACGCACAACAAGCGGAA ATTCAACAAGCGGCGGAAGATAGCTCGCCGGGTGCGGAGGTTTCAGAAATTC GTCCAGAGCACGGACGAGATGTCTGCTGCGTTGGCGCAGTTCOCGCAATOCG CGTGACTATGAGAAAAAGAGCAGCAATTTGTCTAACTCCTTCGAGCGCGTT CTGGAGGACGAGGCACTGCCGAAAAGCGAAGCAGATTCTGAAACTGATCAGC GTGCATGGCCGGTGCCTGGAGGATTTCTGCGTCAGGCGCGCAGCCTGTTC CGGACCCAAGCGATCTGGTGTGGTTCTGCGCGAGCTGTTGCGTCGTAAGG ACCTGGAGGAGATCGTGCCTAAGAAGCTGGAGAGCCTGCTGAAGCACGTGG AGGAACAAACCGACCCGAAAACCTGAAGGCCGGTATTAAGTGCAGCGTGA AGGCGCGTCTGTTTGGCAAGACGCTGTCTCTGAAACCTGGTCTGCTGCGTGC CAGCTACCGCAGTTCATCCAAAGCGAAAAGCCAGAAGTCGAGATTTACAG CGATTGGATCGCCAGCTACGGTTATCAGCGTCCGCTGGTTGTTCTGAGATTC ATTGAAGGCAGCCTGCTGACTGACATCGATGCTAACGATGCAAGCTGCTCC CGTCTGGAGTTTGGCCAACTGCTGCGCCGCTGACCCAGCTGAAAATGTTGC GTAGCGCCGACCTGCTGTTTGTCTCGACGTTGCTGTCTTACAGCTTCACGAA AGCATTTAACGCTGAGGAGAGCAGCTGGCTGTTGCTGATGCTGTCTTTGCTG CAACAGCCGCACGAAGTGGATAGCCTGCTGGCGGACATTATCGGTCTGAAT GCGCTGCTGTTGTCCACAAAGAGCAGCCAGCTTCCTGCAAATCTTCTATC AGGTCTGTAAGGCAATCCCGTCTAGCCTGTTTATGAAGAGTACTGGCAAG AAGAACTGCTGATGGCACTGCGCTCCATGACGGACATTGCTTACAAACACG AAATGGCCGAACAACGTCGTACCATCGAAAAGCTGTCTAA
SEQ ID NO:136	SynInvF_RBS	CTTGGGCACGCGTCCATTAATAAAGGAGTAATTAAGC
SEQ ID NO:137	SynInvF	ATGAGCTTCAGCGAGAGCCGCCACAATGAAAAGTGTCTGATTCAAGAAGGC GCACTGCTGTTTTGTGAGCAAGCAGTCGTGGCGCCTGTCAGCGGTGATCTGG TTTTTCGTCCGCTGAAAATCGAGGTCCCTGAGCAAGCTGCTGGCGTTCATCGA CGGCGCAGGTCTGGTGGATACGACCTACGCGGAGTCGGACAAATGGGTTCT GCTGTCTCCGAGTTCGCTGCTATTTGGCAAGACCGTAAACGTTGCGAATAT TGGTTTTTGCAGCAGATTATACCCCATCTCCGGCGTTCAACAAGGTTCTGG CACTGTTGCGTAAGAGCGAAAAGCTATTGGTTGGTGGCTACTTGCTGGCCCA AAGCACCGCAATACTATGCGTATGTTGGGTGAGGATTACGGTGTTAG CTACACGCATTTCCGCCGTCTGTGCAGCCGCGCTCTGGGCGGTAAGGCGAA AAGCGAGCTGCGCAATTGGCGCATGGCCCAGTCCCTGCTGAATAGCGTGGA AGGTCATGAAAACATCACCCAGCTGGCGGTCAACCACGGTTATAGCAGCCC GTCCCCTTTAGCTCTGAAATCAAGGAGCTGATTGGTGTTCCTCCCGGTAAG CTGTCTAACATCATTACAGCTGGCCGACAAATGA
SEQ ID NO:138	SynInvG_RBS	TGGGCACGCGTCCATTAATGAGGAAAAATTATTAGC
SEQ ID NO:139	SynInvG	ATGAAAACGCACATTCTGTTGGCCCGTGTGCTGGCTTGCAGCAGCTCTGGTGC TGGTCAACCCAGGTTATAGCTCCGAGAAGATCCCGGTTACGGGCAGCGGCT TCGTTGCAAAGGACGATTCTCTGCGCACCTTTTTCGATGCGATGGCACTGCA ATTGAAGGAGCCGGTGATTGTGAGCAAGATGGCGGCTCGCAAAAAGATTAC CGGCAATTCGAGTTCACGATCCAAACGCGCTGCTGGAGAAACTGTCCCT GCAACTGGGTCTGATCTGGTACTTTGATGGTCAGGCGATCTACATCTACGAC GCGAGCGAAATGCGTAATGCGGTTGTGAGCCTGCGTAATGTCAGCCTGAAC GAGTTCAACAATTTCTGAAGCGCAGCGGCTGTACAATAAGAACTACCCT CTGCGTGGTGATAATCGTAAAGGCCACCTTCTATGTCAGCGGTCCGCGGTTG ATGTTGATATGGTTGTAATGCGGCCACCATGATGGACAAACAGAATGATG

SEQ ID NO:	Name	Sequence
		GCATCGAGCTGGGTCGCCAAAAGATCGGTGTTATGCGCCTGAACAACACTT TTGTGGGCGACCGCACCTACAACCTGCGTGATCAAAAAGATGGTCATTCCGG GTATTGCTACGGCAATTGAACGCCTGTTGCAAGGCGAAGAACAACCGCTGG GTAACATTGTAAGCTCCGAGCCTCCGGCCATGCCGGCCTTTAGCGCAAACG GCGAGAAAGGTAAGGCAGCGAATTACGCGGGTGGTATGAGCCTGCAAGAA GCGCTGAAACAGAACGCAGCGGCAGGCAACATCAAAATTGTGGCCTATCCG GACACCAACAGCCTGCTGGTGAAAGGTACGGCGGAGCAGGTGCATTTTCATC GAGATGCTGGTTAAAGCCCTGGACGTGGCGAAACGTACGTTGAATTGAGC CTGTGGATTGTGGATTTGAATAAGAGCGACCTGGAACGCTTGGGCACCAGC TGGAGCGGTAGCATCACCATCGGCGACAAGCTGGGTGTGAGCCTGAACCAG AGCAGCATCTCCACGCTGGACGGTAGCCGCTTTATTGCGGCGGTCAACGCTC TGGAGGAAAAGAAACAGGCCACTGTTGTCAGCCGTCCGGTTCTGCTGACCC AGGAGAATGTCCCGGCGATTTTGGACAATAATCGTACTTTCTATAACCAAAC GATCGGTGAACGTAATGTTGCATTGGAACACGTGACCTATGGCACCATGAT CCGTGTCCTGCCGCGTTTCAGCGCGGACGGCCAGATTGAGATGAGCCTGGA CATCGAAGATGGTAATGACAAAACCCCGCAGTCTGATACGACCACCTCCGT TGATGCGCTGCCAGAAGTGGGTTCGTACCCCTGATCTCGACGATGACAGGTGC CCGCATGGTAAATCTCTGCTGGTTGGTGGCTACACGCGTGATGCAAACACG GACACGGTCCAAAGCATCCCGTTCCTGGGTAAGCTGCCGCTGATTGGCTCGT TGTTTCGCTACAGCAGCAAAAACAAGTCTAATGTCGTCCGTGTCTTTATGAT TGAGCCGAAAGAAATCGTTGACCCGCTGACCCCGGATGCCAGCGAGAGCGT TAACAACATTCTGAAACAGTCCGGTCCGTGGAGCGGTGACGATAAGCTGCA AAAGTGGGTTTCGTGTGTATTTGGACCGTGGTCAGGAGGCCATTAAGTAA
SEQ ID NO:140	SynInvI_RBS	CTTGGGCACGCGTCCATTA AAAAAGGACCAATTAAGC
SEQ ID NO:141	SynInvI	ATGCACTCTCTGACTCGTATCAAGGTCCTGCAACGTCGTTGTACCGTGTCC ATTCTCAGTGCGAGTCCATTCTGTTGCGTTATCAGGACGAAGATCGCGGCTT GCAGGCGGAGGAGGAGGCCATCCTGGAACAGATCGCAGGTCTGAAACTGCT GTTGGACACCCCTGCGTGTGAAAATCGTCAACTGACCCGTGAAGAACTCTA TACGCTGCTGCGCAAACAGAGCATTGTTTCGTCGCCAGATTAAGGATCTGGA GCTGCAAATCATTAGATTCAAGAAAAGCGTAGCGAGCTGGAAAAGAAAC GTGAGGAATTTCAAAGAAAAGCAAATACTGGCTGCGTAAAGAAGGTAAC ACCAGCGCTGGATTATCCGTCAAAAACGCTTCTACATTCAACGTGAGATCCA GCAAGAGGAGGCGGAGAGCGAAGAGATCATTTAA
SEQ ID NO:142	SynInvJ_RBS	AAAAAAAAGCTCTAAAAGATTAAGAGGGGGTAAACAT
SEQ ID NO:143	SynInvJ	ATGGGTGACGTGAGCGCGGTGAGCAGCAGCGGTAACATTCTGCTGCCGCGAG CAGGACGAGGTTGGTGGCTGTCCGAAGCGCTGAAGAAAGCGGTTGAAAA ACACAAAACCGAATACAGCGGTGACAAGAAAGATCGTGATTATGGTGACGCG CTTTGTTATGCACAAGGAAACCGCGCTGCCGTTGTTGCTGGCAGCTTGGCGC CACGGCGCACCGGCGAAAAGCGAGCACCATAACGGTAACGTAAGCGGTCT GCATCACAACGGTAAGAGCGAGCTGCGTATTGCTGAGAACTGCTGAAGGT GACGGCGGAGAAGAGCGTTGGTCTGATTAGCGCTGAAGCGAAGGTGGATA AATCTGCGGCGCTGCTGTCTAGCAAGAATCGTCCGCTGGAATCGGTGACGCG GCAAAAAGTTGTCCGCCGATCTGAAAGCAGTGGAGTCCGTGTCCGAGGTCA CGGATAACGCCACCGGCATTTCCGGATGACAACATCAAAGCATTGCCGGGTG ACAATAAGGCCATCGCCGGTGAAGGTTGTGCGTAAAGAAGGTGCGCCGCTGG CGCGTGACGTGGCTCCGGCACGCGATGGCGGCAGCAAATACGGGCAAGCCGG AGGATAAAGACCACAAGAAGGTCAAGGACGTTAGCCAGCTGCCGCTGCAA CCGACTACCATCGCCGATCTGTCTCAACTGACGGGTGGCGATGAAAAGATG CCGCTGGCAGCGCAGTCCAAAACCGATGATGACCATTTTCCAACCGCGGAC GGCGTTAAAGGTGAGGACAGCTCTCTGACCTATCGTTCCAGCGCTGGGGC

SEQ ID NO:	Name	Sequence
		AACGATTACTCCGTCAATATCCAGGCACGCCAAGCGGGCGAATTTAGCCTG ATTCCTAGCAATACCCAGGTTGAACATCGTCTGCACGACCAATGGCAAAAT GGCAATCCACAACGCTGGCATTGACGCGTGATGACCAGCAAAAACCGCAA CAGCAACAGCATCGTCAGCAGTCCGGTGAAGAGGACGACGCGTAA
SEQ ID NO:144	SynSpaO_RBS	CTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG
SEQ ID NO:145	SynSpaO	ATGAGCTTGCGCGTACGCCAAATTGACCGCCGTGAATGGCTGCTGGCCCAG ACTGCGACCGAATGTCAGCGTCACGGTCGTGAGGCCACCTGGAGTATCCG ACCCGTCAGGGTATGTGGGTCCGCCTGTCTGATGCCGAAAAACGCTGGTCT GCGTGGATCAAACCGGGTGATTGGTTGGAACACGTTAGCCAGCACTGGCG GGTGCCGCGGTCAGCGCAGGCGCGGAGCACCTGGTGGTTCGCTGGCTGGCG GCAACCGAACGCCCGTTCGAGCTGCCGGTCCCGCACTTGAGCTGCCGTGCT TGTGCGTGGAGAACCCGGTTCGGGGTTCGCACTGCCGTGAGGGCAAGCTGC TGCACATCATGTCGGATCGTGGTGGCCTGTGGTTTGAGCATCTGCCGGAGCT GCCGGCTGTTGGCGGTGGCCGTCCGAAGATGCTGCGTTGGCCGCTGCGTTTC GTTATTGGCAGCAGCGACACCCAGCGCAGCCTGCTGGGTGCTATCGGTATC GGTGATGTTCTGCTGATCCGCACCAGCCGTGCTGAGGTTTACTGTTACGCGA AGAAACTGGGCCACTTTAATCGTGTGGAAGGTGGCATTATTGTCGAAACGC TGGACATTCAACATATCGAGGAGGAGAACAACACGACGGAAACGGCGGAA ACCCTGCCGGGTCTGAATCAACTGCCGGTGAAGCTGGAGTTCGTTCTGTATC GTAAAAACGTGACGTTGGCCGAACTGGAAGCAATGGGTGAGCAACAACCTGC TGTCCTTGCCAACCAATGCGGAACTGAACGTGCGAAATCATGGCGAATGGCG TGCTGCTGGGTAACGGCGAACTGGTGCAGATGAATGACACCCTGGGTGTGCG AGATTCATGAGTGGTTGTCCGAGAGCGGTAATGGCCGAGTAG
SEQ ID NO:146	SynSpaP_RBS	CTTGGGCACGCGTCCATGAAAGAAACGACATACTAG
SEQ ID NO:147	SynSpaP	ATGGGTAACGATATTAGCTTGATTGCATTGCTGGCGTTTTCCACCCTGCTGC CGTTCATCATCGCGTCTGGTACTTGGCTTCGTCAAATTCAGCATCGTCTTTGTG ATGGTGCACAACCGCGTGGGTCTGCAACAAATTCGAAGCAATATGACCCCTG AATGGCGTCGCACTGCTGCTGTCGATGTTTGTATGTGGCCGATTATGCACG ACGCGTATGTGTATTCGAGGATGAAGATGTGACCTTAAACGACATCTCCAG CCTGAGCAAGCATGTTGATGAGGGCCCTGGACGGTTATCGCGACTACCTGAT CAAGTATTCCGACCGTGAGCTGGTGCAGTTCTTTGAGAATGCCAGTTGAAA CGTCAGTACGGTGAAGAAACGGAAACCGTTAAACGTGACAAGGACGAGAT TGAAAAGCCGAGCATTTTCGCACTGTTGCCGTACCGCTTGAGCGAGATT AAGAGCGCATTCAAAAATTGGTTTTTACCTGTACCTGCCGTTCTGTTGTGGTGC ATCTGGTTGTCTCCAGCGTTCTGCTGGCCCTGGCATGATGATGATGTGTC GGTTACCATCAGCAGCCGATCAAACCTGGTCTGTTTGTGGCCCTGGATGGT TGGACGCTGCTGTCTAAAGGCTGATCCTGCAATACATGGACATCGCGACCT AA
SEQ ID NO:148	SynSpaQ_RBS	CTTGGGCACGCGTCCATTAAGAAGGAGGAATTAAGC
SEQ ID NO:149	SynSpaQ	ATGGACGATCTGGTTTTCGCCGGCAACAAAGCCCTGTACTTGGTGTGCTGATTC TGTCCGGTTGGCCGACGATTGTCGCAACCATTATCGGTCTGCTGGTGGTCT GTTTCAAACCGTGACGCAAGTTGCAGGAGCAAACCTGCCGTTCCGGTATCAA GCTGCTGGGTGTGTGCTGTGCCTGTTTTGCTGCTGGCTGGTATGGCGAA GTTCTGCTGTCTACGGCCGTCAGGTCATCTTCTGGCTCTGGCGAAAGGTT AA
SEQ ID NO:150	SynSpaR_RBS	CTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG

SEQ ID NO:	Name	Sequence
SEQ ID NO:151	SynSpaR	ATGTTCTATGCATTGTATTTGAGATCCACCATCTGGTGGCGTCCGCGGCTC TGGGTTTTGCGCGTGTGCTCCGATCTTTTTCTTTCTGCCGTTCTGAATAGC GGTGTCTGAGCGGTGCTCCGCGCAACGCGATCATCATTCTGGTGGCGCTGG GTGTATGGCCGCACGCCCTGAATGAGGCGCCACCGTTTCTGTCTGTGGCAAT GATTCCGCTGGTCTGCAAGAGGCGAGCCGTGGGTGTTATGCTGGGTTGCCTG TTGTCCTGGCCGTTTTGGGTTATGCACGCGTTGGGCTGTATCATTGATAACC AACCGGTGCAACCCTGTCCAGCAGCATTGATCCTGCGAATGGCATCGACA CCAGCGAGATGGCCAATTTCTGAACATGTTCCGCGGCTGTCTGTATTTGCA GAACGGTGGCTTGGTCACGATGGTGGACGTGCTGAATAAGTCTTACCAGCT GTGTGATCCGATGAACGAGTGTACGCCGAGCCTGCCTCCGTTGCTGACCTTC ATTAATCAAGTGGCCCAGAACGCCTGGTGTGCGCTCCCGGTCGTTCTGG TTCTGCTGCTGTCGGAAGTTTTCTGGGCTGCTGTCTCGTTTTGCACCGCAA ATGAACGCGTTCCGCAATTAGCCTGACTGTTAAAAGCGGTATTGCGTTTTGA TCATGCTGCTGTATTTAGCCCCGGTCTGCCGGACAATGTTCTGCGTCTGAG CTTTCAGGCGACCGGCTGAGCAGCTGGTCTACGAACGTGGCGCAACGCA TGTGCTGGAATAA
SEQ ID NO:152	SynSpaS_RBS	CTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG
SEQ ID NO:153	SynSpaS	ATGTCCAGCAACAAAACCGAAAAACCGACTAAGAAAACGTCTGGAGGATAG CGCAAAAGAAAAGTCAAGAGCTTCAAGAGCAAGGACCTGATTATCGCGTGCCT GACCCTGGGTGGTATCGCTTATTTGGTGAGCTACGGCAGCTTCAATGAGTTT ATGGGTATCATTAAAGATTATCATCGCTGATAACTTTGATCAGTCGATGGCAG ATTATAGCCTGGCCGTGTTTGGTATTTGGCCTGAAATACCTGATTCCGTTTCA GCTGCTGTGTTTGGTTTGTTCGCACTGCCGGCACTGCTGCAAGCGGGCTTC GTTCTGGCAACCGAGGCCCTGAAGCCGAATCTGTCCGCCCTGAACCCGTT GAAGGCGCAAGAAAAGTGTTCATGCGCACCGTCAAAGACACGGTCAAG ACGCTGCTGTATCTGTGAGCTTTGTGGTTGCGGCAATCATTGTGTTGAAAA AGTATAAAGTCGAGATTTTCAGCCAACCTGAACGGTAAATATCGTGGGTATG CGTTATCTGGCGTGAATTTGCTGCTGGCGTTGGTTCTGACCTGTCTGGCGTG CGCGCTGATCGTGTGCTGTTGCTGGATGCTATTGCCGAGTACTTTCTGACCATG AAAGATATGAAGATGGACAAAGAAGAAGTTAAACGCGAGATGAAAGAGCA GGAGGGTAACCCGGAGGTGAAGTCTAAACGTCGTGAAGTCCACATGGAAAT CCTGAGCGAACAAGTCAAGTCTGACATTGAAAATAGCCGCTCTGATTGTGGC AAACCCTACGCATATTACCATCGGCATCTACTTCAAACCGGAACTGATGCCG ATTCCAATGATTAGCGTCTATGAAACCAATCAACGCGCGCTGGCGGTCCGT GCGTACGCCGAGAAAAGTGGGTGTTCCGGTTATTGTAGACATCAAGCTGGCG CGCAGCCTGTTCAAACGCAACCGTCTGTTACGACCTGGTGAACCTGGAGGAG ATCGACGAGTTTTGCGCCTGCTGGTTTTGGTTGGAAGAGGTGCAAAACGCA GGCAAGGATGTGATCCAACCGCAGGAGAATGAAGTGCCTCATTAA
SEQ ID NO:154	Syn_PrgOrg_Operon	GCATGCATTAAGAGGAGAAATTAAGCATGGCTACTCCGTGGTCTGGTTAT CTGGATGATGTTTTCTGCAAAATTTGACACGGGTGTTGACAACCTGCAAACCC AAGTTACCGAAGCCCTGGACAAGCTGGCTGCGAAGCCGTCGGATCCGATCCGCGC TGCTGGCGGCGTATCAATCCAAACTGTCTGAGTATAACTTGTACCGTAATGC GCAGTCGAACACGGTCAAGGTCTTCAAAGATATTGATGCAGCGATTATTCA GAACCTTCGTTAGTAACCTAGGATTAAGGAGGATAAATTAAGCATGTCTATC GCGACTATCGTGCCTGAAAATGCCGTTATTGGTCAAGCGGTGAATATTCCGA GCATGGAAACGGACATCGTCAGCTTGGACGACCGTTTGGTCAAGCATTTC GGGCAGCGCCATCGCTACCGCCGTCGATAAGCAGACCATTACCAATCGCAT TGAAGACCCTAATCTGGTTACCGATCCGAAGGAGCTGGCGATTAGCCAGGA AATGATTTCCGACTACAATCTGTACGTCAGCATGGTTAGCACCTGACCGCT AAGGGCGTTGGCGCTGTTGAGACTTTGCTGCGTTCTGATAGGCTAGCATTAA AACAGGATAATAAGCATGGAAACTAGCAAGGAGAAAACGATTACGTCCCTT GGTCCGTATATCGTTTCGTTCTGTTGAACTCGTCCCTGAACGGTTGTGAATTC

SEQ ID NO:	Name	Sequence
		<p>CGCTGCTGACTGGTCTACGCTGTTTCGTCGTTGGGTCAGAGCGATGCTCTGAC CGCGTCTGGTCAGCTGCCGGACATTCCTGCCGACTCCTTCTTCATCCCCTG GATCATGGCGGTGTTAATTTTCGAGATTCAAGTGGACACTGACGCGACGGAA ATCATCCTGCACGAACTGAAAAGAGGGCAACTCCGAGAGCCGCTCCGTGCAA CTGAACACCCCGATCCAAGTTGGTGAAC'TGCTGATTTTGATTTCGTCGGAGA GCGAGCCGTGGGTGCCGGAACAGCCGGAGAAGTTGGAAACTTCTGCGAAA AAGAACGAACCGCGCTTTAAAAACGGCATCGTCGCGGCACTGGCGGGTTTC TTTATCCTGGGTATCGGCACGGTTGGCACCCCTGTGGATTCTGAACTCGCCGC AACGTCAAGCAGCCGAATTGGACAGCCTGTTGGGTCAGGAGAAGGAGCGTT TTCAGGTGCTGCCGGGTCGCGACAAGATGCTGTATGTCGCCGCGCAAAAACG AACGCGACACCCTGTGGGCACGTCAAGTCCTGGCACGCGCGGATTACGATA AAAACGCACGCGTTATTAATGAAAATGAGGAGAATAAACGTATCAGCATCT GGCTGGACACGTATTATCCACAACCTGGCATATTACCGTATCCATTTGATGA ACCACGTAAGCCGGTGTFTTGGCTGTCCCGTCAACGCAACACGATGAGCAA GAAAGAGCTGGAGGTGCTGTCCCAGAAATTGCGTGCGCTGATGCCGTACGC CGACAGCGTCAATATTACTCTGATGGATGACGTGACCGCAGCAGGCCAAGC CGAGGCAGGTCTGAAACAACAGGCGCTGCCATACAGCCGCCGTAACCACAA AGGTGGTGTACGTTTCGTTATTCAGGGCGCCTTGGACGACGTTGAGATTCTG CGTGCGCGCCAGTTTGTGACTCCTATTATCGTACCTGGGGTGGTTCGTTACG TTCAATTCGCAATTGAATTGAAAGACGATTGGCTGAAAGGCCGCTCGTTCCA ATACGGTGCAGGAGGCTACATTAAGATGAGCCAGGTCATTGGTATTTTCC GTCTCCTCTGTAATAGAAGCTTATTAACAGGATAATAAGCATGATCCGCGG TTACCTGTATACCTTCTTGTGTTATGACTTTGGCCGGCTGTAAGATAAGC GATCTGCTGAAAGGCTTGGACCAAGAGCAAGCGAATGAGGTCAATTGCGGTT CTGCAAATGCACAACATTGAGGCTAACAAGATTGATAGCGGCAAGCTGGGT TACAGCATTACCGTCGCGGAACCGGATTCACCGCGGCAGTGTATTGGATTA AGACCTACCAGTTGCCGCCTCGCCCGCTGTCGAAATCGCCCAAATGTTTCC GGCAGACAGCCTGGTTAGCTCTCCGCGTGGGAGAAAGCACGTCTGTACTC GGCGATTGAACAGCGCCTGGAGCAGTCGCTGCAAACGATGGAAGGTGTTCT GTCGGCCCGTGTCCACATCAGCTATGATATTGATGCGGGCGAGAACGGTTCG TCCGCCTAAGCCGGTGCACCTGTGCGCTTTGGCGGTGTATGAACGCGTTCC CCTCTGGCCCATCAGATTTTCGGATATTAAGCGCTTCTGAAAAACAGCTTCG CGGATGTTGACTATGATAACATCAGCGTGGTTCTGTCCGAGCGTAGCGACG CACAGTTGCAGGCGCCGGGCACGCCGGTCAAGCGCAATAGCTTCGCTACCT CCTGGATTGTGCTGATTATCCTGTTGTCTGTTATGAGCGCGGGTTTCGGTGT TGGTACTACAAAATCACTATGCGCGTAATAAGAAAGGCATTACTGCCGAT GACAAAGCAAAGTCCAGCAACGAGTAATAAGGTACCGAAAGAAGGGACAG ACTAGATGATTTCGTCGTAACCGTCAAATGAACCGTCAACCACTGCCAATTAT CTGGCAAACGCATTATTTTCGACCCACTGTCTATATTACCCACAACGTTCTG CAAATCGCGCCGGAGATGATCGTTCGTCGCGGACGCGGTGCGGCAGCGAAT GAGCTGATTTTGGCGGCGTGGCGTTTGAAGAACGGTGAGAAGGAGTGCATT CAGAATAGCCTGACGCAGCTGTGGTTGCGTCAATGGCGCCGTCTGCCGAG GTTGCTTACCTGCTGGGTTGCCACAAGTTGCGTGCTGACCTGGCCCCGTCAAG GTGCTTTATTGGGCTGCCGACTGGGCGCAGGCATTCTTGGCGATGCACCA GGGTACGTCCTTGTCCGTTTGTAAATAAGGCGCCGAACACCGTTTCTGCTG TCCGTTGGTTACGCCCAACTGAACGCGTGAATGAGTTCTTCCCGGAGAGCT TGGCCCAACGCTTTCCTCTGCTGTTTCCACCGTTTCATCGAGGAGGACTGAA ACAAGATGCAAGTGGAGATGAGCATCCTGCTGCTGGCCCTGCAATACGCGCA AAAGTATCCTAACACCGTCCCGGCGTTTGGGTGCTAATAAAGATCTGAAAG AAGGGACAGACTAGATGTTGAAAAATATCCCGATTCCATCCCTCTGTCTCC GGTTGAAGGTATCTGATTAAACGCAAGACGTTGGAGCGTTACTTCTCGATT GAGCGCTGGAACAACAGGCGCATCAGCGTGCAAAGCGCATTTTTCGCTGAG GCAGAAGAAGAAGCAAGACCCTGCGCATGTATGCGTACCAGGAGGGGCTA CGAGCAGGGCATGATTGACGCACTGCAACAGGTGGCCGCCCTATTTGACCGA CAACCAGACGATGGCTTGGAAATGGATGGAGAAAATTCAAATCTATGCGCG TGAGTTGTTACGCGCGGCTGTGATCACCCGAAACGTTGTTGACGGTGTCT</p>

SEQ ID NO:	Name	Sequence
		GACGAGTGGCTGCGTGATTTTCGATAAGCCGGAAGGTCAGCTGTTTTTGACC TGCCGGTGAACGCAAAGAAAGATCATCAGAACTGATGGTGCTGCTGATGG AAAATTGGCCGGGCACCTTCAATCTGAAGTATCATCAGGAGCAACGTTTTAT CATGTCCTGTGGCGATCAGATTGCCGAGTTTTCGCCGGAACAATTTGTTGAA ACGGCGGTTGGTGTATCAAGCACCATCTGGATGAGCTGCCTCAGGACTGTC GCACCATTTCCGACAATGCGATTAACGCGCTGATTGATGAATGGAAGACGA AAACCCAAGCTGAAGTTATTCGCTGATAAGGATCCGGCATCAAATAAAACG AAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTTCGGTG AACGCTCTCCTGAGTAGGACAAATCCGCCGCCCTAGAAGTACTAGTccataagaatGC GGCCGC
SEQ ID NO:155	Syn_InvSpa_Operon	CTCGAGAAAAAAAAGCTCTAAAAGATTAAGAGGGGGTAACATATGGGTGA CGTGAGCGCGGTGAGCAGCAGCGGTAACATTCTGCTGCCGAGCAGGACGA GGTTGGTGGCCTGTCCGAAGCGTGAAGAAAGCGGTTGAAAAACACAAAAAC CGAATACAGCGGTGACAAGAAAGATCGTGATTATGGTGACGCCTTTGTTAT GCACAAGGAAACCAGCGCTGCCGTTGTTGCTGGCAGCTTGGCCACGGCCGC ACCGGCGAAAAGCGAGCACCATAACGTAACGTAAGCGGTCTGCATCACAA CGGTAAGAGCGAGCTGCGTATTGCTGAGAACTGCTGAAGGTGACGGCGGA GAAGAGCGTTGGTCTGATTAGCGCTGAAGCGAAGGTGGATAAAATCTGCGGC GCTGCTGTCTAGCAAGAATCGTCCGCTGGAATCGGTCAGCGGCAAAAAGTT GTCCGCCGATCTGAAAGCAGTGGAGTCCGTGTCCGAGGTCACGGATAACGC CACCGGCATTTCCGATGACAACATCAAAGCATTGCCGGGTGACAATAAGGC CATCGCCGGTGAGGGTGTGCGTAAAGAAGGTGCGCCGCTGGCGCGTGACGT GGCTCCGGCACGCATGGCGGCAGCAAATACGGGCAAGCCGGAGGATAAAG ACCACAAGAAGGTCAAGGACGTTAGCCAGCTGCCGCTGCAACCGACTACCA TCGCCGATCTGTCTCAACTGACGGGTGGCGATGAAAAGATGCCGCTGGCAG CGCAGTCCAAACCGATGATGACCATTTCCCAACCGCCGACGGCGTTAAAG GTGAGGACAGCTCTCTGACCTATCGTTTTCCAGCGCTGGGGCAACGATTACTC CGTCAATATCCAGGCACGCCAAGCGGGCGAATTTAGCCTGATTCCCTAGCAA TACCCAGGTTGAACATCGTCTGCACGACCAATGGCAAAAATGGCAATCCACA ACGCTGGCATTGACGCGTGATGACCAGCAAAAACCCGCAACAGCAACAGCA TCGTCAGCAGTCCGGTGAAGAGGACGACGCGTAACTTGGGCACGCGTCCAT TAAGAAGGAGGAATTAAGCATGGACGATCTGGTTTTCCGCCGGAACAAAGC CCTGTAATTGGTGTGATTCTGTCCGTTGGCCGACGATTGTGCAACCATT ATCGGTCTGCTGGTTGGTCTGTTTCAAACCGTGACGCAGTTGCAGGAGCAAA CCCTGCCGTTCCGGTATCAAGCTGCTGGGTGTGTCTGTGCCTGTTTTGCTG TCTGGCTGGTATGGCGAAGTTCTGCTGTCTACGGCCGTCAGGTCATCTTCC TGGCTCTGGCGAAAGGTTAAGACGCTTGGGCACGCGCTCCATTAACACAGG AGTAATTAAGCATGCTGCTGTCCCTGCTGAATAGCGCGCTGTGCTGCTGGA GCTGCTGATTCTGGTTCTGATGGTTATGATCATCAGCATGTTTCGTTATCCCGT TGCCGACCTATTTGGTTGACTTCTTGATCGCTTTGAACATTGTCCTGGCAATT CTGGTGTTTCATGGGCTCCTTCTACATCGACCGCATTCTGAGCTTCAGCACCT TTCCGGCGGTTCTGCTGATCAGACTCTGTTCCGTTTGGCACTGAGCATCAG CACCAGCCGCCTGATCCTGATTGAAGCAGATGCGGGTGAGATCATCGCGAC CTTTGGTCAGTTTGTGATCGGTGACAGCCTGGCGGTTGGTTTCGTCGTATTCT CCATCGTGACGGTGGTGCAGTTTATCGTTATTACCAAGGGCAGCGAACGTTGT GGCGGAGGTCCCGCTCGCTTACGCTGGACGGCATGCCGGGTAACAGAT GTCTATTGATGCAGACCTGAAAGCCGGCATTATTGATGCTGATGCAGCGCG CGAGCGCCGACGCTCCTGGAGCGTGAAAGCCAACGTACGGTTCTTCGA CGGTGCCATGAAGTTCATTAAGGTGATGCGATTGCGGGCATCATTATCATC TTCGTTAACTTCATTGGCGGTATCAGCGTCGGTATGACCCGTCATGGTATGG ATCTGAGCAGCGCCCTGAGCACCTACACCATGCTGACGATTGGTGATGGTCT GGTTGCCCAAATTCGGCGTTGCTGATCGCGATTTCTGCGGGCTTCATCGTT ACCCGCGTCAACGGTGATAGCGATAACATGGGTGCGTAACATTTGACCCAG CTGCTGAATAATCCGTTTGTCTGTTGTAACGGCGATTTTGACCATCAGCA TGGGCACGCTGCCGGGCTTTCCGTTGCCGGTTTTCGTTATTTCTGTCTGTTGTG

SEQ ID NO:	Name	Sequence
		<p>CTGTCCGTCCGTGTTTTACTTTAAGTTCCGCGAGGCGAAACGTAGCGCTGCGA AACCAAAAACGAGCAAGGGCGAGCAACCGTTGTCCATCGAGGAGAAGGAA GGTAGCAGCCTGGGCCTGATTGGCGACCTGGATAAGGTTAGCACGGAAACC GTCCCCTGATTTTTGCTGGTGCCGAAATCGCGTCGTGAGGATCTGGAGAAA GCGCAGCTGGCGGAACGCTGCGCAGCCAATTCTTTATTGATTATGGTGTGC GTCTGCCAGAAGTACTGCTGCGTGACGGTGAGGGTCTGGATGACAACTCTA TCGTCCTGCTGATTAATGAGATTCGCGTTGAACAGTTTACTGTCTATTTTGAC CTGATGCGTGTGGTTAACTACAGCGACGAGGTGGTGAGCTTTGGCATCAAC CCGACCATTACCAGCAAGGTTCCAGCCAGTACTTTTGGGTGACCCATGAG GAAGGCGAAAAGCTGCGCGAGCTGGGCTACGTCCTGCGTAATGCACTGGAC GAACTGTACCACTGTCTGGCGGTGACGCTGGCACGCAATGTGAACGAGTAT TTCGGTATCCAAGAAACGAAACACATGCTGGACCAACTGGAAGCAAAGTTT CCTGACCTGCTGAAGGAGGTTTTCGCGCCACGCCACCGTGCAGCGCATTTCGG AAGTGCTGCAACGCTCTGCTGTCCGAACGCGTGAGCGTCCGTAACATGAAGC TGATCATGGAAGCCCTGGCACTGTGGGCTCCGCGTGAGAAAGATGTGATCA ATCTGGTGGAGCACATCCGTGGTGGATGGCGCGTTATATCTGCCACAAGTT CGCAAATGGTGGTGAAGTTCGCTGCGGTTATGGTTTCCGCCGAAGTTGAGGA TGTCATTCGTAAGGCATTCGTCAAACTTCTGGCTCCACCTTTTTGAGCTTG GACCCGGAGGCTTCGGCAAATCTGATGGACCTGATCACGCTGAAGCTGGAC GACCTGTTGATTGCGCATAAGGACCTGGTCTGTTGACCAGCGTTGACGTGC GTCGTTTTATCAAGAAAATGATTGAAGGTCGTTTTCCGGATCTGGAGGTCTCT GTCTTCGGTGAGATTGCAGATAGCAAAAAGCGTGAATGTCATCAAAAACCAT CTGAATTTAAATCTTGGGCACGCTCCATTAATAAAGGAGTAATTAAGCATG AGTTTCAGCGAGAGCCGCCACAATGAAAACGTCTGATTCAAGAAGGCGCA CTGCTGTTTTGTGAGCAAGCAGTCGTGGCGCCTGTCAGCGGTGATCTGGTTT TTCGTCCGCTGAAAATCGAGGTCTGAGCAAGCTGCTGGCGTTCATCGACG GCGCAGGTCTGGTGGATACGACCTACGCGGAGTCGGACAAATGGGTTCTGC TGTCCTCCGGAGTTCCGTGCTATTTGGCAAGACCGTAAACGTTGCGAATATTG GTTTTTGCAGCAGATTATCACCCATCTCCGGCGTTCAACAAGGTTCTGGCA CTGTTGCGTAAGAGCGAAAGCTATTGGTTGGTTCGGCTACTTGCTGGCCCAA GCACCAGCGGCAATACTATGCGTATGTTGGGTGAGGATTACGGTGTAGCT ACACGCATTTCCGCCGTCTGTGCAGCCGCGCTCTGGCGGTAAGGCGAAAA GCGAGCTGCGCAATTGGCGCATGGCCAGTCCCTGCTGAATAGCGTGGAAG GTCATGAAAACATCACCCAGCTGGCGGTCAACCACGGTTATAGCAGCCCGT CCCCTTTAGCTCTGAAATCAAGGAGCTGATTGGTGTTCCTCCCGGTAAGCT GTCTAACATCATTAGCTGGCCGACAAATGAATTTAAATGCTAGCCTTGGGG ACGCGTCCATTAAGAGGACCAATTAAGCATGAAAACCCACGCTGCTGCTGC AATACCTGGCCTACCCGAGAAAATCACTGGCCCTATCATTGAAGCAGAAC TGCGTGATGTTGCAATTGGTGAATTGTGCGAGATCCGTGCGCGCTGGCACCA GAAGCAGGTTGTGGCCCGTGCAGCAAGTGGTTGGTTGTCAGCGCGAACGTAC CGTCCTGAGCCTGATCGGCAATGCCAAAGGCTGAGCCGTGATGTGGTCTTG TACCCGACCGGCCGTGCTCTGAGCGCGTGGGTTGGTTACAGCGTTCTGGCG CAGTACTGGACCCGACGGGTAATAATCGTTGAACGTTTACCCCGGAAGTCG CACCGATTTCCGAGGAGCGCGTTATCGACGTGGCACCCGCGGACTACGCAA GCCGTGTCGGTGTGCGCGAACCCGCTGATCACGGGTGTCGCGCAATTGATG GTCTGCTGACGTGTGGTGTGGGCCAGCGTATGGGTATTTTCGCAAGCGCGG GTTGTGGTAAGACCATGTTGATGCACATGCTGATCGAGCAAACCGAAGCGG ATGTCTTTGTTATGGCCTGATTGGCGAGCGTGGTCTGTAAGTTACCGAATT TGTAGACATGTTGCGTGATCTCATAAGAAAGAAAAATGTGTGCTGGTTTTT GCCACGTCCGACTTCCCAAGCGTTGACCGCTGCAACGCTGCCAGCTGGCA ACGACGGTCGCCGAGTATTTCCGCGACAGGGTAAACGTTGTCTCTGTTTA TCGACAGCATGACCCGTTATGCACGCGCGTTGCGTGATGTGCGCTGGCGA GCGGCGAGCGTCCGGCTCGCCGTGGTTATCCGGCGTCTGTGTTTCGACAATCT GCCGCGTTTGTGAGCGTCCGGGTGCGACGAGCGAGGGTAGCATTACCGC CTTCTATAACCGTCTGCTGGAGTCGGAAGAAGAAGCGGACCCGATGGCGGA CGAGATCCGTTCTATTCTGGACGGTACCTGTACCTGTCCCGCAAACCTGGCG</p>

SEQ ID NO:	Name	Sequence
		GGTCAAGGTCATTACCCGGCTATCGATGTGCTGAAGAGCGTGAGCCGTGTG TTGGTCAAGTGACCACCCCGACTCACGCGGAGCAAGCGAGCGCGGTCCGT AAGCTGATGACCCGCTCTGGAAGAGCTGCAACTGTTCAATTGACCTGGGCGAG TATCGTCCGGGCGAGAACATTGACAATGATCGTGCGATGCAAATGCGCGAT AGCCTGAAGGCGTGGCTGTGTCAGCCTGTTGCGCAATACAGCAGCTTCGAT GATACGCTGTCCGGCATGAACGCCTTTGCGGATCAGAAGTACTGACTTGGGCAC GCGTCCATGAAAGAAACGACATACTAGATGGGTAACGATATTAGCTTGATT GCATTGCTGGCGTTTTCCACCCTGCTGCCGTTTCATCATCGCGTCTGGTACTTG CTTCGTCAAATTCAGCATCGTCTTTGTGATGGTGCGCAACGCGCTGGGTCTG CAACAAATTCGAAGCAATATGACCCTGAATGGCGTGCAGCTGCTGCTGTCG ATGTTTGTATGTGGCCGATTATGCACGACCGGTATGTGTATTTCCGAGGATG AAGATGTGACCTTTAACGACATCTCCAGCCTGAGCAAGCATGTTGATGAGG GCCTGGACGGTTATCGCGACTACCTGATCAAGTATTCCGACCGTGAGCTGGT GCAGTTCTTTGAGAATGCCAGTTGAAAACGTCAGTACGGTGAAGAAAACGGA AACCGTTAAACGTGACAAGGACGAGATTGAAAAGCCGAGCATTTTCGCACT GTTGCCTGCTTACGCCTTGAGCGAGATTAAGAGCGCATTCAAAAATTGGTTTT TACCTGTACCTGCCGTTTCGTTGTGGTTCGATCTGGTTGTCTCCAGCGTTCTGCT GGCCCTGGGCATGATGATGATGTCCCCGGTTACCATCAGCACGCCGATCAA ACTGGTCTGTTTGTGGCCCTGGATGGTTGGACGCTGCTGTCTAAAGGCCTG ATCCTGCAATACATGGACATCGCGACCTAAGAGCTCTGGGCACGCGTCCAT TAATGAGGAAAAATTATTAGCATGAAAACGCACATTCTGTTGGCCCGTGTG CTGGCTTGGCGAGCTCTGGTGTGCTGGTCAACCCAGGTTATAGCTCCGAGAAGA TCCCGTTACGGGCAGCGGCTTCGTTGCAAAGGACGATTCTCTGCGCACCTT TTTCGATGCGATGGCACTGCAATTGAAGGAGCCGGTATTGTCAGCAAGAT GGCGGCTCGCAAAAAGATTACCGGCAATTTTCGAGTTCACGATCCAAACGC GCTGCTGGAGAACTGTCCCTGCAACTGGGTCTGATCTGGTACTTTGATGGT CAGGCGATCTACATCTACGACGCGAGCGAAATGCGTAATGCGGTTGTGAGC CTGCGTAATGTCAGCCTGAACGAGTTCAACAATTTTCTGAAGCGCAGCGGC CTGTACAATAAGAACTACCCTCTGCGTGGTGATAATCGTAAAGGCACCTTCT ATGTCAGCGGTCCGCCGGTGTATGTTGATAATGGTTGTAATGCGGCCACCAT GATGGACAAACAGAATGATGGCATCGAGCTGGGTCCGCAAAAGATCGGTGT TATGCGCCTGAACAACACTTTTGTGGGCGACCGCACCTACAACCTGCGTGT CAAAAGATGGTCATTCCGGGTATTGCTACGGCAATTGAACGCCTGTTGCAA GGCGAAGAACAACCGCTGGGTAACATTGTAAGCTCCGAGCCTCCGGCCATG CCGGCCTTTCAGCGCAAACGGCGAGAAAGGTAAGGCAGCGAATTACGCGGGT GGTATGAGCCTGCAAGAAGCGCTGAAACAGAACGCAGCGGCAGGCAACAT CAAAATTGTGGCCTATCCGGACACCAACAGCCTGCTGGTGAAAGGTACGGC GGAGCAGGTGCATTCATCGAGATGCTGGTTAAAGCCCTGGACGTGGCGAA ACGTCACGTTGAATTGAGCCTGTGGATTGTGGATTTGAATAAGAGCGACCT GGAACGCTTGGGCACCAGCTGGAGCGGTAGCATCACCATCGGCGACAAGCT GGGTGTGAGCCTGAACCAGAGCAGCATCTCCACGCTGGACGGTAGCCGCTT TATTGCGGCGGTCAACGCTCTGGAGGAAAAGAAAACAGGCCACTGTTGTGAG CCGTCCGGTTCTGCTGACCCAGGAGAATGTCCCAGCGATTTTTGACAATAAT CGTACTTTCTATACCAAATGATCGGTGAACGTAATGTTGCATTGGAACACG TGACCTATGGCACCATGATCCGTGTCCTGCCGCGTTTCAGCGCGGACGGCCA GATTGAGATGAGCCTGGACATCGAAGATGGTAATGACAAAACCCCGCAGTC TGATACGACCACCTCCGTTGATGCGCTGCCAGAAGTGGGTCGTACCCTGATC TCGACGATTGCACGTGTCCCGCATGGTAAATCTCTGCTGGTTGGTGGCTACA CGCGTGATGCAAAACACGGACACGGTCCAAAGCATCCCGTTCTGGGTAAGC TGCCGCTGATTGGCTCGTTGTTTCGCTACAGCAGCAAAAACAAGTCTAATGT CGTCCGTGCTTTATGATTGAGCCGAAAGAAATCGTTGACCCGCTGACCCCG GATGCCAGCGAGAGCGTTAAACAACATTCTGAAAACAGTCCGGTGGTGGAGC GGTGACGATAAGCTGCAAAAAGTGGGTTTCGTGTGATTTGGACCGTGGTCAG GAGGCCATTAAGTAACTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG ATGTCCAGCAACAAAACCGAAAACCGACTAAGAAAACGTCTGGAGGATAG CGCAAAGAAAGGTCAGAGCTTCAAGAGCAAGGACCTGATTATCGCGTGCCT

SEQ ID NO:	Name	Sequence
		<p>GACCCTGGGTGGTATCGCTTATTGGTGAGCTACGGCAGCTTCAATGAGTT ATGGGTATCATTAAAGATTATCATCGCTGATAACTTTGATCAGTCGATGGCAG ATTATAGCCTGGCCGTGTTTGGTATTGGCCTGAAATACCTGATTCCGTTTCAT GCTGCTGTGTTTGGTTTGTTCGCACTGCCGGCACTGCTGCAAGCGGGCTTC GTTCTGGCAACCGAGGCCCTGAAGCCGAATCTGTCCGCCCTGAACCCGTT GAAGGCGCGAAGAAACTGTTTTCCATGCGCACCGTCAAAGACACGGTCAAAG ACGCTGCTGTATCTGTGAGCTTTGTGGTTGCGGCAATCATTGCTGGAAAA AGTATAAAGTCGAGATTTTCAGCCAACCTGAACGGTAATATCGTGGGTATTG CGGTTATCTGGCGTGAATTGCTGCTGGCGTTGGTTCTGACCTGTCTGGCGTG CGCGCTGATCGTGTGCTGGATGCTATTGCCGAGTACTTTCTGACCATG AAAGATATGAAGATGGACAAAGAAGAAGTTAAACGCGAGATGAAAGAGCA GGAGGGTAACCCGAGGTGAAGTCTAAACGTCGTGAAGTCCACATGGAAAT CCTGAGCGAACAAGTCAAGTCTGACATTGAAAATAGCCGTCTGATTGTGGC AAACCCTACGCATATTACCATCGGCATCTACTTCAAACCGGAACTGATGCCG ATTCCAATGATTAGCGTCTATGAAACCAATCAACGCGCGCTGGCGGTCCGT GCGTACGCCGAGAAAGTGGGTGTTCCGGTTATTGTAGACATCAAGCTGGCG CGCAGCCTGTTCAAACGCACCGTCTGTTACGACCTGGTGAGCCTGGAGGAG ATCGACGAGGTTTTGCGCCTGCTGGTTTGGTTGGAAGAGGTGCAAAACGCA GGCAAGGATGTGATCCAACCGCAGGAGAATGAAGTGCCTCATTAACTGCAG GTTTAAACTACTTGGGCACGCGTCCATTAATTAGGATCAATAGCATGATTCC GGGCAGCACCTCCGGTATTTCTTTAGCCGTATCCTGAGCCGTGAGACCTCC CACCAGGATGCAACCCAGCATAACCGACGCAACAACAAGCGGAAATTCACA AGCGGCGGAAGATAGCTCGCCGGGTGCGGAGGTTCAAGAAATTCGTCCAGAG CACGGACGAGATGTCTGCTGCGTGGCGCAGTTCGCAATCGCCGTGACTAT GAGAAAAAGAGCAGCAATTTGTCTAACTCCTTCGAGCGGTTCTGGAGGAC GAGGCACTGCCGAAAGCGAAGCAGATTCTGAAACTGATCAGCGTGCATGGC GGTGCCTGGAGGATTTCTGCGTCAGGCGCGCAGCCTGTTCCCGGACCCA AGCGATCTGGTGCTGGTTCTGCGCGAGCTGTTGCGTCTGTAAGGACCTGGAG GAGATCGTGCCTAAGAAGCTGGAGAGCCTGCTGAAGCACGTGGAGGAACA AACCGACCCGAAAACCCTGAAGGCCGGTATTAAGTGCCTGAAAGGCGCG TCTGTTTGGCAAGACGCTGTCTCTGAAACCTGGTCTGCTGCGTGCAGCTAC CGCCAGTTCATCAAAGCGAAAGCCACGAAGTCGAGATTTACAGCGATTGG ATCGCCAGCTACGGTTATCAGCGTCGCTGGTTGTTCTGGATTTCAATTGAAG GCAGCCTGCTGACTGACATCGATGCTAACGATGCAAGCTGCTCCCGTCTGG AGTTTGGCCAACTGCTGCGCCGTCTGACCCAGCTGAAAATGTTGCGTAGCGC CGACCTGCTGTTTGTCTCGACGTTGCTGTCTTACAGCTTACGAAAGCATT AACGCTGAGGAGAGCAGCTGGCTGTGCTGATGCTGTCTTTGCTGCAACAG CCGCACGAAGTGGATAGCCTGCTGGCGGACATTATCGGTCTGAATGCGCTG CTGTTGTCCACAAAGAGCACGCCAGCTTCCCTGCAAATCTTCTATCAGGTCT GTAAGGCAATCCCGTCTAGCCTGTTTTATGAAGAGTACTGGCAAGAAGAAC TGCTGATGGCACTGCGCTCCATGACGGACATTGCTTACAAACACGAAATGG CCGAACAACGTGCTACCATCGAAAAGCTGTCTAAAGTTTAAACCTTGGGCA CGCGTCCATGAAAGACAGGACCCACTAGATGAGCTTGGCGGTACGCCAAAT TGACCGCCGTGAATGGCTGCTGGCCAGACTGCGACCGAATGTGAGCGTCA CGGTCGTGAGGCCACCCTGGAGTATCCGACCCGTGAGGATGTTGGGTCCG CCTGTCTGATGCCGAAAACCGCTGGTCTGCGTGGATCAAACCGGGTGATTG GTTGGAACACGTTAGCCCAGCACTGGCGGGTGGCGCGTACGCGCAGGCGC GGAGCACCTGGTGGTTCCGTGGCTGGCGGCAACCGAACGCCCGTTTCGAGCT GCCCGTCCCGCACTTGAGCTGCCGTGCTGTGCGTGGAGAACC CGGTTCCG GGTTCCGCACTGCCTGAGGGCAAGCTGCTGCACATCATGTCCGATCGTGGT GGCCTGTGGTTTGGAGCATCTGCCGGAGCTGCCGGCTGTTGGCGGTGGCCGT CGAAGATGCTGCGTTGGCCGTGCGTTTTGTTATTGGCAGCAGCGACACCCA GCGCAGCCTGCTGGGTGCTATCGGTATCGGTGATGTTCTGCTGATCCGCACC AGCCGTGCTGAGGTTTACTGTTACGCGAAGAACTGGGCCACTTTAATCGTG TGGAAGGTGGCATTATTGTGAAACGCTGGACATTCAACATATCGAGGAGG AGAACAACACGACGGAAACGGCGGAAACCCTGCCGGTCTGAATCAACTG</p>

SEQ ID NO:	Name	Sequence
		CCGGTGAAGCTGGAGTTCGTTCTGTATCGTAAAAACGTGACGTTGGCCGAA CTGGAAGCAATGGGTCAAGCAACAACCTGCTGTCTTGCCAACCAATGCGGAA CTGAACGTCGAAATCATGGCGAATGGCGTGCTGCTGGGTAACGGCGAACTG GTGCAGATGAATGACACCCTGGGTGTCGAGATTCATGAGTGGTTGTCCGAG AGCGGTAATGGCGAGTAGAGATCTCTTGGGCACGCGTCCATTAAAAAGGAC CAATTAAGCATGCACTCTCTGACTCGTATCAAGGTCTGCAACGTCGTTGTA CCGTGTTCATTCTCAGTGGGAGTCCATTCTGTTGCGTTATCAGGACGAAGA TCGCGGCTTGCAGGCGGAGGAGGAGGCCATCCTGGAACAGATCGCAGGTCT GAAACTGCTGTTGGACACCCTGCGTGCTGAAAATCGTCAACTGAGCCGTGA AGAAATCTATACGCTGCTGCGCAAACAGAGCATTGTTTCGTCGCCAGATTAA GGATCTGGAGCTGCAAATCATTAGATTCAAGAAAAGCGTAGCGAGCTGGA AAAGAAACGTGAGGAATTTCAAAGAAAAGCAAATACTGGCTGCGTAAAG AAGGTAACCTACCAGCGCTGGATTATCCGTCAAAAACGCTTCTACATTCAAC GTGAGATCCAGCAAGAGGAGGCGGAGAGCGAAGAGATCATTAACTTGGG CACGCGTCCATGAAAGACAGGACCCACTAGATGTTCTATGCATTGTATTTTG AGATCCACCATCTGGTGGCGTCCGCGGCTCTGGGTTTTGCGCGTGTGCTCC GATCTTTTTCTTTCTGCCGTTCTTGAATAGCGGTGTCTGAGCGGTGCTCCG GCAACGCGATCATCATTCTGGTGGCGCTGGGTGTATGGCCGCACGCCCTGA ATGAGGCGCCACCGTTTCTGTCTGTGGCAATGATTCGCTGGTCTGCAAGA GGCAGCCGTGGGTGTTATGCTGGGTTGCCTGTTGTCTGGCCGTTTTGGGTT ATGCACGCGTTGGGCTGTATCATTGATAACCAACGCGGTGCAACCCTGTCCA GCAGCATGATCCTGCGAATGGCATCGACACCAGCGAGATGGCCAATTTCC TGAACATGTTGCGGGCTGTCTGTATTTGCAGAACGGTGGCTTGGTCACGAT GGTGGACGTGCTGAATAAGTCTTACCAGCTGTGTGATCCGATGAACGAGTG TACGCCGAGCCTGCCTCCGTTGCTGACCTTCATTAATCAAGTGGCCAGAAC GCACTGGTGCTGGCGTCCCCGGTCGTTCTGGTCTGCTGCTGTCGGAAGTTT TCCTGGGCCCTGCTGTCTCGTTTTGCACCGCAAATGAACGCGTTCGCCATTAG CCTGACTGTAAAAAGCGGTATTGCGGTTTTGATCATGCTGCTGTATTTACGC CCGGTCTGCCGGACAATGTTCTGCGTCTGAGCTTTCAGGCGACCGGCTGA GCAGCTGGTTCTACGAACGTGGCGCAACGCATGTGCTGGAATAAGGATCCG GCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT CTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCCCTAG AACTAGTccataagaaiGCGGCCGC

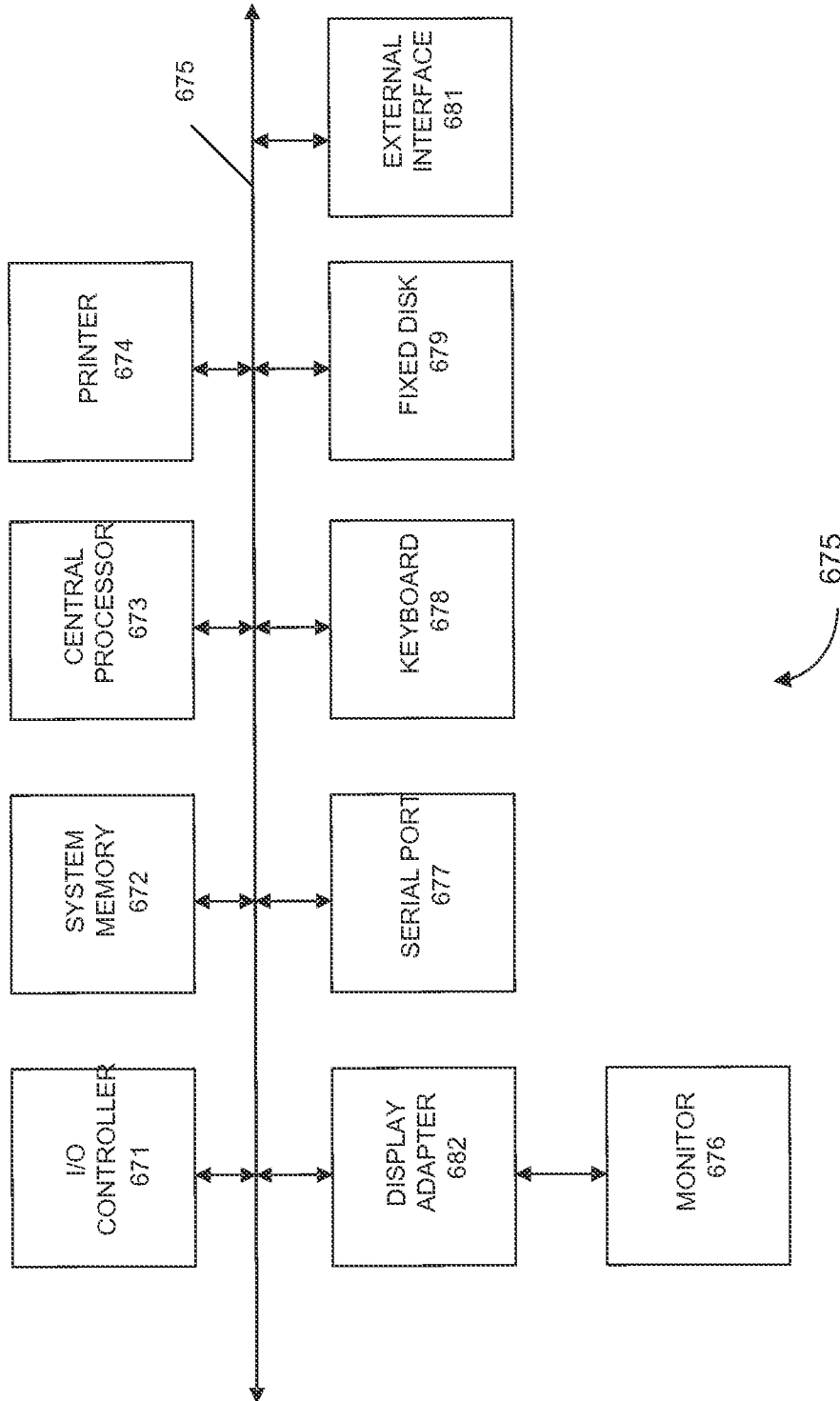


FIG. 25

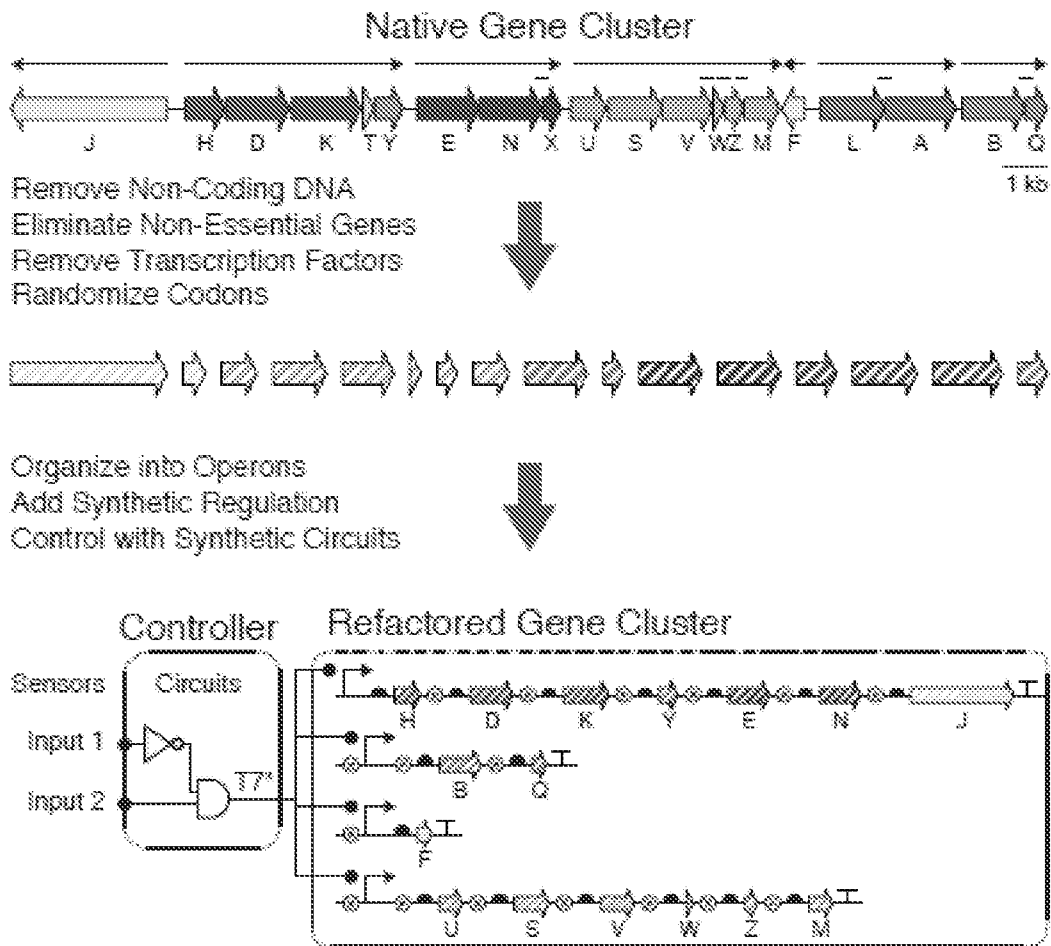


FIG. 26

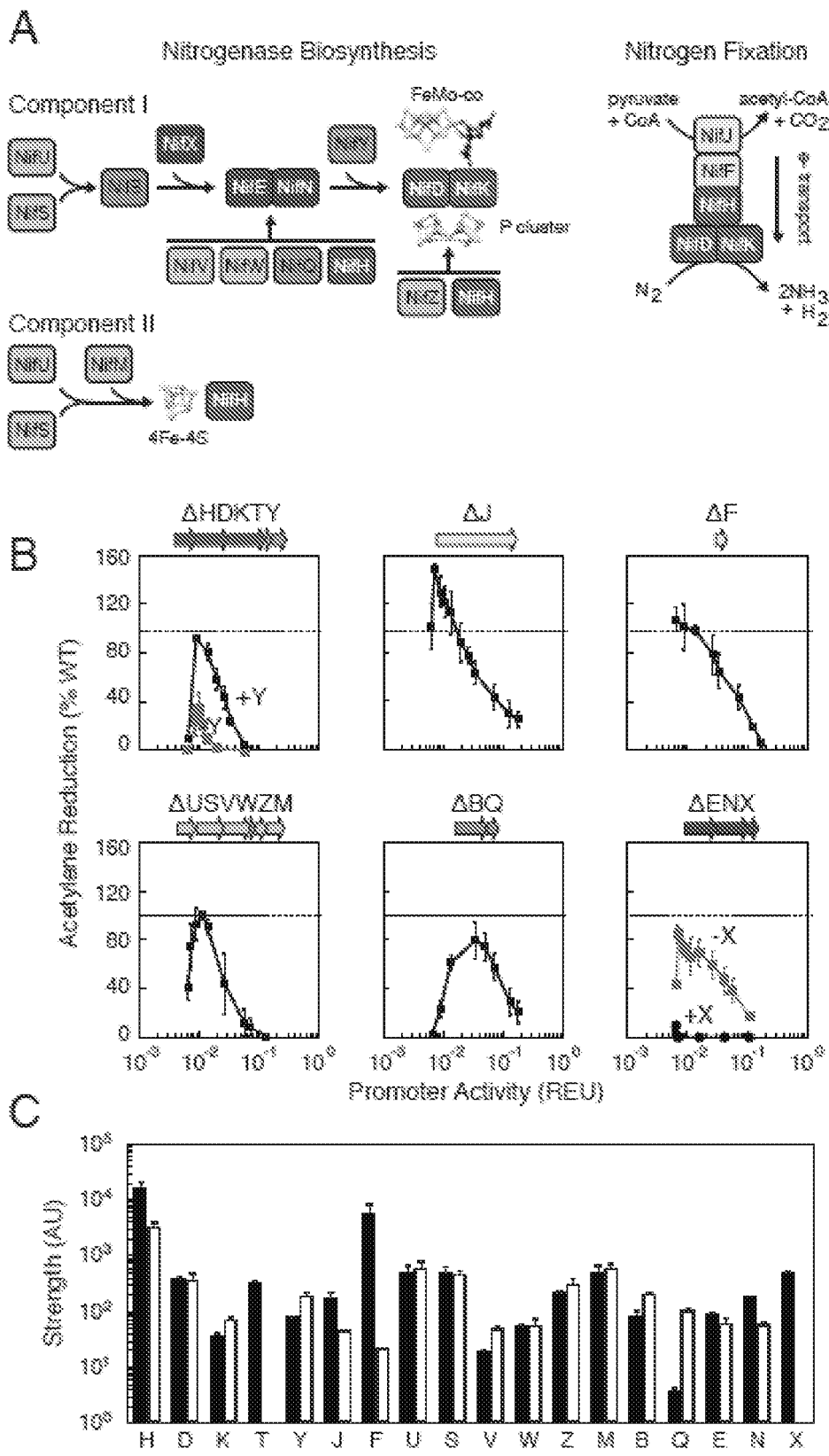


FIG. 27

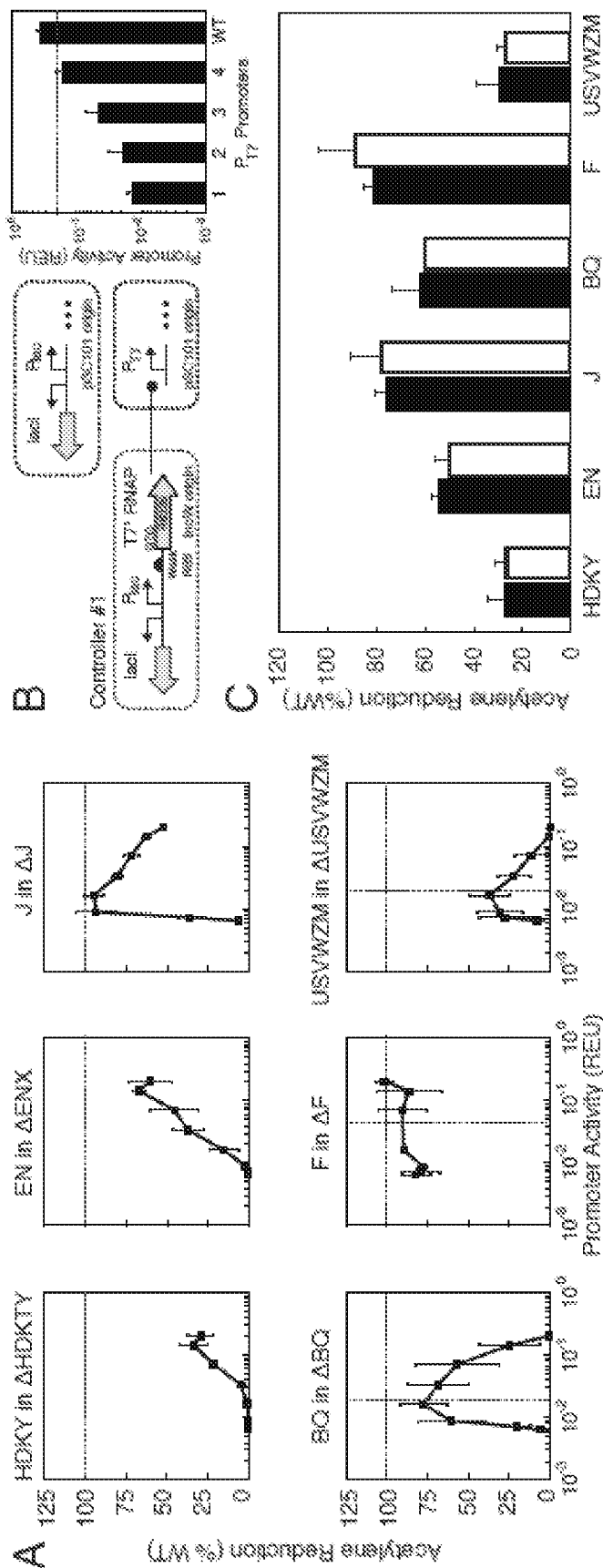


FIG. 28

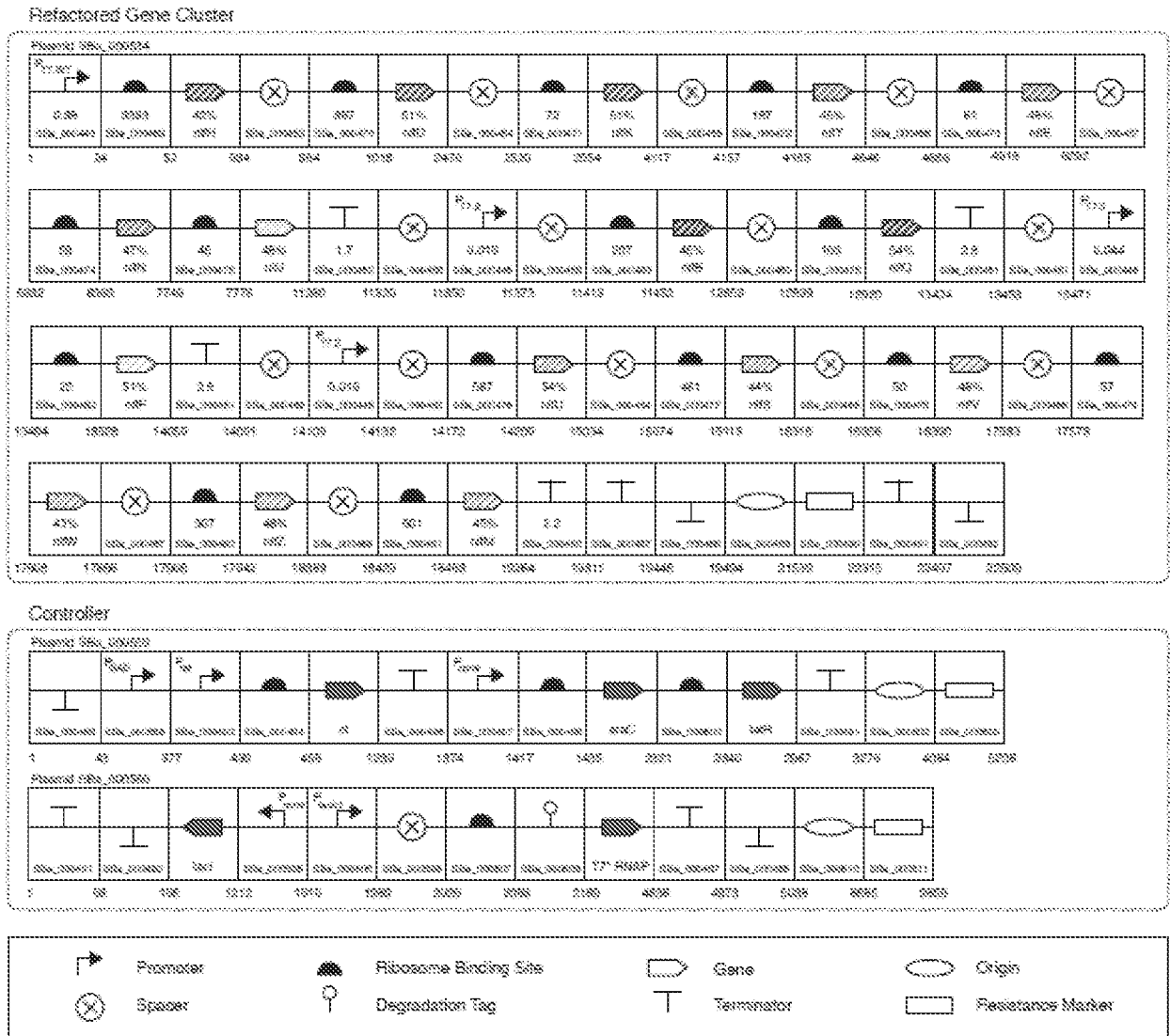


FIG. 29

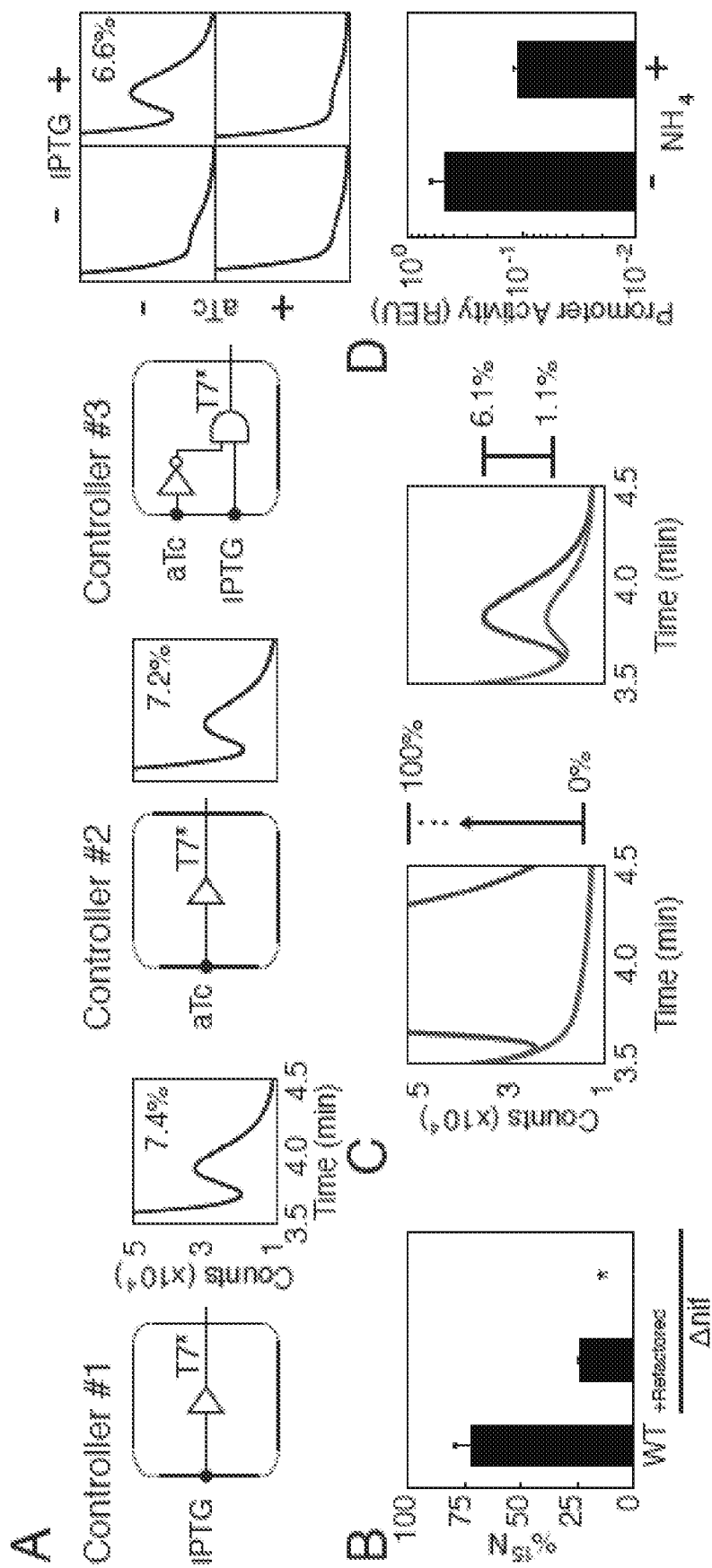


FIG. 30

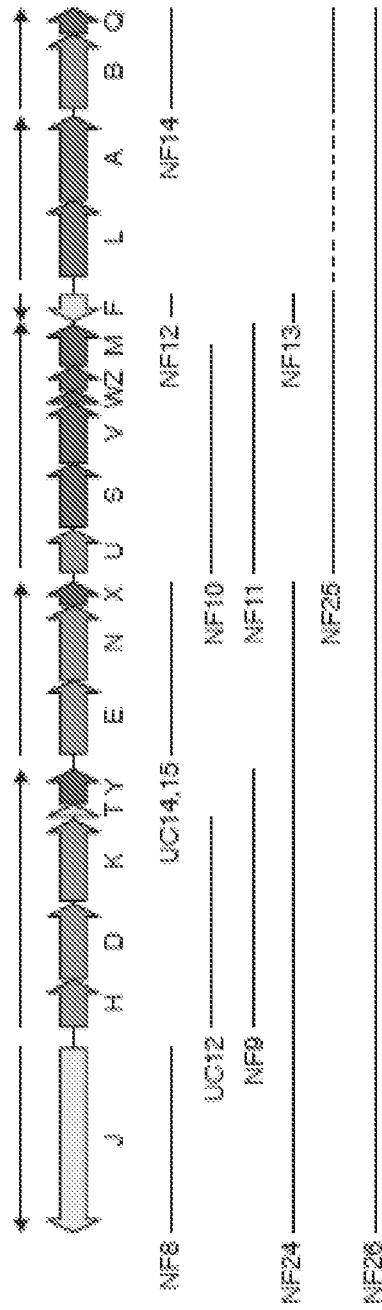


FIG. 31

FIG. 32

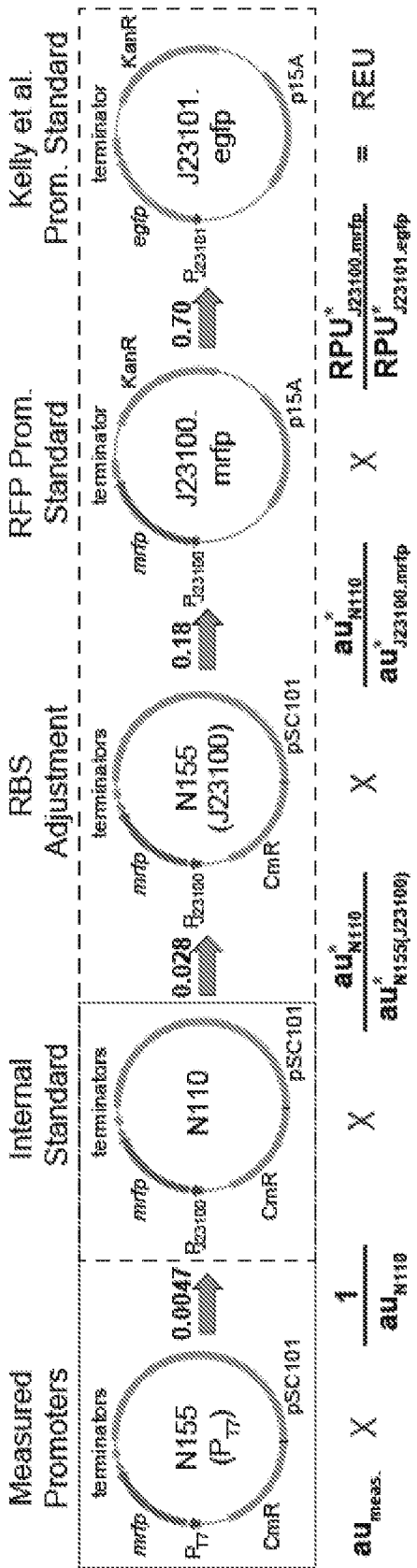
Table. Construction and verification of all <i>K. oxytoca nif</i> gene deletion mutants					
Name	Genotype	Note	Parental Strain	Allelic Exchange Plasmid ⁴	DNA sequence of the deletion region ⁵
NF8	<i>ΔnifJ, Km</i>	operon deletion	M5aL	pNif16	gcaggagaactaaaggcccg GCATGC (Km cassette) GCATGC aaggcccatggcccggca
UC12	<i>ΔnifHDK</i>	in-frame deletion	M5aL	pRHB296	caaaacaggcaaggctgg GGTACC caatcaggtgcccgggctca
NF9	<i>ΔnifHDKTY</i>	operon deletion	M5aL	pNif9	atggcccggcaggcgcaat GCATGC gacgctcttcccacgttac
UC14	<i>ΔnifENX</i>	operon deletion	M5aL	pRHB294	gacgctcttcccacgttac GCATGC gatccggaccggcgccgta
UC15	<i>ΔnifENX</i>	operon deletion	M5aL	pRHB295	ctcttcccacgttaegetc GCATGC gatccggaccggcgccgta
NF10	<i>ΔnifUSVWZM</i>	operon deletion ¹	M5aL	pNif10	cggaaccggcgccgctagccg GCATGC atctttggcagcagagccag
NF11	<i>ΔnifUSVWZM</i>	operon deletion	M5aL	pNif11	cggaaccggcgccgctagccg GCATGC agcctcggcgccgtaaccggtt
NF12	<i>ΔnifF</i>	operon deletion ²	M5aL	pNif12	actggttatacctgatccagc GCATGC ctttgcaactaccggggccca
NF13	<i>ΔnifF</i>	operon deletion	M5aL	pNif13	ccgttaacgctacagcaac GCATGC ctttgcaactaccggggccca
NF14	<i>ΔnifBQ</i>	operon deletion	M5aL	pNif14	atcagggagcggggttggc-----atgtgattatgcgagcttt
NF24	<i>ΔnifJ, HDKTY, ENX</i>	operon deletion ³	M5aL	pNif32	gtcaaacaccagaatattga GAGCTC tgtggttctcgtgacaaagc
NF25	<i>ΔnifUSVWZM, F, BQ</i>	operon deletion	NF14	pNif51	cggaaccggcgccgctagccg GCATGC ctttgcaactaccggggccca
NF26	<i>ΔnifJ, HDKTY, ENX, USVWZM, FLA, BO</i>	whole <i>nif</i> deletion ³	NF24	pNif54	gtcaaacaccagaatattga CTCGAG atgtgattatgcgagcttt

1) 83bp of *nifM* 3' retained2) 82bp of *nifF* 3' retained3) 500bp of *nifJ* 3' retained

4) The backbone for all allelic exchange plasmids is pDS132(1).

5) The bold region shows the restriction site sequence that replaces the deletion region. The 20 nucleotides flanking each site of the deletion region are shown.

A



B

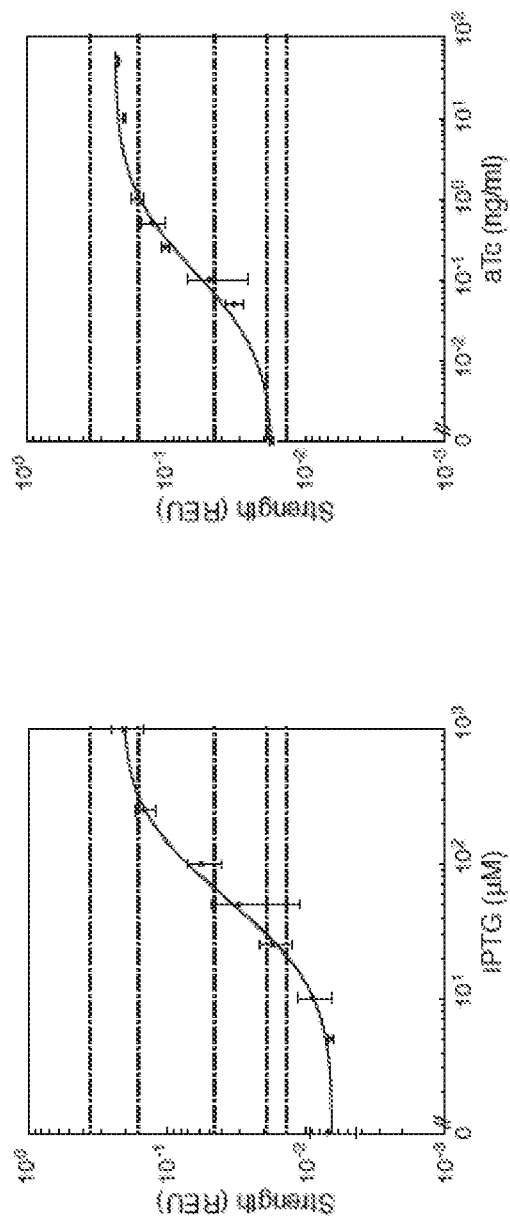


FIG. 33

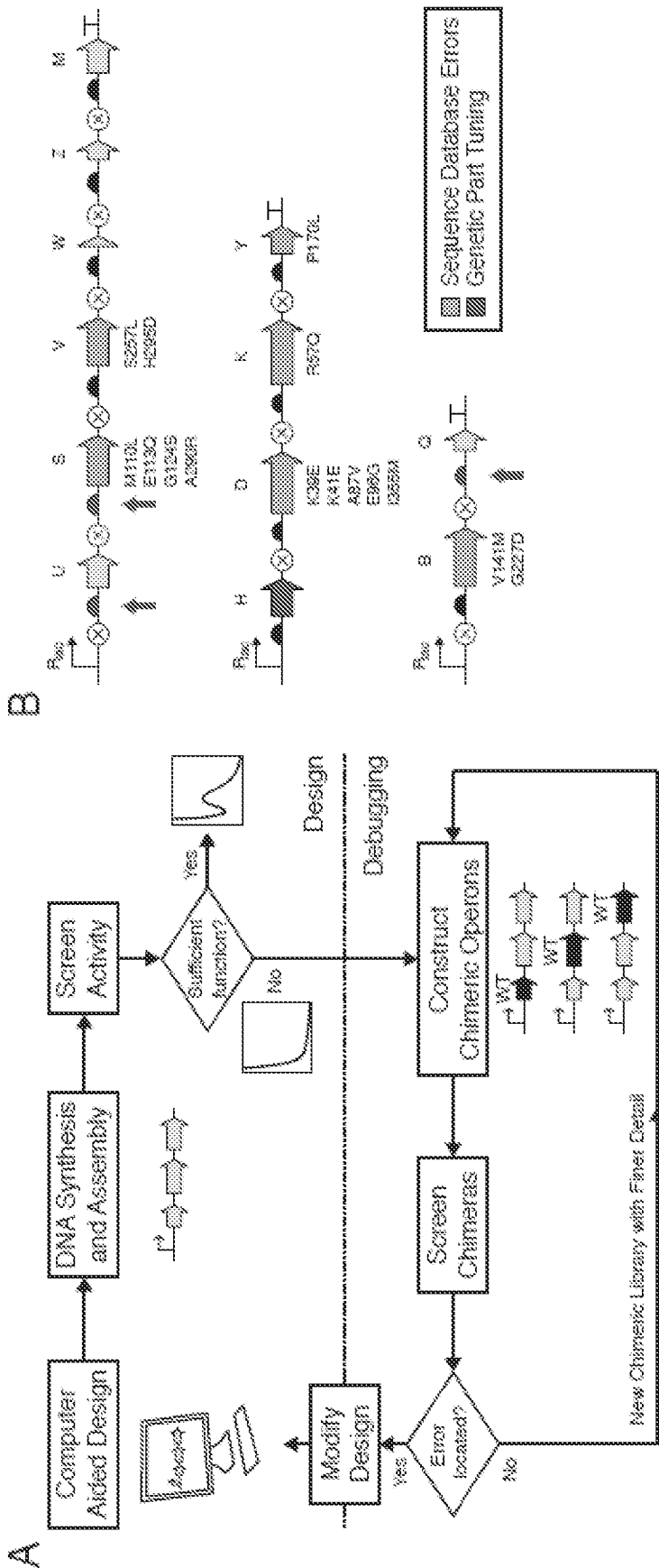


FIG. 34

FIG. 35

Table: DNA sequence errors in <i>nif</i> cluster sequence X13303.1			
Location	X13303.1	This Study	Impact
5187	a	g	NifD K39E
5193	a	g	NifD K41E
5332	ct	tc	NifD A87V
5359	a	g	NifD E96G
6137	c	g	NifD R55M
6140	a	c	silent
6749	ga	ag	NifK R57Q
7371	t	c	silent
7371	t	c	silent
8168	cgg	ggc	<i>nifT</i> 5' UTR
8919	c	t	NifY P170L
9341	-	insert t	<i>nifE</i> 5' UTR
9498	cg	gc	NifE A35G
9697	ac	ct	silent
9891	c	t	NifE P166L
12541	ca	ac	NifX Q131T
12770	t	c	<i>nifU</i> 5' UTR
13646	gc	cg	<i>nifS</i> 5' UTR
13999	a	c	NifS M110L
14008	g	c	NifS E113Q
14037	a	g	silent
14041	g	a	NifS G124S
14539	gc	cg	NifS A290R
15657	tc	ct	NifV S257L
15770	gc	cg	NifV H295D
16739	g	a	silent
16814	c	t	silent
22050	g	a	NifB V141M
22309	g	a	NifB G227D

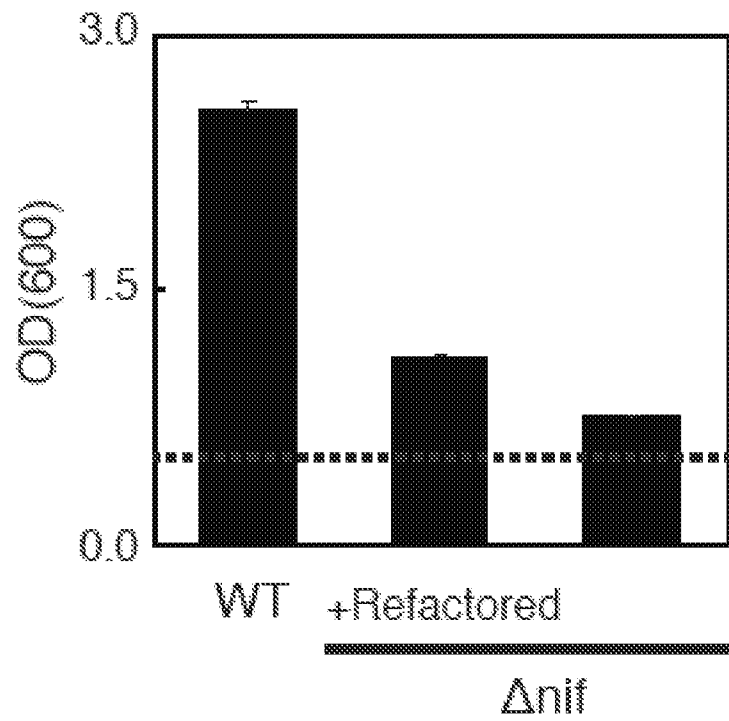


FIG. 36

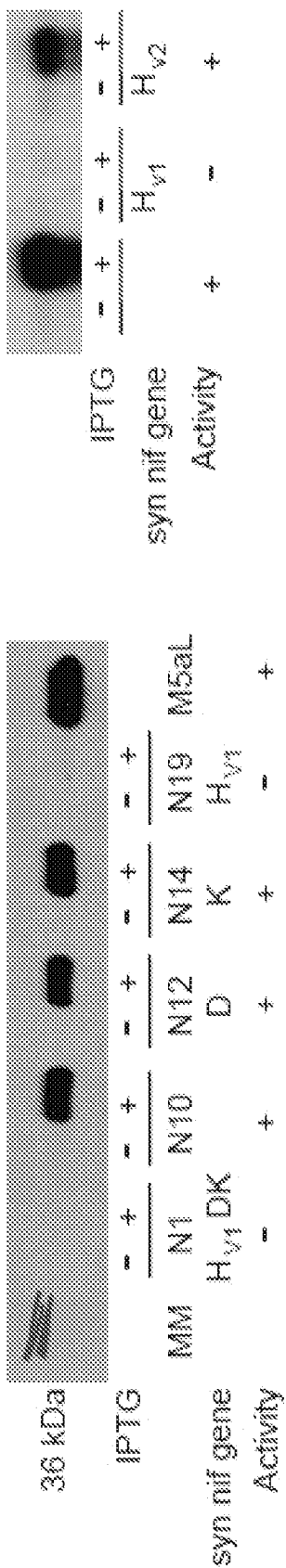


FIG. 37

FIG. 38

Table. DNA sequences of synthetic parts		
Part Number	Part Type	Sequence
SBa_000443	T7 Promoter WT	taatacgactcactatagggaga
SBa_000444	T7 Promoter 1	taatacgactcactacagggaga
SBa_000445	T7 Promoter 2	taatacgactcactagagagaga
SBa_000446	T7 Promoter 3	taatacgactcactaagggaga
SBa_000447	T7 Promoter 4	taatacgactcactatagggaga
SBa_000450	WT T7 Terminator	tageataaaccccttggggcctctaaacgggtcttggaggggttttttgt
SBa_000451	T7 Terminator	tacatatcgggggggtaggggttttttgt
SBa_000452	T7 Terminator	tactotaaaccccatcggccgtcttaggggttttttgt
SBa_000453	Insulator	gccgtgcgaacaccttaatacagccttcaatgctaaaggtcctggctgcactgc
SBa_000454	Insulator	caaacaccccatgttcgatactgaacgaatcgaagcaactcccttccttg
SBa_000455	Insulator	cctgattgtaacccgatctgatgctacagtggttgagtta
SBa_000456	Insulator	catttttgccttgcgacagacctcctacttagatggccac
SBa_000457	Insulator	tgtcaagctaggaggcaattctataagaatgcacactgca
SBa_000458	Insulator	cctgggttgagtcagcgtgcgagcagcggc
SBa_000459	Insulator	cggacttccagagaagaagactactgacttgagcgttcc
SBa_000460	Insulator	actacagatttgaggttaaaccaataagcaagtagtggc
SBa_000461	Insulator	gtctgtagcagctgcactc
SBa_000462	Insulator	ggtcattacaacgggtat
SBa_000463	Insulator	aacatagcgttccatgagggctagaattacctacggcct
SBa_000464	Insulator	catgttaataagccacaaaagagtgatgatagtcattgggt
SBa_000465	Insulator	gagttactggccctgatctctccgcttctaataccgcaca
SBa_000466	Insulator	gactcaacacagctaggagcgtgaagctgattccttcgatg
SBa_000467	Insulator	tccgaaacaagggcagttccgggctgaaagttagcggcggg
SBa_000468	Insulator	acgcccagcgtagtggagacatacagcttcgttgggttcac
SBa_000508	Insulator	tgcagttttatctctcgcagcactgtaataggcaactaa
SBa_000475	Synthetic RBS	ttaaaggagggaatttgaagcATGaaaactatggacggtaacgctgcggctgcattgattageta
SBa_000469	Synthetic RBS	tactagagaacataaactbaactttagggaggataaactATGaacatgcctgagtgccgatttatg
SBa_000470	Synthetic RBS	gcaaaaggttggtattggcaaaaagcagcagcaaccagaacttggggcggccctggccagatg
SBa_000471	Synthetic RBS	ggctgcgtacaccttaatacagcctcaatgctaaaggtcctggctgcattgcaaaacttccactcc
SBa_000472	Synthetic RBS	ctactcagcggaggagcggcATGatgactaatgctactggcgaacgtaaacctggcactgattca
SBa_000473	Synthetic RBS	agaagtaactgggaagtgctccgggaaacggcggcaaaagagcggct
SBa_000474	Synthetic RBS	caaacaccccatgttcgatactgaacgaatcgaagcaactcccttccttgcattctcaactct
SBa_000475	Synthetic RBS	caaaaattagcggaggtagccttccgcaactatcgataaaaactcaactcttggtaaccgctggt
SBa_000476	Synthetic RBS	cgagcaggacgaataatacggaaactgctccgtaacaaaactgagctg
SBa_000477	Synthetic RBS	cctgattgtaaccgactgctgactcagctgggttgagttacactctcaactccggaggtaactt
SBa_000478	Synthetic RBS	ctATGcttgcaaaactgatacctctgttttgggcactgctggcctggttcaagtgcctgcggattt
SBa_000479	Synthetic RBS	gcagcggctcaaatcgtcagattggctg
SBa_000480	Synthetic RBS	caattttgccttgcagacagacctcacttagattgcacacacttcaaatcaactcctgggggt
SBa_000481	Synthetic RBS	tattacaATGaaaggttaacagatcttctctgctggcgaacggcctgtagaacacaaaca
SBa_000482	Synthetic RBS	taaacagaaatccggctgtagcggcccaaaagcgg
SBa_000483	Synthetic RBS	tgtcaagctaggaggcaattctataagaatgcacactgcaactaaactcaacgcaagcagggag
SBa_000484	Synthetic RBS	agtaactATGgcagacattttccgcaactgataagcggctggctgtgctgcagatcaagaacggc
SBa_000485	Synthetic RBS	cagcggctgggtgcagactcctggcctccctgggt
SBa_000486	Synthetic RBS	aacatagcgttccatgagggctagaattactcagcctcagatctgcacaaatcaaacagcgg
SBa_000487	Synthetic RBS	aaggaagtctcaATGtggaactcagcggagaaagtcaaggaacttctcaactccggcgaac
SBa_000488	Synthetic RBS	gcgcgtgttggtaaacgcaaatcggctggggcagctc
SBa_000489	Synthetic RBS	catgttaatagcacccaaaagagtgatgatagctcattgggtgtaacggtagaacatttgaat
SBa_000490	Synthetic RBS	cgaaggggtttttccATGaaacaagtgtactcggcaacaacggcagcaaccggcctggaacgga
SBa_000491	Synthetic RBS	tggttctggaaagcagatgatgcggtttctcagcggatttctat
SBa_000492	Synthetic RBS	gagttactggcctgatctctccgcttctaatacgcacagcggcaggagcctcaactgcacac
SBa_000493	Synthetic RBS	aaggaacactATGagcggcctctgatcaacgataactaccctggctgagggcgaacaaactccg
SBa_000494	Synthetic RBS	gggttagcgttctgtactcaggaagaagtggccatc
SBa_000495	Synthetic RBS	gactcaacacagctaggagcgtgaagctgattctcttcgattgagaagcgaagactagcttcaag
SBa_000496	Synthetic RBS	ggcaattatagATGgagtggttttaacagattccgggtgtagaagaattgcccagcgtgaatc
SBa_000497	Synthetic RBS	cttcttcaagtcttccgggttcaataccagcggaa
SBa_000498	Synthetic RBS	tccgaaacaagggcagttccgggctgaaagttagcggcgggacaagctccggtatttcaaacggcct
SBa_000499	Synthetic RBS	aggggtgtggATGcccggcaaatcaacttctctgaagaggtccggctgagttgcccagattc
SBa_000500	Synthetic RBS	gtaatgatggcagcgtgggggttttgcggcaggtggc
SBa_000501	Synthetic RBS	acgcccagcgtagtggagacatacagcttctgttgggtcaactcagagctggaagttatccggag
SBa_000502	Synthetic RBS	gaggtctatATGcaactcctggcagcggctctgcccgtcaacggccttctcagcggcctggaac
SBa_000503	Synthetic RBS	cgtgatccggctgctctcagcccagcggataccoca
SBa_000504	Synthetic RBS	gcaagagctctcaactttagggaggtttctatATGggcaacatcggcactctcttcttggtaacg
SBa_000505	Synthetic RBS	gatccggcaaaaacggcaagattggcaagatgattcaacaacagctggggagctggcc
SBa_000506	Synthetic RBS	cggacttccagagaagaagactactgacttgagcgttccctctctgttaactcaactcaactcaat
SBa_000507	Synthetic RBS	catagggagggtaaaATGactcttctgtctgctgtttctggggtaaaagcgtggcctccggcgg
SBa_000508	Synthetic RBS	atgactccggcctgactcggctgggtggcggcgaacaggaagct
SBa_000509	Synthetic RBS	aaactacagatttgaggttaaaccaataaagcaagtagtggcaattcaagaggagaactcaagcAT
SBa_000510	Synthetic RBS	Gccgcattggactgggtgctgcttctgtggttactctatcagcggcgaaggcagcttccg
SBa_000511	P _{lac} Promoter and	caatttcgctcaactcaactbaattggcttgcgctcactgcaccgcttccagtcgggaaacctg

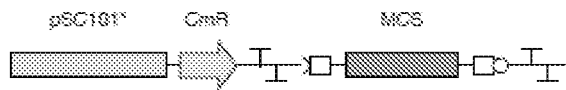
Synthetic <i>nifJ</i>	<p>gcaagggctctgatgttcaactcagccgaatcaggctgctggatcacatggctaaagctgatttggga atctgtgagcgtgacgggtgtagctgggttgaagcaatgaaactggcttaagtctgctgctaaag ctgctggctgctgaggtcaaaagataaagaagactggagagattcttcgcaagcgttgcgctgtgc attgggtaacctcctgatggttccctgtgtggcaggaatacaagaagcctattcagacggcgtt gaaactgatgttccctcggctcagttccgcttacagcctaccatcaacacccaaagataagcag attgatgcaacaacacagggagtctgggatacgtctcaactttgtacacagccaaagcggtagcc accttcgtaagactgtatgtgtggcacaacagagaagtaacggaatcgaatcttttgcaactgatca cgaactcctcggtaagattccggctgacgctgcaaacctgttcaaaagcagtgccgcaaacattg gttgacacataagctctgtgtagtactgctgattctcagcagcagttccgctgacacagttgc acgagctcacaattggcaaaaatgcccagcactccggctaaaggtaacctgaaacctccgtgacat cttagagtcggactccgcttccgctaa atgaaaaactatggacggttaacgctgcccgtcactggaatagctacgcttaccagaagtgctg cgaatcaccgatatacggcagcaccocgagtgccggaaaaatgtgggacgaaatgggctgcgacggg caagaagaaactcttcggccagcgggtgcccctgatggagatgagctcggaaagcgggtgcaaga gggtgctgtgcaatggccttgcgaagctggcgaactgacgaacacactaacccgctgcgagggcc tgttgtgatgatcccaaacatgtacaaaaacgggggtgaaactgctccgggtgctcttccatgt ttcggcagcggcactggccacaaatagcctcaacatctttggcgatcactcaggatgttaaggcg gtgcccacaaacgggctgcgctgctgtgtggcagagaataacgtccagcaacacccgctcactccg cgttagcccaacttggcagcgaatacaaggtcgcactccgcttcgtagaacttctcctgagcttccg caccagccacgaataccagaagatcagaggttctgggaatatagaacagctggccacacttctgtgat cgtccggcctggacagcttccgctgtaacgcttcaaccggacacacccgctcactccgctggca ccgcccagaaacccggacactcacttccaggaaactgagggcggtaaacgcttctcactcagggcct ccggatattgtggaactctacatgacccagattctcgcctgactggtccgagatacactg tttaactacactgctgtgctgggtagcggagcggctgactcagcagatgggctctgctgtagaca ccgtccaaagaggtggttgcacacgctgaaatgacaggggtgagaaggtgggtctgctctccgctca tctttccgcccgtttctgttagcgaactctcgcacactgcgcaaaaactgtaacagcgtatc gcagatattggaactgacgaagaagcagggctcagaagcagcagcagctgctcagatgtggaaga atgctcttaaccaactgacagatgcccctgtgattgtgggtggctcgtatgcttggcgggttaa ggaaggtgtgcccgaacgataatggcggcctgtttgataaacctgaacaaacccgctcggatggac ggcttcaagctgggtatcgtggacagatgttacctcaactctctccgcccagcagcagacaacc tggcgggtttctcagcagcggcactcagggcactgtaagttttggggcactgggctccgacggcaagg tggctgcgaaacaggtcccgatcaagaattatcggcgaacaaacccgctcgtatgctgcgcaagctac tttctcactagcactcagaagaagaggggtggtatcaactcagcactctcggcttttggtagctgccc cgtacaactcccgctattgatcactcggcggatcttcaactcctcgtgcaagcagcaaaagctatgt tgaacgctacgactcgtgctgagtgcccttaaacccgggtggcaacttctcgtgcaactcctcgg agcagatgcccgaactggagcaacactcggcggctgggttcaaacgcttctcggcagcagagaata tcaactctcaactctcaacgctgtggacatacggcctgagcttgggttgggtggcggcttcaaa cactcgtgatcagcggcctctctcaaacctggcggcagatcattgacccgcaactcgtgcccagc tatctgaagcagctgttgagaaaagctatggcagcaaaaggtgcccgggtcactcagagatgaacc agcgtgcaactcagacttggcattggcagcagcctgcaacagggtagcagatcccgcaacattgggcccac cctggatgagccagcggcggcagggctcggcagatgagcaggacttctccgcaactcctcgcaca ccgatgaaacgctcagtgccggcagccagcttccggctgctggcttttctcggcagtggaagatggca ccttccgctccggcagcggcagcagggagaaagcctggcactcggccttagaggtgcccagcttggca gcccgaaggtgcaagcagtgcaacagctgcccctcaatttctcggcagcagcagcagctcgtccg ggctgttgaatggcgaagagcactgatgctgcccggctggcctcgtgagcaaacggccacaag gcccataaagaatacactatcactcggcagattagcccggctggactgctccggctgctggcaactc cgttgaactttgtccgctcgtggcaagcgttgaagatgcaactctcggatgcccagcagcagcag atggctccggctgtgggattatgcccggcctgaccccgaagctcaacccgcttctgtaaaaaca ccgtcaaaagctcgcagctcgaaaacccgctgctggagtttagcggctgctgcccgtggttggctg cgaaaacccgctatgcccggcctcaatcccaagctgttttggcagcagcagcagcagcagcagcag accggctgttccagcactcggggcagctcgtgcccagcagcactccgctataccacaaactcactgctg gtcactgctccggcctggggcaatagcctgtttgaggaacaaatgcccgaatttgggttttaggtatgat gctggggcggctcaagctcgtgctcaacagatcggcgaagatatacagcctgctgcttagcctccg gcttccgatgagctgagcagcagcagtgcccagctggttgggaaacaggaagcgggtgagagca cgcgtgagcgtgcccagcctctgagcagcagccttagcccgggaaagagagggccttccgctgtt agagcagctgtggcaaatcgtgatctcttggctgctgcaagcagctggattttcggcggctgac ggctgggctatgatattggcttccggtggcctggaccagctcctccacagcgggtgagagatgta acattctggtatttgacacccgaagctcactcgaacacccggcggctcaaaagcagcaaatcagcccc ggctcggcggcactcgcacagcttccgctcagggcagcagcagcagcagcagcagcagcagcagcag atggcagatgagctacggcaacgctcactgtagccacaggtggcagtggttggcggataaagataaaa ctctgcccggcactgcccgaagctgaagcgtggcccagggcctcctggtgatgctgctatgccc ctgcaataatcagcctggaagcggctatgctgtgcaagcacaagctgagggcgaagcagcagcagc gagggcgggctactggcaacctgtggcgttatcaacggcagcggcgaagcggaaagcagcagcagc ttatgttagatagcgaagaacccggaagagctggttccgtgacttctcgttgggtgaggtgcccata cgcactcctgcaacaagaaccccgcaactcggcagctgcccctttcagcggtaacgaagaagatg ggcggctgcccagcttggcgaatacagctcggcctggtggcgaagagtaaa</p>
Synthetic <i>nifH</i>	<p>atgacctgctcagctcggcgtattatggcaaaaggtggtattggcaaaagcagcagcagccaga acttgggtggcggcctcggcggagatgggtaaaaaggttatgattgtgggttgcgacccgaaagcc cgaacagcagcagcagctgattctgcaagcgaagcaaaaacagcttatggagatggctgcccag gttggtagcgtggaggtatctggagctggaggaagcttctgcaaatgggttaagctgatgtctgct gcccagagagcgggtggctcggaaacaggtgtcggctgtgcccggctcgtgggtgatcaagcctat caatttctcggaaagagagggctgctgacgaagatgatctggatttctgcttctcactagctgctg ggtagctgctgctgctgcttctgcaatggcgaatccgaggaataagggcacaagaaatctaca ttgtctgtgcccggagatgatggcaatgtacgctgctcaacacactcagcaagggatactgttaa atacgcacaaagcggtaaggttccgttgggtgggtttgatttgcaacagcagcagcagcagcagc gaggaagcaactgactcagcctggctgagaaactgggcaacacaaatgactcactcctgctgcccac cggataaatattgtccacgtgcaagaatccgctgtagcagctcactgagctgtagcaagggatactgt caagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag ccgacccag atgatgactaatgctactggcgaacgtaacctggcactgattcaagaagctactggaaggtgttcc</p>
Synthetic <i>nifD</i>	

	<p>cgaaaacccgcccgaagagcgcgtaaacacatgatgggttctgaccgGaaatgGaatctgtg gggtaaaatgcaatcctctaaatccgaaatctccagccgggtgctcagcctgttctcgtggctgtg taocgaggttctaaaggtgtggtatccggcccgatcaaaagatagggccatctctctcaaggcc cggTaggctgtggccagtaactctccgcccgaagcctgtaactacacagggcgttctcgtggct tgaactctccggccagctgaaactcaccctctgactccagggaacgugadaacttctcgtggc gataaaaagctgtccaaactgatcgaagaaatggaactgctgttcccgctgactaaaggcatta ctatccaaagcgaatgtccggtgggtctgactcgggtgatgacacagccgctgcccacacgac ttccaaagccctggataagccgggtgatccgggttctgctgagggcttccggggcgttctcag tctctgggtcactacatccgcaaacgactgttctgctgactggatctctgaaacaaacggtgaaagc agccttctgaaacacccctctatgactgctgcttattggcgaatataaacctggcggcgaagc ctggccatcccgcatctgctggaggagatgggtctcgtgctgtgtcgcacagtggtctcggcag ggcaccctggttgaatggaaaacaccccttctgtaaacctgaaacccctggttccactgtaacgct ccatgaaactacttgcctcctacatggaagaaaacatcagatcccttggatggaatacacaact cttccgtccgactaaaatccgagaaacccctgctgtaaaatccgcccagctggttctgataccat cgcgcgaacgctgaaagcagtaattcgcgcctacgaaggccagatggcagcactcaatgctgaag accctccgcccctggaaggctgtaagctgctgctgacatgggtggctcctcgtccacgctcagct gatcgggtccctacagggactcggccatggagatcactcgcagccggttccgaatctgcaacac gaagctatgatcctgactgcccagacctgaaagaggtacgctgctgttctgacagcagcagct cttatgaaactggaagcctctcgtgaaagcctgaaacacagacctgactcggctccggcatcaagga aaaatcacttctccagaaaatggcgcgtcccttccgcagatgactcctggacatactccggt ccgtacccagcctacagcgggttctcgtactctcctcctcctgactggatcctgactgaaatacc cagcgtggatgaaactgaccgcaacgctgggtgaaactcctgataa</p>
<p>Synthetic <i>nifK</i></p>	<p>atgtctcaaaactatogataaaaactcaactctgttaccctcgtctcagcagggacgaaataccag aactgttccgttaacaaacgtcagctggaagaaagccacgacgcaacgcccgtgcaaggaggtgt cgcactggaccaccccccgggaatccgaaagctctgaaactccagcgcgaagccctgacggttct ccggcgaagcgtgcccagcctctgggtcgggtctcgtgacgctgggttctgcccacacccctgc cgtatgtccacgggttccagggctcggctagcctcactccgttacttccacacccctcttaa agaacacactcgcctgctgtccgacagcctgacggaggacgcccagctgttccgttggtaacaa aacatgaaactcggcctgcaaaaactgcttccgactgtacaaacccggaactcactcagctgtcta ccaactgactggcagaggttattggtgatgatctgcaagcatttattgcccacgcaaaagaa cggcttctgtgacagctctatccgggttccgcaacgctcactccacgcttctccactctccac ghaactgttgggcaacatgttccgaggttccgcaaaaacttctccgcaagcctatcaagggc aacgggttaaacctgcccagctgaaactgggtgacggcttgaacacactccctggcaacttctc tgtcctgaaagcctgatggagcagatggcgggttccgtgttctcctgctgctgacccgctctgag gttctggaacctccagggcagccactatccgactgtattctgggtggcaccactccagcagaa tgaagagggcccagacggatgaaacccctgctgctgcaaacctggcagctgtgaaagcaa gaaagttgtccagaaatgtgaaacccagccggcaacgggaagtgcactcccgctgggtctgcca gctactgacgaactgctgatgacgctgtcccaactgagccgcaaaccaactcagcagctgctgga ccctggaacggcgtccctggtggacatgatgctggacagccacagctggctgcatgcaagaa atttggcctgtaccggtgacccggacttctgaaatggcctgacccgcttctcctgctgaaactggc tgcagcogactgttactcctgtctccaaacgctaacaaaacttggcagaaagccactgaaacaaa tgcctggtgagcagccacacagccgctgtagcgaaggttccactcaactgacacctggtgcaatt ccgctctctgatgttctcgcctcagccggatctcactgatcggtaactctcagcgaactcactc cagcgtgacactctggccaaaggcaagcgttgaagtgccgctgactctcctcggcttccgc tgtccagcctcaaacctgcacccgcaagcaaacctgggttacgaaggcggatgaaacactcgt aactactctggtaaacgactgaaagcgtggcagcagcagactccacagctgggcaaaaacc gactattcttccgactcgggtcgttaa</p>
<p>Synthetic <i>nifY</i></p>	<p>atgtctgacaaatgataccctgttctggcagctgctggcgtgttctcagctcctgcccgaatttgc agccggcccaaatcgtcagatggctggcgcaggaatccggcgaacacccctgacgcccggagcct tgcacccctgacccaaacccgcaactccggcgtcgttcccaacccgcccagcagctgtagccccc gctcgtggagccggcttatggcttctctgcaaggccctccagcccaacttggcagatcgtac gtccggcgcagcgttcccccacactgctcgcgcgcttctgagcacaagcagccctgttctca tggctcatttccgcccaggttctctgcttctcatttaccgcttctgacgagcagcagcagcagctg tactgacttggcgcgctatccgagccgcaacccgcaacagcaggaagcgaatgaggtgctgctcgtc tgaattgaaagtgcagcctgctgttctgcccaggaatggcggctccggcagcagcagcagctgat ccgcccacgcaatccatccgatgaaaggcagccgggtactcagatccagggcagctgtaagcct atcaacacccctgctggcggctcctcgcgcgctggctcgcgaacagcttgaacccgctgataacc cgtcggagagcgtgtgtttaa</p>
<p>Synthetic <i>nifE</i></p>	<p>atgaaaggtaacagagatcttctgctggaagaaacggcctgtgaaacaaacatataacaga aatccgctgttagcgcacccaaagccgggtgagcagccggctggctgagcttctcagatggtgccc gatccactgctccgattgaggagcttgcacacactcgtgactggcccaactgggttggcaggt agctcttgggaacacccgtggcagcctccagcggctccgacccctgactcgttgggttatacca ctgactgaaatgaaacagatgtgatcagggctcgcgcgagcagctcctcgttccacactgctg ccaatttctcaccgcttaccacccagcggcagcttctcactcaactacgtcgtgcccggctatg gaaaggatgaaactggaggccgctgtgacagcagcagcagcagcagcagcagcagcagcagcagc caatttgatcggctggcttctcagcttccgaagaaactgggcaacccctccggcagccagatgct ggttaaacgcgtctatggcccaacgtgagccagcgcctggcggcagagcaacccctgttgcaccc gagcaacgtctgacatctgcttgaacgggtgagctcaacatctggggcaggttgggcaacttcc agcctgcttctgagcctgggtatccgcttctgggttctcctcagcggcagatggtcgttctgc cgagattcaaacactgacccgtgcccagcgaacatgctggtgtgagccgctgctctgatcaat gttgcgcgtgctcttggaaacagcctatggcaccocgctgggttgaaggctcgttctatggtaacc ggcagaccagcagcagcctcgcgcagcttagcggcctgctggggcagtagcaacacccctcagc caccgagggcctgatccgctggaagaaacggcggctgagctggcctgcaacccgtggcgtgaa cagctcgtggcagcagccctgctctacacgggtgggtcctcaaaaggctggcggcagcagcagc cgttccaggtatcgggtatgacccgtgtgcccagggcaccgctaaagagccaggaagagataa acagcccatccgcaatgtagggcgaagagggcctgactgctgaaagagggcaacagcagctacc tattggatgtagttatcgtctcaagcagacctgatgattgcccgtggcggcagcagcagcagc ccgctcaaaagcgcgcttccgcttctggacatcaaacaggaacccgagcagcagcagcagcagc ctaaccaggaactcgtgactttagcgcagcagcagcagcagcagcagcagcagcagcagcagc cagactcattccggcgaacgggtggcgttaa</p>
<p>Synthetic <i>nifN</i></p>	<p>atggcagacatttccgcaactgataagccgttggctgtgctgcccagtaaaagccggcagcagc</p>

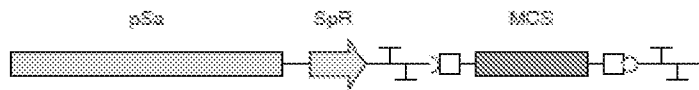
		<p>tgggtggatcctggcgtccctgggtatcgagcaactcgattccgctggtaaatggcggcaggg ctglttcggcttttggcaaggttttttattatccagcaactccacgatccgggtcccgctgcaaac acggcaatggaccgacccagcaaatcaatggggcgtgatggtaaactctccacggctggaca ctctctgccaacgcaataacccggcaagcaattgtgctgctgagaccggcctctccgaggcgca gggcagcgaatattcccgctgtatggcgtcagttccgtgaaagaatcccgctcaataaaggcgtg gcgattctgactgttaacaccccggaactttaccgtagcaaggagaccgctttccgctgtcc tggagctctgtgattgaaacagtgggtcccgccagccccacgtccggcgcagcgaatcgtccgct caatcttttggtagccatctctgtagcccaggcgaattttagtggctggcggcttggctcgag gccctccgctccgcaacgcaatctctgcccggatctggctccagagcaaggacggcaactctc agggtagctttccgctgctgacgacggggcgccacggcttggcccaatccagcagaaggcga gagcctttgctctttggcattggcgtcagctccacccgtggagcagccctgtggctccgct tgcgtggcgaagtcaatccgcttggcccaactcaatgaactggaaacgctggcagcccttctcc atcagttggcgaataatcagcggctccgcccgtccggagtggtctggaaacgcccagccgctcag gcaagaccgcaatgatcgattggcaacatgtggctgcaaggccagcgcattggcattggccggaa ggcaactgctggcagcgtggctggatttccgcaactctcaaggtatgacagccggctccatgg ttgctccgacgggtcaatccgagcctggctcagttggcgggtggagccgctggctggcgggtgat ggaggatctccagaccctcttatggcgaatccggccgacttactggtggcgaactcccaacc cgtgatctccagcaaatccgctccgctggctggcggcaggtctccgctgtttggaacaac tggcgaatcttccgctgttccgcaagggttatagcggatagcgtgatccctgtccaggttggc gaaactgatccgtgaaacgcaatcaatctggctcaattatccgaccccgctggccagcaacca gaatccctcgtttgtctaccgggtggcggctaccgacggattaa</p>
	<p>Synthetic <i>nifU</i></p>	<p>atgtggaaactacagcggaaagtcaaggaccatttctcaatccgcccacacggcgtgtgtgg ataaccgaaatggcggcggcgaagctccgacgcttatcttggcgaagctctccgcttgatgct gccgctggaaccgagagcgaataatcgaagaagcgggcttcaagaactgggctggcggcagc gcgattgctgctccagcgcactgacggagctgatcattcgtccacccctggcgaagcgggtc agatcccaacccagcagatccgacactatctggacggcttaccggcgaagaagatgcactgtctc tgraatggccaggaagctctctcgtggcgaatctgtaacttccgctggaatcgcctggaaag gagcaatgacagggtaagctgatctgcaagtgcttccgctgtagaaggcaatattccgctg ctgtccagcaacaacggctctaccgactctggcgaaggctcaatcaacgaaggcaggtggcgg ttgtaccagctgcccagaaaaatcagactggcctggcggagattctccgcccacagccgca accaccccgccagttccgctccggtaaaagatccgcaactggcagagcctcgtggaataccatccgctg aactggctcccaatacccaaggcagcggctgtgcaatggcctgtttgtccgctggaaccaca agtgaactgtttccgctgctggcagcgtgtctggctgcaatgacagcaatcgaactggcgtgg ctgcaacagaaattgatggagcgtaccggctgctatattggaagtctgttggcggcttaa</p>
	<p>Synthetic <i>nifS</i></p>	<p>atgaaacagctgtaccctggcaacaacagcgaaccaaccgctgggaaccgctggatggtaagcga tgaagcgtttctccagatttctatggcaactccgtccagcaatccatgacttggcaaccggc accagcggcctggaaacgtggcaccagcaagctggcgaactgctggcgcagagataccgctct gaaactcaatttccagagctgtggcagcggcgaactgcaaacggcaatggcctccgctcagct tatggcggaaacggcgaataatcaactcggtagtggagcaaccggctaccgctggcggcgtg cagacacctggaaacgcaagctatccatccatccgatttggcgtggatagcgaaggctgctgctg gcaatggcgaactccgctccgctcagcctctccgctgctgctccgctggctgctgctggctgca accaagcaaacggcgtgctgttccgatggcgaataggcagacttggccacagcagggcgc tctgttccactggcagctccgctcaggtcgttggcaaaatcccaattggcttggcgaagcgc atcgaatgctgtctcctccgcaacaagtctcaatggctccgaaggctgttggctgtctgact taactgctggcagcggcttccgctccgctgcttccgctggcgaatcaagaataatgctccgctg cggcaatgagaatctctggcactcctcggcaatggcggctcctggcgaactggcgaacatccat ctgcccgttatgacccatattggcagctaccgaatccgctggcgaacggctcctgctccagcgt tgcctccgctgatggttatggcggctgttcagccgctgtaccgggtatgctcaactggcgtt cgaatttatcgaaggctgaagcgaatccgctctctgctgaaacagcctggcaatggcgaagctcc ggctccgctgtgtaactctggcagcttggagccagcaatgtagtggcggcagcaatccatccat acccggcctcagcgaacatctgttttggctgagcggcttatccgcccagaaagagatcga ctacgtcgttggcaaccctccgcaactcaatgatcgtctgctgctccttggcccaatctggcag aattggtaagccggctccggcagatggcagctttaaaccggctttaaaccggctaa</p>
	<p>Synthetic <i>nifV</i></p>	<p>atggagcggctctgatacaacgatacctccctggctgtagggcaacaatcccgggcgttagcgt ttctgactcccgagaaagtggcaatccggcagcgaactgtacgctccgggtatccacccgctgga agtcggcaactccggcgaatgggtgtagaagagatccgcccgaatccagctggcggcgtcgtcaactg ccggaacggcagcttatgacccctggctgctggaacgctctggaactccgtccagagcggcgaatc tgggtatgactgggtggaatactcagatccagcaatccgaacagctggctcagtaacagctggc tgaagcctggcggctgctgctggagcggcttggcagtttatccatctggcccacaagctgaagc ctcaagatctgatttggcttggcaggttggagccgctggctctggctcagacccctggcggcaatg ccgaggtggcccagcaatggcggctggcggcttggcttaactgacacccggggcctgctgga cccgttcaaccaaccagccagatcagcggcctcgtgacgtttggctggcggagatcgagatg catgctcaaatgatctggcctggcctaccgcaacaacgctggcggcagcttccgctggcggca cgtcgtgaaacactaccgctccctggctgggtgaaacgtggcggcaacggcggcctggaaacgg tgcctggcggcctggaacgctgctggcggctggaaacggcggctccattccagcggcctccagcgg ctctgtccagcggctccgagggctggcaacagcggcaatccgcccgaacagcggcctggctgggtg aatggctttcaaccagagctcctggttccagcttggcggcggctgctggcggcagcgaactca tcaatctatggcccagcctcaatggcggctgactaccgctcgtgctggcggcggcggcggcggc cgtccagctgtcaacgggtgtttccgacagatgggttaccacctggaatggcggcggcagatcaatc agttgctccggcgaactccgcttccgcccagaaattggaaacgctctccgaagaactcaagaact ggttggatctatgacgaatttggcggctggaatccgccccttggctgctccggcggctaa</p>
	<p>Synthetic <i>nifW</i> Synthetic <i>nifZ</i></p>	<p>atggagtggttttaccagattccgggtgtagcaaatggcgaacgctggaatccctcttccagtt tctccggttccatccagccggaactgctggcggcggctgctcgttccgctgtttagcagcgtt ccaaccgcaactggctggcggaggtcccgctgcaaaaacggctccggaacaatgatcgtggcgg tggctcttggcggcggcggcctcctggccgaatcttaccagcgaatccaggaagcggcggcaact aa atggcccgaatcccaactctctgaaagaggtccggctagttccgcccattcgttaactgagggca ccgtggcgggttttggcggcaggtggcggctgctgcttccgctggcggctccagggccttggcggcga ctgggggtggttctccagagaccagatcaatctcaaatccacttccggaaacggaacggcatt atcgtgctccgagcagggttaaccagatcccaagcggctggctggcggtaactccagtt atcgtgacagcgtccagtcgaatggcaactggctgcaacggctgagctggttctggagcggcgg</p>

Synthetic <i>nifM</i>	<p>tcaacgtggccgtgtggagccactgatcgtggcgaacttggcgatbcctacaacgtggacttcc agcggccgttgggttcccggttccgggtccaggccactccggcctgatctgaagagcggcgaagaataaa atgaaatccgtggcagcgcctttgcccgtcaacgccttgcctccagcccgctggaaacgggtgatccgg ctgctctcgaaccagccgatcccccagcgttcgagcagcgttggcagcgtcaatgcccataatgga acaacccatcgtagccggtgtccgggaaggcgatattccggctgccttactggaaaaatccgg gccagcctggcgaatctggctggacgagggtgacttcgctccggagcgcgctcggatctgtgc gtcactcatgcaagctcgtggagctggcgtttgcccgaatattgcccgcagcgaacggcaacgggatct gagcaaggttcaaacgctggtaattcgcctcaaccagactcaatccatgctccggagcagcgcgtctg accgctcaactgctccgagcgttgataatgactggcgaaggcgtgcaatcaacgcaatccctggcc tgtatcgtcagatcaacgcgagcgcgtgacgccttcccccactggcaacagcgcactctcattg ccgctccgcttggagaaggcgcctctggctggatctccgctggctctgctgtaaccgcagctc gaaacgcgcttggtttagcctggcgaacacgcactgctcgtcgcgattgctcggaaattgggtt ggcaactgttattgctggagggccatctcggcagcccccagctggagcggcaaacggcccttga atctgcgcgcgactacttgtggcagcagaccagcgcgcgaacagcgtcaatggctggagcag atgatttcccggccaaacgggcctctgtgggttaa</p>
Synthetic <i>nifF</i>	<p>atggcgaacatccgcaatctcttggtaacggataccggcaaaaacccgcaagattggcaagatga ttcacaacacagctggcgagctggccgatgcccgggttaacatcaatcgtaccactttggatga ctttatggcttaccagctccgttgcctggcaacgcagcgccttggctcaactcctccggcc ttagagcggcgtcgcagagcgaagcgtggctgagtttatctccggtctggatgacgcttccc tgaaggcgaacacgcctggcgccttggcctggcgcagaccagcgtggttaccgggacaactcgt gtcgggtatggctccgctgttcgacgcctgagcgcgcgctggcgcgcagatgattggtagctgg ccgaaacgaaggcttatgagtttagcgcctcgtccgcgctggaaaggcgaacgccttccctcggcttgg tgcggatcaagaatcagttcgcaccagaaccgaagcgcgcctggcgtcttggcttgaagat caaacgcaccgcttctgtaa</p>
Synthetic <i>nifB</i>	<p>atgacatcttcttgcctgcttctggcggtaaacgcgtgcccgtccggccgcagatccgcgctga ctccgctggcggcgaacggcagctggcgaacccgctgctatagccgcacggccatccacgctt cgcgcgtatgcactgcacgtcgtccggcctgcaactcaaatgcaactcctgcaacgcgaag ttcgattgacgcaatgaaagccgtccggcgtgtcctctacccctgtagcgcggaaacaggctg tggtagaggtgcgcagcgtgcgccagctatccgcagcctgctcggctggctggatctgctggctg ggcggatccgcttgcgaatattgcgccaccttcgctaccttggagcttatccgcaacagcttgc ccggaactgaaactgctgcctgagcaaccaacggcttggctgctgcagatgcgcttgcctgctgc tcgactggcgcgtggatcacgttaacgtcaacttaaacacctggacgcgaacatccgcagcga aacctacgcctgggttctggctggatggcgaacgcctacccgctgcgcaagcggcgaatctcc attgcccgcagctggaaaggctacgtcgcctgacgcgcaaggtgtgctcgtcaagatcaaca ggctattgattccggcactcaatgacagcgcgatggcgggtgttagcgcgtgcctgcgcgcgctc tggctgcgttcatcccaacatcatgcaactgattggcgcgtccggagcaatggcaactgtttccggt ctgaaacggccagcggaaacgggacgcggaaacccctggcggcgaacgcgctcccgctgcggcagg ttatgcccacaaatgacccactgcaaccagtgccgctgcccgaacggatggcaatgcttggtagga tcgctgcgaacagtttagcaataccggctccggagctccctccggcctggctgcgactcctg catcagcgtgctcagttgcctgcgagcaatgcgcaacgcgcgctgagagcgaagcggatgacgcct gctggctggcgttgcctgcgacccgtggcgtatgtaattgactgcactttggcgaatgcgaacg tttctatatctatagcctgtcgcggctggatggctctggctaaacgaacgcttcccccgaaa tactgcccaggctgcgactgcagcgcgcgaggaacaatgcgcacgcttctgctccatccttg agctgctggcgaagctcaaacgcgctgtttctgtgtgcctatcggcgaacacccgctggcaacgct ggagcaggaaggaatcgaacgcgtgcctggatggcgcctggcgtccggtatccgaggtcctgcg gcattggctggcagcagcgcctggctgctgctggcgcctgcaattgcccgaacaaaggcgttgcgtaa atgcccacatggactggctgctgctgcttctggttaactctatcagccggcaaacggcagcttcc cgcctcgtatggcttctgcgcgcgtgactggcaagctctgcgcctgcctggcggcaggtgga aacgcgcctggatggcgaacccctgaccgcctgcgcctgctgctgctggcggagctgaatggcgaacgc gaaagaaacggccagcagctgggtgcctggcctggcggctggatgcaacaggaatgcgcgctcga tggcgcagatattcgcagaggtgagcctggcgttaacccaatctctggcaggaacttgcgcctgc gagccgcctgaaatgcctgctgctgctgctgctgcttccgcagctggctgttatgaaacgag caacaatgcgctggaaagaaattctttaccgcagcgttgcctgctgcaacagggcgaagctca tctgctgcagccctcttgcgatgaatgctgggaacgtctcgcctgcttgcgttaa</p>
Synthetic <i>nifQ</i>	<p>atgcccacatggactggctgctgctgcttctggttaactctatcagccggcaaacggcagcttcc cgcctcgtatggcttctgcgcgcgtgactggcaagctctgcgcctgcctggcggcaggtgga aacgcgcctggatggcgaacccctgaccgcctgcgcctgctgctgctggcggagctgaatggcgaacgc gaaagaaacggccagcagctgggtgcctggcctggcggctggatgcaacaggaatgcgcgctcga tggcgcagatattcgcagaggtgagcctggcgttaacccaatctctggcaggaacttgcgcctgc gagccgcctgaaatgcctgctgctgctgctgctgcttccgcagctggctgttatgaaacgag caacaatgcgctggaaagaaattctttaccgcagcgttgcctgctgcaacagggcgaagctca tctgctgcagccctcttgcgatgaatgctgggaacgtctcgcctgcttgcgttaa</p>

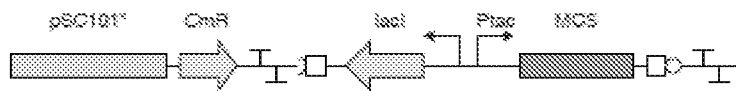
pSB4C5



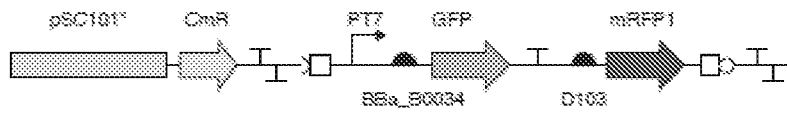
pInoW



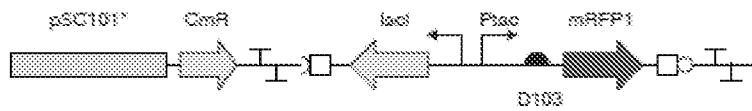
N58: pSB4C5 with tac promoter



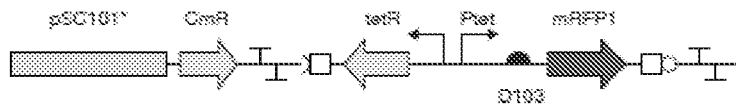
N292: T7 terminator characterization



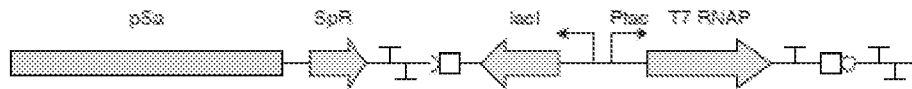
N149: tac promoter characterization



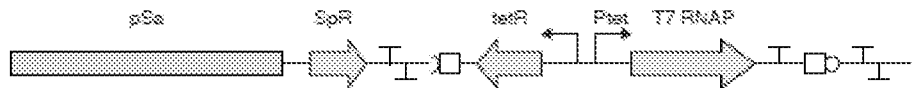
N505: tet promoter characterization



N249: Controller #1



N504: Controller #2



N639: Controller #3

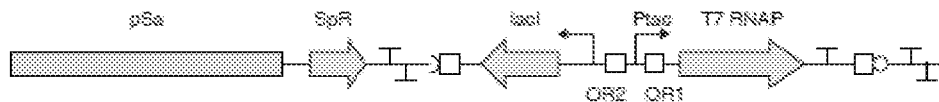


FIG. 39

