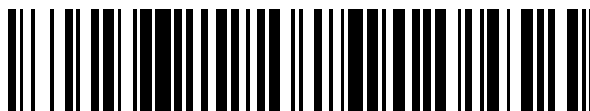


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 752 081**

51 Int. Cl.:

C12N 15/09 (2006.01)

C12N 15/31 (2006.01)

C12N 15/63 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **14.06.2012** **PCT/US2012/042502**

87 Fecha y número de publicación internacional: **20.12.2012** **WO12174271**

96 Fecha de presentación y número de la solicitud europea: **14.06.2012** **E 12800054 (4)**

97 Fecha y número de publicación de la concesión europea: **28.08.2019** **EP 2721153**

54 Título: **Complejos génicos sintéticos**

30 Prioridad:

16.06.2011 US 201161497781 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
02.04.2020

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF
CALIFORNIA (100.0%)
1111 Franklin Street, 12th Floor
Oakland, CA 94607, US**

72 Inventor/es:

**MIRSKY, ETHAN;
TEMME, KARSTEN;
VOIGT, CHRIS y
ZHAO, DEHUA**

74 Agente/Representante:

PONS ARIÑO, Ángel

ES 2 752 081 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Complejos génicos sintéticos

- 5 Esta invención se realizó con el apoyo del gobierno con las subvenciones números CFF0943385 y EEC0540879 otorgadas por la Fundación Nacional para la Ciencia. El gobierno tiene ciertos derechos en la invención.

Antecedentes de la invención

- 10 Las células con programación genética requieren sensores para recibir información, circuitos para procesar las entradas y actuadores para vincular la salida del circuito a una respuesta celular (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)). En este paradigma, la detección, la integración de la señal y la actuación están codificadas por distintos 'dispositivos' compuestos por genes y elementos reguladores (Knight TK, Sussman GJ Unconventional Models of Computation 257-272 (1997); Endy D Nature 438: 449-453 (2005)). Estos dispositivos se comunican entre sí a través de cambios en la expresión y actividad génica. Por ejemplo, cuando se estimula un sensor, esto puede conducir a la activación de un promotor, que luego actúa como la entrada a un circuito. Voigt, CA, describió la refactorización de complejos génicos en "Gaining Access: Rebuilding Genetics from the Ground Up", recuperado de: <http://www.iom.edu/~media/Files/ActivityFiles/PublicHealth/MicrobialThreats/2011-MAR-14/Voigt.pdf>.

Breve resumen de la invención

- 25 La presente invención proporciona un método para reemplazar la regulación nativa de un conjunto de genes asociados colectivamente con una función con regulación sintética, comprendiendo el método proporcionar secuencias codificantes de un conjunto de polipéptidos codificados por genes asociados colectivamente con una función; cambiar la identidad del codón en al menos una secuencia codificante, eliminando así al menos una secuencia reguladora dentro de la secuencia codificante, en el que eliminar la al menos una secuencia reguladora comprende el reemplazo de codones nativos en la secuencia codificante con codones sinónimos no nativos que tienen una distancia máxima de los codones de la secuencia codificante nativa; organizar las secuencias codificantes en uno o más operones sintéticos, en el que la organización comprende ordenar secuencias codificantes dentro de los operones de modo que el gen de mayor expresión (basado en la expresión nativa) ocurra primero y el gen de menor expresión (basado en la expresión nativa) ocurra el último;
- 30 el reemplazo de codones nativos en la secuencia codificante con codones sinónimos no nativos que tienen una distancia máxima de los codones de la secuencia codificante nativa; organizar las secuencias codificantes en uno o más operones sintéticos, en el que la organización comprende ordenar secuencias codificantes dentro de los operones de modo que el gen de mayor expresión (basado en la expresión nativa) ocurra primero y el gen de menor expresión (basado en la expresión nativa) ocurra el último;
- 35 unir operativamente una o más secuencias reguladoras transcripcionales heterólogas al operón u operones, controlando así la magnitud de la expresión génica del operón u operones; y expresar el uno o más operones sintéticos en una célula bajo el control de un polipéptido que se une directa o indirectamente a la secuencia reguladora transcripcional heteróloga.
- 40 En el presente documento se describe un polinucleótido que comprende un operón sintético, en el que el operón comprende al menos dos secuencias codificantes bajo el control de una secuencia reguladora transcripcional heteróloga, en el que cada secuencia codificante está operativamente unida a un sitio de unión al ribosoma (RBS) heterólogo. En algunos aspectos, las secuencias codificantes son del mismo operón nativo y los RBS heterólogos regulan la traducción de las secuencias codificantes en una proporción que es sustancialmente similar a la proporción de traducción nativa del operón nativo. En algunos aspectos, las secuencias codificantes son de diferentes operones nativos y los RBS heterólogos regulan la traducción de las secuencias codificantes en una proporción que es sustancialmente similar a la proporción de traducción nativa del operón nativo. En algunos aspectos, las secuencias codificantes son del mismo operón nativo y las secuencias codificantes en el operón comprenden uno o más codones alterados en comparación con el operón nativo. En algunos aspectos, los codones de una o más secuencias codificantes se han seleccionado para la distancia máxima del uso de codón de una secuencia codificante correspondiente en el operón nativo.

- En algunos aspectos, al menos dos secuencias codificantes codifican diferentes proteínas codificadas por el complejo génico *nif* de *Klebsiella pneumoniae*. En algunos aspectos, las proteínas se seleccionan del complejo que consiste en *nifJ*, *nifH*, *nifD*, *nifK*, *nifY*, *nifE*, *nifN*, *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, *nifM*, *nifF*, *nifB*, y *nifQ* (por ejemplo, en el que las secuencias codificantes son sustancialmente idénticas a las enumeradas en la Fig. 18). En algunos aspectos, el operón comprende secuencias codificantes de *nifH*, *nifD*, *nifK* y *nifY* de *Klebsiella pneumoniae*. En algunos aspectos, el operón comprende secuencias codificantes de *nifE* y *nifN* de *Klebsiella pneumoniae*. En algunos aspectos, el operón comprende secuencias codificantes de *nifU*, *nifS*, *nifV*, *nifW*, *nifZ* y *nifM* de *Klebsiella pneumoniae*. En algunos aspectos, el operón comprende secuencias codificantes de *nifB* y *nifQ* de *Klebsiella pneumoniae*.

- En algunos aspectos, al menos dos secuencias codificantes codifican diferentes proteínas del sistema de secreción tipo III de *Salmonella typhimurium*. En algunos aspectos, las proteínas se seleccionan del complejo que consiste en *PrgH*, *PrgI*, *PrgJ*, *PrgK*, *OrgA*, *OrgB*, *InvA*, *InvC*, *InvE*, *InvF*, *InvG*, *InvI*, *InvJ*, *SpaO*, *SpaP*, *SpaQ*, *SpaR*, y *SpaS* (por ejemplo, en el que las secuencias codificantes son sustancialmente idénticas a las enumeradas en la Fig. 24). En

algunos aspectos, el operón comprende secuencias codificantes de PrgH, PrgI, PrgJ, PrgK, OrgA y OrgB de *Salmonella typhimurium*. En algunos aspectos, el operón comprende secuencias codificantes de InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR y SpaS de *Salmonella typhimurium*.

- 5 También se describe en el presente documento una célula hospedadora (opcionalmente aislada) que comprende un polinucleótido como se describe anteriormente o en otro lugar del presente documento. En algunos aspectos, la célula hospedadora es una célula procariota o eucariota (que incluye, pero sin limitación, una célula de mamífero, vegetal o fúngica).
- 10 También se describe en el presente documento un sistema que comprende un conjunto de dos o más operones sintéticos diferentes, comprendiendo cada uno del dos o más operones al menos dos secuencias codificantes bajo el control de una secuencia reguladora transcripcional heteróloga, en el que cada secuencia codificante está operativamente unida a un sitio de unión al ribosoma (RBS) heterólogo, en el que la secuencia reguladora transcripcional de cada operón en el conjunto está controlada por el mismo polipéptido o polipéptidos activadores o represores transcripcionales.
- 15

En algunos aspectos, el sistema comprende además un casete de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica el polipéptido o polipéptidos activadores o represores transcripcionales. En algunos aspectos, el promotor del casete de expresión es un promotor inducible. En algunos aspectos, el polinucleótido en el casete de expresión codifica un represor transcripcional. En algunos aspectos, el polinucleótido en el casete de expresión codifica un activador transcripcional. En algunos aspectos, el activador transcripcional es una ARN polimerasa (ARNP). En algunos aspectos, la ARNP es la ARNP de T7 o es sustancialmente similar a la ARNP de T7.

20

- 25 En algunos aspectos, las secuencias reguladoras de la transcripción de al menos dos de los operones son diferentes.

En algunos aspectos, las secuencias codificantes en los operones están organizadas de tal manera que las secuencias codificantes que tienen una expresión nativa sustancialmente similar se agrupan en el mismo operón. En algunos aspectos, la secuencia reguladora transcripcional de al menos dos operones tiene diferentes promotores que están regulados diferencialmente por la ARN polimerasa de T7 y en el que la diferente fuerza de los promotores corresponde a la fuerza relativa de los promotores nativos de las secuencias codificantes.

30

En algunos aspectos, el casete de expresión y los operones sintéticos se expresan en una célula. En algunos aspectos, la célula es de una especie diferente a la especie de la que se aisló el operón nativo. En algunos aspectos, la célula es de la misma especie de la que se aisló el operón nativo.

35

En algunos aspectos, el sistema codifica una nitrogenasa. En algunos aspectos, el sistema comprende un primer operón que comprende secuencias codificantes de nifH, nifD, nifK y nifY de *Klebsiella pneumoniae*; un segundo operón que comprende secuencias codificantes de nifE y nifN de *Klebsiella pneumoniae*; un tercer operón que comprende secuencias codificantes de nifU, nifS, nifV, nifW, nifZ y nifM de *Klebsiella pneumoniae*; y un cuarto operón que comprende secuencias codificantes de nifB y nifQ de *Klebsiella pneumoniae*. En algunos aspectos, el primer, segundo, tercer y cuarto operón que comprende un promotor de ARN polimerasa (ARNP) de T7 y el sistema comprende además un casete de expresión que comprende un promotor unido operativamente a un polinucleótido que codifica una ARNP sustancialmente idéntica a la ARN polimerasa de T7 (ARNP).

40

45

En algunos aspectos, el sistema codifica un sistema de secreción tipo III. En algunos aspectos, el sistema de secreción tipo III es un sistema de secreción tipo III de *Salmonella typhimurium*. En algunos aspectos, el sistema comprende un primer operón que comprende secuencias codificantes de PrgH, PrgI, PrgJ, PrgK, OrgA y OrgB de *Salmonella typhimurium* y un segundo operón que comprende secuencias codificantes de InvA, InvC, InvE, InvF, InvG, InvI, InvJ, SpaO, SpaP, SpaQ, SpaR y SpaS de *Salmonella typhimurium*.

50

Las realizaciones de la presente invención también proporcionan un método para reemplazar la regulación nativa de un conjunto de genes asociados colectivamente con una función con regulación sintética. En algunas realizaciones, el método comprende proporcionar secuencias codificantes de un conjunto de polipéptidos codificados por genes asociados colectivamente con una función; cambiar la identidad del codón en al menos una secuencia codificante, eliminando así al menos una secuencia reguladora dentro de la secuencia codificante; organizar las secuencias codificantes en uno o más operones sintéticos; unir operativamente una o más secuencias reguladoras transcripcionales heterólogas al operón u operones, controlando así la magnitud de la expresión génica del operón u operones; y expresar el uno o más operones sintéticos en una célula bajo el control de un polipéptido que se une directa o indirectamente a la secuencia reguladora transcripcional heteróloga.

55

60

En algunas realizaciones, el polipéptido es heterólogo de la célula.

- 65 En algunas realizaciones, la provisión comprende obtener las secuencias de nucleótidos del gen y eliminar secuencias no codificantes.

En algunas realizaciones, el conjunto de genes es de un complejo génico. En algunas realizaciones, el conjunto de genes son de un procariota. En algunas realizaciones, los genes son de un operón nativo.

- 5 En algunas realizaciones, la al menos una secuencia reguladora se identifica usando el cálculo. En algunas realizaciones, el cálculo comprende búsquedas de secuencias codificantes de sitios de unión a ribosomas, terminadores y/o promotores.

- 10 En algunas realizaciones, la eliminación de la al menos una secuencia reguladora comprende el reemplazo de codones nativos en la secuencia codificante con codones sinónimos no nativos. En algunas realizaciones, la eliminación comprende seleccionar codones no nativos que tienen una distancia máxima de los codones de la secuencia codificante nativa. En algunas realizaciones, la eliminación comprende seleccionar codones no nativos para una expresión óptima en una célula hospedadora.

- 15 En algunas realizaciones, el método comprende además identificar y eliminar uno o más de los sitios de inserción de transposones, sitios que promueven la recombinación, sitios para la escisión por endonucleasas de restricción y sitios que están metilados.

- 20 En algunas realizaciones, la organización comprende agrupar secuencias codificantes en operones basados en un nivel de expresión nativa sustancialmente similar.

- 25 En algunas realizaciones, la organización comprende ordenar secuencias codificantes dentro de los operones de modo que el gen de mayor expresión (basado en la expresión nativa) aparece primero y el gen de menor expresión (basado en la expresión nativa) ocurre el último. En algunas realizaciones, la organización se basa en la expresión temporal nativa, la función, la facilidad de manipulación del ADN y/o el diseño experimental. En algunas realizaciones, la magnitud de la expresión de las secuencias codificantes corresponde sustancialmente a la proporción de proteínas codificadas por las secuencias codificantes medidas en el sistema nativo. En algunas realizaciones, la magnitud de la expresión de las secuencias codificantes se determina mediante cálculo. En algunas realizaciones, el cálculo comprende un algoritmo de optimización numérica.

- 30 En algunas realizaciones, el algoritmo de optimización numérica es un algoritmo de Nelder-Mead, un método de Newton, un método cuasi-Newton, un método de gradiente conjugado, un método de punto interior, un descenso de gradiente, un método de subgradiente, un método de elipsoide, un método de Frank-Wolfe, un método de interpolación y métodos de búsqueda de patrones, o un modelo de colonia de hormigas.

- 35 En algunas realizaciones, la secuencia o secuencias reguladoras transcripcionales heterólogas comprenden un promotor o promotores ANRP de T7.

- 40 En algunas realizaciones, la secuencia o secuencias reguladoras transcripcionales heterólogas comprenden un promotor inducible.

- 45 En algunas realizaciones, el método comprende además unir operativamente un sitio de unión al ribosoma (RBS) heterólogo a una o más secuencias codificantes en el operón sintético. En algunas realizaciones, diferentes RBS están unidos operativamente a diferentes secuencias codificantes. En algunas realizaciones, los RBS regulan la traducción de las secuencias codificantes en una relación que es sustancialmente similar a la proporción de traducción nativa del operón nativo.

- 50 En algunas realizaciones, el método comprende además unir operativamente una secuencia de terminación transcripcional heteróloga a una o más secuencias codificantes en el operón sintético. En algunas realizaciones, los terminadores son terminadores ARNP de T7. En algunas realizaciones, los terminadores para diferentes operones son diferentes.

- 55 En algunas realizaciones, el método comprende además unir operativamente secuencias amortiguadoras entre dos secuencias funcionales en un operón en el que las secuencias funcionales se seleccionan del complejo que consiste en un promotor, sitio de unión al ribosoma, secuencia codificante y terminador. En algunas realizaciones, la secuencia amortiguadora se selecciona del complejo que consiste en una secuencia aleatoria, una región UP de un promotor, una secuencia 5-UTR extendida y un sitio de escisión de ARNasa.

- 60 En algunas realizaciones, los operones se expresan a partir de un plásmido. En algunas realizaciones, el plásmido tiene un origen de replicación con número de copias bajo.

- 65 En algunas realizaciones, el polipéptido que se une directa o indirectamente a la secuencia reguladora transcripcional heteróloga se expresa a partir de un casete de expresión de control, comprendiendo el casete de expresión un promotor de control operativamente unido a una secuencia polinucleotídica que codifica el polipéptido. En algunas realizaciones, el casete de expresión está contenido en un plásmido de control separado de un plásmido que contiene los operones. En algunas realizaciones, el promotor de control es un promotor inducible.

En algunas realizaciones, el polipéptido heterólogo comprende una ARN polimerasa (ARNP). En algunas realizaciones, la ARNP es la ARNP de T7. En algunas realizaciones, el casete de expresión es un sensor ambiental.

También se describe en el presente documento un método para determinar un punto de experimentación para controlar la magnitud de la expresión de dos o más genes (por ejemplo, dentro de un operón sintético). En algunos aspectos, el método comprende: recibir uno o más puntos de datos de entrada, en el que los puntos de datos de entrada proporcionan información sobre uno o más elementos reguladores y una propiedad del sistema; y determinar, con un ordenador, un siguiente punto de datos usando un método de cálculo, en el que el siguiente punto de datos proporciona información sobre uno o más elementos reguladores.

En algunos aspectos, el método comprende además usar el siguiente punto de datos para una experimentación adicional para optimizar la expresión de los dos o más genes. En algunos aspectos, los elementos reguladores incluyen, por ejemplo, sitios de unión al ribosoma y/o elementos reguladores de la transcripción.

En algunos aspectos, el método de cálculo es una técnica de análisis numérico. En algunos aspectos, el método de optimización numérica es el algoritmo de Nelder-Mead, el método de Newton, el método cuasi-Newton, un método de gradiente conjugado, un método de punto interior, un descenso de gradiente, un método de subgradiente, un método de elipsoide, el método de Frank-Wolfe, un método de interpolación y métodos de búsqueda de patrones, o un modelo de colonia de hormigas. En algunos aspectos, el método de optimización numérica utilizado para determinar el siguiente punto de datos para una mayor experimentación requiere considerar el punto de reflexión, el punto de expansión o el punto de contracción en función de uno o más puntos de datos de entrada.

En algunos aspectos, el método de cálculo es un método de diseño de experimentos (DoE).

También se describe en el presente documento un producto de programa informático que comprende un medio tangible legible por ordenador que almacena una pluralidad de instrucciones para controlar un procesador para realizar una operación para determinar un punto de experimentación para controlar la magnitud de la expresión de dos o más genes, comprendiendo las instrucciones recibir uno o más puntos de datos de entrada, en los que los puntos de datos de entrada proporcionan información sobre uno o más elementos reguladores y una propiedad del sistema; y determinar, con un ordenador, un siguiente punto de datos usando un método de cálculo, en el que el siguiente punto de datos proporciona información sobre uno o más elementos reguladores.

DEFINICIONES

Cuando se cita “un(o)”, “una”, “el” o “la” pretende significar “uno o más” a menos que se indique específicamente lo contrario.

Una secuencia de polinucleótidos o polipéptidos es “heteróloga de” un organismo o de una segunda secuencia si procede de una especie extraña o, si es de la misma especie, está modificada respecto a su forma original. Por ejemplo, un promotor unido operativamente a una secuencia codificante heteróloga se refiere a una secuencia codificante de una especie diferente de la que se deriva el promotor o, si es de la misma especie, una secuencia codificante que no está naturalmente asociada con el promotor (p.ej., un promotor de la ARN polimerasa de T7 operativamente unido a un operón *nif* sintético).

La expresión “operativamente unido” se refiere a una unión funcional entre una secuencia de control de la expresión de un ácido nucleico (tal como un promotor, o conjunto de sitios de unión al factor de transcripción) y una segunda secuencia de ácido nucleico, en la que la secuencia de control de la expresión dirige la transcripción del ácido nucleico correspondiente a la segunda secuencia. En el contexto de un sitio de unión a ribosomas (RBS) y secuencias codificantes, la expresión se refiere a la unión funcional del RBS con la secuencia codificante en la que el RBS recluta ribosomas para la traducción de la secuencia codificante en un ARN.

Un “par afín”, como se usa en el presente documento, se refiere a un polipéptido de unión a ADN específico de la secuencia y una secuencia de ADN diana que está unida por el polipéptido de unión al ADN específico de la secuencia particular. Para los polipéptidos de unión a ADN específicos de la secuencia que se unen a más de un ácido nucleico diana, el par afín puede formarse con el polipéptido de unión a ADN específico de la secuencia y una cualquiera de las secuencias de ADN diana a las que se une el polipéptido.

Los sistemas transcripcionales “ortogonales” se refieren a sistemas (p. ej., uno, dos, tres o más) de elementos reguladores de la transcripción que comprenden secuencias de ADN diana reguladas por su polipéptido de unión a ADN específico de la secuencia afín de tal manera que los polipéptidos de unión a ADN específico de la secuencia en el sistema no “interaccionan”, es decir, los polipéptidos de unión a ADN específicos de la secuencia no interfieren ni regulan elementos reguladores de la transcripción en el sistema que no sean los elementos reguladores de la transcripción que contienen la secuencia de ADN diana afín del polipéptido de unión a ADN específico de la secuencia.

Los “polipéptidos de unión a ADN específicos de la secuencia” se refieren a polipéptidos que se unen al ADN de una manera específica de la secuencia de nucleótidos. Los ejemplos de polipéptidos de unión a ADN específicos de la secuencia incluyen, pero sin limitación, factores de transcripción (por ejemplo, activadores transcripcionales), ARN polimerasas y represores transcripcionales.

5 Un “activador transcripcional” se refiere a un polipéptido, que cuando se une a una secuencia promotora, activa o aumenta la transcripción de un ARN que comprende la secuencia codificante unida operativamente. En algunas realizaciones, el activador transcripcional unido a una secuencia objetivo en un promotor puede ayudar al reclutamiento de ARN polimerasa al promotor. Un “represor transcripcional” se refiere a un polipéptido, que cuando
10 se une a una secuencia promotora, bloquea o disminuye la transcripción de un ARN que comprende la secuencia codificante unida operativamente. En algunas realizaciones, el represor transcripcional bloquea el reclutamiento de la ARN polimerasa al promotor o bloquea el movimiento de la ARN polimerasa a lo largo del promotor.

15 La expresión “secuencia codificante” como se usa en el presente documento se refiere a una secuencia de nucleótidos que comienza en el codón para el primer aminoácido de una proteína codificada y termina con el codón para el último aminoácido y/o termina en un codón de parada.

La expresión “célula hospedadora” se refiere a cualquier célula capaz de replicar y/o transcribir y/o traducir un gen heterólogo. Por lo tanto, una “célula hospedadora” se refiere a cualquier célula procariota (que incluye pero sin limitación *E. coli*) o célula eucariota (incluidas, pero sin limitación, células de levadura, células de mamíferos, células
20 de aves, células de anfibios, células vegetales, células de peces y células de insectos), ya sea localizadas *in vitro* o *in vivo*. Por ejemplo, las células hospedadoras pueden estar localizadas en un animal transgénico o una planta transgénica. célula procariota (que incluye pero sin limitación *E. coli*) o célula eucariota (incluidas, pero sin limitación, células de levadura, células de mamíferos, células de aves, células de anfibios, células vegetales, células de peces y
25 células de insectos).

Los “elementos reguladores de la transcripción” se refieren a cualquier secuencia de nucleótidos que influye en el inicio y la velocidad de la transcripción, o la estabilidad y/o movilidad de un producto de transcripción. Las secuencias reguladoras incluyen, pero sin limitación, promotores, elementos de control del promotor, secuencias de unión a proteínas, UTR 5' y 3', sitios de inicio de la transcripción, secuencias de terminación, secuencias de poliadenilación, intrones, etc. Tales secuencias reguladoras de la transcripción pueden ubicarse en 5'-, 3'-, o dentro de la región codificante del gen y puede promover (elemento regulador positivo) o reprimir (elemento regulador negativo) la transcripción del gen.

35 La expresión “ácido nucleico” o el término “polinucleótido” se refiere a desoxirribonucleótidos o ribonucleótidos y polímeros de los mismos en forma monocatenaria o bicatenaria. A menos que esté específicamente limitado, el término abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares a las del ácido nucleico de referencia y que se metabolizan de manera similar a los nucleótidos naturales. A menos que se indique lo contrario, una secuencia de ácido nucleico particular también abarca
40 implícitamente variantes modificadas conservadoramente de la misma (p.ej., sustituciones de codones degenerados) y secuencias complementarias, así como la secuencia indicada explícitamente. Específicamente, las sustituciones de codones degenerados se pueden lograr mediante la generación de secuencias en las que la tercera posición de uno o más codones seleccionados (o todos) se sustituye con restos de bases mixtas y/o desoxiinosina (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)). La expresión ácido nucleico se usa indistintamente con gen, ADNc y ARNm codificado por un gen.

Los términos “polipéptido”, “péptido” y “proteína” se usan indistintamente en el presente documento para referirse a un polímero de restos de aminoácidos. Los términos se aplican a los polímeros de aminoácidos en los que uno o
50 más restos de aminoácidos es un imitador químico artificial del correspondiente aminoácido natural, así como a los polímeros de aminoácidos naturales y los polímeros de aminoácidos no naturales. Como se usa en el presente documento, los términos abarcan cadenas de aminoácidos de cualquier longitud, incluidas proteínas de longitud completa, en las que los restos de aminoácidos están unidos por enlaces peptídicos covalentes.

55 Se dice que dos secuencias de ácido nucleico o polipéptidos son “idénticas” si la secuencia de nucleótidos o de restos de aminoácidos, respectivamente, en las dos secuencias es la misma cuando se alinea para una correspondencia máxima como se describe a continuación. La expresión “complementario a” se usa en el presente documento para indicar que la secuencia es complementaria a toda o una parte de una secuencia de polinucleótidos de referencia.

60 Ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al., Nucleic Acids Res. 25: 3389-3402 (1997) y Altschul et al., J. Mol. Biol. 215:403-410 (1990), respectivamente. El software para realizar análisis BLAST está disponible públicamente en la Web a través del Centro Nacional de Información Biotecnológica (www.ncbi.nlm.nih.gov/). Este algoritmo implica primero identificar pares de secuencia de alta puntuación (HSP)
65 mediante la identificación de la longitud de palabra (W) corta en la secuencia de consulta, que coinciden o satisfacen

alguna puntuación de umbral (T) de valor positivo cuando se alinean con una palabra de la misma longitud en una base de datos secuencia. T se conoce como el umbral de puntuación de palabras vecinas (Altschul et al., *supra*). Estas coincidencias iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSP más largos que las contengan. Las coincidencias de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta que se pueda aumentar la puntuación de alineación acumulativa. Las puntuaciones acumulativas se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos coincidentes; siempre > 0) y N (puntuación de penalización para restos que no coinciden; siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabras en cada dirección se detiene cuando: la puntuación de alineación acumulativa se reduce en la cantidad X de su valor máximo alcanzado; la puntuación acumulativa llega a cero o menos, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, M = 5, N = -4 y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el programa BLASTP utiliza por defecto una longitud de palabra de 3 y una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (ver Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA 89:10915, (1989)) alineaciones (B) de 50, expectativa (E) de 10, M = 5, N = -4, y una comparación de ambas cadenas.

El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (ver, por ejemplo, Karlin et Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5787, (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad de que se produzca una coincidencia entre dos secuencias de nucleótidos o aminoácidos por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es menos de aproximadamente 0,2, más preferiblemente menos de aproximadamente 0,01, y lo más preferiblemente menos de aproximadamente 0,001.

El "porcentaje de identidad de secuencia" se determina comparando dos secuencias alineadas de manera óptima sobre una ventana de comparación, en la que la porción de la secuencia de polinucleótidos en la ventana de comparación puede comprender adiciones o deleciones (es decir, huecos) en comparación con la secuencia de referencia (que no comprenden adiciones o deleciones) para una alineación óptima de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que se produce la misma base de ácido nucleico o resto de aminoácido en ambas secuencias para producir el número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones en la ventana de comparación y multiplicando el resultado por 100 para obtener el porcentaje de identidad de secuencia.

La expresión "identidad sustancial" de secuencias de polinucleótidos significa que un polinucleótido comprende una secuencia que tiene al menos un 25 % de identidad de secuencia con una secuencia de referencia designada. Como alternativa, el porcentaje de identidad puede ser cualquier número entero del 25 % al 100 %, por ejemplo, al menos: 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 % en comparación con una secuencia de referencia que usa los programas descritos en el presente documento; usando preferiblemente BLAST parámetros estándar, como se describe a continuación. Un experto reconocerá que los valores de porcentaje de identidad anteriores se pueden ajustar de manera apropiada para determinar la identidad correspondiente de proteínas codificadas por dos secuencias de nucleótidos teniendo en cuenta la degeneración de codones, la similitud de aminoácidos, el posicionamiento del marco de lectura y similares. La identidad sustancial de las secuencias de aminoácidos para estos fines normalmente significa una identidad de secuencia de al menos 40 %. El porcentaje de identidad de los polipéptidos puede ser cualquier número entero del 40 % al 100 %, por ejemplo, al menos 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 %. En algunas realizaciones, los polipéptidos que son "sustancialmente similares" comparten secuencias como se indicó anteriormente, excepto que las posiciones de restos que no son idénticas pueden diferir por cambios conservadores de aminoácidos. Las sustituciones conservadoras de aminoácidos se refieren a la intercambiabilidad de restos que tienen cadenas laterales similares. Por ejemplo, un complejo de aminoácidos que tienen cadenas laterales alifáticas es glicina, alanina, valina, leucina e isoleucina; un complejo de aminoácidos que tienen cadenas laterales alifáticas-hidroxilo es serina y treonina; un complejo de aminoácidos que tienen cadenas laterales que contienen amida es asparagina y glutamina; un complejo de aminoácidos que tienen cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un complejo de aminoácidos que tienen cadenas laterales básicas es lisina, arginina e histidina; y un complejo de aminoácidos que tienen cadenas laterales que contienen azufre es cisteína y metionina. Los ejemplos de complejos de sustitución conservadora de aminoácidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, ácido aspártico-ácido glutámico y asparagina-glutamina.

Breve descripción de los dibujos

La Fig. 1 representa un esquema que ilustra los genes *nifE* y *nifN* bajo el control de promotores de T7 únicos.

La Fig. 2 ilustra la medición cuantitativa de la capacidad del operón sintético para complementar una cepa con *nifEN* inactivado y recuperar la capacidad de fijar el nitrógeno.

La Fig. 3 ilustra una biblioteca de promotores de T7 de tipo silvestre y mutante y su fuerza para controlar la expresión génica.

5 La Fig. 4 ilustra las fuerzas de los promotores T7 para controlar los genes *nifE* y *nifN* en cepas mutantes seleccionadas. La Fig. 4A representa las fuerzas de tres cepas. La Fig. 4B representa las coordenadas de reflexión calculadas.

La Fig. 5 ilustra la fijación de nitrógeno en la cepa Reflexión y las cepas iniciales.

10

La Fig. 6 ilustra el método de refactorización de la fijación de nitrógeno.

La Fig. 7 ilustra el complejo génico *nif* de *Klebsiella oxytoca*.

15 La Fig. 8 representa un esquema de un plásmido indicador fluorescente en el que los 150 pb que rodean el codón de inicio de un gen (de -60 a +90) se fusionaron con el gen mRFP y se expresaron bajo el control del promotor Ptac.

La Fig. 9 ilustra la fluorescencia medida por citometría de flujo.

20

La Fig. 10 ilustra los múltiples clones utilizados para identificar el sitio de unión al ribosoma sintético que mejor coincide con el sitio de unión al ribosoma nativo.

La Fig. 11 ilustra los operones quiméricos.

25

La Fig. 12 enumera los errores en los operones totalmente sintéticos.

La Fig. 13 muestra que cada operón sintético requería diferentes niveles de concentración de IPTG para una función óptima. También muestra el rendimiento de operones individuales en el sistema de conexiones de T7 bajo el control del promotor Ptac.

30

La Fig. 14 muestra una tabla del control de los operones sintéticos en el sistema.

La Fig. 15 muestra la fijación de nitrógeno de un complejo sintético completo expresado en una cepa completa con *nif* inactivado.

35

La Fig. 16 ilustra el uso del controlador N.^o 1 o del controlador N.^o 2 para producir el mismo rendimiento del complejo sintético completo.

La Fig. 17 representa un esquema detallado del complejo sintético completo.

40

La Fig. 18 muestra secuencias de ADN para genes nativos y genes sintéticos, así como el porcentaje de identidades comunes de nucleótidos y codones entre cada par.

La Fig. 19 muestra los nombres y secuencias de partes del controlador sintético.

45

La Fig. 20 enumera los nombres, secuencias y fuerzas de cada componente del complejo completo.

La Fig. 21 muestra un diagrama del vector de prueba RBS.

50

La Fig. 22 representa esquemas de los operones *inv-spa* y *prg-org* y los plásmidos utilizados. La Fig. 22A muestra un esquema de la cepa de *Salmonella* SL1344 con inactivación Δ *prg-org*. Los operones *inv-spa* y *prg-org* se muestran en un recuadro. La Fig. 22B muestra un esquema del vector de prueba del operón *prg-org* y el plásmido indicador. El plásmido de control y el plásmido indicador están a la derecha.

55

La Fig. 23 muestra una transferencia Western de una proteína secretada expresada a partir del operón sintético *prg-org* en la cepa con inactivación Δ *prg-org*. La Fig. 23A muestra que la cepa con inactivación Δ *prg-org* no expresa el operón *prg-org*. La Fig. 23B muestra que el operón *prg-org* refactorizado sintético en células de *Salmonella* Δ *prg-org* puede controlarse mediante la adición de IPTG.

60

La Fig. 24 muestra las secuencias sintéticas de RBS y del operón sintético de T3SS.

La Fig. 25 muestra un diagrama de bloques de un sistema informático.

65 La Fig. 26 ilustra el proceso de refactorización de un complejo génico. El complejo génico de fijación de nitrógeno de *K. oxytoca* de tipo silvestre se muestra en la parte superior. Los genes están coloreados por función: azul

(nitrógenasa), verde (biosíntesis de cofactores, el sombreado corresponde a los operones), amarillo (transporte electrónico) y gris (desconocido). Las flechas delgadas muestran la longitud y la orientación de los siete operones y una barra horizontal indica el solapamiento de genes. Los genes recodificados se muestran como líneas discontinuas. Los símbolos utilizados para definir el complejo y el controlador refactorizados se definen en las Figs. 29 y 30, respectivamente.

La Fig. 27 ilustra la robustez de la vía de fijación de nitrógeno a los cambios en la expresión de las proteínas componentes. (A) Se muestra la vía para la maduración de nitrógenasas y las proteínas están coloreadas por función (Fig. 26). Los complejos metálicos se sintetizan por la vía biosintética (23, 24). La fijación de nitrógeno catalizada por la nitrógenasa madurada se muestra con su cadena de transporte electrónico *in vivo*. (B) Se muestra la tolerancia de la actividad de nitrógenasa a los cambios en la expresión de las proteínas componentes. La actividad se mide mediante un ensayo de reducción de acetileno y se presenta el % en comparación con *K. oxytoca* de tipo silvestre. Los operones de tipo silvestre se expresan a partir de un promotor P_{tac} en un plásmido de bajo número de copias. La actividad del promotor se calcula como la producción del promotor P_{tac} a una concentración dada de IPTG y en comparación con un promotor constitutivo. El efecto de no incluir *NifY* (-Y) y *NifX* (-X) se muestra en rojo. (C) Se muestra la comparación de la fuerza de los sitios de unión al ribosoma (RBS) de tipo silvestre (negro) y sintético (blanco). Los RBS se midieron a través de una fusión transcripcional en marco (-60 a +90) con mRFP. La fuerza se mide como el promedio geométrico de una distribución de células medida por citometría de flujo. Los RBS sintéticos de *nifF* y *nifQ* no están destinados a coincidir con la medición de tipo silvestre. Las barras de error representan la desviación estándar de al menos tres experimentos realizados en días diferentes.

La Fig. 28 ilustra la conversión a ARNP de T7* Control. (A) La actividad nitrógenasa se muestra en función de la fuerza del promotor para cada operón refactorizado en cada cepa inactivada de *K. oxytoca* ($\Delta nifHDKTY$, $\Delta nifENX$, $\Delta nifJ$, $\Delta nifBQ$, $\Delta nifF$, y $\Delta nifUSVWZM$). Las líneas discontinuas verticales indican la fuerza del promotor de T7 mutante que controla cada operón en el complejo génico refactorizado completo. (B) Un plásmido controlador desacopla la expresión del operón del promotor inducible. Se diseñó una variante de ARNP de T7 (T7*) para reducir la toxicidad. Se usó un conjunto de 4 promotores de T7 mutados para controlar la expresión de cada operón (los números de parte y las secuencias para los mutantes 1-4 se enumeran en la sección *Materiales y Métodos*). La actividad de P_{tac} bajo la inducción IPTG 1 mM se indica mediante una línea horizontal discontinua. (C) La actividad nitrógenasa se compara para cada operón refactorizado bajo el control del promotor P_{tac} a la concentración óptima de IPTG (negro) y el controlador (parte D) con IPTG 1 mM y expresión controlada por diferentes promotores de T7 (blanco). Los promotores de T7 utilizados son WT para operones HDKY, EN y J; el promotor 2 para operones BQ y USVWZM; y el promotor 3 para F. Las barras de error representan la desviación estándar de al menos tres experimentos realizados en días diferentes.

La Fig. 29 muestra un esquema completo para el complejo génico y el controlador refactorizados completos. Cada una de las 89 partes está representada de acuerdo con el estándar visual SBOL (www.sbolstandard.org), y se muestran el número de parte del Registro SynBERC (register.synberc.org) y la actividad de la parte. Las fuerzas del promotor de T7 se miden con proteína fluorescente roja (mRFP) y se expresan en REU (ver *Materiales y Métodos*). Las fuerzas del terminador se miden en un plásmido indicador y se expresan como el factor de reducción en la expresión de mRFP en comparación con un indicador sin un terminador. La fuerza de RBS se expresa como unidades de expresión arbitrarias del promotor P_{tac} inducido (IPTG 1 mM) y un gen de fusión entre los primeros 90 nucleótidos del gen y la proteína fluorescente roja. Se muestran los números de nucleótidos para los plásmidos que contienen el complejo y el controlador refactorizados. La identidad del codón de cada gen recodificado en comparación con el tipo silvestre se muestra como un porcentaje.

La Fig. 30 muestra la regulación del complejo génico refactorizado completo. (A) Se muestra la actividad nitrógenasa para los tres controladores: inducible por IPTG, inducible por aTc y lógica IPTG Y aTc. El registro de la cromatografía de gases se muestra para cada uno, así como el porcentaje calculado de actividad de tipo silvestre, (7,4 % $\pm$ 2,4 %, 7,2 % $\pm$ 1,7 % y 6,6 % $\pm$ 1,7 % respectivamente). La desviación estándar se calcula utilizando datos de al menos dos experimentos realizados en días diferentes. (B) Se muestra la incorporación de ^{15}N en la biomasa celular. La fijación de nitrógeno del gas N_2 por el complejo génico refactorizado fue rastreado usando $^{15}N_2$ y medido usando la espectrometría de masas de relaciones isotópicas (IRMS). Los datos se representan como la fracción de nitrógeno celular que es ^{15}N . La desviación estándar representa dos experimentos realizados en días diferentes. (C) Se muestra el efecto del amoníaco en la regulación de la expresión de la nitrógenasa. Las trazas de reducción de acetileno se muestran con (rojo) y sin adición (azul) de acetato de amonio 17,5 mM para las células de tipo silvestre (izquierda) y las células que llevan el sistema *nif* sintético (derecha). El sistema sintético fue inducido por el Controlador N.º 1 usando IPTG 1 mM y exhibió una actividad nitrógenasa de 1,1 % $\pm$ 0,5 % y 6,1 % $\pm$ 0,4 % con y sin acetato de amonio respectivamente. (D) Se muestra la expresión de la ARNP de T7* del Controlador N.º 1 correspondiente a la Parte C. Las cepas que portaban el Controlador N.º 1 y un plásmido indicador RFP se caracterizaron por inducción con IPTG 1 mM con o sin adición de acetato de amonio.

La Fig. 31 muestra las delecciones del operón *nif* utilizadas en este estudio. Las líneas continuas muestran la región de los operones *nif* eliminados. La línea discontinua en NF25 muestra el operón *nifLA* retenido.

La Fig. 32 muestra una tabla de construcción y verificación de todos los mutantes de delección del gen *nif* de *K. oxytoca*.

La Fig. 33 muestra la caracterización del promotor utilizando Unidades de Expresión Relativa. (A) Conversión de unidades arbitrarias a Unidades de Expresión Relativa (REU). Los promotores se caracterizaron usando la proteína indicadora fluorescente mRFP1 en N155 (Promotores medidos). Los datos se normalizaron primero por la fluorescencia de N110 (patrón interno) y después se escalan por la fluorescencia de N155 (J23100) para tener en cuenta las diferencias de RBS entre N155 y N110 (ajuste de RBS). Para comparar directamente nuestras mediciones con los niveles de expresión de los patrones de Kelly et al., multiplicamos además por la proporción entre la fluorescencia de N110 y la fluorescencia de un plásmido patrón de Kelly que expresa mRFP1 (patrón de promotor de RFP). Se aplica un factor de conversión final para comparar todas las mediciones con el patrón del promotor J23101-EGFP de Kelly et al. basado en una fuerte correlación lineal de la fuerza del promotor (RPU) entre construcciones que expresan mRFP y EGFP. Se dibujaron recuadros rellenos y rayados para indicar qué plásmidos se midieron en diferentes instalaciones. Las unidades con asterisco y sin asterisco se midieron en diferentes instalaciones y corresponden a los factores de conversión anteriores. (B) Caracterización del promotor para el promotor P_{lac} (izquierda) y el promotor P_{tet} (derecha). Las fuerzas del promotor en el caso del promotor P_{lac} y del promotor P_{tet} se midieron a concentraciones variadas de inductores (IPTG o aTc). Las fuerzas de los promotores de T7 (WT y mutantes, Fig. 28B) se muestran como líneas de puntos horizontales.

La Fig. 34 ilustra la depuración de los operones refactorizados. (A) El proceso se muestra para la identificación de secuencias problemáticas dentro de un operón refactorizado. Después del diseño y la síntesis, el ADN problemático se cruza con el tipo silvestre para crear una biblioteca quimérica, la cual se analiza. Esto se hace de forma iterativa para reducir el tamaño de la región problemática hasta que se identifiquen los errores específicos. (B) El proceso de depuración condujo a la corrección de las fuerzas de RBS (flechas rojas), la secuencia recodificada de *nifH*, y los numerosos errores de nucleótidos encontrados en el complejo secuenciado en la base de datos. Se muestran mutaciones de aminoácidos para corregir errores en la secuencia sintética.

La Fig. 35 representa una tabla de errores de secuencia de ADN en la secuencia del complejo *nif* X13303.1.

La Fig. 36 muestra el crecimiento celular soportado por la fijación de nitrógeno. La línea de puntos indica la densidad de siembra inicial de OD600 0,5. La *Klebsiella* de tipo silvestre creció a una OD600 de $2,57 \pm 0,07$ después de 36 horas de incubación en condiciones de depresión. La eliminación del complejo *nif* completo inhibió fuertemente el crecimiento celular (Δnif , OD600 $0,76 \pm 0,02$). La complementación de la cepa inactivada con el complejo refactorizado y el Controlador N.º 1 bajo la inducción de IPTG 1 mM produjo un crecimiento de OD600 $1,10 \pm 0,03$.

La Fig. 37 muestra la expresión de las variantes *nifH* sintéticas. Transferencia Western para detectar la expresión de *nifH*_{V1} sintética (izquierda) y *nifH*_{V2} sintética (derecha). Todas las construcciones llevaban P_{lac} -*nifHDK* con el gen sintético indicado. Los cultivos se indujeron con IPTG 50 μ M.

La Fig. 38 representa una tabla de secuencias de ADN de partes sintéticas.

La Fig. 39 muestra mapas de plásmidos clave. La notación gráfica SBOL se usa para describir partes genéticas: el prefijo y el sufijo BioBrick son cuadrados en blanco y los terminadores tienen la forma de una T.

Descripción detallada de los dibujos

1. Introducción

La presente invención se refiere a la ingeniería de complejos génicos. Se ha descubierto cómo manipular y seleccionar de manera recombinante y mediante cálculo las secuencias codificantes de complejos génicos nativos y las secuencias reguladoras heterólogas de modo que las secuencias codificantes estén bajo el control de la regulación heteróloga y produzcan el producto funcional del complejo génico (por ejemplo, un operón nativo). Al eliminar elementos reguladores nativos fuera y dentro de las secuencias codificantes de complejos génicos, y luego agregar sistemas reguladores sintéticos, los productos funcionales de operones genéticos complejos y otros complejos génicos se pueden controlar y/o mover a células heterólogas, incluidas células de diferentes especies distintas de las especies de las cuales se derivaron los genes nativos.

Como se demuestra a continuación, los inventores han rediseñado el complejo génico *Nif* de *Klebsiella oxytoca*, así como un sistema de secreción de proteínas tipo III de *Salmonella*, que genera productos funcionales (por ejemplo, enzimas fijadoras de nitrógeno y complejos de secreción de péptidos, respectivamente) bajo el control de un sistema regulador heterólogo. Una vez rediseñados, los complejos génicos sintéticos pueden controlarse mediante circuitos genéticos u otros sistemas reguladores inducibles, controlando así la expresión de los productos según se desee.

II. Generación de complejos génicos sintéticos

Se cree que los métodos descritos en el presente documento pueden usarse y adaptarse para rediseñar la regulación de esencialmente cualquier operón u otro complejo génico. En general, los operones nativos o los complejos génicos que se diseñarán tendrán el mismo producto funcional en el hospedador nativo. Por ejemplo, en algunas realizaciones, al menos la mayoría de los productos génicos dentro del operón nativo o complejo génico para ser rediseñados funcionarán cada uno para producir un producto o función específica del hospedador nativo. Los productos funcionales pueden incluir, por ejemplo, enzimas multicomponentes, complejos asociados a membranas, incluidos, pero sin limitación, complejos que transportan moléculas biológicas a través de membranas u otros complejos biológicamente activos. Por ejemplo, en algunas realizaciones, los productos funcionales son, por ejemplo, un sistema de secreción de proteínas Tipo III, un microcompartimento bacteriano, una vesícula de gas, un magnetosoma, un celulosoma, una vía de degradación de alcanos, un complejo de fijación de nitrógeno, un complejo de degradación de polibifenilo, una vía para la biosíntesis de poli(3-hidroxbutirato), enzimas para la biosíntesis de péptidos no ribosomales, productos de complejos génicos para la biosíntesis de policétidos, una vía de biosíntesis de terpenoides, una vía de biosíntesis de oligosacáridos, una vía de biosíntesis de indolocarbazol, un complejo fotosintético para captación de luz, un estresosoma, o un complejo de percepción de quórum. Véase, Fischbach y Voigt, *Biotechnol. J.*, 5: 1277-1296 (2010) para obtener una descripción detallada y ejemplos de cada uno.

Los operones nativos o complejos génicos utilizados en las realizaciones de la presente invención pueden derivarse (originarse) de procariotas o eucariotas.

Como se usa en el presente documento, "nativo" pretende referirse a la célula hospedadora o al genoma del hospedador del cual se deriva originalmente un operón o complejo génico (por ejemplo, como el operón que se encuentra en la naturaleza). Por lo tanto, la "expresión nativa" de un operón se refiere a los niveles y patrones de expresión específicos de un conjunto de genes en un operón o complejo génico en un hospedador nativo.

Un operón se refiere a una unidad de ADN que comprende múltiples secuencias codificantes separadas bajo el control de un único promotor. Las secuencias codificantes separadas se expresan normalmente dentro de una única molécula de ARN y posteriormente se traducen por separado, por ejemplo, con niveles de traducción variables debido a la fuerza de los sitios de unión a ribosomas (RBS) asociados con las secuencias codificantes particulares. Los operones se encuentran más normalmente en las células procariotas.

Los complejos génicos se refieren a conjuntos de genes que tienen una función o producto funcional común. Los genes se encuentran normalmente en proximidad física entre sí dentro del ADN genómico (por ejemplo, dentro de un centiMorgan (cM)). Los complejos génicos pueden existir en células procariotas o eucariotas.

A. Secuencias codificantes

Una vez que se ha identificado un operón nativo o complejo génico para rediseñarlo, se pueden identificar las secuencias codificantes que se van a rediseñar. En general, será deseable comenzar solo con las secuencias codificantes del operón nativo o complejo génico, eliminando así los promotores nativos y otras secuencias reguladoras no codificantes. Dependiendo de la función de los diversos productos génicos del operón nativo o complejo génico, en algunas realizaciones, todas las secuencias codificantes de un operón nativo o complejo génico se rediseñan.

Como alternativa, una o más secuencias codificantes pueden omitirse del proceso de rediseño. Por ejemplo, se puede saber que uno o más de los productos génicos en un operón nativo o complejo génico no contribuyen al producto funcional del operón o pueden no ser necesarios para la generación del producto del operón o complejo. Por ejemplo, como se describe en los ejemplos a continuación, en el rediseño del operón *Nif*, el gen *nifT* no tenía una función conocida y, en particular, se sabía que la eliminación de *nifT* no afectaba significativamente a la función última del operón, es decir, la fijación de nitrógeno. Así, *nifT* no se incluyó en el proceso de rediseño.

En algunas realizaciones, el operón o complejo génico incluirá secuencias codificantes de proteínas reguladoras que regulan la expresión o actividad de uno o más de los otros productos del operón o complejo génico. En tales realizaciones, puede ser deseable omitir tales proteínas reguladoras del proceso de rediseño porque en su lugar se empleará la regulación sintética. Por ejemplo, como se describe en los ejemplos a continuación, en el rediseño del operón *nif*, se sabía que *nifL* y *nifA* actuaban como genes reguladores para el operón *nif* y, por lo tanto, se omitieron de modo que se pudo utilizar en su lugar la regulación sintética.

Una vez que se ha identificado el conjunto de productos génicos a rediseñar, se puede comenzar con la secuencia codificante nativa, o las secuencias de aminoácidos de los productos génicos. Por ejemplo, en algunas realizaciones, las secuencias de aminoácidos de los productos génicos pueden usarse para producir una secuencia codificante sintética para la expresión en la célula hospedadora en la que los productos rediseñados deben expresarse en última instancia.

En algunas realizaciones, las secuencias codificantes nativas del conjunto de productos génicos a rediseñar se usan

como punto de partida. En este caso, en algunas realizaciones, se eliminan las secuencias que no son esenciales para la producción de los productos génicos. Por ejemplo, se pueden eliminar los sitios de unión a ribosomas, terminadores o promotores dentro de las secuencias codificantes. En algunas realizaciones, las secuencias de nucleótidos de las secuencias codificantes se analizan usando un algoritmo (es decir, en un ordenador) para

5 identificar sitios de unión a ribosomas, terminadores o promotores dentro de la secuencia o secuencias.

Las secuencias reguladoras no esenciales dentro de las secuencias codificantes pueden reducirse o eliminarse alterando los codones de las secuencias codificantes nativas. Las secuencias reguladoras que comprenden codones se pueden interrumpir, por ejemplo, cambiando los codones por codones sinónimos (es decir, que codifican el

10 mismo aminoácido) dejando de ese modo la secuencia de aminoácidos codificada intacta mientras se cambia la secuencia codificante. Se pueden alterar uno o más codones de una o más secuencias codificantes.

En algunas realizaciones, se reemplazan al menos 5 %, 10 %, 15 %, 20 % o más codones de una o más secuencias codificantes nativas para insertar en un operón sintético. En algunas realizaciones, se reemplazan al menos 5 %, 10

15 %, 15 %, 20 %, 30 %, 40 %, 50 % o más codones de cada una de las secuencias codificantes nativas a insertar en un operón sintético.

En algunas realizaciones, los codones de reemplazo pueden seleccionarse, por ejemplo, para ser significativamente divergentes de los codones nativos. Los cambios en el codón pueden dar como resultado la optimización del codón para la célula hospedadora, es decir, la célula en la que se va a expresar el polinucleótido para la prueba y/o para la expresión final. Se conocen métodos de optimización de codones (p.ej., Sivaraman et al., *Nucleic Acids Res.* 36:e16 (2008); Mirzahoseini, et al., *Cell Journal (Yakhteh)* 12(4):453 Winter 2011; patente US-6.114.148) y puede incluir referencias a codones de uso común para una célula hospedadora particular. En algunas realizaciones, uno o más codones se asignan al azar, es decir, un codón nativo se reemplaza con un codón aleatorio que codifica el mismo

20 aminoácido. Este último enfoque puede ayudar a eliminar cualquier secuencia de acción cis involucrada en la regulación nativa del polipéptido. En algunas realizaciones, los codones se seleccionan para crear una secuencia de ADN que esté a la máxima distancia de la secuencia nativa. En algunas realizaciones, se usa un algoritmo para eliminar secuencias transcripcionalmente funcionales en un gen que codifica el polipéptido. Por ejemplo, en algunas realizaciones, los sitios de unión a ribosomas, elementos reguladores de la transcripción, terminadores u otras

25 secuencias de ADN unidas por proteínas se eliminan de la secuencia codificante nativa. En particular, las secuencias funcionales eliminadas pueden ser funcionales en las especies nativas (de las cuales se deriva originalmente la secuencia), en la célula hospedadora heteróloga, o en ambas. En algunas realizaciones, la optimización comprende la eliminación de secuencias en la secuencia codificante nativa que son funcionales para activadores o represores transcripcionales heterólogos que se usarán para regular los operones sintéticos que se

30 generarán.

En muchos casos, la generación de secuencias codificantes sintéticas, así como las porciones restantes del operón sintético, se realizarán *de novo* a partir de oligonucleótidos sintéticos. Por lo tanto, en algunas realizaciones, los codones se seleccionan para crear una secuencia de ADN que no genera dificultades para la producción o

40 combinación de oligonucleótidos. Por lo tanto, en algunas realizaciones, se evitan las secuencias de codones que darían como resultado la generación de oligonucleótidos que forman horquillas.

En algunas realizaciones, como se indicó anteriormente, la alteración del codón dependerá de la célula hospedadora utilizada. Las células hospedadoras pueden ser cualquier célula procariota (incluidas, sin limitación, *E. coli*) o células eucariotas (incluidas, sin limitación, células de levadura, células de mamíferos, células de aves, células de anfibios,

45 células vegetales, células de peces y células de insectos).

Las secuencias reguladoras no esenciales dentro de las secuencias nativas pueden identificarse, en algunas realizaciones, usando un algoritmo realizado por un procesador que ejecuta instrucciones codificadas en un medio de almacenamiento legible por ordenador. Por ejemplo, en algunas realizaciones, los sitios de unión a ribosomas se identifican usando un modelo termodinámico que calcula la energía libre de la unión de ribosomas al ARNm. En algunas realizaciones, los promotores se identifican con un algoritmo que usa una matriz ponderada por posición. En algunas realizaciones, los terminadores transcripcionales se identifican mediante un algoritmo que identifica las horquillas y/o pistas de poli-A dentro de las secuencias. En algunas realizaciones, un algoritmo identifica otras

50 secuencias transcripcionalmente funcionales, que incluyen pero sin limitación sitios de inserción de transposones, sitios que promueven la recombinación, sitios para la escisión por endonucleasas de restricción y/o secuencias que están metiladas.

En vista de las alteraciones descritas anteriormente, en algunas realizaciones, una secuencia codificante en un operón sintético descrita en el presente documento es menos del 90, 85, 80, 75 o 70 % idéntica a la secuencia codificante nativa. En algunas realizaciones, la secuencia codificante codifica una secuencia de proteína que es idéntica a la proteína nativa o es al menos 80, 85, 90 o 95 % idéntica a la proteína nativa. En algunas realizaciones, menos del 70 %, 60 % o 50 % de los codones en una, dos o más secuencias codificantes en un operón sintético son idénticos a los codones en la secuencia codificante nativa.

60

B. Organización de secuencias codificantes en operones sintéticos

Una vez que se han seleccionado las secuencias codificantes (por ejemplo, y están sustancialmente “limpias” de secuencias reguladoras nativas o espurias), las secuencias codificantes se organizan en uno o más operones sintéticos. La organización del operón u operones sintéticos incluye la inserción de varias secuencias transcripcionales y traduccionales heterólogas entre, antes y/o después de las secuencias codificantes, de modo que la expresión de cada secuencia codificante se controla como se desee. Así, por ejemplo, las secuencias del promotor 5' se pueden seleccionar para conducir la expresión de un ARN de operón que comprende las secuencias codificantes del operón. La selección de uno o más terminadores de fuerza apropiada también afectará a los niveles de expresión. Además, el orden de las secuencias codificantes dentro de un operón sintético y/o la selección de RBS para las secuencias codificantes permite el control de las tasas de traducción relativas de cada secuencia codificante, permitiendo así varios niveles de control para los niveles absolutos y relativos de los productos proteicos finales.

Debido a que cada operón sintético puede tener su propio promotor, se pueden expresar diferentes operones sintéticos con diferentes potencias. Por lo tanto, en algunas realizaciones, las secuencias codificantes se organizan en diferentes operones en función de los niveles relativos de expresión nativa. Dicho de otra manera, en algunas realizaciones, las secuencias codificantes se organizan en operones agrupando secuencias codificantes expresadas sustancialmente en el mismo nivel nativo en un operón sintético particular.

Además, debido a que las secuencias codificantes en el extremo 5' (frontal) de un ARN pueden expresarse a un nivel más alto que las secuencias codificantes más hacia 3', en algunas realizaciones, las secuencias codificantes se ordenan dentro de un operón sintético de manera que la secuencia codificante de mayor expresión (en el contexto nativo) aparece primero y el gen de expresión más baja aparece en último lugar. En algunas realizaciones, la organización de genes dentro de los operones se basa en la expresión temporal nativa, la función, la facilidad de manipulación del ADN y/o el diseño experimental.

Al diseñar los controles transcripcionales (p.ej., promotores) y traduccionales (p.ej., RBS) de los operones sintéticos, se puede considerar la proporción de proteínas medida en el sistema nativo. Por lo tanto, en algunas realizaciones, dos o más secuencias codificantes que se expresan en un contexto nativo sustancialmente al mismo nivel y/o que se expresan deseablemente en una proporción de aproximadamente 1:1 para lograr la funcionalidad (por ejemplo, donde dos o más miembros son parte de un complejo funcional en una relación 1:1) se colocan cerca uno del otro dentro de un operón sintético. “Proximidad” generalmente significará que las secuencias codificantes son adyacentes entre sí en el operón sintético.

En algunas realizaciones, los niveles de expresión relativa de las secuencias codificantes dentro y, en algunas realizaciones, entre operones sintéticos se determinan probando uno o más operones de prueba en cuanto a la expresión deseada y/o funcionalidad deseada y luego mejorando la expresión en base a los resultados iniciales. Si bien este método puede realizarse en una base de “prueba y error”, en algunas realizaciones, se emplea un método de optimización numérica para guiar la selección de elementos reguladores para alterar la expresión génica y mejorar las propiedades deseadas del sistema. Tales métodos, por ejemplo, pueden ser realizados por un procesador ejecutando instrucciones codificadas en un medio de almacenamiento legible por ordenador (descrito más adelante). Los ejemplos de métodos de optimización numérica incluyen, entre otros, un algoritmo de Nelder-Mead, un método de Newton, un método cuasi-Newton, un método de gradiente conjugado, un método de punto interior, un descenso de gradiente, un método de subgradiente, un método de elipsoide, un método de Frank-Wolfe, un método de interpolación y métodos de búsqueda de patrones, o un modelo de colonia de hormigas. En algunas realizaciones, se emplea un método de diseño mediante cálculo de experimentos (DoE) para alterar la expresión génica y mejorar las propiedades deseadas del sistema en los operones sintéticos.

Los elementos reguladores de la transcripción, los sitios de unión al ribosoma, los terminadores y otras secuencias que afectan a la transcripción o la traducción pueden seleccionarse de colecciones existentes de tales secuencias, y/o pueden generarse mediante el rastreo de bibliotecas generadas por diseño o por mutación aleatoria. Los ejemplos de secuencias reguladoras incluyen secuencias de nucleótidos que actúan en cis unidas por un polipéptido de unión a ADN específico de la secuencia, por ejemplo, un activador transcripcional o un represor transcripcional. Los ejemplos de activadores transcripcionales incluyen, pero sin limitación, factores sigma, ARN polimerasas (ARNP) y activadores asistidos por chaperona. En algunas realizaciones, el par afín activador transcripcional/secuencia de acción en cis será ortogonal a la célula hospedadora. Dicho de otra manera, la secuencia reguladora no estará unida por otras proteínas de las células hospedadoras, excepto por el activador transcripcional heterólogo que se une a la secuencia de acción en cis.

i. Factores sigma

En algunas realizaciones, el polipéptido de unión a ADN específico de la secuencia es un factor sigma (σ) y la secuencia reguladora del operón sintético comprende la secuencia de nucleótidos de acción en cis afín al factor sigma. Los factores sigma reclutan ARN polimerasa (ARNP) para secuencias promotoras específicas para iniciar la transcripción. La familia σ 70 consiste en 4 grupos: el Grupo 1 son los σ constitutivos y son esenciales; los grupos 2-4 son σ alternativos que dirigen la transcripción celular para necesidades especializadas (Gruber y Gross, Annu.

Rev. Microbiol., 57:441-466 (2003)). El Grupo 4 de σ (también conocido como ECF σ ; función extracitoplasmática) constituye el grupo más grande y diverso de σ , y se han clasificado en 43 subgrupos (Staron et al., Mol Microbiol 74(3): 557-81 (2009)).

- 5 En algunas realizaciones, el conjunto de polipéptidos de unión a ADN específicos de la secuencia comprende múltiples factores sigma. En algunas realizaciones, el conjunto comprende factores sigma del Grupo 1, Grupo 2, Grupo 3 y/o factores Sigma del Grupo 4. Se cree que el subgrupo ECF del Grupo 4 reconoce diferentes secuencias promotoras, lo que hace que estos σ sean particularmente útiles para construir sistemas ortogonales de promotores σ . Sin embargo, se apreciará que cualquier complejo de factores sigma se puede usar de acuerdo con los métodos
- 10 de las realizaciones de la invención para desarrollar pares afines.

Tabla 1

Grupo N.º ^a	ID ^b	GI ^c	ESPECIE ^d	CLASE ^d	FILO ^d
ECF01	>3473	109899616	<i>Pseudoalteromonas atlantica</i> T6c	Gammaproteobacteria	Proteobacteria
ECF01	>4085	114562024	<i>Shewanella frigidimarina</i> NCIMB 400	Gammaproteobacteria	Proteobacteria
ECF02	>2817	16130498	<i>Escherichia coli</i> K12	Gammaproteobacteria	Proteobacteria
ECF02	>915	119774011	<i>Shewanella amazonensis</i> SB2B	Gammaproteobacteria	Proteobacteria
ECF03	>1198	29350055	<i>Bacteroides thetaiotaomicron</i> VPI-5482		Bacteroidetes
ECF03	>1244	34541012	<i>Porphyromonas gingivalis</i> W83		Bacteroidetes
ECF04	>1609	21673117	<i>Chlorobium tepidum</i> TLS		Chlorobi
ECF04	>1617	68549683	<i>Pelodictyon phaeoclathratiforme</i> BU-1		Chlorobi
ECF05	>965	28868416	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	Gammaproteobacteria	Proteobacteria
ECF05	>1054	67154316	<i>Azotobacter vinelandii</i> AvOP	Gammaproteobacteria	Proteobacteria
ECF06	>3576	15595669	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF06	>853	26987094	<i>Pseudomonas putida</i> KT2440	Gammaproteobacteria	Proteobacteria
ECF07	>980	67154823	<i>Azotobacter vinelandii</i> AvOP	Gammaproteobacteria	Proteobacteria
ECF07	>1134	15598606	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF08	>3580	15595872	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF08	>3627	70730114	<i>Pseudomonas fluorescens</i> Pf-5	Gammaproteobacteria	Proteobacteria
ECF09	>3581	15597622	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF09	>1009	70730971	<i>Pseudomonas fluorescens</i> Pf-5	Gammaproteobacteria	Proteobacteria
ECF10	>3486	77360766	<i>Pseudoalteromonas haloplanktis</i> TAC125	Gammaproteobacteria	Proteobacteria
ECF10	>2914	88706154	gamma proteobacterium KT 71	Gammaproteobacteria	Proteobacteria
ECF11	>3726	28868260	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	Gammaproteobacteria	Proteobacteria
ECF11	>987	28899132	<i>Vibrio parahaemolyticus</i> RIMD 2210633	Gammaproteobacteria	Proteobacteria
ECF12	>807	86158800	<i>Anaeromyxobacter dehalogenans</i> 2CP-C	Deltaproteobacteria	Proteobacteria
ECF12	>808	108762328	<i>Myxococcus xanthus</i> DK 1622	Deltaproteobacteria	Proteobacteria
ECF13	>1146	33152898	<i>Haemophilus ducreyi</i> 35000HP	Gammaproteobacteria	Proteobacteria
ECF13	>1025	37524103	<i>Photobacterium luminescens</i> subsp. laumondii TTO1	Gammaproteobacteria	Proteobacteria
ECF14	>3200	15608361	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF14	>1324	21223516	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria

ECF15	>436	77464848	<i>Rhodobacter sphaeroides</i> 2.4.1	Alphaproteobacteria	Proteobacteria
ECF15	>524	16127705	<i>Caulobacter crescentus</i> CB15	Alphaproteobacteria	Proteobacteria
ECF16	>3622	104782321	<i>Pseudomonas entomophila</i> L48	Gammaproteobacteria	Proteobacteria
ECF16	>973	161378140	<i>Pseudomonas putida</i> KT2440	Gammaproteobacteria	Proteobacteria
ECF17	>1691	15607875	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF17	>1458	21221399	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF18	>4451	21230791	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	Gammaproteobacteria	Proteobacteria
ECF18	>4438	21242133	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	Gammaproteobacteria	Proteobacteria
ECF19	>3197	15607586	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF19	>1315	21219164	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF20	>992	70731405	<i>Pseudomonas fluorescens</i> Pf-5	Gammaproteobacteria	Proteobacteria
ECF20	>2913	88706222	<i>gamma proteobacterium</i> KT 71	Gammaproteobacteria	Proteobacteria
ECF21	>1280	29350128	<i>Bacteroides thetaiotaomicron</i> VPI-5482		Bacteroidetes
ECF21	>2825	89889680	<i>Flavobacterium bacterium</i> BBFL7		Bacteroidetes
ECF22	>4450	21232074	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	Gammaproteobacteria	Proteobacteria
ECF22	>1147	21243541	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	Gammaproteobacteria	Proteobacteria
ECF23	>231	15895043	<i>Clostridium acetobutylicum</i> ATCC 824		Firmicutes
ECF23	>1851	30261806	<i>Bacillus anthracis</i> str. Ames		Firmicutes
ECF24	>69	16079737	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		Firmicutes
ECF24	>1034	32470052	<i>Escherichia coli</i>	Gammaproteobacteria	Proteobacteria
ECF25	>1645	170078575	<i>Synechococcus</i> sp. PCC 7002		Cyanobacteria
ECF25	>1643	17230772	<i>Nostoc</i> sp. PCC 7120		Cyanobacteria
ECF26	>4464	58581966	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	Gammaproteobacteria	Proteobacteria
ECF26	>837	77459110	<i>Pseudomonas fluorescens</i> PfO-1	Gammaproteobacteria	Proteobacteria
ECF27	>4265	21222299	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF27	>1331	31795084	<i>Mycobacterium bovis</i> AF2122/97		Actinobacteria
ECF28	>1088	114563849	<i>Shewanella frigidimarina</i> NCIMB 400	Gammaproteobacteria	Proteobacteria
ECF28	>1040	15641058	<i>Vibrio cholerae</i> O1 biovar <i>eltor</i> str. N16961	Gammaproteobacteria	Proteobacteria
ECF29	>371	13476734	<i>Mesorhizobium loti</i> MAFF303099	Alphaproteobacteria	Proteobacteria
ECF29	>2688	71281387	<i>Colwellia psychrerythraea</i> 34H	Gammaproteobacteria	Proteobacteria
ECF30	>35	16079766	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		Firmicutes
ECF30	>83	18309341	<i>Clostridium perfringens</i> str. 13		Firmicutes
ECF31	>2963	85713274	<i>Idiomarina baltica</i> OS145	Gammaproteobacteria	Proteobacteria
ECF31	>34	16080921	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		Firmicutes
ECF32	>1122	4581629	<i>Erwinia amylovora</i>	Gammaproteobacteria	Proteobacteria
ECF32	>3724	28868612	<i>Pseudomonas syringae</i> pv. <i>tomato</i> str. DC3000	Gammaproteobacteria	Proteobacteria
ECF33	>375	27378153	<i>Bradyrhizobium japonicum</i> USDA 110	Alphaproteobacteria	Proteobacteria
ECF33	>423	39934888	<i>Rhodospseudomonas palustris</i> CGA009	Alphaproteobacteria	Proteobacteria
ECF34	>3302	77164965	<i>Nitrosococcus oceani</i> ATCC 19707	Gammaproteobacteria	Proteobacteria
ECF34	>1384	21218750	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF35	>3582	15598092	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF35	>1119	24375055	<i>Shewanella oneidensis</i> MR-1	Gammaproteobacteria	Proteobacteria
ECF36	>3196	15609206	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF36	>1595	21219385	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF37	>3390	89094252	<i>Oceanospirillum</i> sp. MED92	Gammaproteobacteria	Proteobacteria
ECF37	>2513	83718468	<i>Burkholderia thailandensis</i> E264	Betaproteobacteria	Proteobacteria
ECF38	>1322	21222029	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF38	>1442	152967344	<i>Kineococcus radiotolerans</i> SRS30216		Actinobacteria
ECF39	>1438	21223369	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria
ECF39	>2973	84494624	<i>Janibacter</i> sp. HTCC2649		Actinobacteria
ECF40	>3198	15610550	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria
ECF40	>1380	62389491	<i>Corynebacterium glutamicum</i> ATCC 13032		Actinobacteria
ECF41	>491	16127496	<i>Caulobacter crescentus</i> CB15	Alphaproteobacteria	Proteobacteria

ECF41	>1141	77459658	<i>Pseudomonas fluorescens</i> PfO-1	Gammaproteobacteria	Proteobacteria
ECF42	>3583	15596548	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria
ECF42	>4454	77747962	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	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

Además de los factores sigma nativos, también se pueden usar factores sigma quiméricos u otras variantes en el método de la invención. Por ejemplo, en algunas realizaciones, uno o más factores sigma se someten a mutación para generar una biblioteca de variantes del factor sigma y la biblioteca resultante se puede explorar en busca de nuevas actividades de unión al ADN.

En algunas realizaciones, se pueden usar factores sigma quiméricos formados a partir de porciones de dos o más factores sigma. Por consiguiente, los aspectos de la divulgación proporcionan la generación de una biblioteca de polinucleótidos que codifican factores sigma quiméricos, en el que los factores sigma quiméricos comprenden un dominio de al menos dos factores sigma diferentes, en el que cada uno de los dominios se une a la región -10 o -35 de un elemento regulatorio; y la expresión de factores sigma quiméricos de la biblioteca de polinucleótidos, generando así una biblioteca de factores sigma quiméricos. Por ejemplo, en algunas realizaciones, se generan factores sigma quiméricos que comprenden una "Región 2" de un primer factor sigma y una "Región 4" de un segundo factor sigma, generando así factores sigma quiméricos con actividades de unión de ADN novedosas. La "Región 2" de factores sigma es un dominio conservado que reconoce -10 regiones de promotores. La "Región 4" es un dominio conservado de factores sigma que reconoce -35 regiones de promotores. Se apreciará que los factores sigma quiméricos se pueden generar a partir de dos factores sigma nativos que se unen a diferentes secuencias de ADN diana (por ejemplo, diferentes secuencias promotoras). Se ha encontrado que los factores sigma quiméricos formados a partir de los subgrupos ECF2 y ECF11 tienen actividades de unión de ADN únicas útiles para generar conjuntos ortogonales como se describe en el presente documento. Los ejemplos de factores sigma quiméricos incluyen, pero sin limitación, ECF11_ECF02 (que contiene los aminoácidos 1-106 de ECF02_2817 y 122-202 de ECF11_3726) y ECF02_ECF11 (que contiene los aminoácidos 1-121 de ECF11_3726 y 107-191 de ECF02_2817).

La secuencia de aminoácidos ECF11_ECF02 es la siguiente:

```

1  MRITASLRTFCHLSTPHSDSTTSRLWIDEVTAVARQDRDSFMRIYDHFAPRLLRYLTGL
61  NVPEGQAEELVQEVLLKLWHKAESFDPKASLGTWLFRIARNLYIDSVRKDRGWVQVQNS
121  LEQLERLEAISNPENLMLSEELRQIVFRTIESLPEDLRMAITLRELDGLSYEEIAAIMDC
181  PVGTVRSRIFRAREAIDNKVQPLIRR*

```

La secuencia de aminoácidos ECF02_ECF11 es la siguiente:

```

1  MSEQLTDQVLVERVQKGDQKAFNLLVVRYQHKVASLVSRYPVPSGDVPDVPVQEAFFIKAYRA
61  LDSFRGDSAFYTWLYRIAVNTAKNYLVAQGRRPPSSDVDAIEAENFEQLERLEAPVDRTL
121  DYSQRQEQQLNSAIQNLPTDQAKVLRMSYFEALSHREISERLDMPLGTVKSCRLRAFQKL
181  RSRIEES*

```

ii. ARN polimerasas

En algunas realizaciones, el polipéptido de unión a ADN específico de la secuencia es un polipéptido que tiene actividad de unión a ADN y que es una variante de la ARN polimerasa (ARNP) de T7 y la secuencia de acción en cis afín de la ARNP (por ejemplo, un promotor reconocido por la ARNP) está operativamente unido al operón sintético para controlar la expresión del operón. La secuencia de aminoácidos de la ARNP de T7 es la siguiente:

1 mntiniaknd fsdielaaip fntladhyge rlareqlale hesyemgear frkmferqlk
 61 agevadnaaa kplittllpk miarindwfe evkakrgkrp tafqflqeik peavayitik
 121 ttlacletsad nttvqavasa igraiedear fgrirdleak hfkknveeq l nkrvghvykk
 181 afmqvveadm lskgllggea wsswhkeds i hvgvrcieml iestgmvs l h rqnagvvvgqd
 241 setielapey aeaiatraga lagispmfqp cvvppkpwtg itgggywang rrplalvrth
 301 skkalmryed vympevykai niaqntawki nkkvlavanv itkwkhcpve dipaiereel
 361 pmkpedidmn pealtawkra aaavyrkdka rksrrislef mlegankfan hkaiwfpynm
 421 dwrgrvyavs mfnpggndmt kglitlakgk pigkegyywl kihgancagv dkvpfperik
 481 fieenhenim acaksplent wwaeqdspfc flafcfeyag vqhghlsync slplafdgsc
 541 sqiqhfsaml rdevggravn llpsetvqdi ygivakkvne ilqadaingt dnevvvtvde
 601 ntgeisekvk lgtkalagqw laygvtrsvt krsvmtlayg skefgfrqqv ledtiqpaid
 661 sgkglmftqp nqaagymakl iwesvsvtvv aaveamnwlk saakllaaev kdkktgeilr
 721 krcavhwvtp dgfpvwqeyk kpiqtrlnlm flgqfrlqpt intnkds eid ahkqesgiap
 781 nfvhsqdgsh lrktvvwahe kygiesfali hdsfgtipad aanlfkavre tmvdtyescd
 841 vladfydqfa dqlhesql dk mpalpakgnl nlr dilesdf afa

El promotor de la ARNP de T7 también se ha caracterizado (ver, por ejemplo, Rong et al., Proc. Natl. Acad. Sci. USA, 95(2):515-519 (1998)) y es bien conocido.

Se han divulgado métodos para generar pares ortogonales de variantes de la ARNP y variantes del promotor diana. Debido a la toxicidad de la expresión de la ARNP de T7 nativa, se puede diseñar una serie de mutaciones y modificaciones de modo que se pueda expresar y analizar una biblioteca de variantes de ARNP para determinar su actividad en células sin toxicidad excesiva. Por consiguiente, los aspectos de la divulgación proporcionan una o más de las siguientes modificaciones (y, por lo tanto, un aspecto de la divulgación proporciona células hospedadoras que comprenden casetes de expresión, o ácidos nucleicos que comprenden casetes de expresión, en el que el casete de expresión codifica una variante de ARNP sustancialmente idéntica a la ARNP de T7, en el que el casete de expresión comprende uno o más de los siguientes):

La expresión de la variante de la ARNP de T7 se puede expresar a partir de un plásmido con bajo número de copias. La expresión de la ARNP puede controlarse mediante una proteína codificada por separado de un vector separado, bloqueando así la expresión de la ARNP hasta que se agrega un segundo vector a las células que promueven la expresión de la ARNP;

Control de la traducción: un codón de inicio GTG; se pueden usar sitios de unión a ribosomas débiles y/o espaciadores de ADN aleatorios para aislar la expresión de la ARNP;

Una etiqueta molecular para promover la degradación rápida de la ARNP. Por ejemplo, una etiqueta Lon N-terminal dará como resultado una degradación rápida de la ARNP etiquetada por el sistema de proteasa Lon.

Un sitio activo de la ARNP mutado (por ejemplo, entre los aminoácidos 625-655 de la ARNP de T7). Por ejemplo, se ha descubierto que una mutación de la posición correspondiente al aminoácido 632 (R632) de la ARNP de T7 se puede mutar para reducir la actividad de la ARNP. En algunas realizaciones, la ARNP contiene una mutación correspondiente a R632S.

Además, se ha descubierto una variedad de promotores de T7 mutantes que pueden usarse en un circuito genético. Por lo tanto, en algunas realizaciones, la secuencia reguladora de un operón sintético comprende una secuencia mutante como se expone en la tabla a continuación.

Nombre del promotor	Secuencia	Fuerza (2009.10.02 a 2009.10.09)
WT	TAATACGACTCACTANNNNNAGA	
Mut1	TAATACGACTCACTATAGGGAGA	5263
Mut2	TAATACGACTCACTACAGGCAGA	365
Mut3	TAATACGACTCACTAGAGAGAGA	366
	TAATACGACTCACTAATGGGAGA	577

(continuación)

Nombre del promotor	Secuencia	Fuerza (2009.10.02 a 2009.10.09)
	TAATACGACTCACTANNNNNAGA	
Mut4	TAATACGACTCACTATAGGTAGA	1614
Mut5	TAATACGACTCACTAAAGGGAGA	1018
Mut6	TAATACGACTCACTATTGGGAGA	3216

- 5 También se han descubierto varias estructuras diferentes de horquilla que funcionan como terminadores para la ARNP de T7. Ver, la siguiente Tabla. Por consiguiente, un aspecto de la divulgación proporciona un operón sintético que comprende un promotor funcional para una ARNP de T7 nativa o una ARNP sustancialmente idéntica a la misma, en el que el polinucleótido unido operativamente comprende un terminador seleccionado de la siguiente tabla. Se pueden seleccionar terminadores con diferentes secuencias para diferentes transcritos para evitar la recombinación homóloga.

Nombre del terminador	Secuencia	Fuerza (Ensayo 2009.12.16)
	TANNNNAACSSWWSSSSSSTCWWW WCGSSSSSWSSGGTTTTTTGT	
52	TATAAACGGGGGGCTAGGGGTTTTT GT	107
23	TACTCGAACCCCTAGCCCGCTCTTATC GGGCGGCTAGGGGTTTTTTGT	714
72	TAGCAGAACCGCTAACGGGGGCGAAG GGGTTTTTTGT	1051
48	TACTCGAACCCCTAGCCCGCTCTTATC GGGCGGCTAGGGGTTTTTTGT	1131
1	TACATATCGGGGGGGTAGGGGTTTTTT GT	1297
2	TACATATCGGGGGGGTAGGGGTTTTTT GT	1333
WT	TAGCATAACCCCTTGGGGCCTCTAAAC GGGTCTTGAGGGGTTTTTTGT	1395
31	TACCCTAACCCCTTCCCCGGTCAATCG GGGCGGATGGGGTTTTTTGT	1586
58	TAGACCAACCCCTTGGGGCCTCAATCG GGGGGGATGGGGTTTTTTGT	1608
25	TACTCTAACCCCATCGGCCGTCTTAGG GGTTTTTTGT	1609

(continuación)

Nombre del terminador	Secuencia	Fuerza (Ensayo 2009.12.16)
17	TACCTCAACCCCTTCGCGCCTCATATC GCGGGGCATGCGGTTTTTGT	1887

5 En algunas realizaciones, las variantes de ARNP pueden diseñarse con un bucle de especificidad alterado (correspondiente a las posiciones entre 745 y 761). Por lo tanto, en algunas realizaciones, se proporciona una ARNP que es idéntica o sustancialmente idéntica a la ARNP de T7 o T3 pero que tiene una secuencia de bucle seleccionada de las de las tablas siguientes entre las posiciones 745 y 761.

A

Familia ARNP	Andamio	Plásmido ARNP	Plásmido promotor	Secuencia del huésped	Secuencia del promotor
T7	N249	N249	N155	YMQEYKKPIQIRLNLMLFGLQFRLOPTINTNKDSEI DAHK	TAATAGGACTCACTATA GGGAGA
T3	N115	N377:115	N352	YMQEYKKPIQIRLNLMLFGLQFRLOPTINTNKDSEI DAHK	TAATAAGCCCTCACTATA GGGAGA
K1F	N115	N421:115	N353	YMQEYKKPIQIRLNLMLFGLSPNLQPTVNTNKDSEI DAHK	TAATAACTATCACTATA GGGAGA
N4	N77	W78	W74	YMQEYKKPIQIRLNLMLFGLSPNLQPTVNTNKDSEI AHK	TAATAAGCCCTCACTATA GGGAGA

B

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

iii. Activadores que requieren chaperonas

- 5 En algunas realizaciones, el conjunto de polipéptidos de unión a ADN específicos de la secuencia comprende polipéptidos que tienen actividad de unión a ADN y que requieren una proteína chaperona separada para unirse al polipéptido de unión a ADN específico de la secuencia para que el polipéptido de unión a ADN específico de la secuencia sea activo. Los ejemplos de activadores transcripcionales que requieren una chaperona para la actividad incluyen, pero sin limitación, el activador que es sustancialmente similar a InvF de *Salmonella typhimurium*, MxiE de *Shigella flexneri*, y ExsA de *Pseudomonas aeruginosa*. Estos activadores enumerados requieren la unión de SicA de *Salmonella typhimurium*, IpgC de *Shigella flexneri*, o ExsC de *Pseudomonas aeruginosa*, respectivamente, para la activación.

La información de secuencia para los componentes anteriores se proporciona a continuación:

Nombre	Tipo	Secuencia de ADN que codifica el polipéptido nombrado	Mutación opcional
<i>sicA</i>	Gen	atggattatcaaaataatgtcagcgaagaacgtgttgcggaaatgattgggaatgccgttagtgaag ggccacgctaaaagacgttcattgggacccctcaagatatgatggacgggttatatgctcatgctta tgagtttataaccaaggacgactggatgaagctgaagacgttcttctgttcttattgcaatttatgatttt tacaatcccgaattacacattgggactggcggcagtatgcccaacgaaaaacaatttcagaagc atgtgacctttatgcagtagcgtttacttaaaatgattatgcccccgtttttaccgggcagt gtcaatratattatgcgttaaggcagcaaaagccagacaggttttgaactgtcaatgaacgtactga agatgagctctctcggcgaagcgttggctatctggaagcgttaaaacgcgcgagacagag cagcagatggaacagaagaaggaataa	
<i>sicA*</i>	<i>sicA</i> mutante	atggattatcaaaataatgtcagcgaagaacgtgttgcggaaatgattgggaatgccgttagtgaag ggccacgctaaaagacgttcattgggacccctcaagatatgatggacgggttatatgctcatgctta tgagtttataaccaaggacgactggatgaagctgaagacgttcttctgttcttattgcaatttatgatttt ttacaatcccgaattacacattgggactggcggcagtatgcccaactgaaaaacaatttcagaagc atgtgacctttatgcagtagcgtttacttaaaatgattatgcccccgtttttaccgggcagt gtcaatratattatgcgttaaggcagcaaaagccagacaggttttgaactgtcaatgaacgtactga agatgagctctctcggcgaagcgttggctatctggaagcgttaaaacgcgcgagacagag	La "t" grande de la secuencia de <i>sicA</i> anterior se mutó a "a" por PCR propensa a errores. Esta mutación se hizo para reducir la interferencia entre <i>SicA</i> y <i>MxiE</i> .
		cagcacagtgaacaagaaaggaataa	
<i>invF</i>	Gen con nuevo codón de inicio	atgctaaatagcaggaagtacttaagaaggagagaagaacggaaatccgagcccggaagca tggtttatagacagctgttcgcgcaaaagctgcattatgcttctgaaagccgacacaatga aaattgcttgattcaggaaggcgcgtctttttgcgagcagccggtgtgcacagtatcagg agacctgtgttttcagcgttaaaaatgaagctacagcaaaatctgcatattatgatggcgcag gattagtggaacagacatatgtgaatccgataaatgggttttgcgtgagctctgagtttcgcgtattt ggcaagatcgtaaacgctgcgagctactgtttttgcagcaaaattattacgcttctccgcttcaat aaggtaactggcgtgttacgaaaagcgagagttactgtgtgtgtgtgtatttactcgtcagcaca ccagcggcaacacgtatgagatgctgggagaagactatggcggttcttaacccatttctgcgttt gtgcagcagagcgttggcggaaaagcgaagaggaattacgaaacgtgcgtatggcgcacatc gctgctgaatagtgtagaaggccagcagaacatcccaattagccgttaatactggttactcatic ccttcaattttctagtgatcaagagcgtatcggttttccgcgggaattatcaaatattat tcaattggcagacaatga	Se determinó que el codón de inicio aceptado (el "atg" grande) era incorrecto y se encontró un codón de inicio correcto en dirección 5'.
<i>psicA</i>	Promotor	ccacaagaacgaggtagcgcattgagccgtaaggcagtagcgtatgattcattggcggtttt tgaatgttcaataaccacgtcgggttttaaaactgcatcagataaacgagrcgttaagttctaca aagtcggtgacagataacaggaatgaatga	

(continuación)

Nombre	Tipo	Secuencia de ADN que codifica el polipéptido nombrado	Mutación opcional
<i>ipgC</i>	Gen	atgctttaaataaccgaaatgaagcatctctactgcagtaattgatgcaalttaactctggcgct acactgaaagatafaatgcaatctctgatgatafatgagatgacattttatcatatgcttatgactttta caacaaaggaaatafagagggaagctgaagttttctcaggrtttatgatatcagacttttacaatg agactacattatgggactcgcagctatllatcagataaaagaacagtlccacaagcagcagacct ttatgctgctgctttgcatfaggaaaaaargactatacaccagttatccatactggaacatgctacgct tcgggtgaagccccccttaaaagctaaagaagctcgcgaactcgtatllcaacacagcaatgatga aaanttaaaataaagcacaatcatactggagcgaatcaggatatacgaaggatga	
<i>mxiE</i>	Gen con optimización de codones	A tagaataatataaaggcctgaacaccagcaacatgctctacatctacagctctgctgcaagc ggfagaacgtfgaactggtgaaagataaagaacgttaacatcatcgaactggcaccggcggtgaaa gg ctttttcttt gtgcglaaccagaacatcaaatlcagcgataaacgttaactaccactacc gctcaacatcaactcttggcgaatactctggcggtttgggaatttttagcgccgccctggtgnaac attctacacgaaataatgcatcactttctaccagaaacagctcgtgatatgctgatafacgga atctatctgctggataaacgtagctgctgcttcatlltagtagcgatcagaaacgtgctaaagccctggc aatgatccgtaigaccgaaagtratactctggttctgacatgctgctgacgattgaaaaaaa gaagtgccgcatcaaaagccctgaccgaacatctatggcgtllctcgaagcglaattcgtatgctgct cnaagagcgtgggtgccaaatggaagaacagctgaacacgtggcgccctggtgnaatggcctgc tggatggttctgcafaaccagaccattacgagcgccggccatgaacaatggttatgctgctaccag tcactcagcaatgaatlaaaacgctgctggcgctttagtgcgcggaactgagcaacatcaccttc ctggtgaaggaaataatgaaaaaatctaa	El gen de tipo silvestre tiene “tttttttt” en esta región de secuencia ampliada. Se agregó una “t” más para hacer “ttttttttt” y luego con GenScript se optimizó todo el gen. Se agregó la “t” adicional para hacer este ORF en marco. Además, el gen de tipo silvestre comienza con “g” y este gen sintético comienza con “a”.
<i>pipaH9.8</i>	Promotor	gcgaaatgacatcaaaaacgccatlaacctgatgltctggggaaataaataatgicaggctagggc aaaaatcgtggcggtgacaaaatgctgcgttacgctcattgagcatalccaggacatggccggcaa accgggtacgcgatctgttgccttggaaggtgactgacctcagtaaatcaatacgggtctga cggcgcgttacggtcaaatatgaagtacgatgtttaactaacggaaacaagaacaatacgggt gcaaacaggccattcacgggttaactgaacagatcgttttllacagccaatttgtttalcttatta t A ataaaaaagrgct	
<i>pipaH9.8*</i>	Promotor con mutación	gcgaaatgacatcaaaaacgccatlaacctgatgltctggggaaataaataatgicaggctagggc aaaaatcgtggcggtgacaaaatgctgcgttacgctcattgagcatalccaggacatggccggcaa accgggtacgcgatctgttgccttggaaggtgactgacctcagtaaatcaatacgggtctga cggcgcgttacggtcaaatatgaagtacgatgtttaactaacggaaacaagaacaalacgggt gcaaacaggccattcacgggttaactgaacagatgctgttttllacagccaatttgtttalcttatta ag ataaaaaagtgct	La “ta” ampliada arriba de <i>pipaH9.8</i> fue mutada a “ag” por mutagénesis de saturación. Esta mutación se hizo para reducir la expresión parcial de <i>pipaH9.8</i> .
<i>exsC</i>	Gen	atggaattaacgagcaaggltcaaccgactgcttggcgagltcggagccgfatcgggttgcctcc ctgtccctcgacgagggagggatggcgagccctctgttcgacgaacaggtggcgctacccctgt tgcctctgccgagcgagcgtctgttgcctggaggccgatggtggcgagcgcgatgctgctggg cgaggggatctllcggcagctcggcagcttcaaccggcatlggcagcggtllcgaltcgtatllcggc ttcgacgagctgaccggcaggttccagttgatatggcgaattctcgagcgcaactgaccctcga atgcttcgagggcgacctggccaatctctcgatcacggcgagttctggcagcgctgctgctgct ggcacagtgatcgagggcggtcgtcgtcggtcgcatgagggttga	
<i>exsD</i>	Gen	atggagcagggaagcagataagcagtlactccgagaagcgggtgttcgctggcagggcggtatcc gtgtgtgggctcggacggccgctcgcggggcgggtgcccgggttacgcatgacaggtllgtatc gtgagttccggaatcatcagtgctgcggcaactggcgtgtgctgacagcgatgctgcccgcctcgc gctggagcaactgttccgtgcgagtggtltagcagcagcgctggcgcgccggcctggcgctggg ggcggaaggagtgccggcagattctctctcgcggcgcgagagcagcagcggcgtgtgtgtcggc actggcgaccgggtcaacctcggcgtgcgcagtgatgatgactgggtcctgctgcccggct tatgctgctgtggaaaacgtctcgcgacacggcgatccccggcgtggcgctgctgctgctgagc tggagacccagltcccggcaatcgagagtcnagtlccgaattctggatcccggcggcagagctggga ggcgagcagcgcccgcgaggaaactggccagggctgccaaagtgcagggcgccgacccagga acaggtgcccgaactggccgcaagctggagacggccttcggcactgcggaagagcgccctggc gnaactggcagcggggcagltggcgacgtgctcgcagcggcggtggcgtggcgttgcagccg atccccgaggtcctgaatgctcgtggcaacctctctgcccgtcgtgagcagcagcgtcgcgcg cggagcgcgtccagggcgtgctgacaggaacgcaactgtgccaaggcagcagatcactctactg gcagagctga	

(continuación)

Nombre	Tipo	Secuencia de ADN que codifica el polipéptido nombrado	Mutación opcional
<i>exsA</i>	Gen	atgcaaggagccaatctcttggccgaagcagataacgtctgtcattggaaacatccaacttgcg aatacagggtataaacaaggagggcgatatagtctgtcgaaggcgaactgaccgtccaggga catcgattccacttttgcctggcgcctggcgagttgcttttcgtccgccggaagctatgtcgtaa	
		gtaccaagggaaggacagccgaatacttggatccallatctgccagtttctacaaggcttcgt ccagcgcttccgcgcgtgttggatgaagtcgagcgttgcgacgagcccgccggcgtatc gcgttgcgtgccacgcctctgtctggccgggtgcgtcaagggttgaaaggatgcttggcatgag catccgccgagtcctgcctgcctgaagcagcaggttgcgtcttcgcgtcagtcgcga ggccgcctgtgatgtcgtcctgcgcgaactgaacacggcagtcgagcgtcgtcagcgtat carggaagcactacccaacagagtggaagctgcccgaactctcccgcagttcgcagtgagg ctgaccaccccaaggagcttgcgcaggtctatgggggttccgcgcgcctggatcagcga gcggagaatcctctatgccatcagttgctgctcaacagcgacatgagcagtcgacatccat ggagcggggcttttcagtgatgctatllcaccagagctatcgccgccgttccgctgcacgcc gagccgctcgcgcaggggaaggacgaatgccggcgtaaaataactga	
<i>pexsD</i>	Promotor	gaaaggacgaatccggcgtaaaaaaactgacgtttttgaaagcccgtagcggctgcatgag agaatcggcccaaat	
<i>pexsC</i>	Promotor	gatgtggctttttcttaaaagaaagtctctcagtgacaaaagcagatgcatagcccgggtgtagca tgccgtgagcttt	
<i>rfp</i>	Gen	atggcttccctccgaagacgttatcaagagttcatgcgtttcaagttcgtatggaaggttccgttaa cggtcacgagttcgaatcgaagtggaagtggaagtcgtccgtacgaaggtacgcagaccgt aaactgaaagttaaccaagggtggtccgtcgtcgttgggacatccgtgcccgagttccag tacggttccaaggcttactgtaaacacccggctgacatccggactacgtgaaactgtccttcccg gaaggtttcaaatgggaacgtttatgaactcgaagacggtggtgtgttaccgttaccaggact ctccctgcaagacgggtgagtcacatcacaaggttaactgcgtgactaactcccgccgacgc gtccgggttatgcagaanaaaacatgggttgggaagcttccaccgaacgtatglaaccgggaagac ggtgctctgaagggtgaatcaaatgcgtctgaaactgaaagacggtggtcactacgacgtga agttaaaccacctacatggctlaaaaaaccgggtcagctgccgggtgcttacaanaaccgacatca aactggacatcacctccacaacgaagactacacatcgttgaacagctacgaacgtgctgaaggt cgtaactccacgggtgctgcagcaaacgacgaatactacgtttaa	

C. Control de la expresión del operón

Como se señaló anteriormente, el uno o más operones sintéticos están controlados por elementos reguladores que responden a un polipéptido de unión a ADN específico de la secuencia (por ejemplo, un activador transcripcional). Cuando se usa más de un operón, puede ser deseable que cada operón responda al mismo activador transcripcional, aunque con una secuencia reguladora diferente que controla la “fuerza” de expresión de un operón particular. Como se señaló anteriormente, en algunas realizaciones, el activador transcripcional es una ARNP de T7 o una variante de la misma.

La expresión del polipéptido de unión al ADN específico de la secuencia se puede controlar en un casete de expresión separado, comprendiendo el casete de expresión un promotor unido operativamente a un polinucleótido que codifica el polipéptido de unión al ADN específico de la secuencia. En algunas realizaciones, el promotor es inducible, impartiendo así el control de la expresión del operón basado en el inductor. Los ejemplos de promotores inducibles (con inductor entre paréntesis) incluyen, por ejemplo, Ptac (IPTG), Ptrc (IPTG), Pbad (arabinosa), Ptet (aTc), Plux (AI-1). Como alternativa, en algunas realizaciones, el promotor es constitutivo.

En algunas realizaciones, se insertan secuencias de nucleótidos “amortiguadores” adicionales entre los promotores y los sitios de unión al ribosoma, entre las secuencias codificantes y los terminadores, y/o entre las secuencias codificantes y un sitio de unión al ribosoma posterior. Estas secuencias actúan como “amortiguadores” en el sentido de que reducen o eliminan la interferencia reguladora entre diferentes secuencias codificantes. En algunas realizaciones, el espaciador forma una horquilla, es una secuencia nativa de una ruta metabólica, o es de una 5'-UTR, por ejemplo, obtenida de un fago. En algunas realizaciones, la horquilla es una ribozima. En algunas realizaciones, la ribozima es RiboJ. En algunas realizaciones, la secuencia amortiguadora se selecciona de secuencias de una longitud dada con nucleótidos seleccionados al azar. En algunas realizaciones, la secuencia amortiguadora es una región UP de un promotor. Las regiones UP pueden influir positivamente en la fuerza del promotor y generalmente están centradas en la posición -50 de un promotor (como se mide desde el inicio de la transcripción). Ver, por ejemplo, Estrem, et al., PNAS, 95 (11): 9761-9766 (1988) En algunas realizaciones, la secuencia amortiguadora es una secuencia extendida de 5-UTR.

Los ejemplos de secuencias amortiguadoras incluyen las enumeradas en la tabla a continuación:

Fuentes	Secuencias
Fago T5	agttcgatgagagcgataaccctctacaaataattttgttaa
Fago T5	ataaattgataaacaacacccctctacaaataattttgttaa
Fago T5	ataaattgagagaggaggtcctctacaaataattttgttaa
Fago T5	attaaaggaggagaattaaccctctacaaataattttgttaa
Fago T5	aaacctaatggatcgacctcctctacaaataattttgttaa
Fago T7	atcgagagggacacggcgacctctacaaataattttgttaa
Fago T7	gctaggtaacactagcagccctctacaaataattttgttaa
Fago T7	atgaaacgacagtgcacctctacaaataattttgttaa
Fago T7	agggagaccacaacgggttcctctacaaataattttgttaa
Escape con niveles altos de transcripción	attaaacacctgctaggatcctctacaaataattttgttaa
Escape con niveles altos de transcripción	ataaaggaaaacggtcaggtcctctacaaataattttgttaa
Escape con niveles altos de transcripción	atagggtaaaagcctgtcatcctctacaaataattttgttaa
Utilización de carbono	acaataaaaaatcattacatgttccctctacaaataattttgttaa
Utilización de carbono	agaagcagcgcgcaaaaatcagctgcctctacaaataattttgttaa
Utilización de carbono	atgagttcatttcagacaggcaaatcctctacaaataattttgttaa
Utilización de carbono	aactgcagttatttactgtgattacctctacaaataattttgttaa
Utilización de carbono	agccacaaaaaaagtcagtgttggtcctctacaaataattttgttaa
Utilización de carbono	acacagtcacttatcttttagttaaagggtcctctacaaataattttgttaa
Antisecuencias de escape	atccggaatcctcttcccggcctctacaaataattttgttaa
	aacaaaaataaaaaggagtcgctcacccctctacaaataattttgttaa
Fago T5	agttcgatgagagcgataacagttccagattcaggaactataa
Fago T5	ataaattgataaacaacaaaagttccagattcaggaactataa
Fago T5	ataaattgagagaggagttagttccagattcaggaactataa
Fago T5	attaaaggaggagaattaacagttccagattcaggaactataa
Fago T5	aaacctaatggatcgaccttagttccagattcaggaactataa
Fago T7	atcgagagggacacggcgaagtccagattcaggaactataa
Fago T7	gctaggtaacactagcagcagttccagattcaggaactataa
Fago T7	atgaaacgacagtgcagtcaagtccagattcaggaactataa
Fago T7	agggagaccacaacgggttcagttccagattcaggaactataa
Escape con niveles altos de transcripción	attaaacacctgctaggatagttccagattcaggaactataa
Escape con niveles altos de transcripción	ataaaggaaaacggtcaggtagttccagattcaggaactataa
Escape con niveles altos de transcripción	atagggtaaaagcctgtcatagttccagattcaggaactataa
Utilización de carbono	acaataaaaaatcattacatgtttagttccagattcaggaactataa
Utilización de carbono	agaagcagcgcgcaaaaatcagctgagttccagattcaggaactataa
Utilización de carbono	atgagttcatttcagacaggcaaatagttccagattcaggaactataa
Utilización de carbono	aactgcagttatttactgtgattagttccagattcaggaactataa
Utilización de carbono	agccacaaaaaaagtcagtgttggttagttccagattcaggaactataa
Utilización de carbono	acacagtcacttatcttttagttaaagggttagttccagattcaggaactataa
Antisecuencias de escape	atccggaatcctcttcccggagttccagattcaggaactataa
	aacaaaaataaaaaggagtcgctcacagttccagattcaggaactataa
Horquillas	gatcaccagggggatcccccggtgaaggat
Horquillas	gatcgcccaccggcagctgcgggtggcgatcaaggat
Horquillas	gatcatcggttagagtaatttgagcagatcccccggtgaaggat
Horquillas	attgatctggttattaaaggtaaatcggttcattta
Horquillas	gttctccacgggtgggatgagccctctggtggaaatgcg
Horquillas	agcatgagggtaaagtgtcatgcacca
Horquillas	acgtcgacttatctcgagtgagatattgtgacgggtac
Horquillas	acgtcgacttatctcgagtgagataagttgacgggtac
Horquillas	acgtcgacttatctcgagactgcagttcaatagagatattgtgacgggtac
Horquillas (ribozima)	gactgtcaccggatgtgcttccggctgatgagtcggtaggacgaaacag
Horquillas	gatcaccagggggatcccccggtgaaggatcctctacaaataattttgttaa
Horquillas	Gatcgccacggcgagctgcgggtggcgatcaaggatcctctacaaataattttgttaa

(continuación)

Fuentes	Secuencias
Horquillas	gatacgcggtagaggttaattgagcagatccccggggaaggatcctacaaa taatttggtttaa
Horquillas	attgatctgggttttaaaggtaatcgggtcattttacctctacaaataatttggtttaa
Horquillas	Cttctccacgggtgggatgagccctcgtggtagaaatgcgcctctacaaataa tttggtttaa
Horquillas	agcatgagggttaaagtgtcatgcaccaacctctacaaataatttggtttaa
Horquillas	Acgtcgacttatctcgagtgagataattgtgacgggtacctctacaaataatttgg tttaa
Horquillas	Acgtcgacttatctcgagtgagataattgtgacgggtacctctacaaataatttgg tttaa
Horquillas	acgtcgacttatctcgagtgagataattgtgacgggtacctctacaaataatttgg tttaa
Horquillas (ribozima)	gactgtcacccggatgtgttcgggtcgtgatgagtcgtgaggacgaacagcc ctacaaataatttggtttaa

Los operones sintéticos y/o el casete de expresión para expresar el polipéptido de unión a ADN específico de la secuencia se pueden transportar en uno o más plásmidos, por ejemplo, en una célula. En algunas realizaciones, el operón y el casete de expresión están en diferentes plásmidos. En algunas realizaciones, el plásmido con el casete de expresión y/o el o los plásmidos con el operón son plásmidos con bajo nivel de copias. Los plásmidos con bajo nivel de copias pueden incluir, por ejemplo, un origen de replicación seleccionado de PSC101, PSC101*, plásmido F, R6K o IncW.

III. Operones sintéticos

También se describen operones sintéticos, por ejemplo, los generados por los métodos descritos en el presente documento.

IV. Sistemas de operones sintéticos

También se describen sistemas que comprenden operones sintéticos y uno o más casetes de expresión de control, en los que el casete de expresión codifica un polipéptido de unión a ADN específico de la secuencia que controla la expresión del operón o los operones sintéticos. En algunas realizaciones, el casete o casetes de expresión de control son circuitos genéticos. Por ejemplo, los casetes de expresión pueden diseñarse para actuar como puertas lógicas, generadores de impulsos, osciladores, interruptores o dispositivos de memoria. En algunas realizaciones, el casete de expresión de control está unido a un promotor de tal manera que el casete de expresión funciona como un sensor ambiental. En algunas realizaciones, el sensor ambiental es un sensor de oxígeno, temperatura, tacto, estrés osmótico, estrés de membrana o redox.

Como se explicó anteriormente, en algunas realizaciones, el casete de expresión codifica la ARNP de T7 o una variante funcional de la misma. En algunas realizaciones, la ARNP de T7 es la salida del circuito o circuitos genéticos.

Los operones y los casetes de expresión se pueden expresar en una célula. Así, en algunos aspectos, una célula contiene los sistemas descritos en el presente documento. Cualquier tipo de célula hospedadora puede comprender el sistema.

V. Cálculo

En el presente documento se describe un producto de un programa informático que determina valores experimentales para controlar la magnitud de la expresión de dos o más genes. Esto puede usarse, por ejemplo, para optimizar una propiedad del sistema (por ejemplo, niveles de fijación de nitrógeno). En una realización, el código del programa recibe uno o más puntos de datos de entrada, en el que los puntos de datos de entrada proporcionan información sobre uno o más elementos reguladores y una propiedad del sistema. A continuación se un método de cálculo para determinar el siguiente punto de datos. En un aspecto, el método de cálculo puede ser un

método de diseño de experimentos (DoE).

En algunas realizaciones, el siguiente punto de datos generado por el código del programa puede usarse después para experimentación adicional, por ejemplo, para ver si el siguiente punto de datos sugerido da como resultado un nivel de expresión optimizado para dos o más genes, lo que conduce a una mejora en una propiedad del sistema deseada. En un aspecto, la generación de los siguientes puntos de datos se repite hasta que se obtiene un nivel de propiedad del sistema deseado. En otro aspecto, los siguientes puntos de datos se generan iterativamente hasta que la magnitud de expresión de dos o más genes alcanza el nivel deseado.

En algunas realizaciones, el código del programa informático puede usar un método de cálculo que emplea análisis numéricos o algoritmos de optimización. En algunos aspectos, los métodos de optimización numérica pueden usar el algoritmo de Nelder-Mead, el método de Newton, el método cuasi-Newton, el método de gradiente conjugado, un método de punto interior, un descenso de gradiente, un método de subgradiente, un método de elipsoide, el método de Frank-Wolfe, un método de interpolación y métodos de búsqueda de patrones, o un modelo de colonia de hormigas.

En una realización específica, el programa informático para generar el siguiente punto de datos para la experimentación usa el algoritmo de Nelder-Mead. El método implementado por ordenador recibirá uno o más puntos de datos de entrada y calculará el punto de reflexión, el punto de expansión o el punto de contracción para determinar mediante cálculo el siguiente punto de datos con el que experimentar, en base a los puntos de datos de entrada.

En una implementación del algoritmo de Nelder-Mead, el código del programa tomará los puntos de datos de entrada recibidos como los vértices simples de un espacio n-dimensional, que tiene n+1 vértices simples. Luego, se evaluará la función objetivo para cada vértice del simplex, y el algoritmo usa esta información para proponer una secuencia de nuevas coordenadas para la evaluación. Las nuevas coordenadas serán determinadas por el código del ordenador de acuerdo con la siguiente lógica algorítmica:

1. Ordenar los vértices simples: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
2. Calcular x_0 , el centro de gravedad de todos los puntos excepto x_{n+1} .
3. Calcular una coordenada de reflexión: $x_r = x_0 + \alpha(x_0 - x_{n+1})$
4. Calcular una coordenada de expansión: $x_e = x_0 + \rho(x_0 - x_{n+1})$
5. Calcular una coordenada de contracción: $x_c = x_{n+1} + \rho(x_0 - x_{n+1})$
6. Calcular coordenadas de reducción: $x_i = x_1 + \sigma(x_i - x_1)$ para todos $i \in \{2, \dots, n+1\}$

La función objetivo se evalúa en estos puntos y se utiliza para determinar un nuevo simplex de acuerdo con los siguientes criterios:

1. Si las coordenadas de Reflexión, Expansión o Contracción son mejores que el peor punto simplex, x_{n+1} , definir un nuevo simplex reemplazando el peor punto simplex por el mejor de los tres (Reflexión, Expansión o Contracción).
2. De lo contrario, definir un nuevo simplex combinando el mejor punto simplex con las coordenadas de reducción.

En un aspecto, se proporciona un producto de un programa informático que comprende un medio tangible legible por ordenador que almacena una pluralidad de instrucciones para controlar un procesador para realizar una operación para determinar un punto de experimentación para controlar la magnitud de expresión de dos o más genes, comprendiendo las instrucciones recibir uno o más puntos de datos de entrada, en el que los puntos de datos de entrada proporcionan información sobre uno o más elementos reguladores y una propiedad del sistema; y determinar, con un ordenador, un siguiente punto de datos usando un método de cálculo, en el que el siguiente punto de datos proporciona información sobre uno o más elementos reguladores.

La Figura 25 muestra un diagrama de bloques de un ejemplo de sistema informático 600 utilizable con un sistema y métodos de acuerdo con las realizaciones de la presente invención. El sistema informático 600 puede usarse para ejecutar el código de programa para diversas reivindicaciones del método de acuerdo con realizaciones de la presente invención.

Cualquiera de los sistemas informáticos mencionados en el presente documento puede utilizar cualquier número adecuado de subsistemas. En la Figura 25 se muestran ejemplos de tales subsistemas en el aparato informático 600. En algunas realizaciones, un sistema informático incluye un único aparato informático, donde los subsistemas pueden ser los componentes del aparato informático. En otras realizaciones, un sistema informático puede incluir múltiples aparatos informáticos, cada uno de los cuales es un subsistema, con componentes internos.

Los subsistemas mostrados en la Figura 25 están interconectados a través de un bus de sistema 675. Se muestran subsistemas adicionales tales como una impresora 674, teclado 678, disco duro 679, monitor 676, que está acoplado al adaptador de pantalla 682, y otros. Los dispositivos periféricos y de entrada/salida (E/S), que se acoplan

al controlador de E/S 671, se pueden conectar al sistema informático por cualquier medio conocido en la técnica, como el puerto en serie 677. Por ejemplo, el puerto en serie 677 o la interfaz externa 681 se puede usar para conectar el sistema informático 600 a una red de área amplia como Internet, un dispositivo de entrada de ratón o un escáner. La interconexión a través del bus del sistema 675 permite que el procesador central 673 se comuniquen con cada subsistema y controle la ejecución de instrucciones desde la memoria del sistema 672 o el disco duro 679, así como el intercambio de información entre subsistemas. La memoria del sistema 672 y/o el disco duro 679 pueden incorporar un medio legible por ordenador. Cualquiera de los valores mencionados en el presente documento puede enviarse de un componente a otro y puede enviarse al usuario.

Un sistema informático puede incluir una pluralidad de los mismos componentes o subsistemas, por ejemplo, conectados entre sí por una interfaz externa 681 o por una interfaz interna. En algunas realizaciones, los sistemas informáticos, subsistemas o aparatos pueden comunicarse a través de una red. En tales casos, un ordenador puede considerarse un cliente y otro ordenador un servidor, donde cada uno puede ser parte de un mismo sistema informático. Un cliente y un servidor pueden incluir múltiples sistemas, subsistemas o componentes.

Debe entenderse que cualquiera de las realizaciones de la presente invención puede implementarse en forma de lógica de control usando hardware y/o software informático de manera modular o integrada. Según la divulgación y las enseñanzas proporcionadas en el presente documento, una persona de habilidad ordinaria en la técnica conocerá y apreciará otras formas y/o métodos para implementar realizaciones de la presente invención usando hardware y una combinación de hardware y software.

Cualquiera de los componentes o funciones de software descritos en esta solicitud puede implementarse como código de software para ser ejecutado por un procesador utilizando cualquier lenguaje informático adecuado, como, por ejemplo, Java, C++ o Perl, utilizando, por ejemplo, técnicas convencionales u orientadas a objetos. El código de software puede almacenarse como una serie de instrucciones o comandos en un medio legible por ordenador para almacenamiento y/o transmisión, los medios adecuados incluyen memoria de acceso aleatorio (RAM), una memoria de solo lectura (ROM), un medio magnético como un disco duro o un disquete, o un medio óptico como un disco compacto (CD) o DVD (disco versátil digital), memoria flash y similares. El medio legible por ordenador puede ser cualquier combinación de tales dispositivos de almacenamiento o transmisión.

Dichos programas también pueden codificarse y transmitirse utilizando señales portadoras adaptadas para la transmisión a través de redes cableadas, ópticas y/o inalámbricas que se ajusten a una variedad de protocolos, incluido Internet. Como tal, se puede crear un medio legible por ordenador de acuerdo con una realización de la presente invención usando una señal de datos codificada con dichos programas.

Los medios legibles por ordenador codificados con el código del programa pueden empaquetarse con un dispositivo compatible o proporcionarse por separado de otros dispositivos (por ejemplo, mediante descarga de Internet). Cualquier medio legible por ordenador puede residir en o dentro de un solo producto de programa informático (por ejemplo, un disco duro, un CD o un sistema informático completo), y puede estar presente en o dentro de diferentes productos de programas informáticos dentro de un sistema o red. Un sistema informático puede incluir un monitor, impresora u otra pantalla adecuada para proporcionar cualquiera de los resultados mencionados en el presente documento a un usuario.

Cualquiera de los métodos descritos en el presente documento puede realizarse total o parcialmente con un sistema informático que incluye un procesador, que puede configurarse para realizar las etapas. Por lo tanto, las realizaciones pueden dirigirse a sistemas informáticos configurados para realizar las etapas de cualquiera de los métodos descritos en el presente documento, potencialmente con diferentes componentes que realizan las etapas respectivas o un grupo de etapas respectivas. Aunque se presentan como etapas numeradas, las etapas de los métodos en el presente documento se pueden realizar al mismo tiempo o en un orden diferente. Además, partes de estas etapas se pueden usar con partes de otras etapas de otros métodos. Además, todas o partes de una etapa pueden ser opcionales. Además, cualquiera de las etapas de cualquiera de los métodos se puede realizar con módulos, circuitos u otros medios para realizar estas etapas.

La descripción anterior de ejemplos de realizaciones de la invención se ha presentado con fines ilustrativos y descriptivos. No pretende ser exhaustiva o limitar la invención a la forma precisa descrita, y son posibles muchas modificaciones y variaciones a la luz de las enseñanzas anteriores. Las realizaciones se eligieron y describieron para explicar mejor los principios de la invención y sus aplicaciones prácticas para permitir a otros expertos en la técnica utilizar mejor la invención en diversas realizaciones y con diversas modificaciones que sean adecuadas para el uso particular contemplado.

Ejemplos

Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar la invención reivindicada.

Ejemplo 1: Uso del método de Nelder-Mead para optimizar la eficiencia del descubrimiento de operones

Este ejemplo ilustra cómo manipular y seleccionar de manera recombinante y mediante cálculo secuencias codificantes de complejos génicos nativos y secuencias reguladoras heterólogas. Hemos denominado a este proceso "refactorización", que comprende la optimización de múltiples genes, la regulación del complejo génico y el establecimiento del contexto genético para el circuito biológico. La refactorización de complejos génicos complejos y la ingeniería de rutas metabólicas requiere numerosas iteraciones entre el diseño, la construcción y la evaluación para mejorar la propiedad deseada del sistema, p.ej. títulos más altos del producto, menor toxicidad o mejor fijación de nitrógeno.

Una forma común de afectar estas propiedades es modificar los niveles de expresión génica dentro del sistema, incluso si se desconoce la relación directa entre la expresión génica y la propiedad del sistema. Se pueden lograr cambios cuantitativos en la expresión génica mediante el uso de elementos reguladores, p.ej. promotores y sitios de unión a ribosomas, que exhiben un comportamiento racionalmente predecible.

Es posible utilizar métodos de optimización numérica para guiar la selección de elementos reguladores para alterar la expresión génica y mejorar las propiedades deseadas del sistema. Un algoritmo relevante es el método de Nelder-Mead, un algoritmo de optimización no lineal que minimiza una función objetivo en el espacio multidimensional. Utilizamos el método de Nelder-Mead para optimizar una propiedad del sistema donde cada dimensión en el espacio algorítmico corresponde a la expresión de un gen en el sistema diseñado. Los puntos en este espacio representan una combinación particular de niveles de expresión para los genes en el sistema. Como resultado, cada punto puede considerarse una cepa de diseño único. El algoritmo se utiliza para sugerir nuevas coordenadas en el espacio que mejoran la propiedad del sistema. Se pueden diseñar nuevas cepas modificando elementos reguladores para lograr los niveles sugeridos de expresión génica. Después de evaluar el rendimiento de las nuevas cepas, el algoritmo se puede usar para predecir modificaciones posteriores. Este proceso se repite hasta que la propiedad del sistema se haya mejorado la cantidad deseada.

El método de Nelder-Mead se basa en el concepto de un simplex, que es un objeto en el espacio N dimensional que tiene N+1 vértices. La función objetivo se evalúa en cada vértice del simplex, y el algoritmo usa esta información para proponer una secuencia de nuevas coordenadas para la evaluación. Se proponen nuevas coordenadas de acuerdo con el siguiente proceso:

1. Ordenar los vértices simples: $f(x_1) \leq f(x_2) \leq \dots \leq f(x_{n+1})$
2. Calcular x_0 , el centro de gravedad de todos los puntos excepto x_{n+1} .
3. Calcular una coordenada de reflexión: $x_r = x_0 + \alpha(x_0 - x_{n+1})$
4. Calcular una coordenada de expansión: $x_e = x_0 + \rho(x_0 - x_{n+1})$
5. Calcular una coordenada de contracción: $x_c = x_{n+1} + \rho(x_0 - x_{n+1})$
6. Calcular coordenadas de reducción: $x_i = x_1 + \sigma(x_i - x_1)$ para todos $i \in \{2, \dots, n+1\}$

La función objetivo se evalúa en estos puntos y se utiliza para determinar un nuevo simplex de acuerdo con los siguientes criterios:

1. Si las coordenadas de Reflexión, Expansión o Contracción son mejores que el peor punto simplex, x_{n+1} , definir un nuevo simplex reemplazando el peor punto simplex con el mejor de los tres (Reflexión, Expansión o Contracción).
2. De lo contrario, definir un nuevo simplex combinando el mejor punto simplex con las coordenadas de reducción.

Estas etapas constituyen una iteración del algoritmo. El simplex recién definido se convierte en la semilla para generar nuevas coordenadas durante la siguiente iteración del algoritmo. Las iteraciones generalmente continúan hasta que una de las coordenadas en el simplex cruza un umbral deseado para la evaluación de la función objetivo. Hemos optimizado el rendimiento de un operón de fijación de nitrógeno variando la selección de promotores que controlan la expresión de genes individuales. Inicialmente refactorizamos el operón *nifEN* para que cada gen se exprese bajo el control de un promotor de T7 único (Fig. 1). Para evaluar el impacto de refactorizar el operón *nifEN*, medimos cuantitativamente la capacidad del operón sintético para complementar una cepa con *nifEN* inactivado y recuperar la capacidad de fijar nitrógeno (Fig. 2). Nuestro sistema refactorizado mostró capacidad limitada para fijar nitrógeno (20 % de la actividad de tipo silvestre).

Posteriormente aplicamos el método de Nelder-Mead para optimizar la expresión de los genes *nifE* y *nifN* con el objetivo de mejorar las tasas de fijación de nitrógeno. Nuestro espacio algorítmico consistía en dos dimensiones, la expresión de *nifE* y *nifN*. Nuestro sistema de coordenadas se ajustó a la fuerza de los promotores que controlan estos genes. Para permitir niveles variados de expresión génica, generamos y caracterizamos una biblioteca de promotores de T7 mutantes (Fig. 3). Nuestra biblioteca cubre tres órdenes de magnitud de expresión génica (esta es la misma biblioteca que se describe en la Solicitud de patente provisional de los Estados Unidos N.º 61/493.733. Aquí, se caracteriza por el comportamiento en *Klebsiella oxytoca*. A continuación, seleccionamos al azar mutantes de la biblioteca de promotores de T7 para generar dos cepas adicionales con niveles racionalmente alterados de expresión de *nifE* y *nifN*. La fuerza de los promotores de T7 utilizados en estas tres cepas definió nuestro simplex inicial. Evaluamos la fijación de nitrógeno para cada cepa en el simplex (cepa 1: 20 %, cepa 2: 9 %, cepa 3: 12 %) y

utilizamos el algoritmo para calcular las coordenadas de Reflexión (Fig. 4). Para construir una cepa que coincidiese con las coordenadas de Reflexión, elegimos los promotores de nuestra biblioteca más cercanos a las coordenadas en fuerza. Evaluamos la fijación de nitrógeno en esta cepa Reflexión y descubrimos que superó significativamente (52 %) a nuestras cepas iniciales (Fig. 5).

Nuestra cepa mejorada tuvo resultados sorprendentes y superó las expectativas, y tuvo un rendimiento suficiente para las aplicaciones posteriores. Para alcanzar niveles más altos de expresión génica, pueden diseñarse y usarse promotores más fuertes en los métodos de la invención. Como alternativa, para lograr los niveles de expresión deseados se pueden usar cambios complementarios en múltiples elementos reguladores, por ejemplo, el promotor y el sitio de unión al ribosoma para un gen dado. Esto implica describir las fuerzas de cada tipo de elemento en unidades comunes de expresión. Este ejemplo demuestra que pueden diseñarse nuevas cepas modificando elementos reguladores para alcanzar los niveles deseados de expresión génica. El ejemplo también ilustra el uso de métodos de optimización numérica, tales como, entre otros, el método de Nelder-Mead, para guiar la selección de elementos reguladores a fin de alterar la expresión génica y mejorar las propiedades deseadas del sistema.

Ejemplo 2: Refactorización de la fijación de nitrógeno

Este ejemplo demuestra el método de refactorización del complejo génico de fijación de nitrógeno. El método incluye etapas que comprenden: 1) eliminar la regulación del hospedador e implementar la regulación ortogonal sintética; 2) rastrear la contribución de cada parte reguladora a la función del complejo génico; 3) promover la modularidad y la integración con circuitos sintéticos; y 4) crear una plataforma susceptible de optimización racional. En ciertas realizaciones, el método para refactorizar la fijación de nitrógeno comprende reducir el complejo a genes característicos y ensamblar el complejo sintético.

El complejo génico *nif* de *Klebsiella oxytoca* ha sido uno de los principales modelos para el estudio de la enzima nitrogenasa (Fig. 7; ver, Rubio y Ludden, Maturation of Nitrogenase: a Biochemical Puzzle, J. Bacteriology, 2005). Es un complejo génico conciso, que abarca 20 genes en 7 operones dentro de 25kb de ADN. La enzima nitrogenasa se compone de dos unidades principales, Componente I y Componente II, que interactúan para facilitar la reducción de gases con enlaces múltiples como el N₂. Dentro del complejo enzimático, múltiples complejos de Fe-S son responsables de la química del sitio activo y la transferencia de electrones al sitio activo. La mayoría de los genes en el complejo génico están involucrados en la biosíntesis del complejo Fe-S, el acompañamiento y la inserción en el complejo enzimático final.

Casi todos los genes *nif* producen una proteína con una función que se sabe que es esencial para el ensamblaje o función de la nitrogenasa (ver, Simon, Homer and Roberts, Perturbation of *nifT* expression in *Klebsiella pneumoniae* has limited effect on nitrogen fixation, J. Bacteriology, 1996 y Gosink, Franklin y Roberts, The product of the *Klebsiella pneumoniae nifX* gene is a negative regulator of the nitrogen fixation (*nif*) regulon, J Bacteriology, 1990). Dos genes *nifL* y *nifA*, codifican las proteínas reguladoras maestras. El gen *nifT* no tiene una función conocida, y su eliminación tiene poco efecto sobre la fijación de nitrógeno. Además, aunque la eliminación de *nifX* tiene un efecto menor en la fijación de nitrógeno, su sobreexpresión reduce perjudicialmente la actividad enzimática. Por estas razones, elegimos eliminar *nifL*, *nifA*, *nifT* y *nifX* de nuestro complejo génico refactorizado.

Diseñamos genes sintéticos mediante la codificación aleatoria del ADN que codifica cada secuencia de aminoácidos. Las secuencias codificantes de proteínas se basaron en la secuencia depositada en la base de datos NCBI (X13303; ver, 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) La selección de codones se realizó mediante DNA2.0 usando un algoritmo interno y dos criterios de guía. Especificamos que nuestros genes se expresan razonablemente bien en *E. coli* y *Klebsiella*. Además, especificamos que nuestro uso de codones fuese lo más divergente posible del uso de codones en el gen nativo. Al diseñar genes sintéticos, rastreamos cada secuencia propuesta en busca de una lista de características no deseadas y rechazamos cualquier en la que se encontró una característica. La lista de características incluye sitios de reconocimiento de enzimas de restricción, sitios de reconocimiento de transposones, secuencias repetitivas, promotores sigma 54 y sigma 70, sitios de unión a ribosomas crípticos y terminadores independientes de rho. La Figura 18 muestra secuencias de ADN para genes nativos y genes sintéticos, así como el porcentaje de identidades comunes de nucleótidos y codones entre cada par.

Se eligieron los sitios de unión a ribosomas sintéticos para que coincidan con la fuerza de cada sitio de unión a ribosomas nativo correspondiente. Para caracterizar la fuerza de un sitio de unión a ribosomas nativo dado, construimos un plásmido indicador fluorescente en el que los 150 pb que rodean el codón de inicio de un gen (de -60 a +90) se fusionaron con el gen mRFP (Fig. 8). La quimera se expresó bajo el control del promotor Ptac, y la fluorescencia se midió mediante citometría de flujo (Fig. 9). Para generar sitios de unión al ribosoma sintéticos, construimos una biblioteca de plásmidos indicadores utilizando 150 pb (-60 a +90) de un casete de expresión sintético. Brevemente, un casete de expresión sintético consistía en un espaciador de ADN aleatorio, una secuencia degenerada que codifica una biblioteca RBS y la secuencia codificante para cada gen sintético. Examinamos múltiples clones para identificar el sitio de unión al ribosoma sintético que mejor coincidía con el sitio de unión al ribosoma nativo (Fig. 10).

Construimos operones sintéticos que consistían en los mismos genes que los operones nativos. Esta estrategia nos permitió inactivar un operón nativo de *Klebsiella* y complementar la delección usando la contraparte sintética.

5 Cada operón sintético consistía en un promotor Ptac seguido de casetes de expresión génica sintéticos (espaciador de ADN aleatorio, rbs sintético, secuencia codificante sintética) y un terminador de la transcripción. El espaciador de ADN aleatorio sirve para aislar la expresión de cada secuencia codificante sintética de los casetes anteriores. Cada operón sintético se examinó para eliminar secuencias reguladoras no deseadas (similar al proceso utilizado durante el diseño y síntesis de genes sintéticos).

10 En dos casos, encontramos operones sintéticos que no mostraron complementación funcional en la cepa inactivada correspondiente (nifHDKTY y nifUSVWZM). Para depurar los operones sintéticos, dividimos el operón en casetes de expresión génica constituyente. Luego construimos operones quiméricos, en los que algunos casetes tenían componentes sintéticos y otros eran genes nativos y sus sitios de unión al ribosoma (Fig. 11). Esta estrategia nos permitió probar cada operón quimérico para la complementación funcional e identificar rápidamente los casetes de expresión sintética problemáticos. Con un análisis más detallado de los casetes de expresión problemáticos, pudimos diagnosticar y corregir errores en los operones totalmente sintéticos. La Figura 12 ilustra una lista de errores en los dos operones.

20 Cada operón sintético se diseñó inicialmente para ser controlado por un promotor inducible por Ptac. Al valorar la concentración de IPTG, podríamos especificar con precisión la fuerza del promotor y la correspondiente expresión del operón sintético. Esto nos permitió variar el nivel de expresión para identificar la función óptima del operón. Descubrimos que cada operón sintético requería diferentes niveles de concentración de IPTG para una función óptima (Fig. 13).

25 Utilizamos el sistema de conexiones de T7 para desacoplar el promotor Ptac de cada operón sintético. Al insertar la conexión entre el promotor y la unidad transcripcional, logramos dos hitos importantes. Primero, obtuvimos la capacidad de modular la señal transcripcional mediante el uso de varios promotores de T7 mutantes. Esto nos permitió cambiar la función óptima del operón a una concentración de inductor único seleccionando los promotores de T7 mutantes correspondientes. En segundo lugar, modulamos el control del operón sintético (Fig. 14). Es decir, cualquier circuito genético puede controlar el operón sintético siempre que pueda producir la concentración necesaria de ARNP de T7 para guiar cada conexión.

30 Adoptamos un enfoque jerárquico para ensamblar operones individuales en un complejo completamente sintético. Primero, ensamblamos tres operones en semi-complejos ((nifJ-nifHDKY-nifEN y nifUSVWZM-nifF-nifBQ) y demostramos la capacidad de cada semi-complejo sintético para complementar la función en una cepa inactivada correspondiente. A continuación, combinamos los dos semi-complejos en un complejo sintético completo y demostramos la fijación de nitrógeno en una cepa con *nif* completo inactivado (Fig. 15).

40 Hemos demostrado que el uso de conexiones de T7 produce un complejo génico sintético modular. Hemos demostrado que el uso del controlador N.º 1 o del controlador N.º 2 produce el mismo rendimiento funcional del complejo sintético (Fig. 16). En el controlador N.º 1, la ARNP de T7 está bajo el control del promotor Ptac. En el controlador N.º 2, ARNP de T7 está bajo el control del promotor Ptet.

45 La Figura 17 muestra un esquema del complejo biológico completo, con cada parte detallada. La Figura 19 muestra la lista de piezas de los controladores sintéticos. La Fig. 20 muestra nombres, secuencias y fuerzas de cada componente del complejo completo.

50 Además, hemos demostrado que pueden usarse circuitos genéticos complejos para producir el rendimiento funcional del complejo génico sintético. Construimos un circuito genético que codifica la lógica "A y no B" y usamos este circuito para controlar la ARNP de T7. En este circuito, la lógica "A y no B" corresponde a la presencia o ausencia de los inductores, IPTG y aTc, de modo que la célula calcula "IPTG y no aTc". El circuito se construyó modificando el controlador N.º 1 para incluir los sitios de unión al represón cl OR1 y OR2 en el promotor Ptac para producir el controlador N.º 3. Además, el plásmido pNOR1020 (ver, por ejemplo, Tamsir y Voigt Nature 469:212-215 (2011)) codifica el represor cl bajo control del promotor Ptet. Cuando pNOR1020 y el controlador N.º 3 se transforman conjuntamente, producen el circuito lógico "IPTG y no aTc".

Secuencia promotora del Controlador Ptac N.º 1:

tattctgaaatgagctgttgacaattaatcatcgctcgataatgttggaattgtgagcgataacaatt

60 Secuencia promotora del Controlador N.º 3:

tattaacacgctgctgttgacagctataacctctggcggtataatgctagcggaattgtgagcgataacaatt

Entrada	Salida lógica esperada	Rendimiento de fijación de nitrógeno sintético (% en peso)
Sin inductor	0	<0,5 %
IPTG 1 mM	1	9 %

(continuación)

Entrada	Salida lógica esperada	Rendimiento de fijación de nitrógeno sintético (% en peso)
a Tc 50 ng/ml	0	<0,5 %
IPTG 1 mM y a Tc 50 ng/ml	0	<0,5 %

En este experimento, también incluimos el controlador N.º 1 como referencia de rendimiento. En condiciones inductoras (IPTG 1 mM), el controlador N.º 1 exhibe el 12 % de la fijación de WT.

Ejemplo 3: refactorización del sistema de secreción bacteriana tipo III (T3SS)

Este ejemplo ilustra el uso del método descrito en el presente documento para refactorizar completamente el sistema de secreción bacteriana tipo III (T3SS). Este ejemplo también ilustra que los operones sintéticos refactorizados de T3SS son controlables y funcionan independientemente de todo control y regulación nativos.

Los sistemas de secreción bacteriana tipo III (T3SS) son valiosos porque, a diferencia de las vías Sec y Tat utilizadas convencionalmente, translocan polipéptidos a través de las membranas internas y externas. Esto permite la administración de proteínas directamente a los medios de cultivo, que pueden ser uno de los requisitos críticos en la tecnología de manipulación bacteriana. Por ejemplo, las proteínas tóxicas se pueden eliminar del citoplasma sin permitir que entren en el periplasma y las enzimas funcionales (p.ej., celulasas) que necesitan trabajar fuera de la célula, se pueden administrar directamente a los medios.

Sin embargo, la dificultad para utilizar T3SS en sistemas bacterianos diseñados es doble. T3SS generalmente existe en bacterias patógenas que utilizan estos mecanismos para la invasión de las células hospedadoras. Por lo tanto, los T3SS están muy estrechamente regulados en la célula y son difíciles de controlar de forma independiente. Debido a esto, elegimos usar los métodos de la presente invención para refactorizar completamente el T3SS y probar la función de los operones refactorizados en las células inactivadas.

El término "refactorización" se refiere a un proceso que implica la optimización de múltiples genes, la regulación de un complejo génico y el establecimiento del contexto genético para un circuito biológico. La refactorización de complejos génicos complejos y el diseño de rutas biológicas requiere numerosas iteraciones entre el diseño, la construcción y la evaluación para mejorar la propiedad deseada del sistema. Brevemente, la refactorización incluye descomponer un sistema biológico en sus componentes y reconstruirlo sintéticamente. También implica eliminar todo el control nativo y la regulación del sistema biológico para reemplazarlo con un mecanismo que proporcione un control independiente.

Este ejemplo ilustra un método de recodificación de 18 genes de los sistemas de secreción bacteriana tipo III. El término "recodificación" se refiere a un método para eliminar o reemplazar la secuencia de un gen con el fin de reducir o eliminar cualquier elemento de regulación nativo, al tiempo que preserva la secuencia de proteína codificada por el gen. Los genes del sistema de secreción tipo III se recodificaron utilizando un algoritmo proporcionado por DNA2.0 (Menlo Park, CA) en el que los codones individuales de cada gen se vuelven a seleccionar de modo que el gen codifique la misma proteína, pero con la máxima diferencia con la secuencia nativa.

Los 18 genes están dispuestos en dos operones bacterianos. Cada gen es una versión recodificada de un gen nativo de *Salmonella typhimurium*. Cada gen está acoplado a una secuencia de sitio de unión al ribosoma (RBS) sintético que establece un nivel de expresión apropiado para cada gen individual. Los detalles de la selección sintética de RBS se describen a continuación. Los operones se pueden inducir con cualquier promotor deseado. En este ejemplo, se usan promotores inducibles simples. Los operones T3SS recodificados se pueden conectar a cualquier circuito de control genético según sea necesario.

Para seleccionar una secuencia de RBS sintética que coincida mejor con el nivel de expresión nativa de cada uno de los 18 genes de los sistemas de secreción bacteriana tipo III, medimos la expresión de cada gen en el sistema natural. Clonamos la región de 36 bases en dirección 5' en el codón de inicio, junto con las 36 bases de la región de codificación fusionadas a una RFP (proteína fluorescente roja). Esto se clonó en un plásmido con un promotor constitutivo.

Esta construcción se transformó en *Salmonella typhimurium* SL1344 y creció durante la noche a 37 °C en medio inductor de PI-1 (LB con 17 g/l de NaCl). El cultivo se subcultivó en medios de inducción frescos hasta una OD₂₆₀ de 0,025, se cultivó durante 2 horas a 37 °C hasta que las células alcanzaron la fase logarítmica. La fluorescencia se midió en un citómetro. La media geométrica de la fluorescencia de RFP de al menos 10.000 células se utilizó como medida de la expresión de proteínas.

Para encontrar secuencias de unión al ribosoma para probar, utilizamos la Calculadora de sitios de unión a ribosomas (<http://www.voigtlab.ucsf.edu/software/>), identificamos secuencias RBS conocidas del Registro de partes biológicas estándar (http://partsregistry.org/Main_Page), y se generó una serie de secuencias aleatorias. Las secuencias aleatorizadas comprenden los siguientes formatos:

CTTGGGCACGCGTCCATTAANNAGGANNAATTAAGC;
 TGGGCACGCGTCCATTAANNAGGANNAATTATTAGC;
 TACTTGGGCACGCGTCCATTAANNAGGANNAATAGC;
 CTTGGGCACGCGTCCATTAANAAGGAGNAATTAAGC;
 5 CTTGGGCACGCGTCCATTANTAAGGAGGNATTAAGC.

Todas las secuencias de RBS se clonaron en el vector de prueba de RBS (Fig. 21) junto con las primeras 36 bases del gen sintético que se generaron como guía. Seguimos el mismo procedimiento experimental utilizado para medir la expresión de cada gen en el sistema natural. De los RBS aleatorizados, se analizaron 12 - 48 colonias de cada
 10 secuencia aleatorizada. La construcción sintética que mejor se ajustaba al nivel de expresión nativa fue seleccionada y secuenciada. Esta secuencia se usó después en la construcción de los operones refactorizados.

Se ensamblaron dos operones. El primero, "prg-org" contiene 6 genes, y el segundo "inv-spa" contiene 13 genes. Estos genes se asignan a cada operón de la misma manera que en el sistema de tipo silvestre. Sin embargo, el
 15 orden de los genes en cada operón se organiza en función del nivel de expresión medido de más fuerte a más débil. Los operones se ensamblaron colocando el RBS sintético seleccionado frente a su secuencia genética sintética correspondiente. Se agregaron sitios de unión de enzimas de restricción entre genes o pares de genes para facilitar la manipulación futura. La secuencia completa fue sintetizada por DNA2.0. El operón sintético se clonó en un vector de prueba de bajo número de copias y se colocó bajo el control de un promotor inducible (p.ej., pTac o pBad - IPTG
 20 o inducción de arabinosa). Se creó un plásmido indicador que contiene una proteína efectora secretable de *Salmonella* nativa que se fusionó con una etiqueta de epítipo FLAG para identificación. Este indicador se puso bajo el control de un promotor constitutivo fuerte. La Figura 22B muestra un esquema del vector de prueba de operón prg-org y un plásmido indicador.

También generamos líneas celulares SL1344 de *Salmonella* con dos operones inactivados (prg-org e inv-spa) usando el método descrito en Datsenko, Wanner, Proc. Natl. Acad. U.S.A., 2000. La Figura 22A muestra un esquema de una cepa de SL1344 de *Salmonella* Δprg-org inactivada. Los operones inv-spa y prg-org se muestran en un recuadro. La Figura 23A muestra que la cepa Δprg-org inactivada no expresa el operón prg-org.

El plásmido de prueba (o el plásmido de control) y el plásmido indicador se transformaron en la cepa inactivada apropiada. Las cepas se cultivaron a partir de una colonia durante la noche en medio con bajo contenido de sal (LB con NaCl 5 g/l) a 37 °C. Los cultivos fueron subcultivados hasta una OD₂₆₀ de 0,025 en medio con bajo contenido de sal fresco y se cultivaron durante 2 horas. Los cultivos se diluyeron 1:10 en medios de inducción con alto contenido de sal (LB con 17 g/l de NaCl) en matraces sin deflectores de 50 ml y se cultivaron durante 6-8 horas. Se centrifugó
 30 1 ml de cada cultivo a 3000xg durante 5 minutos, y a continuación el sobrenadante se filtró a través de un filtro de 0,2 µM. Este cultivo se analizó después en un gel SDS-PAGE y se realizó una transferencia Western con un anticuerpo anti-FLAG.

La Figura 23B muestra que el operón prg-org refactorizado sintético en células de *Salmonella* Δprg-org puede controlarse mediante la adición de IPTG. El nivel de expresión es comparable al generado por el promotor natural PprgH.
 40

Ejemplo 4: Refactorización del complejo génico de fijación de nitrógeno a partir de *Klebsiella oxytoca*

Los genes bacterianos asociados con un único rasgo a menudo se agrupan en una unidad contigua del genoma conocida como complejo génico. Es difícil manipular genéticamente muchos complejos génicos debido a la compleja, redundante e integrada regulación del hospedador. Hemos desarrollado un enfoque sistemático para especificar por completo la genética de un complejo génico mediante la reconstrucción de abajo hacia arriba utilizando solo partes sintéticas y bien caracterizadas. Este proceso elimina toda la regulación nativa, incluida la que
 45 no se ha descubierto. Primero, se elimina todo el ADN no codificante, las proteínas reguladoras y los genes no esenciales. Los codones de genes esenciales se cambian para crear una secuencia de ADN tan divergente como sea posible del gen de tipo silvestre. Los genes recodificados se exploran mediante cálculo para eliminar la regulación interna. Se organizan en operones y se colocan bajo el control de partes sintéticas (promotores, sitios de unión al ribosoma y terminadores) que están funcionalmente separados por partes aisladoras. Finalmente, un
 50 controlador que consiste en sensores y circuitos genéticos regula las condiciones y la dinámica de la expresión génica. Aplicamos este enfoque a un complejo génico agrícolamente relevante de *Klebsiella oxytoca* que codifica la vía de fijación de nitrógeno para convertir el N₂ atmosférico en amoníaco. El complejo génico nativo consiste en 20 genes en 7 operones y está codificado en 23,5kb de ADN. Construimos un complejo génico refactorizado que comparte poca identidad de secuencia de ADN con el tipo silvestre y para el cual se define la función de cada parte
 55 genética. Este trabajo demuestra el potencial de las herramientas de biología sintética para reescribir la genética que codifica funciones biológicas complejas para facilitar el acceso, la manipulación y la transferibilidad.

INTRODUCCIÓN

Muchas funciones de interés para la biotecnología están codificadas en complejos génicos, que incluyen vías metabólicas, nanomáquinas, mecanismos de eliminación de nutrientes y generadores de energía (1). Los complejos
 65

generalmente contienen regulación interna que está integrada en la red reguladora global del organismo. Los promotores y las 5'-UTR son complejos e integran muchas entradas reguladoras (2, 3). La regulación es altamente redundante, conteniendo, por ejemplo, bucles de proalimentación y retroalimentación integrados (4). La regulación también puede ser interna a los genes, incluidos los promotores, los sitios de pausa y los ARN pequeños (5, 6). Además, los genes a menudo se superponen físicamente y las regiones de ADN pueden tener múltiples funciones (7). La redundancia y el alcance de esta regulación dificultan la manipulación de un complejo génico para romper su control mediante estímulos ambientales nativos, optimizar su función o transferirlo entre organismos. Como consecuencia, muchos complejos son crípticos, lo que significa que no se pueden identificar las condiciones de laboratorio en las que están activos (8).

Los complejos génicos se han controlado de arriba a abajo manipulando la regulación nativa o agregando regulación sintética en una constitución por lo demás de tipo silvestre (9). Por ejemplo, inactivando un represor o sobreexpresando un activador se han activado complejos que codifican rutas biosintéticas (10-14). Cuando el complejo es un operón único, se ha demostrado que se puede insertar un promotor en dirección 5' para inducir la expresión (15). Todo el complejo biosintético de equinomicina se transfirió a *E. coli* colocando cada gen nativo bajo el control de un promotor sintético (16).

En ingeniería, un enfoque para reducir la complejidad de un sistema es "refactorizarlo", un término prestado del desarrollo de software donde el código subyacente a un programa se reescribe para lograr algún objetivo (por ejemplo, estabilidad) sin cambiar la funcionalidad (17). Este término se aplicó por primera vez a la genética para describir la simplificación de arriba a abajo de un genoma de fago mediante el rediseño de elementos genéticos conocidos para que se puedan cambiar individualmente por digestión de restricción estándar (18). Aquí, lo usamos para referirnos a un proceso integral de abajo a arriba para eliminar sistemáticamente la regulación nativa de un complejo génico y reemplazarlo con partes y circuitos genéticos sintéticos (Fig. 26). El producto final es una versión del complejo génico cuya secuencia de ADN ha sido reescrita, pero que codifica la misma función. El proceso de diseño tiene lugar en un ordenador, y después la secuencia de ADN resultante se construye usando síntesis de ADN (19). La primera etapa del proceso es eliminar todo el ADN no codificante y los genes reguladores. A continuación, cada gen esencial se recodifica seleccionando codones que producen una secuencia de ADN que está lo más distante posible de la secuencia de tipo silvestre. La intención es introducir mutaciones en todo el gen para eliminar la regulación interna (incluida la que no se ha descubierto), como operadores, promotores, estructura secundaria de ARNm, sitios de pausa, sitios de metilación y regulación de codones. Las secuencias recodificadas se analizan adicionalmente por métodos de cálculo para identificar secuencias funcionales putativas, que luego se eliminan. Los genes recodificados se organizan en operones artificiales y los niveles de expresión están controlados por sitios de unión a ribosomas sintéticos (RBS), y las secuencias aislantes separan físicamente los genes. El resultado final es un complejo génico refactorizado cuya regulación nativa se ha eliminado y se ha organizado en un conjunto de partes genéticas discretas y bien caracterizadas.

Una vez que se ha eliminado la regulación nativa, se puede volver a agregar la regulación sintética para controlar la dinámica y las condiciones bajo las cuales se expresa el complejo. La construcción de dicha regulación ha sido un gran impulso de la biología sintética e implica el diseño de sensores y circuitos genéticos y la comprensión de cómo conectarlos para formar programas (20). En nuestro diseño, separamos genéticamente la detección/circuitería de la ruta refactorizada al transportarlo en diferentes plásmidos de bajo número de copias (Fig. 26). El plásmido que contiene los sensores y circuitos se conoce como el "controlador" y la salida de los circuitos conduce a la expresión de una polimerasa T7 manipulada (T7*). El complejo refactorizado está bajo el control de los promotores de T7. Una ventaja de esta organización es que la polimerasa T7 es ortogonal a la transcripción nativa y los promotores de T7 están muy apagados en ausencia del controlador. Además, cambiar la regulación se simplifica para intercambiar el controlador por uno que contenga diferentes sensores y circuitos, siempre que el rango dinámico de T7* sea fijo.

Como demostración, hemos aplicado este proceso para refactorizar el complejo génico que codifica la fijación de nitrógeno en *Klebsiella oxytoca* (21). La fijación de nitrógeno es la conversión de N₂ atmosférico en amoníaco (NH₃), para que pueda entrar en el metabolismo (22). La fijación industrial de nitrógeno a través del proceso de Haber-Bosch se utiliza para producir fertilizantes. Muchos microorganismos fijan el nitrógeno y los genes necesarios generalmente se presentan juntos en un complejo génico, incluidas las subunidades de la nitrogenasa, las enzimas biosintéticas del metalocomplejo y las chaperonas, el transporte electrónico y los reguladores (Fig. 27A) (23, 24). El complejo génico de *K. oxytoca* ha sido un sistema modelo para estudiar la fijación de nitrógeno y consiste en 20 genes codificados en 23,5 kb de ADN (Fig. 26, arriba) (25). La biosíntesis de nitrogenasa está estrechamente regulada por una cascada transcripcional de dos capas en respuesta al nitrógeno, oxígeno y temperatura fijos (26). El complejo completo se ha transferido a *E. coli*, demostrando así que tiene todos los genes necesarios para la fijación de nitrógeno (27). La codificación de esta función es compleja, muchos de los genes se superponen, los operones están orientados en direcciones opuestas y hay muchos posibles elementos reguladores ocultos, incluidos los promotores internos y las horquillas (25). El propósito de la refactorización es reorganizar el complejo, simplificar su regulación y asignar una función concreta a cada parte genética.

RESULTADOS

Tolerancia del complejo génico nativo a los cambios en la expresión

Antes de refactorizar un complejo, se realiza un análisis de robustez para determinar las tolerancias de un gen o conjunto de genes a los cambios en el nivel de expresión (Fig. 27B). Esto aporta información sobre la agrupación de genes en operones y la selección de partes sintéticas para obtener los niveles de expresión deseados. En la constitución de tipo silvestre, los genes son inactivados y complementados bajo control inducible. La tolerancia se obtiene midiendo la actividad de la nitrogenasa en función de la actividad del promotor inducible.

La función de la nitrogenasa es notablemente sensible a los cambios de expresión y cada tolerancia tiene un óptimo claro (Fig. 27B). La chaperona NifY es necesaria para lograr una actividad completa y amplía la tolerancia a los cambios en el nivel de expresión. NifT no tuvo ningún efecto sobre la actividad, como se observó anteriormente (28), y con frecuencia está ausente de los complejos homólogos (29). Los genes que controlan el transporte de electrones (*nifJ* y *nifF*) deben expresarse a niveles bajos, y la actividad cae rápidamente a medida que aumenta la expresión. Los valores óptimos para los genes que participan en las rutas biosintéticas de los complejos metálicos varían. El operón *nifUSVWZM*, que codifica proteínas para la formación temprana de complejos Fe-S y proteínas para la maduración de componentes, debe expresarse a niveles bajos, mientras que *nifBQ*, que codifica proteínas para la síntesis del núcleo de FeMo-co y la integración de molibdeno deben expresarse 10 veces más. NifEN es tolerante a niveles de expresión variados. Sin embargo, la actividad se pierde con la inclusión de *nifX*, que se ha caracterizado como un regulador negativo (30). El complejo nativo también incluye las proteínas reguladoras NifL y NifA, que integran señales ambientales (26). Los genes *nifT*, *nifX*, y *nifLA* no están incluidos en el complejo refactorizado.

El complejo génico refactorizado completo

Las actividades de nitrogenasa de los operones refactorizados se midieron en función del promotor P_{tac} inducible por IPTG (Fig. 28A). Cada operón tiene un óptimo diferente. Para combinar los operones, los promotores P_{tac} fueron reemplazados por promotores de T7 que tienen una fuerza cercana al óptimo medido (Fig. 28B y sección Materiales y Métodos). Los genes de nitrogenasa (*nifHDK*) están altamente expresados en *Klebsiella* en condiciones de fijación (hasta 10 % de proteína celular) (31), por lo que se utilizó el promotor más fuerte para controlar este operón (T7.WT, 0,38 REU) (32). Se construyó un operón largo para incluir los genes *nifEN* y *nifJ*, donde la menor expresión requerida para *nifJ* se logró mediante atenuación transcripcional. El gen *nifF* se codificó por separado bajo el control de un promotor de fuerza media (T7.3, 0,045 REU). Finalmente, los operones *nifUSVWZM* y *nifBQ* fueron controlados por promotores débiles (T7.2, 0,019 REU). Cada uno de los operones refactorizados individuales bajo el control de un promotor de T7 pudo recuperar la actividad observada del promotor P_{tac} y la correspondiente concentración óptima de IPTG (Fig. 28C).

La transición del control a los promotores T7* y T7 facilita el ensamblaje del complejo completo a partir de operones refactorizados. Primero ensamblamos semicomplejos usando Gibson Assembly (33) y verificamos su función en cepas con los genes correspondientes eliminados. El primer semicomplejo consistió en el operón *nifHDKYENJ*. El segundo semicomplejo se ensambló a partir de los operones *nifBQ*, *nifF*, y *nifUSVWZM*. Los semicomplejos pudieron recuperar el 18 % $\pm$ 0,7 % y el 26 % $\pm$ 8,4 % de la actividad de tipo silvestre, respectivamente. El complejo sintético completo se ensambló a partir de ambos semicomplejos (Fig. 29), y su actividad se midió en una cepa donde se ha eliminado el complejo completo. El complejo génico sintético recupera la actividad de la nitrogenasa a 7,4 % $\pm$ 2,4 % del tipo silvestre (Fig. 30A). Las cepas que portaban el complejo génico sintético utilizaron el ambiente de N_2 como fuente de nitrógeno, creciendo 3,5 veces más lento que la cepa de tipo silvestre (Fig. 37) e incorporando nitrógeno marcado con ^{15}N en 24 % $\pm$ 1,4 % de su contenido de nitrógeno celular, medido por la espectrometría de masas con relaciones isotópicas (IRMS) (Fig. 30B).

El complejo refactorizado completo consiste en 89 partes genéticas, incluido un controlador, y la función de cada parte está definida y caracterizada. Por lo tanto, la genética del sistema refactorizado está completa y se define en el esquema de la Fig. 29. Sin embargo, el proceso de simplificación y modularización reduce la actividad (18). Este es un resultado esperado de la refactorización de un sistema altamente evolucionado.

Intercambio de controladores para cambiar la regulación

La separación del controlador y el complejo refactorizado simplifica el cambio de la regulación del sistema. Esto se puede lograr transformando un plásmido controlador diferente, siempre que se mantenga el rango dinámico de la expresión de la ARNP de T7*. Para demostrar esto, construimos dos controladores adicionales (Fig. 30A). El controlador N.º 2 cambia la química que induce el sistema al colocar la expresión de la ARNP de T7* bajo el control del promotor P_{tet} inducible por aTc. Cuando se induce, el Controlador N.º 2 produce una actividad nitrogenasa idéntica a la inducida con el Controlador N.º 1 (7,2 % $\pm$ 1,7 %). El controlador también puede servir como plataforma para codificar circuitos genéticos para controlar la dinámica reguladora o para integrar múltiples sensores. Para este fin, el Controlador N.º 3 contiene dos sistemas inducibles (IPTG y aTc) y una puerta AND (34, 35). En presencia de IPTG y en ausencia de aTc, la fijación de nitrógeno es 6,6 % $\pm$ 1,7 % de la actividad de tipo silvestre. Estos controladores representan la simplicidad por la cual se puede cambiar la regulación del complejo refactorizado.

Además de permitir agregar nueva regulación, el proceso de refactorización elimina la regulación nativa del

complejo. Esto se demuestra mediante el desacoplamiento de la actividad nitrogenasa de las señales ambientales que normalmente regulan su actividad. Por ejemplo, el amoníaco es un regulador negativo que limita la sobreproducción de nitrógeno fijo (26). En presencia de amoníaco 17,5 mM, no se observa actividad nitrogenasa para el complejo de tipo silvestre (Fig. 30C). Por el contrario, el complejo génico refactorizado mantiene la actividad en presencia de amoníaco ($1,1 \% \pm 0,5 \%$). Curiosamente, esta reducción de actividad de 7 veces no se debe a la regulación residual presente en el sistema. Más bien, ocurre porque la adición de amoníaco a los medios reduce el rendimiento del controlador en 4,5 veces (Fig. 30C). En teoría, esto podría solucionarse aumentando el nivel de expresión de ARNP de T7*, pero habla de la necesidad de crear circuitos genéticos que sean robustos para el contexto ambiental.

DISCUSIÓN

El objetivo de la refactorización es facilitar la ingeniería avanzada de sistemas de múltiples genes codificados por genética compleja. Los complejos génicos nativos son producto de procesos evolutivos; por lo tanto, exhiben alta redundancia, eficiencia de la codificación de información y capas de regulación que dependen de diferentes mecanismos bioquímicos (36-38). Estas características inhiben la alteración cuantitativa de la función por sustitución de parte, porque el efecto puede incrustarse en una red de interacciones. Aquí, la modularización del complejo, la separación física y el aislamiento de las partes, y la simplificación de su regulación han guiado la selección y el análisis de las sustituciones de partes. La información obtenida de la detección de las permutaciones en un sistema refactorizado puede retroalimentarse limpiamente en el ciclo de diseño.

El complejo refactorizado también puede servir como plataforma para abordar preguntas en biología básica. Primero, permite que el impacto de las interacciones reguladoras se cuantifique de forma aislada. Por ejemplo, en el sistema natural, un circuito de retroalimentación podría integrarse en muchos otros circuitos reguladores. La eliminación sistemática de dicha regulación proporciona un sistema de referencia limpio (potencialmente menos activo y robusto que el tipo silvestre) a partir del cual se pueden cuantificar las mejoras como resultado de agregar una nueva regulación. También sirve como base para la comparación de programas reguladores radicalmente diferentes o principios organizativos; por ejemplo, para determinar la importancia del control temporal de la expresión génica (4, 39) o la necesidad de que los genes se codifiquen con una estructura de operón particular (40, 41). En segundo lugar, el proceso de reconstrucción y depuración es un mecanismo de descubrimiento que probablemente revele nuevos modos genéticos y reguladores. En este trabajo, la mejora del 0 % al 7 % reveló solo cambios menores: anotaciones erróneas en los genes y niveles de expresión inadecuados. Sin embargo, el proceso de depuración en sí mismo es ciego al mecanismo: simplemente identifica regiones problemáticas de ADN.

Una de las aplicaciones inmediatas de la refactorización es el acceso a complejos génicos a partir de la información de la secuencia genómica. Esto podría ser necesario porque el complejo es silente, lo que significa que no se puede activar en el laboratorio, o porque el complejo deseado proviene de una muestra metagenómica o una base de datos de información y el ADN físico no está disponible (42). Ha habido muchos métodos ingeniosos para activar un complejo génico, incluida la colocación de promotores inducibles en dirección 5' de los operones naturales y la división del complejo en cistrones individuales, que luego se pueden volver a ensamblar (43). Con los avances en la tecnología de síntesis de ADN, es posible construir complejos génicos completos con un control completo sobre la identidad de cada nucleótido en el diseño. Esta capacidad elimina la dependencia del ADN físico natural para la construcción y permite el rediseño simultáneo de componentes en el sistema completo. Aprovechar al máximo esta tecnología requerirá la combinación de métodos de cálculo para seleccionar partes y examinar diseños, bibliotecas de partes caracterizadas y métodos para reducir su dependencia del contexto.

MATERIAL Y MÉTODOS

Cepas y medios

Se usó la cepa S17-1 de *E. coli* para la construcción y propagación de todos los plásmidos utilizados en la construcción del mutante desactivado de *Klebsiella oxytoca*. La cepa M5al de *K. oxytoca* (Paul Ludden, UC Berkeley) y los mutantes derivados de M5al se usaron para experimentos de fijación de nitrógeno. Se usó medio de Luria-Bertani (LB)-Lennox para la propagación de la cepa. Todos los ensayos se llevaron a cabo en un medio mínimo que contenía (por litro) 25 g de Na_2HPO_4 , 3 g de KH_2PO_4 , 0,25 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g de NaCl, 0,1 g de $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 2,9 mg de FeCl_3 , 0,25 mg de $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ y 20 g de sacarosa. Los medios de crecimiento se definen como medios mínimos complementados con 6 ml (por litro) de 22 % de NH_4Ac . Los medios de desrepresión se definen como medios mínimos suplementados con 1,5 ml (por litro) de serina al 10 %. Los antibióticos utilizados fueron $34,4 \mu\text{g ml}^{-1}$ Cm, $100 \mu\text{g ml}^{-1}$ Spec, $50 \mu\text{g ml}^{-1}$ Kan, y/o $100 \mu\text{g ml}^{-1}$ Amp.

Aleatorización de codones

Las secuencias génicas iniciales fueron propuestas por DNA2.0 para maximizar la distancia de Hamming de la secuencia nativa a la vez que se busca un equilibrio óptimo entre el uso de codones de *K. oxytoca* y las preferencias de codones de *E. coli* determinadas experimentalmente por la empresa (44). Los codones raros (<5 % de aparición en *K. oxytoca*) se evitaron y se suprimió la estructura de ARNm en la región de inicio de la traducción. El algoritmo

DNA2.0 eliminó los motivos de secuencia conocidos, incluidos los sitios de restricción, los sitios de reconocimiento de transposones, las secuencias de Shine-Dalgarno y los terminadores transcripcionales.

Eliminación de la regulación no deseada

Cada operón sintético se examinó antes de la síntesis de ADN para identificar y eliminar la regulación no deseada. Se identificaron múltiples tipos de regulación utilizando software disponible al público. Se utilizó la calculadora RBS (Ingeniería inversa, ARN 16S: ACCTCCTTA) para identificar sitios de unión a ribosomas a lo largo de la secuencia de ADN del operón propuesta (45). El servidor Prokaryotic Promoter Prediction se utilizó para identificar supuestos sitios promotores $\sigma 70$ (valor de corte de 5, base de datos sigma.hmm) (46). El algoritmo PromScan se usó para identificar supuestos sitios promotores $\sigma 54$ utilizando las opciones predeterminadas (47). El software TransTermHP se utilizó con parámetros predeterminados para identificar secuencias de terminación tanto en la dirección directa como inversa (48). Los RBS superiores a 50 UA y todos los promotores y terminadores identificados se consideraron significativos.

Ensayo de actividad de nitrogenasa

La actividad *in vivo* de la nitrogenasa se determina mediante reducción de acetileno como se describió previamente (49). Para el ensayo de actividad de nitrogenasa de células enteras en *K. oxytoca*, las células que albergan los plásmidos apropiados se incubaron en 5 ml de medio de crecimiento (suplementado con antibióticos, 30 °C, 250 rpm) en tubos cónicos de 50 ml durante 14 horas. Los cultivos se diluyeron en 2 ml de medio de desrepresión (suplementado con antibióticos e inductores) hasta una OD₆₀₀ final de 0,5 en botellas de 14 ml, y las botellas se sellaron con tapones de goma (Sigma Z564702). El espacio de cabeza en las botellas fue evacuado repetidamente y enjuagado con N₂ y pasado a una trampa de catalizador de cobre usando un colector de vacío. Después de incubar los cultivos durante 5,5 horas a 30 °C, 250 r.p.m, el espacio de cabeza se reemplazó por 1 atmósfera de Ar. El acetileno se generó a partir de CaC₂ usando una botella de Burris, y se inyectó 1 ml en cada botella para comenzar la reacción. Los cultivos se incubaron durante 1 hora a 30 °C, 250 rpm antes de que el ensayo se detuviera por inyección de 300 µl de solución de NaOH 4M en cada botella. Para cuantificar la producción de etileno, se extrajeron 50 µl de espacio de cabeza de cultivo a través del tapón de goma con una jeringa hermética a los gases y se inyectaron manualmente en un cromatógrafo de gases HP 5890. La actividad de la nitrogenasa se expresa como un porcentaje de la actividad de tipo silvestre. Brevemente, la producción de etileno por las cepas se cuantificó integrando el área bajo el pico utilizando el software HP Chemstation y dividiendo la producción de etileno de cepas experimentales por la producción de etileno de un control de tipo silvestre incluido en cada ensayo.

Crecimiento dependiente de N₂ y ensayo de incorporación de ¹⁵N₂

La fijación de nitrógeno por el complejo *nif* sintético en *K. oxytoca* queda demostrada por el crecimiento dependiente de N₂ y la incorporación de ¹⁵N₂. Las células se diluyen como se describe en el ensayo de reducción de acetileno. El espacio de cabeza de las botellas se reemplaza por gas N₂ normal o por el isótopo estable del nitrógeno, ¹⁵N₂ (átomo ¹⁵N 99,9 %, Icon Isotopes, N.º Cat.: IN 5501). Después de incubar los cultivos durante 36 horas a 30 °C, 250 r.p.m, el crecimiento dependiente de N₂ de las células se determina midiendo la densidad óptica a 600 nm (OD₆₀₀). Para realizar el ensayo de incorporación de ¹⁵N₂, las células enriquecidas en ¹⁵N con los cultivos de control correspondientes bajo gas nitrógeno normal se recogen por centrifugación, los sedimentos celulares se secan en un horno de laboratorio a 100 °C durante 12 horas. Se analiza la relación ¹⁵N/¹⁴N de los gránulos secos en el Centro de Biogeoquímica de isótopos estables de la Universidad de California, Berkeley, utilizando el espectrómetro de masas de relaciones isotópicas Finnigan MAT Delta plus.

Cepas de *K. oxytoca* inactivadas

Todos los mutantes de *K. oxytoca* se construyen a partir de M5al mediante intercambio de alelos usando el plásmido suicida pDS132 que lleva la correspondiente delección del gen *nif* (pDS132 fue gentilmente proporcionado por el laboratorio Paul Ludden en UC Berkeley como obsequio del Dr. Dominique Schneider en la Universidad Joseph Fourier) (49). Hicimos una ligera modificación a un protocolo publicado previamente (50). Aquí, un casete de resistencia a la kanamicina se clonó en el plásmido suicida en dirección 5' del fragmento de intercambio homólogo izquierdo. Estas delecciones de operón en el complejo génico *nif* abarca el promotor y las secuencias completas de codificación de aminoácidos, excepto cuando se designa específicamente. Todos los mutantes se verificaron mediante secuenciación de ADN del producto de PCR de la región génica correspondiente para confirmar la eliminación física del ADN y mediante un ensayo de reducción de acetileno de células completas para confirmar la falta de actividad de nitrogenasa.

Caracterización de promotores

Como se describe en este ejemplo, la función de los promotores se expresa como unidades de expresión relativa (REU). Esto es simplemente un factor lineal que se multiplica por las unidades arbitrarias medidas por el citómetro de flujo. El objetivo de normalizar a REU es estandarizar las mediciones entre laboratorios y proyectos. El factor lineal es 1,66x10⁻⁵ y la división por este número se convierte en unidades arbitrarias sin procesar. Este número se

calculó como un proxy de las RPU (unidades promotoras relativas) descritas por Kelly y sus colaboradores (51). Nuestras medidas estandarizadas originales se realizaron antes del artículo de Kelly e incluyeron un promotor de referencia diferente, proteína fluorescente (mRFP), RBS y esqueleto del plásmido. Debido a estas diferencias, no se puede calcular las RPU según lo definido por Kelly, et al. En cambio, se realizó una serie de plásmidos (Fig. 33A) para estimar la expresión relativa de la proteína indicadora a partir de construcciones experimentales en comparación con la construcción estándar en Kelly, et al. Los factores de conversión entre construcciones se midieron y multiplicaron para obtener el factor lineal anterior. Cambiamos el nombre de la unidad a REU (unidades de expresión relativa) porque pretende ser una normalización simple de unidades fluorescentes (similar a un cordón fluorescente) y no una medición directa de la actividad de un promotor (por ej., el flujo de polimerasa).

Las células se cultivaron como en el ensayo de reducción de acetileno con dos modificaciones. La descarga inicial del espacio de cabeza con N₂ no se realizó, y el ensayo se detuvo después de la incubación de 5,5 horas. Para detener el ensayo, se transfirieron 10 µl de células de cada botella a una placa de 96 pocillos que contenía solución salina tamponada con fosfato suplementada con 2 mg ml⁻¹ de kanamicina. Los datos de fluorescencia se recogieron utilizando un citómetro de flujo BD Biosciences LSRII. Los datos fueron controlados por dispersión frontal y lateral, y cada conjunto de datos consistió en al menos 10.000 células. FlowJo se utilizó para calcular las medias geométricas de las distribuciones de fluorescencia. El valor de autofluorescencia de las células de *K. oxytoca* que no albergan plásmido se sustrajo de estos valores para dar los valores dados en este estudio. Las fuerzas de los promotores de T7 mutantes se caracterizaron por intercambiarlos in situ del promotor P_{tac} en el plásmido N149 (SBa_000516), co-transformando con el controlador N° 1 (plásmido N249), y midiendo la fluorescencia mediante citometría de flujo bajo inducción de IPTG 1 mM.

Para reemplazar el promotor P_{tac} por un promotor de T7 en cada operón sintético, seguimos un proceso simple. Primero, identificamos la concentración de IPTG correspondiente a la actividad funcional máxima de cada operón sintético. En segundo lugar, traducimos esta concentración de IPTG a REU en función de la caracterización del promotor P_{tac} (Fig. 33B, izquierda). En tercer lugar, seleccionamos el promotor de T7 mutante con la fuerza más cercana en REU. Para el operón sintético *nifF*, observamos una fijación amplia y robusta bajo el promotor P_{tac}. Encontramos que T7 mut 3 producía actividad funcional inducible con un máximo con inducción IPTG 1 mM de la ARNP de T7. Para el operón *nifJ* sintético, nuestro método sugiere que usemos un promotor de T7 mutante débil. Sin embargo, descubrimos que un promotor de T7 WT producía actividad inducible con un máximo de IPTG 1 mM. Atribuimos esta desviación a un cambio en la fuerza de RBS debido a diferencias contextuales entre P_{tac} y el promotor de T7.

Depuración de operones sintéticos

Algunos de los diseños iniciales para los operones refactorizados mostraron poca o ninguna actividad. Cuando esto ocurre, es difícil identificar el problema porque se han realizado tantos cambios genéticos simultáneamente hasta el punto que casi no hay identidad de ADN con la secuencia de tipo silvestre. Para identificar rápidamente el problema, se desarrolló un método de depuración que se puede generalizar al refactorizar diferentes funciones (Fig. 34A). Los operones quiméricos se crean al reemplazar una región de ADN de tipo silvestre con su contraparte sintética. La función de cada quimera en esta biblioteca se evalúa para identificar qué región de ADN sintético causó una pérdida de actividad. A continuación se construyen nuevas quimeras con cambios de resolución cada vez más finos entre el ADN sintético y el de tipo silvestre. Este enfoque “se acerca” a la región problemática del ADN, que luego se puede solucionar. El problema más común se debe a errores en la secuencia de ADN de referencia (Genbank, X13303.1) (52). Los genes refactorizados se diseñaron utilizando solo la información de la secuencia de aminoácidos de la base de datos; por lo tanto, eran sensibles a los errores de secuencia que conducen a mutaciones sin sentido que reducen o eliminan la actividad. De hecho, 18 de tales mutaciones se identificaron y confirmaron mediante una nueva secuenciación cuidadosa del complejo de tipo silvestre (Fig. 35). Quince de las 18 mutaciones ocurrieron en operones refactorizados que requirieron depuración y fueron corregidos (Fig. 34B). Esto demuestra la dificultad de reconstituir las funciones biológicas utilizando solo información de la base de datos y la síntesis de ADN (55).

La modificación de la fuerza sintética del RBS también fue importante para la depuración. La función del operón *nifUSVWZM* sintético se mejoró significativamente al cambiar RBS para que coincidiese con una relación 1:1 de NifU:NifS. La selección inicial de RBS condujo a una relación observada de 10:1 en sus respectivas fuerzas de RBS. Después de la depuración, la fuerza de RBS de *nifU* y *nifS* estaba mejor equilibrada (1,25: 1) y esto mejoró la actividad. Para un RBS, el método de medición resultó ser inexacto. Observamos que la fuerza medida del RBS de *nifQ* de tipo silvestre era extremadamente baja (Fig. 27C), y el operón *nifBQ* sintético mostró baja actividad cuando el RBS de *nifQ* sintético se comparaba con la resistencia medida. Por el contrario, el análisis de robustez mostró un requisito para un alto nivel de expresión del operón *nifBQ* (Fig. 27B). Por lo tanto, se usó un RBS sintético con una fuerza cercana a la del RBS de *nifB* y se mejoró significativamente la actividad del operón de *nifBQ*. En un caso, nuestro gen *nifH* recodificado inicial no se expresó bien utilizando la regulación de tipo silvestre ni la sintética (Fig. 37). Diseñamos un nuevo gen sintético, lo que requería la divergencia en la secuencia de ADN de las secuencias de ADN nativas y las primeras sintéticas y descubrimos que el nuevo gen sintético expresaba bien y recuperaba la actividad.

Crecimiento por fijación de nitrógeno

Las células capaces de fijación de nitrógeno deben exhibir un crecimiento medible en medios que carecen de nitrógeno utilizando N₂ atmosférico como fuente de nitrógeno. Por el contrario, las células incapaces de fijación de nitrógeno no deben crecer en medios libres de nitrógeno.

En paralelo al ensayo de incorporación de ¹⁵N₂, se controló el crecimiento de la cepa en condiciones de medios limitados en nitrógeno y 100 % de atmósfera de ¹⁵N₂ (Métodos, Ensayo de crecimiento dependiente de N₂). Las células se cultivaron en medios de desrepresión como se usa en el ensayo de actividad de nitrogenasa. Los medios de desrepresión no están estrictamente libres de nitrógeno, contienen serina 1,43 mM para promover la producción de ARN ribosómico y acelerar la biosíntesis de nitrogenasa (54).

Las cepas que contienen el controlador N.^o 1 y el complejo génico refactorizado crecieron casi un 30 % tanto como las cepas de tipo silvestre. Por el contrario, se observó un crecimiento mínimo en cepas Δnif , consistente con el limitado nitrógeno disponible de la serina y productos de la lisis celular (55). La Fig. 37 ilustra el crecimiento celular soportado por la fijación de nitrógeno.

Ensayo de transferencia Western para la expresión de *nifH* sintético

El primer *nifHDK* sintético no exhibió actividad nitrogenasa bajo la inducción que variaba de IPTG 0 a 1 mM, y el gen *nifH* (*nifH*_{v1} sintético) se identificó como una parte problemática utilizando el protocolo de depuración que se muestra en la Fig. 34. Sin embargo, no se encontró ninguna mutación. Las transferencias Western se usaron para confirmar la expresión problemática del *nifH* sintético.

Una transferencia Western para la proteína NifH en la Fig. 37 (izquierda) mostró que el *nifH* tipo silvestre se expresaba bien con cualquier *nifD* o *nifK* sintético (construcción N10, N12, N14), mientras que el *nifH*_{v1} sintético no se expresaba independientemente del contexto de *nifDK* (construcción N1 y N19). Se utilizó un segundo *nifH* sintético (*nifH*_{v2} sintético) para reemplazar el *nifH*_{v1} sintético. La transferencia Western en la Fig. 37 (derecha) mostró que el *nifH*_{v2} sintético (construcción N38) se expresaba bien.

Las muestras para las transferencias Western se prepararon mediante ebullición de las células de *K. oxytoca* recogidas en tampón de carga SDS-PAGE y desarrolladas en geles de SDS-poliacrilamida al 12 % (Lonza Biosciences). Las proteínas en los geles se transfirieron a membranas de PVDF (BioRad N.^o Cat: 162-0177) usando Trans-Blot SD Semi Dry Transfer Cell (BioRad N.^o Cat: 170-3940). El bloqueo de la membrana y la unión de anticuerpos se realizaron usando SNAP i.d. Protein Detection System (Millipore N.^o Cat WBAVDBA). Las membranas fueron bloqueadas por TBST-1 % BSA (TBS-Tween20). Los anticuerpos anti-NifH y anti-NifDK (proporcionados amablemente por Paul Ludden Lab en UC-Berkeley) se usaron como anticuerpos primarios. El anticuerpo anti-NifH era un anti-NifH universal hecho preparado contra una mezcla de proteínas NifH purificadas de *Azotobacter vinelandii*, *Clostridium pasteurianum*, *Rhodospirillum rubrum* y *K. oxytoca*. El anticuerpo anti-NifDK se preparó contra la proteína NifDK purificada de *Azotobacter vinelandii*. Los anticuerpos anti-NifH y anti-NifDK se usaron a 1:500 y 1:2000 respectivamente. El anticuerpo secundario (anti-IgG de conejo de cabra-HRP, Sigam N.^o Cat: A0545) se usó a 1:10.000. El desarrollo se realizó usando un sustrato quimioluminiscente mejorado para HRP (Pierce N.^o Cat: 32209) y se capturó en una película (Kodak: N.^o Cat: 178-8207).

Construcción de plásmidos y partes

Los plásmidos fueron diseñados *in silico*. Las partes sintéticas (promotores, RBS, terminadores y aisladores) se combinaron con las secuencias génicas sintéticas iniciales propuestas por DNA2.0 en ApE (A Plasmid Editor, <http://biologylabs.utah.edu/jorgensen/wayned/apel/>) y GeneDesigner (56) para crear operones sintéticos. Los operones sintéticos se analizaron mediante cálculo para eliminar la regulación no deseada (Métodos, "Eliminación de la regulación no deseada"), y las partes que contienen dicha regulación fueron reemplazadas. Este proceso reiterativo continuó hasta que los operones sintéticos incluyeron solo la regulación diseñada.

El ADN físico se construyó utilizando técnicas de manipulación estándar. Los métodos de ensamblaje siguieron los protocolos publicados e incluyeron BioBrick (57), Megawhop (58), mutagénesis dirigida al sitio de Phusion o métodos de ensamblaje Gibson (59). Descubrimos que Gibson Assembly era el método de ensamblaje de ADN más eficiente, excepto cuando se realizaban pequeños cambios (<10 pb) en plásmidos de menos de 10 kb de tamaño. Observamos que los fallos de ensamblaje eran poco frecuentes, más comunes en ensamblajes por encima de 15 kbps, y estaban vinculados a la presencia de homología dentro de ~500 pb de los extremos de la parte. En estos casos, observamos la hibridación de partes inesperadas para crear uniones no deseadas.

El plásmido plncW (pSa, SpR) se generó a partir de pEXT21 (pSa, SpR) mediante la eliminación de *osa*, *nuc1*, el gen integrasa Tn21 y ORF18 (60). El plásmido pSB4C5 (pSC101, CmR) se obtuvo del Registro de partes biológicas estándar y sirve como el vector base para la complementación de tipo silvestre, la caracterización de RBS y los operones sintéticos (57). El plásmido N58 (pSC101, CmR) se generó insertando el casete Ptac (Registro SynBERC, SBa_000561) entre el prefijo BioBrick y el sufijo BioBrick de pSB4C5. El plásmido N292 (SBa_000566) se generó insertando un casete de caracterización del terminador entre el prefijo BioBrick y el sufijo BioBrick de pSB4C5. El

casete consiste en el promotor PT7, RBS (SBa_000498), GFP, el terminador de T7 de tipo silvestre, RBS D103 (SBa_000563) de Salis et. al. (13) y mRFP (SBa_000484). El plásmido N149 (SBa_000516) se construyó insertando el casete promotor P_{tac} (SBa_000563), RBS D103 (SBa_000563) de Salis et. al. (13) y mRFP (SBa_000484) entre el prefijo BioBrick y el sufijo BioBrick de pSB4C5. El plásmido N505 (SBa_000517) se construyó insertando el casete promotor P_{tet} (SBa_000562), RBS D103 (SBa_000563) y mRFP (SBa_000484) entre el prefijo BioBrick y el sufijo BioBrick de pSB4C5. El plásmido N110 (SBa_000564) se construyó insertando un promotor constitutivo (SBa_000565), un RBS fuerte (SBa_000475) y mRFP (SBa_000484) entre el prefijo BioBrick y el sufijo BioBrick de pSB4C5. El plásmido N573 (SBa_000559) se construyó insertando el marcador de resistencia AmpR en pNOR1020 (14).

Se ha demostrado que la expresión multicopia de algunos genes de fijación de nitrógeno puede eliminar la maduración y función de la nitrogenasa (es decir, la inhibición multicopia) (63, 64). Una incertidumbre adicional es que el reemplazo del promotor nativo con un promotor inducible podría interrumpir su función. Para examinar estos efectos, construimos plásmidos para complementar las actividades de las cepas inactivadas (Fig. 31) y probamos su actividad bajo control inducible. Estos plásmidos también son la base de los experimentos para cuantificar la robustez de los cambios en la expresión (Fig. 27).

Los plásmidos de complementación se construyeron insertando el ADN que codifica cada operón de tipo silvestre entre el promotor P_{tac} y el sufijo BioBrick del plásmido N58 (pSC101, CmR). Una excepción fue el plásmido Nif18, que se construyó clonando el operón nifHDKTY en el sitio de clonación múltiple de pEXT21 (60). Las secuencias de operones de tipo silvestre se definieron mediante sitios de inicio de la transcripción publicados (65).

Los vectores de caracterización de RBS de tipo silvestre se construyeron insertando la región de -60 pb a + 90 pb para cada gen nativo y mRFP (SBa_000484) entre el casete P_{tac} (SBa_000561) y el sufijo BioBrick del plásmido N58 (pSC101, CmR). La secuencia del gen nativo de + 1pb a +90 pb formó una fusión en marco con mRFP. En los casos en que la transcripción del gen no se extiende a -60 pb, se clonó un casete más corto en N58. La fuerza de RBS se caracterizó usando el Ensayo de Caracterización del Promotor descrito en el presente documento.

Los RBS sintéticos de longitud suficiente para capturar la huella completa del ribosoma (~35 pb) se generaron con la calculadora RBS (61). La fuerza de cada uno se midió usando un vector de caracterización de RBS sintético. Estos vectores se construyeron de manera similar a los vectores de caracterización de RBS de tipo silvestre usando -60 pb a + 90 pb del gen sintético diseñado. Esta región incluye parte de una secuencia amortiguadora, el RBS sintético y la región de +1 pb a + 90 pb del gen sintético. Si los RBS sintéticos y de tipo silvestre diferían en más de 3 veces en la expresión, se generaban y analizaban nuevas secuencias de RBS. Las partes aislantes que consisten en ~50 pb de ADN aleatorio preceden a cada RBS sintético (66).

Los operones sintéticos se clonaron en el esqueleto de pSB4C5 (pSC101, CmR) entre el prefijo BioBrick y el sufijo BioBrick.

40 Generación de partes sintéticas

ARN polimerasa de T7*: La ARN polimerasa de T7 se modificó para que no fuese tóxica tanto para *Klebsiella* como para *E. coli* a altos niveles de expresión. La ARNP se expresó a partir de un origen de bajo número de copias (pSa) bajo el control de un RBS débil (SBa_000507, TATCCAAACCAGTAGCTCAATTGGAGTCGTCTAT) y etiqueta de degradación N-terminal (SBa_000509, TTGTTTATCAAGCCTGCGGATCTCCGCGAAATTGTGACTTTCCGCTATTTAGCGA TCTTGTTTCAGTGTGGCTTTTCCTTCACCGGCAGCAGATTACGTTGAACAGCGCATC GATCTGGGTGGC). El codón de inicio se cambió de ATG a GTG, y el sitio activo contenía una mutación (R632S).

Promotores de T7: los promotores de T7 se generaron a partir de una biblioteca aleatoria. La secuencia de semilla del promotor de T7 fue TAATACGACTCACTANNNNNAGA. Para las secuencias de promotores individuales, véase la Figura 38.

Terminadores de T7: los terminadores de T7 se generaron a partir de una biblioteca aleatoria y se insertaron en el vector de caracterización del terminador N292 (SBa_000566). La secuencia semilla del terminador T7 fue TANNNAACSSWWSSNSSSSSTCWWCGSSSSSSWWSSGTTT. Los plásmidos terminadores se transformaron conjuntamente con el plásmido N249 y se caracterizaron (Métodos, Caracterización de fluorescencia) bajo inducción de IPTG 1 mM de la ARNP de T7*. La expresión de RFP se midió para cada terminador, y los datos se expresan como la reducción de veces en la fluorescencia medida cuando se compara con un derivado de N292 que no lleva terminador. Para las secuencias de terminadores individuales, véase la Fig. 38

Sitios de unión al ribosoma: la calculadora RBS se utilizó para generar un RBS que coincidiese con la fuerza medida del RBS de tipo silvestre. En tres casos, se seleccionaron RBS sintéticos de las partes existentes (SBa_000475 para nifJ y nifQ, y SBa_000469 para nifH). En los casos en que la fuerza del RBS sintético inicial difería del RBS WT en más de 3 veces (nifV, nifZ y nifM), se construyó una biblioteca de RBS sintético reemplazando los 15 pb en dirección 5' del codón de inicio con NNNAGGAGGNNNNNN. Seleccionamos mutantes en cada biblioteca para identificar RBS

sintéticos con tres veces la fuerza de RBS WT. La fuerza del sitio de unión al ribosoma se expresa en unidades de fluorescencia arbitrarias medidas usando el ensayo de caracterización de fluorescencia.

Secuencias de aislador (secuencias espaciadoras): las secuencias de aislador se generaron usando el generador de ADN aleatorio usando un contenido de GC aleatorio del 50 % (66).

Lógica AND: construimos un circuito genético que codifica la lógica A AND B y usamos este circuito para controlar la ARNP de T7* en el Controlador N.º 3. En este circuito, la lógica A AND B corresponde a la presencia o ausencia de los inductores, IPTG y aTc, de modo que la célula calcula IPTG AND aTc. El circuito se construyó modificando el promotor Ptac en el Controlador N.º 1 (SBa_000520) para incluir los sitios de unión al represor cl OR1 y OR2 para producir el plásmido N639 (SBa_000560). Además, el plásmido pNOR1020 codifica el represor cl bajo control del promotor Ptet (62). Modificamos pNOR1020 cambiando el marcador de resistencia para conferir resistencia a ampicilina para producir N573 (SBa_000559). Cuando N639 y N573 se transforman conjuntamente, producen el circuito lógico IPTG AND aTc.

Secuencia de Ptac (SBa_000512):

tattctgaaatgagctgttgacaattaatcatcggtctgtataatgtgtggaattgtgagcggataacaatt

Secuencia de Ptac más OR1 y OR2 (SBa_000506):

tattaacaccgtgctgttgacagctatacctctggtgtataatgctagcgggaattgtgagcggataacaatt

La Figura 39 ilustra mapas para plásmidos clave.

El complejo génico nif en *K. oxytoca* Ma5L se volvió a secuenciar a partir de fragmentos de PCR. La secuencia de ADN re-secuenciada se comparó con la secuencia de referencia de Genbank, X13303.1 (52). Las diferencias de secuencia se enumeran en la Fig. 35. Las ubicaciones de los nucleótidos están numeradas en relación con X13303.1. Se muestran mutaciones de aminoácidos para corregir errores en el registro X13303.1 (Impacto).

Referencias

1. Fischbach M, Voigt, C.A. (2010) Prokaryotic gene clusters: A rich toolbox for synthetic biology. *Biotechnol. J.* 5:1277-1296.
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 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 is beyond the RNA horizon? *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.
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 *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, Breitling, R., Takano, E. (2011) Synthetic biology in *Streptomyces* 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.* 53:1299-1304.
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.
17. Fowler M & Beck K (1999) Refactoring : improving the design of existing code (Addison-Wesley, Reading, MA) pp xxi, 431 p.

18. Chan LY, Kosuri S, & Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* 1:2005 0018.
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.
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.
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.
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.
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.
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.
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. Tamsir A, Tabor JJ, & Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469(7329):212-215.
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.
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.
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 (MEGAWHOP). *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.

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,

<http://www.faculty.ucr.edu/~mmaduro/random.htm>.

REIVINDICACIONES

1. Un método para reemplazar la regulación nativa de un conjunto de genes asociados colectivamente con una función con regulación sintética, comprendiendo el método
- 5 proporcionar secuencias codificantes de un conjunto de polipéptidos codificados por genes asociados colectivamente con una función;
cambiar la identidad de los codones dentro de al menos una secuencia codificante, eliminando así al menos una secuencia reguladora dentro de la secuencia codificante, en el que eliminar la al menos una secuencia reguladora comprende el reemplazo de codones nativos en la secuencia codificante con codones sinónimos no nativos que
- 10 tienen una distancia máxima de los codones de la secuencia codificante nativa;
organizar las secuencias codificantes en uno o más operones sintéticos, en el que la organización comprende ordenar secuencias codificantes dentro de los operones de modo que el gen de mayor expresión (basado en la expresión nativa) ocurra primero y el gen de menor expresión (basado en la expresión nativa) ocurra el último;
unir operativamente una o más secuencias reguladoras transcripcionales heterólogas al operón u operones
- 15 controlando así la magnitud de la expresión génica del operón u operones; y
expresar el uno o más operón u operones sintéticos en una célula bajo el control de un polipéptido que se une directa o indirectamente a la secuencia reguladora transcripcional heteróloga.
2. El método de la reivindicación 1, en el que el conjunto de genes es de un procariota.
- 20 3. El método de la reivindicación 1, que comprende además identificar y eliminar uno o más sitios de inserción de transposones, sitios que promueven la recombinación, sitios para escisión por endonucleasas de restricción y sitios que están metilados.
- 25 4. El método de la reivindicación 1, en el que la organización comprende agrupar secuencias codificantes en operones basados en un nivel de expresión nativa sustancialmente similar.
5. El método de la reivindicación 1, en el que la magnitud de la expresión de las secuencias codificantes corresponde sustancialmente a la proporción de proteínas codificadas por las secuencias codificantes tal como se mide en el sistema nativo.
- 30 6. El método de la reivindicación 1, en el que la secuencia o secuencias reguladoras de la transcripción heterólogas comprenden un promotor o promotores de la ARNP de T7.
- 35 7. El método de la reivindicación 1, que comprende además unir operativamente un sitio de unión al ribosoma (RBS) heterólogo a una o más secuencias codificantes en el operón sintético.
8. El método de la reivindicación 7, en el que diferentes RBS están unidos operativamente a diferentes secuencias codificantes.
- 40 9. El método de la reivindicación 8, en el que los RBS regulan la traducción de las secuencias codificantes en una proporción que es sustancialmente similar a la proporción de traducción nativa del operón nativo.
10. El método de la reivindicación 1, que comprende además unir operativamente una secuencia de terminador transcripcional heterólogo a una o más secuencias codificantes en el operón sintético.
- 45 11. El método de la reivindicación 10, en el que los terminadores son terminadores de la ARNP de T7.
12. El método de la reivindicación 1, en el que el polipéptido que se une directa o indirectamente a la secuencia reguladora transcripcional heteróloga se expresa a partir de un casete de expresión de control, comprendiendo el casete de expresión un promotor de control unido operativamente a una secuencia polinucleotídica que codifica el polipéptido.
- 50 13. Un producto de programa informático que comprende un medio tangible legible por ordenador que almacena una pluralidad de instrucciones para controlar un procesador para realizar una operación para reemplazar la regulación nativa de un conjunto de genes asociados colectivamente con una función con regulación sintética, comprendiendo las instrucciones las etapas del método de cualquiera de las reivindicaciones 1-12.
- 55

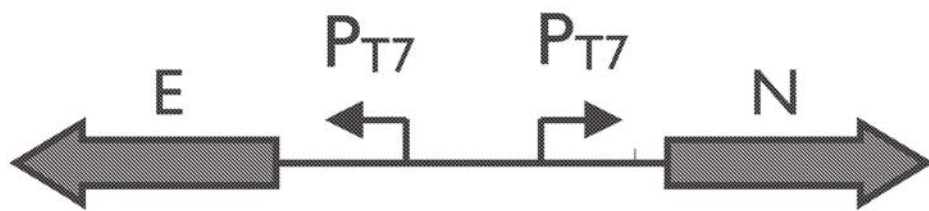


FIG.1

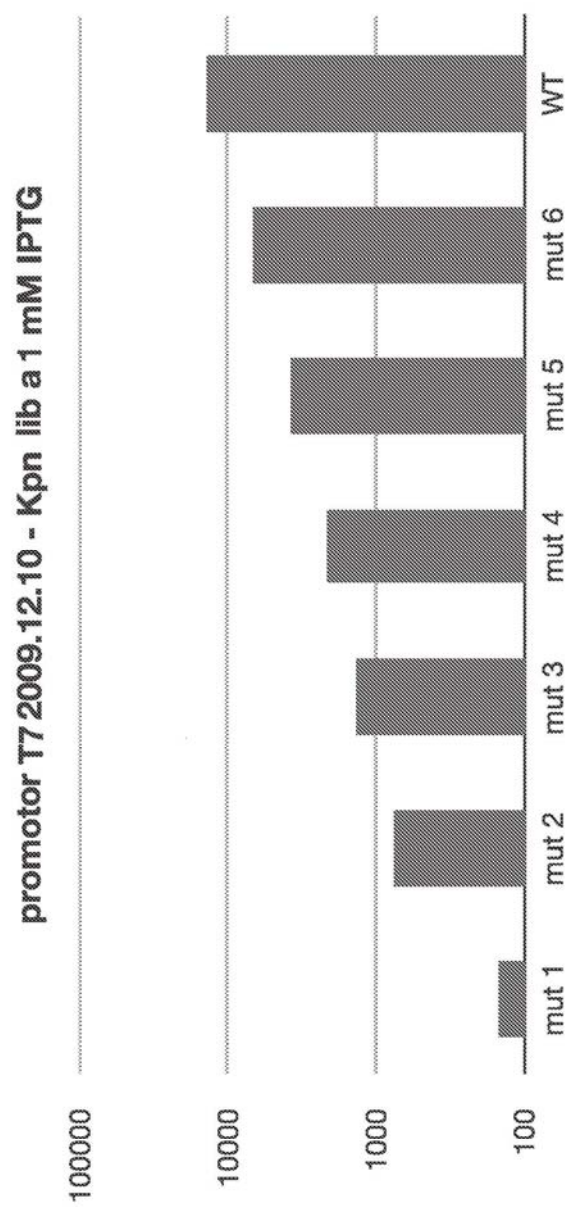
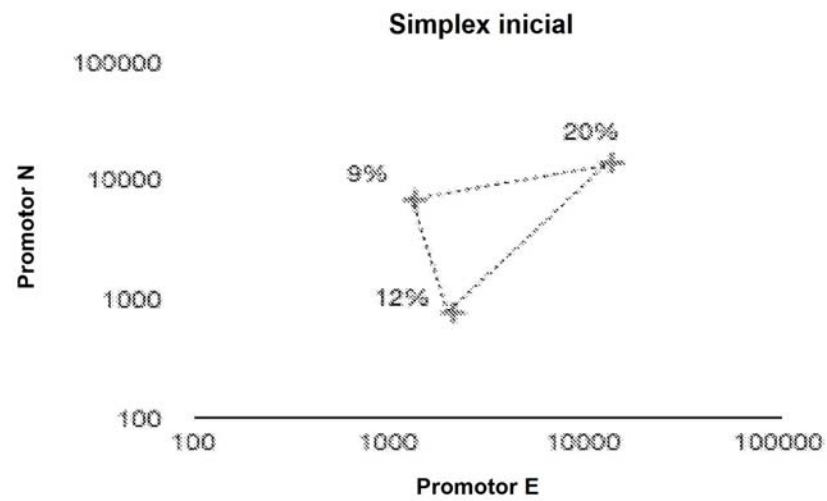


FIG. 2

Nombre del promotor	Secuencia TAATACGACTCACTANNNNNAGA	Fuerza
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

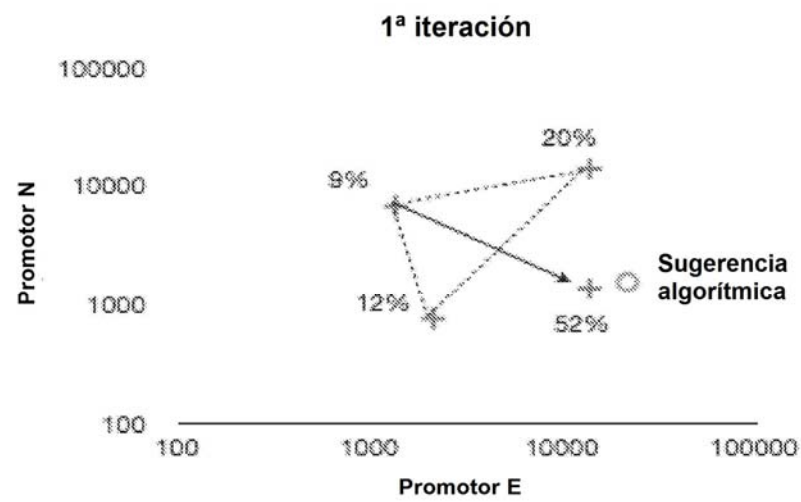


FIG. 4

Nombre de la cepa	Promotor E	Promotor N	Fijación
N.º1 (N259)	WT	WT	20%
N.º2 (N438)	Mut 4	Mut 5	9%
N.º (N307)	Mut 4	Mut 2	12%
Reflexión (N319)	WT	Mut 3	52%

FIG. 5

Refactorización de la fijación de nitrógeno

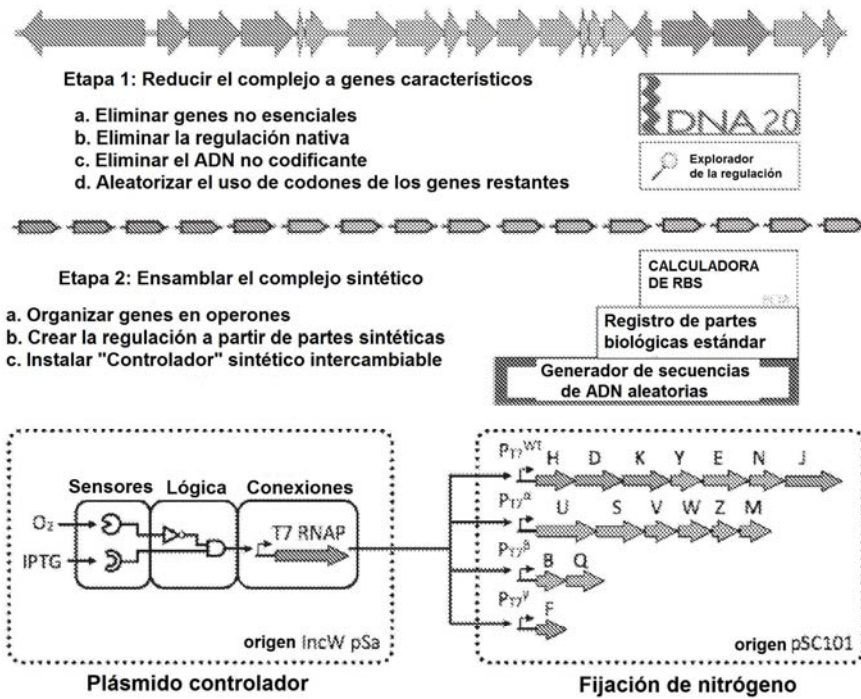
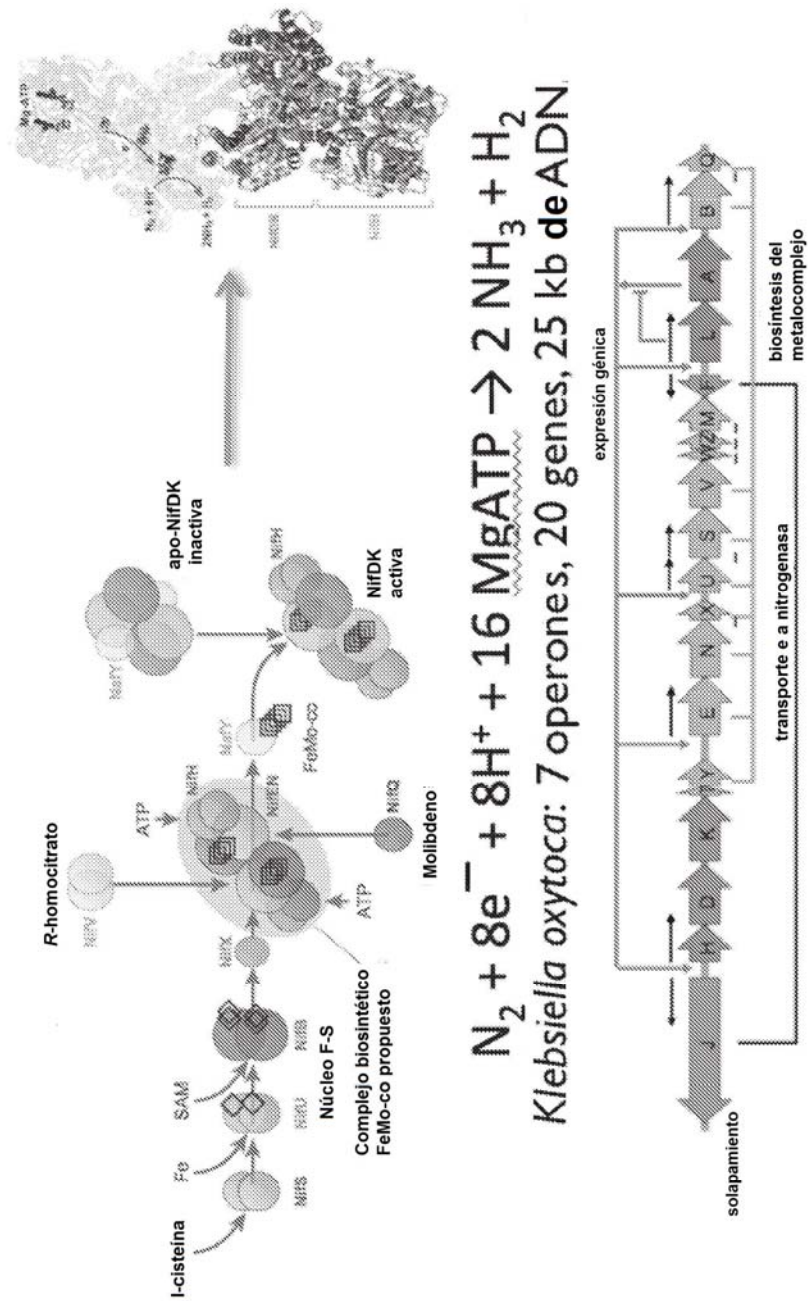


FIG. 6



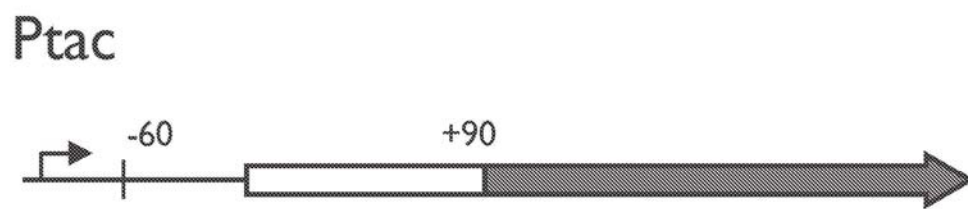


FIG. 8

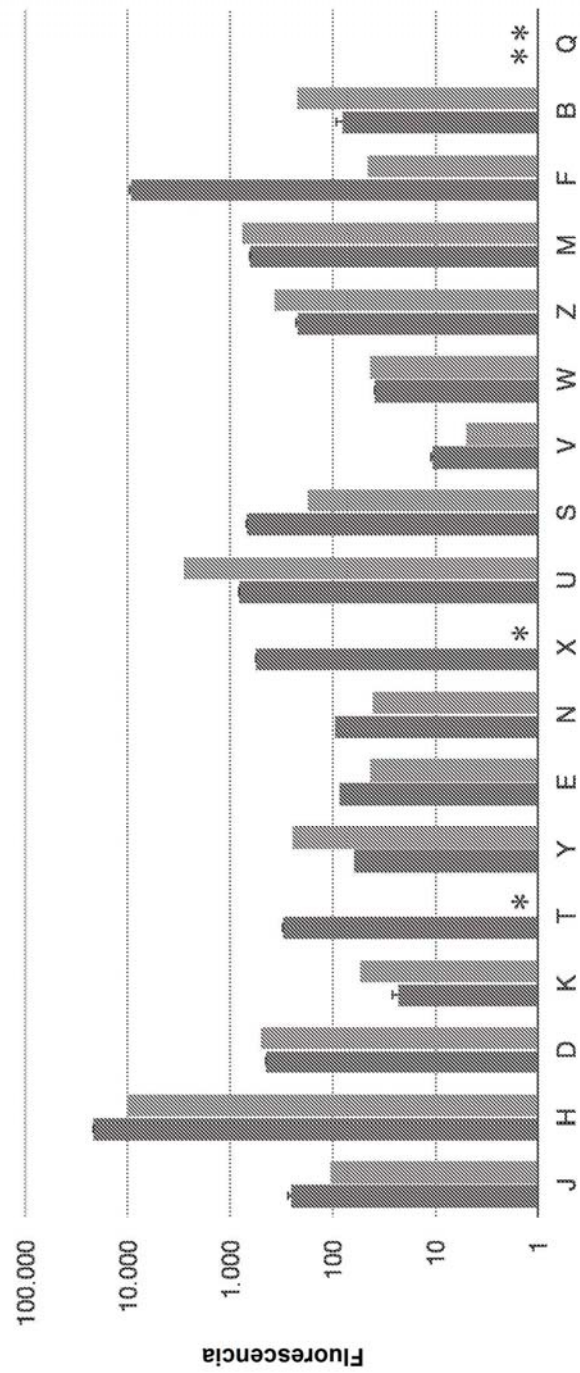


FIG. 9

Gen	Fuerza del RBS nativo	Fuerza del RBS sintético
J	254	104
H	21641	9903
D	446	496
K	23	54
T	303	Eliminado
Y	62	243
E	85	43
N	95	41
X	558	Eliminado
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	No Detectable	No Detectable

FIG. 10

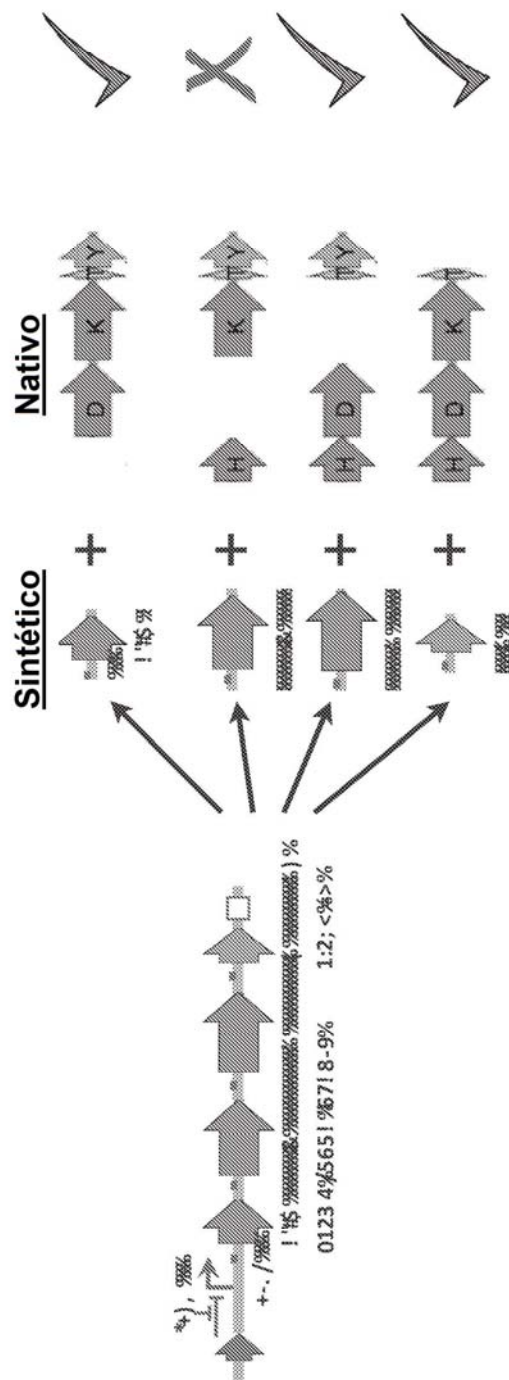


FIG. 11

Operón	Localización del error	Error
HDKY	H	Inserción de transposón
USVWZM	S	Errores de secuenciación en entrada NCBI Arnold (L110M, Q113E, S124G, R290A)
USVWZM	V	Errores de secuenciación en entrada NCBI Arnold (L257S, D295H)
USVWZM	M	Mutación de clonación
USVWZM	Ptac	Error de clonación que condujo a promotor constitutivo

FIG. 12

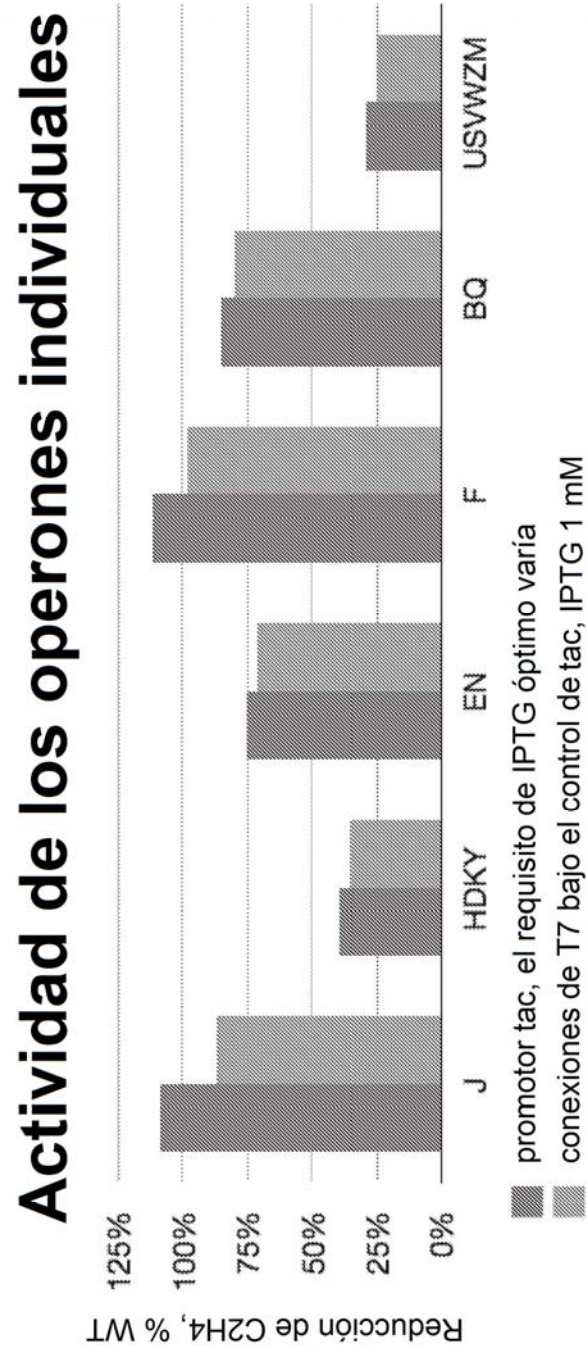


FIG. 13

Operón	Función con promotor Ptac	Promotor T7	Función con promotor T7
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

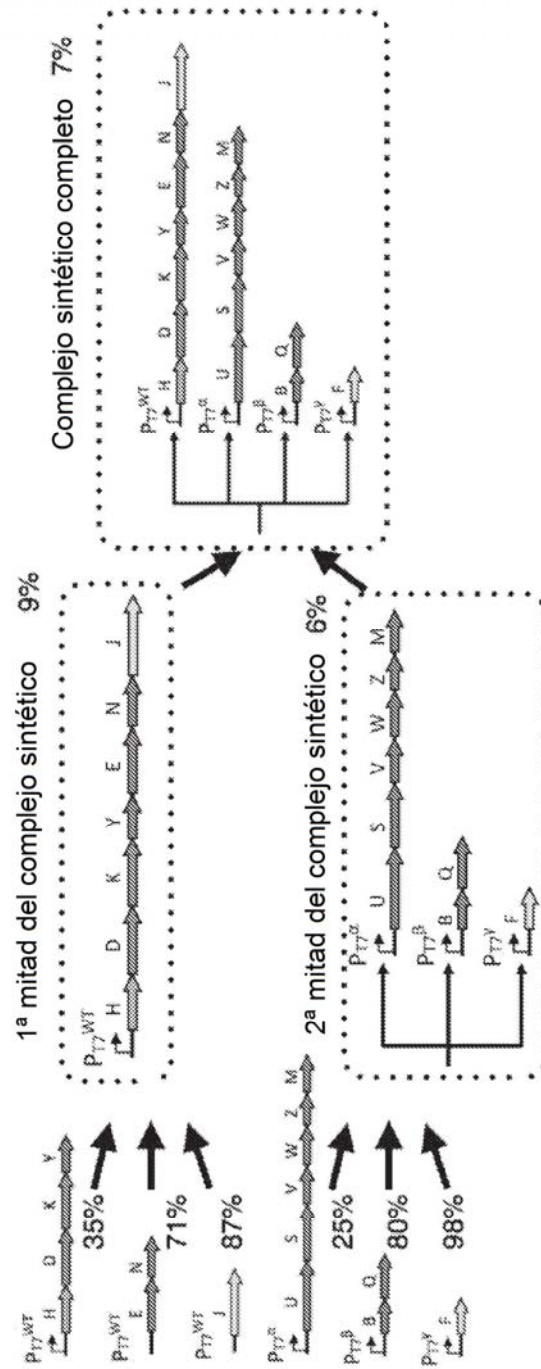


FIG. 15

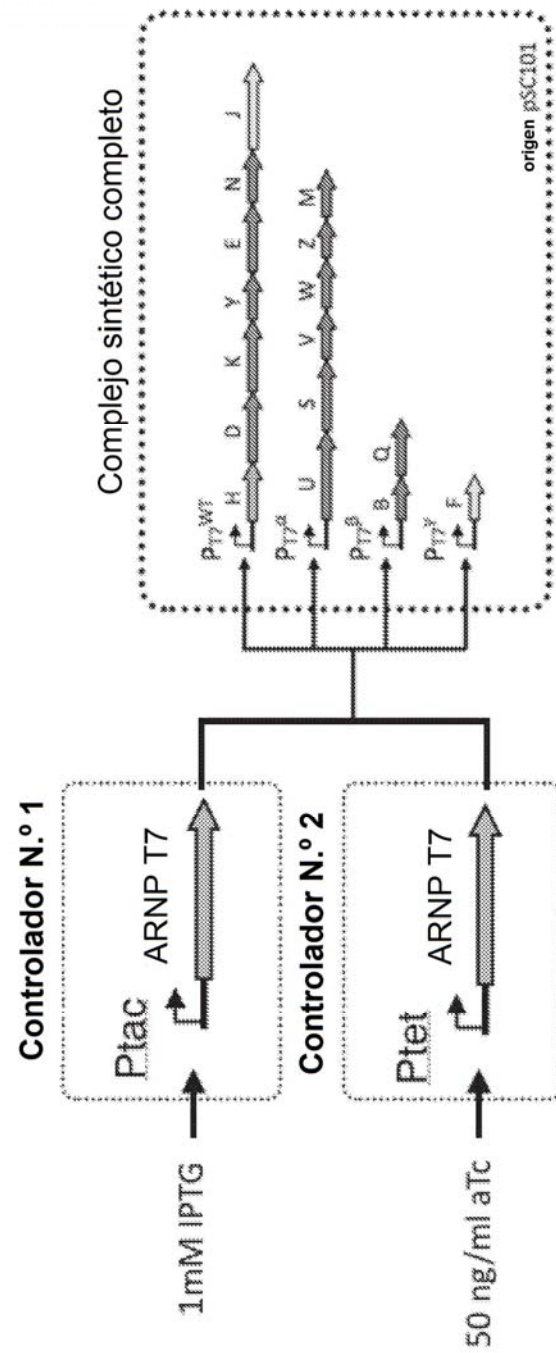


FIG. 16

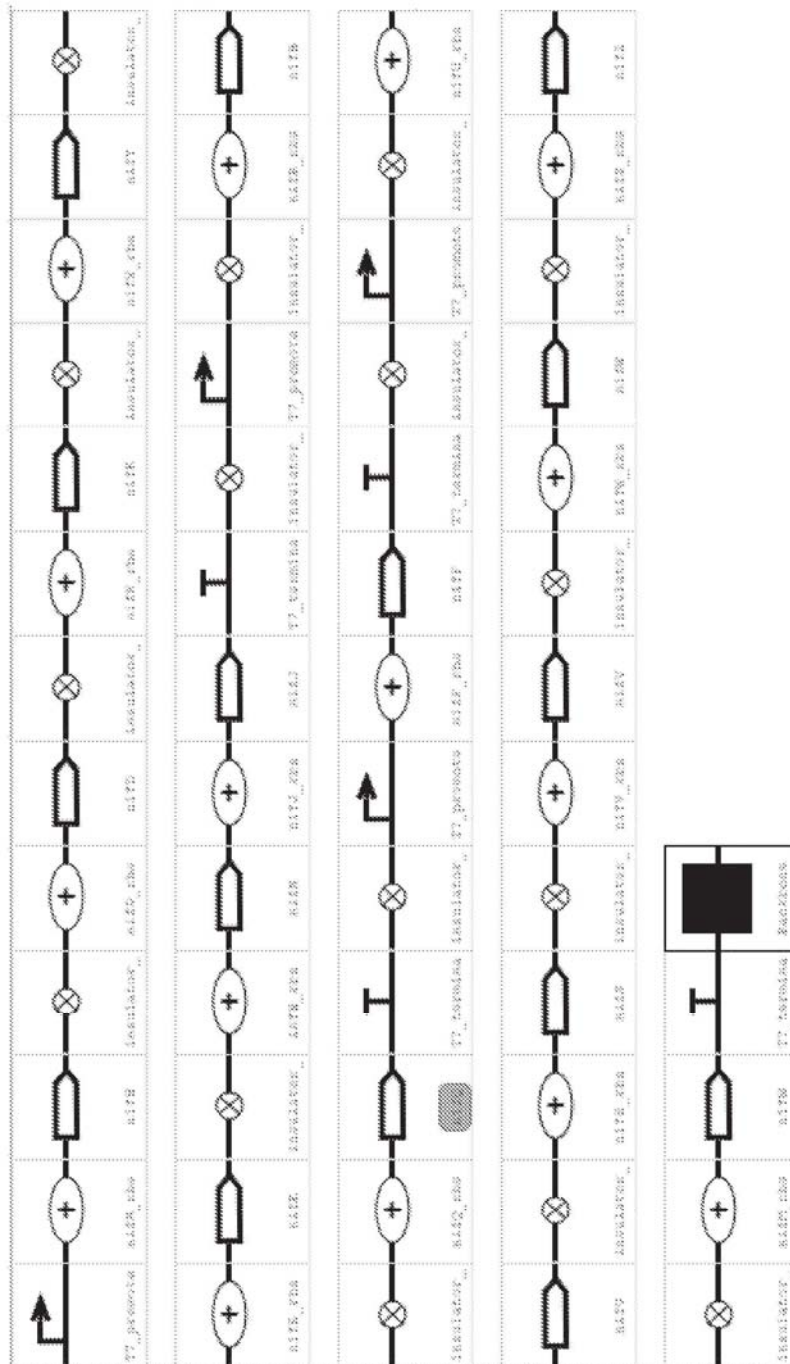


FIG. 17

Fig. 18

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
J	El gen WT es SEQ ID NO:1 El gen J sintético es SEQ ID NO:2	atgaaacaatggatggcaacgcgcgcgcgcgc tggatctttagccittaccgaggtcgcgcgcgatit acccataaacccctccacgcgcgatggcggaaa acgtcgacgagtgggcggcgacggggaanaag aacctttttggccagccgggtgcgttaatggagat gcagtcggagggcggcgcgagggcgcgcgtcc acggcgcgcgtcgagggcgggcggtcaccacc acctatacggcctccagggcggtgctgatgat cccccaatgatcaaaaatcgccgtgaaactgctg ccggggcgtcttcaagtcagcccgccggcgctgg cgaccaatcgctgaatatittggcgatcaccagg atgtgatggcggtccgcgcagacggcgcgcgat gctggcggaacaacgtgcagcaggtgatgga ttgtcgcggtggcgcatctggcgcggttaagg gacgcateccgtttgtaacttttcacgggttttc cactctgcacgaattcaaaaatcgaggtgttgg aatatgagcagctggcgacgctggtgaccggc ccggcgtcgacagcttccgcgtaacgcgcgtgca tccggatcaccggtcgtccgcggacggcgcgag aaaccggatatacttccagggcggggagggcg ggcaaccgtttttatcagggcggtcccgacattgtc gaaagctatagacgaaatcagcgcgctcaccg gcccgggagtaicattgtttaactataccggcgcg gcggatggcggaacgggtgattatcgcaatgggat cggctcgcgataccgttcaggaagtgggtgatac gctgaacgcggcgggagagagggcggtcgct cagcgtgcatcgttttgccttttgcgtggccac ttcttgcgcagctggcgaaaaccgtgcagcgat tgcgtcttgaaccgtacaaagagccggcgct caggctgaaccgctgtgcttggatgtgaaaacg cctttaccaccagcagatgcgcgcgtatcgtc ggcgccgctatgctgttggcggaaggtgttgc tgcacacgatacggcggtctttgataacctea ataaacggctgcggatggaggggtttaccctggg gattgtcgatgatgtaacctttactcgttaaccgg gcccagcagacgtggcggtctcgatgacggc atcaccgctgttaattttggggcatgggctcggg cggcaccgtcggggccaataaaagcgcgatcaa aatatcggcgataaaacgcgcgtacgcgcag gcctacttttctatgactcgaaaaatcggcgggc attacggtttacatttgcgttcggcgacggcg atcaattcgccctacctgacatcgggcggtttt atctctgttgcagcgtctacgttgagcgctac gatctgtggcaggtatgaagccggcggggacct ttttactaacgtcagctgggtcgatgcggagctg gagcagcatctgcccgttgcgtttaaacgctatct ggcgcggggaataatcattttataccctgaacg ccgtggatcgcggcgagctcgggtggggcg ggcgcttttaatgctgtagcagggcggttttta agctggcgggcgattatgacccgcagacgcgg cggatfaccctaagcagggcggttgaanaagcta cggcagcaaggggcgggcggtgattgagatgaa	ATGAAAACATATGGACGGTAA CGCTGCGGCTGCATGGATTA GCTACGCCTTTACCGAAGTGG CTGCGATCTACCCGATTACGC CGAGCACCCCGATGGCGGAA AATGTGGACGAATGGGCTGC GCAGGGCAAGAAGAACCTCT TCGGCCAGCCGGTGCCTCTG ATGGAGATGCAGTCGGAAGC GGGTGCAGCAGGTGCTGTGC ATGGCGCCTTGCAAGCTGGC GCACGTGACGACCACCTACAC CGCGTCGACGGGCTGTTGCT GATGATCCCAACATGTACA AAATCGCGGGTGAACCTGCTG CCGGGTGTCTTTTATGTTTCG GCACGCGCACTGGCCACCAA TAGCCTCAACATCTTTGGCGA TCATCAGGATGTAATGGCGG TGCGCCAAACGGGCTGCGCG ATGTTGGCCGAGAATAACGT CCAGCAAGTTATGGATTGTG CGCGGTAGCCCACTTGGCAG CGATCAAAGGTGCGATTCCGT TCGTGAACCTTCTCGATGGCT TTCGCACCAGCCACGAAATC CAGAAGATCGAGGTCTTGA ATATGAACAGCTGGCCACCTT GTTGGATCGTCCGGCCCTGGA CAGCTTCCGCGCTAACGCCCT TCACCCGGACACCCGGTCA CCGTGGCACCGCCAGAAC CGGACATCTACTTCCAGGAA CGTGAGGCCGGTAACCGTTTC TATCAGGCGCTCCCGGATATT GTGGAATCTTACATGACCCA GATTCTGCGCTGACTGGTTCG CGAGTATCACCTGTTTAACTA CACTGGTGTGCGGATGCGG AGCGCGTGATCATCGCGATG GGCTCTGTCTGTGACACCGTC CAAGAGGTGGTTGACACGCT GAATGCAGCGGGTGAGAAAG TTGGTCTGCTCTCCGTTTATC TTTCCGCCCCGTTTCGTTAG CGCACTTCTTCGCCCCACTGC CGAAAACGTGACAGCGTATC GCAGTATTGGACCGTACGAA AGAGCCAGGTGCTCAAGCAG AGCCGCTGTGCTCGATGTGA AGAATGCCTTTTACCACCATG ACGATGCCCCGTTGATTGTGG	80,8	48

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
		ccagcggcgatcagctgggcatggcctcgt gcatcaggtgacgattccggcgactggcgac gctggatgaacccggcgcaagcagccat gatgccgatttcaccccgatattctgcagccgat gaacccgagtgccggaccagctgcccgtgag cgcttctgctggtatggagagggacatttctt cgggcaccggcgctgggagaacggcgatc gctctggaagtgcggctgctgagccggagggc tgcacgagtgtaacagtgccctttatctccc gcatcgccgcatcccgccgctgctcaacgg cgaagagcatgacggcgccgcttgggctgct gagcaaacccggcgagggagcgaaggagatc actaccatctggctatctcccgtgattgttccg gctcgggcaactgctggatattctcccggcg cggaaggcgtaaaatgacgtctctgatagcc agcgtcagatggcgccgctgctgactatgct ggcgctgacggcggaagagcaatccgttctgaag acgacggcgaaggcgagccagttgaaacccg ctgctggagtttctggcgccctgcccggatgct gtgaacgccttatgcccgcctgataaccagctg ttggcgaccggatgctgctgaacggcaccg gctgctcttctatctggggagccagcgccgctg attccctataccaccaaccacggcgccatggc cgcatggcggaactgctgttggagataatgc ggagtttggcctcgcatgctgctggcgccag cgctgctgcaacaatcgccgatgatgacgg ccgctgctgctacccggctgacggagcaactag cgacgcaatgcccagtgctggcggaagcagga tgaaggcgaggcgaccggcgagcgccggacc ggctcagcggaacggctggcgccggaaaaagag ggcgctgcccgtgttggagcagctgtggcagaacc ggcactatttctgctgctgctgagtgatcttgg cgccgacggctggcctacgatalcggttccggc ggctcgtacacgtgctggcgagcggggaagac gtcaatattctgtgttgacacggaggtttactcca ataccggcgccagctgcttaaatcgaccgggt ggcgccatcgcggaagtgtccggcgaggcgaa acgcacggcgaaaaaagatcggcatgatggc gatgagctacggcaatgtgctggtggccaggtc gcatggcgccgtgataaagatcagaccctacgg gcatcgccgagggcggaagccggcgggacc gtcgtggtgattgctacggcgctgcatgaacc acggcgctgaagggcggtatgctgagccagc gcgagggcgaacggcggttggagcgggatact ggcaccgtgtggcgtatcatcggcgagggaag cggaaggtaagacggcgtttatgctgattccgaa gagccggaggagagcttccggacttttcttgg cgaagtgcgtacggctgctgcacaaacgac gcccgtctggcggtgctctttagccgaacc gaggaggagcggcgggcccgcttggccagat cgccggctggcgccggaggagtag	GTGGTCGCTATGCCTTGGGGCG GTAAGGACGTGTTGCCGAAC GATATTGCGCCCGTGTGAT AACCTGAACAAACCGCTGCC GATGGACGGCTTCACGCTGG GTATCGTGGACGATGTTACCT TCACCTCTCTCCCGCCAGCGC AGCAGACCCTGGCGGTTTCTC ACGACGGCATACGGCATGT AAGTTTGGGGCATGGGCTCC GACGGCACGGTTGGTGCGAA CAAGTCCGCGATCAAGATTA TCGGCGACAAAACGCCACTG TATGCGCAAGCGTACTTTTCC TACGACTCGAAGAAGAGCGG TGGTATTACCGTCAGCCATCT GCGTTTGGTGATCGCCCGAT CAACTCCCCGTATTTGATCCA TCGCGCGGATTTCTCTCGTG CAGCCAGCAAAAGCTATGTTG AACGCTACGATCTGCTGGAT GGCCTTAAACCGGGTGGCAC CTTTCTGCTGAACGTCTCTG GAGCGATGCCGAACCTGGAGC AACATCTGCCGGTTCGGTTTCA AACGTTATCTGGCACGCGAG AATATCCACTTCTACACTCTC AACGCTGTGGACATCGCCCG TGAGCTTGGTTTGGGTGGCCG TTTCAACATGCTGATGCAGGC TGCCTTCTTAAACTGGCCGC GATATTGACCCGCACTG CTGCGGACTATCTGAAGCAG GCTGTTGAGAAAAGCTATGG CAGCAAAAGGTGCGGCGGTCA TCGAGATGAACCAGCGTGCC ATCGAGCTTGGCATGGCCAG CCTGCACAGGTGACGATCC CGGCACATTGGGCCACCTG GATGAGCCAGCGCGCAGGC GTCCGCGATGATGCCGACTT TATCCGCGACATCTGCAACC GATGAACCGTCAGTGCGGCG ACCAAGCTTCCGGTGTGCGCTT TTGTGGCATGGAAGATGGC ACCTTCCCGTCCGGCACGGCC GCATGGGAGAAAAGTGGCAT CGCCCTTGAGGTGCCAGTCTG GCAGCCGGAAGGCTGCACGC AGTGCAACCAGTGCGCCTTC ATTTGTCCGCACGCCGCGATT CGTCCGGCGTTGTTGAATGGC GAAGAGCATGATGCTGCCCC GGTTGGCTGCTGAGCAAAC		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			CGGCACAAGGCGCTAAAGAA TATCACTATCATCTGGCGATT AGCCCCTGGACTGCTCCGG CTGTGGCAACTGCGTTGACAT TTGTCCAGCTCGTGGCAAAGC GTTGAAGATGCAGTCTCTGG ATAGCCAACGCCAGATGGCT CCGCTGTGGGATTATGCGCTG GCGCTGACCCGAAAGTCTAA CCCGTTTCGTAAAACCACCGT CAAAGGCTCGCAGTTCGAAA CCCCGCTGCTGGAGTTAGCG GTGCGTGCCTGGTTGTGGCG AAACGCCGTATGCGCGCTC ATTACCCAGCTGTTTGGCGAC CGCATGCTGATTGCCAATGCC ACCGGCTGTTCCAGCATCTGG GGCGCATCTGCGCCGAGCAT CCCGTATACCAACAATCATCG TGGTCATGGTCCGGCCTGGGC GAATAGCCTGTTTGAGGACA ATGCCGAATTTGGTTTAGGTA TGATGCTGGGCGGTCAAGCT GTGCGTCAACAGATCGCGGA CGATATGACGGCTGCGTTAG CGCTCCCGGTTTCCGATGAGC TGAGCGACGCGATGCGCCAG TGGTTGGCGAAACAGGACGA GGGTGAAGGCACGCGTGAGC GTGCGGACCGTCTGAGCGAG CGCTTAGCCGCGGAGAAAGA GGGCGTTCCGCTGTTAGAGC AGCTGTGGCAAAATCGTGAT TACTTTGTGCGTCGCAGCCAG TGGATTTTCGGCGGTGACGGC TGGGCCTATGATATTGGCTTC GGTGGCCTGGACCACGTCTC GCCAGCGGTGAGGATGTGAA CATTTCTGGTATTTGACACCGA AGTCTACTCGAACACCGCG GTCAAAGCAGCAAAATCGACC CCGGTCGCGGCCATCGCAA GTTGCGGGCTCAGGGCAAGC GCACCCGCAAGAAAGACCTG GGTATGATGGCGATGAGCTA CGGCAACGTCTATGTAGCCC AGGTGGCGATGGGTGCGGAT AAAGATCAAATCTGCGCGC CATTTGCGGAAGCTGAAGCGT GGCCAGGCCCGTCGCTGGTG ATTGCGTATGCGGCCTGCATC AATCATGGCTGAAGGCCGG TATGCGTTGCAGCCAACGTG AGGCGAAGCGCGCTGTTGAG		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			GCGGGCTACTGGCACCTGTG GCGTTATCACCCGCAGCGCG AAGCGGAAGGCAAGACGCCG TTTATGTTAGATAGCGAAGA ACCGGAAGAGTCGTTCCGTG ACTTCTGTGGGTGAGGTGC GCTACGCATCCCTGCACAAG ACCACCCCGCACCTCGCCGAT GCCCTTTTCAGCCGTACCGAA GAAGATGCGCGTGCGCGCTT TGCGCAATACCGTCGCCTGGC TGGCGAAGAGTAA		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
H	El gen H WT es SEQ ID NO:3 El gen H sintético es SEQ ID NO:4	atgaccatgcgtcaatgcgtatttacggtaagg cggatcggtaaatccaccaccagcagaacctc gtcggcgctggcgagatgggtaagaaagt atgatcgtcggctgcgaicgaaggcggactcca cccgctgatctgcacgccaaagcacagaaac caattatggagatggcgaggagtcgctcggtc gaggacctcgaactcgaagacgtgctgcaattg gctacggcgatgtcgtcgcgggaatcggcg gcccgagccaggcgctggctgcgggacgc ggcgtgatcacggcgatcaactttctgaagaaga aggcgctacgaggacgatcgaattcgtgttct tgacgtcgtcggcgacgtgctcgcggcggttc gcatgccgatccgcgaacaaagccaggag atctacatcgtctgctcggcgaaatgatgggat gtacgcggccaacatctcgaaggatcggtta aatacgccaatccggcaaggtgcgctcggcg gctgatctgtaactcacgtcagaccgacgtgan gacgaactgattatggcctggcgaaaagctcg gtaccagatgatccactttgtgcccgcgacaac atcgtgcagcgcgggagatccgcgcgatgacg gttatcgagtagaccccgctgtaaacaggcca acgaataccgcaccctggcgagaagatcgtaaa caacaccatgaagtggtgcggacgcccgcacc atggatgagctggaatcgtcgtgatggagtcgg catcatggaagggaagacaccgcatcatggc aaaaccgcccgaagaaaacggcgctga	ATGACCATGCGTCAGTGCGC GATTATGGCAAAGGTGGTA TTGGCAAAAGCACGACGACC CAGAACTTGGTGGCGGCCCT GGCCGAGATGGGTAAAAAGG TTATGATTGTGGGTTCGACACC CGAAGGCCGACAGCACGCGC CTGATTCTGCACGCGAAAGC ACAAAACACGATTATGGAGA TGGCTGCCGAGGTTGGTAGC GTGGAGGATCTGGAGCTGGA GGACGTTCTGCAAATTGGTTA CGGTGATGTTCTGTCGACAG GAGCGGTGGTCCGGAACAG GTGTCGGCTGTGCGGGTCGTG GTGTGATTACCGCTATCAATT TCCTGGAAGAAGAGGGTGCG TACGAAGATGATCTGGATTTC GTTTTCTACGATGTGCTGGGT GATGTCGTGTGCGGTGGTTTT GCAATGCCGATTTCGCGAGAA TAAGGCACAAGAAATTTACA TTGTCTGTAGCGGCGAGATG ATGGCAATGTACGCTGCTAA CAACATCAGCAAGGGTATTG TTAAATACGCAAAAAGCGGT AAGGTTTCGCTTGGGTGGTTTTG ATTTGCAACAGCCGTCAGAC CGACCGTGAGGACGAACTGA TCATCGCCCTGGCTGAGAAA CTGGGCACCCAAATGATCCA CTTCGTGCCACGCGATAATAT TGTTCAACGTGCAGAAATCC GCCGTATGACCGTCATTGAGT ATGACCCGGCATGCAAGCAA GCGAACGAGTACCGCACCTT GGCACAGAAAATCGTGAACA ACACCATGAAGGTTGTTCCG ACGCCGTGTACGATGGACGA GCTGGAGAGCCTGCTGATGG AGTTCCGGCATTATGGAGGAG GAGGACACCAAGCATTATCGG TAAGACCGCAGCGGAGGAGA ATGCGGCATAA	79	41,8

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
D	El gen D WT es SEQ ID NO:5 El gen D sintético es SEQ ID NO:6	atgatgaccacgcaacggcggaacgtaacttgg cgtgacacaggaagtcctggagggttcccgga aacggcgcaaaagagcgcaaaagcacatgat ggtcagcagatccgaaatgaagagcgtcggcaa ggtcattatcttaaccgcaaatcacaaccggcg taatgacggtagcggcgtgcgctacgcccgttc caaaggggtggtatttggccgattaaggatagg cccatatttcgacggacggcgtggtgctggcca gtatcccgcgccgaacgacgcaactactacacc ggagtcagcggcgtcgtatgcttggcagcgtga acttcacctgatttgcagggcgacatcgtctt cggcgccgataaaaagctcagcaagctgattgaa gagatggagtgctgttccgctcacaaggggat caccattcagtcggaatcccgggtggcgtgac gggtgatatacagcggcggtggccaacggcagc agcaaggcgctggataaacgggtgacccggtag gctcgggaaggctttcggcggtgctcgcagctcgt gggcacatacgcgaacgagctggtgctgagct ggatcctgaacaatcgcgaagcagacgggttga aacaccaccttaccgagtggtgacatcggcgact acaacatcggcgcgacgcttggccctcgcgca ttctgtcggaaagagatgggctacgggtagtcgc gcagtggttcggcgacggcagcgtgggtggagat ggagaatccccattcgtcaagctgaacctgggtc actgctaccgttgcgtgaactatcgcggccata tggaggagaacatcagatccgtggtgagga caacttcttggcgccgacaaatcggcgaatcg ctgcgcaaatcggcagcaggttcgacgatacca ttcgcggcaacggcgaagcgggtgacgcccgtt atgaggggcagatggcgccgattatcgcgaata tcgcccgcgcctggaggggcgtaagggtgctgct tatatcgaggccctgcggccgcgcacgtatttgg cgctatgaggatcgggatggagatcgcgc ggcggtcagagtttgcataacgatgattacga ccgacacctcgggacgtgaagggggcagcgt gctgttcgtagcggcagcagctacgagctggaa gcgttcgtcaaggcgctgaagcccgacattacg gctccggcatcaaggaaaaatatcttcagaaa atggcggtgcccgttcggcagatgactcgtggg actatttcggccgtacacggctacgatgggttc gccatttccggccgataatgatatgacccigaa caaccggcggtggaacgaactgacgcctcgtg gctgaagctcgtgga	atgatgactaatgctactggcggaacgtaacctgg cactgaticcaagaagtagtggagggttcccgga aacggcgcaaaagagcggcgttaaacacatg atgggttctgacccgaaaatgaatctgtgggta aatgcatcatcttaacgcaaatctcagcgggg tgtcatgactgtcgtggtgctgtagcagcagggt ctaaagggtgctgtatcggccgacaaagatag ggcgcatatctcgtgcccggcagggctgtgg ccagtactctcggcggaacgtcgtactacta cacggcggttctggcgttgactcttccggcagc ctgaactcaccctcgtactcagggaacgtgaca tcgtttcgggtggcgataaaaagctgtccaaact gatcgaagaatggaactgtgttcccgctgact aaaggcattactatcgaagcgaatgtccgggtg ggctgcatcgggtgagatcagcggcggtcga aacgcatcttccaaagccctggataagccgggtg atcccggttctgtcggagggttccgcggcggtt ctcagctctgggtcgtacatcgaacagatgtt gtcgtgactggatctgaacacccgtgaagggt cagccctttgaaccaccccttatgacgttgcgat tatggcgcatataacatcggcgcgacgacctg ggcatcccgcatcctgtcgtgaggagatgggtct gcgtgttctcgcacagtggtctggcgatggcac cctgggtgaaatggaaaacacccgttltgtaaa ctgaacctgggtcactcgtacccgtccatgaact acattggccgtcacatggaagaaaaatcaga tcccgtgagtggaatacaacttcttgggtcgact aaaatcgcagaatccctcgttaaaatcggcgtat cagtttgatgatacattcgcgcgaacgtgaag cagtaattgcgcgtacgaaggccagatggca gcaatcattgctaaatgacgttccggccttggaa ggctgtaaaagtgtcgtgtacatgggtgtgtcgt gtccacgtcatgtgacgtggtcgtacgaggacc tggcagtgagatcagcagcgggttacgaa ttgcacacaacgacgacatgactgtacgtgc cagacctgaagaaagggtacgtgctgttggacg acggcagctcttgaactgggaagccttctgtgaa agcgctgaaccagacgtgacggcctccggca tcaggaaaaatacatcttcagaaaaatggcggt ggcggtcccgagatgcactcgtgggactatc cggtccgtaccacggctacgacgggttccgtatc ttcgtctgtgacatggatatgacctgaataacc cagcgtggaatgaactgacgcacccgtggtcgt aaatctgcataa	82,6	51,2

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
K	El gen K WT es SEQ ID NO:7 El gen K sintético es SEQ ID NO:8	atgagccaaacgatigataaaaataatagctgttat ccgctatlcgaacaggatgaataccaggagctgtt ccgcaataaaggcagctggaagagcgacag atgcgcagcgcgtgagaggcttctgctggac caccaccgccgagtagaagcgctgaattccga cgcgagggcgagcgtgagccggcgaaagcc tgccagccgcttggcgcggtgcttctgctggg atttcccaacacctgccgtatgtgcacgctctca gggggtgcgtggcctacttgcacattttaaccg ccatttcaaaagagcgcgtgcgttccgact cgatgaccgaagacgcggcggtcttggcgga acaacaataigaacctggcgctgcagaacggcag cgcgctgtacaacccggagatcattgggtgtcc accacctgcattggcggaagttaicggcgatgacc tgaggcgcttattcgccaacgctaaaaaagatggc ttcgtgcagacgagcgtccgtgcccacggcc atagcccaagcttattcgagccagctcaccgg ctgggataacatgttgaaggcttgcgaacacatt cactgcgactacagggcgagccggcgcaatt ggcggaagctcaatctgggtgacggcttgaacct atctcggaacttccggtatttaaacgggatgatg gaacagatggcggtgcgtgcagcctgctctccg atccgtcggaggtctgcacacggccggacgg tcactatcggtatctcggcgccaccacgcagc aggagatgaagaggccctgacgcatcgata cgctgctctgcagccgtggcagctgctgaagag caaaaaaagtggtgcaggaatgtggaacagcc cgccaccgaggtcggcattccgtggggctggc cgccaccgagtaactgctgaagaccgtcagccag cttagcggaagccgattggcgacggcctcacc ttgagcgcgccggctgggtgacatgatgctga ctccacacctggctgcacggcaagagttggc ctgtacggcgatccggacttctgtatgggctcac ccgcttctgctggagctgggtgcgagccaacg gtgactcggagccataacgccaacaaacgtggc aaaaagcgatgaacaaatgctcgtatgctcggc gtacggcgcgatagcgaagtgttaacacgtcg atttgggacttccgttgcgtgatgttaccgtca ggcggaattatgatcggcaactctacggcaagt ttatcagcgcgataccctggcgaaagggtlaaagc ctttgaagtgcgcttatccgctcggcttccgct gttcgaccgccaccatctgaccgccagacaacc tggggtatgaagggcgatgaacattgtgacga cgctgtgtgaacggcgctgaggaacatggatag cgataccagccagctgggcaaacggattacagc ttcgtatcgtcgttaa	atgtctcaactatcgataaaatcaactctgttac ccgctgttcgagcaggacgaataicaggaaactg ttccgttaacaaacgtcagctggaagaagcgac gacgcacagcgcggtgcaggaagtgttgcattg gaccaccacggcggaatacgaagctctgaactt ccgtcgcgaagccctgacgggtgatccggcga aagcgtgcccagcctctgggtgcgtctgtgca ggcctgggtttggcaaacacctggcgtatgtca cggttccaggcgctgcgtagcctacttccgtacc tatttcaacggcactttaagaacaaatcgct gcgtgtcgcagacatgacggagacgcggc agtttctgggtgtgaacaacaacatgaactgggc ctgcaaaatgcttccgactgtacaacggga atcactgcaggtctacacacgtgagcagag gttatgggtgatgctgcgaagcatttattgccaac gcaaaagaagacgggttctgttgcagcgtctatcg cggttccgacgctataccccgtctctatcgg ttctacgttaactgggtggacaacatgttgcga ggcttgcgaacaaactttaccgagactatcaag ggcaacgggttaaacgtccgaagctgaacctg gtgacggcttgaacactacctgggcaacttct gtgtctgaagcgcatgagtgagcagatggcg gttccgtgttctctgtgtctgacccgtctgaggtt ctgacacaccagcgagcgccactatcgcgt gtattctggggcaccactcagcagggaatga agaggcccccagacgcgattgacacctgtctgc tgcaaccgtggcagctgctgaagaacgaagaa gttctcaggaaatgtggaacagccggcaacg gaagtgtcaatcccgctgggtctggcagctact gacgaactgctgatgaccgtgttcccaactgagc ggcaaaccaatcgggtatgcttgcaccttga acgcggctgcctgtggacatgatgtggaaca ggcacacgtggctgcatggcaagaatttggcc tgtacgggtgaccggacttctgaatgggcttga cccgttctgtctggaactgggtctcgagccga ctgtatctgtctcaaacgctaaacaaacttgg cagaaggccalgaacaaatgctggatgcgag ccatagggcggtgatagcgaagtgttcatcaa ctgcgacctgtggcatttccgtctctgtatgttac gcgtcagccggatttcatgatcggtaactttac ggcaaatcatccagcgtgacactctggccaaa ggcaagcgtttgaagtccgctgattctctgg gcttccgtgttgcacgtaccacctgcacgg ccagaccaccttgggggtacgaagcgcgatga acatcgtlaactacttgggtaaacgcagttactga aaagctggacagcgatacttccagctgggca aaaccgactattcttgcattggttctgtaa	82,1	51,2

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
Y	El gen Y WT es SEQ ID NO:9 El gen Y sintético es SEQ ID NO:10	atgtccgacaacgataccctattctggcgtatgctg gcgctgtttcagtcctcctggacctacagccggc gcaaatcgtcactggtggcgcagagagcggg cgagacgctgagccagagcgtctggcgacct gacccagccgagctggcgcagctttccctcc gcgacggcgggtgatgccccgcctgctggtgc gggtgatggcgagcctgcagggcgcgctgccc ccatttacgcatcggtgcctgcccagcgcaag ccgagctgctggcgccattttgctcccaggatgg gctgggtgattaacggccatttcggccagggacga ctgttttatctacggttcgatgaacaaggcggt ggtgtgacgatctggcgcctatccctccgcccc caccagcaggaggccaacgaagtgcgcgcccg gcttatgaggactgicagctgctgtttggcagga gataggcggggcccgccgcgcggctgatcc gccaatgcatacccccgatgaagcgagcccg ggacgacgattcaggccacagtgcgaggcgatca atagcctgctggccggccgtttgcgcgtgggt ggcgaagcgggttaacagggaacctctggaa gaacgcgttttttaa	ATGTCTGACAATGATACCCCTG TTTTGGCGCATGCTGGCGCTG TTTCAGTCGCTGCCGGATTTG CAGCCGGCTCAAATCGTCGA TTGGCTGGCGCAGGAATCCG GCGAAACCTGACGCCGGAG CGCCTTGCCACCCTGACCCAA CCGCAACTCGCGGCGTCGTTT CCATCCGCGACGGCAGTGAT GAGCCCGGCTCGCTGGAGCC GCGTTATGGCTTCTCTGCAAG GCGCCCTCCAGCCCACTTGC GCATCGTACGTCCGGCGCAG CGTACCCCGCAACTGCTCGCC GCGTTTTGCAGCCAAGACGG CCTTGTTATCAATGGTCATTT CGGCCAGGGTCGTCTGTCTT CATTTACGCCTTTGACGAGCA GGGCGGCTGGCTGTATGACTT GCGCCGCTATCCGAGCGCAC CGCACCAGCAGGAAGCGAAT GAGGTGCGTGCTCGTCTGATT GAAGATTGCCAGCTGCTGTTT TGCCAGGAGATTGGCGGTCC GGCAGCAGCGCTCCGATCC GCCACCGCATCCATCCGATG AAGGCGCAGCCGGTACTAC GATTCAGGCGCAGTGTGAAG CTATCAACACCCTGCTGGCCG GTCGCTGCCGCGTGGCTCG CCAAACGTTTGAACCGTGAT AACCCGCTGGAAGAGCGTGT GTTTTAA	79,8	45,2

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
E	El gen E WT es SEQ ID NO:11 El gen E sintético es SEQ ID NO:12	atgaagggaatgaatctggcgtcgtggatga accggcctgfgaacacaaccataaacaataacc ggcgtcagcggcccaaacccggcgccaccgc cgccggcgtcgcgttcgacggcgccagataac cctgctgcccacgcgcgacgtggcgcatctggte cacggcccccacgctgcccgggaagctcatgg gataaccgcggcagcgccagctccggccccacc cttaacggcctcgggttcaccaccgatcaacga acaggacgtgattatggccggcggaacgcgg actgtttacgcgcgtggccatctgacccgcta tcacccggcggtgtttatctacaacaccgtgt accggccatggagggcgatgacctggaagcgg atgccagggcgcgacacggccacggcgatcc gggtatcgtatgacggcggtttctacggca gtaaaaatctcggtaacggcgccggcgacg tcattgtcaaacgggtcgcggccagcgccggcc cgccccctggcggaagacacgctcttggcccg gagcagcgtcacgatatggcctgattggcgaatt caattatggcgcgatgttctgcatattcagcgct gctcgacgaactggggatccgctgctcgcaag cctctccgtggtggcggttcgcggagatccag accatgacccggcgcgagcccaatgtctgtct gctcgccggcggttaattaacgtgccagagccct ggagcagcgtacggcagcgctgttgcgaagg cagcttttacgggtacggcgccacctgtgacccc tgcggcagctggcgcgctgtggtggcgacgacg accitgccagcgccaccgaagcgctgattgcgcg ggaggaacagcgccgggaactggcgctacagc cgtggcggaacagctgcccggcgcaaaagcg ctgctctataccggcggtgaaatcctggtcggt ggatcggcgctgaggtatgggcatgaccgtg gtggcaaccggcagcgctaaatccaccgaagag gataaacagcggtatccggagctgattggcgaa gagcggttaattgctggaagagggcaacggccg cacgctgctggtatggtctatcgtatcaggccg accigtatgattggcgcgacgcaatatgtacacc gctataaagccaggctgcccgttctcgtatcaat caggagcggaacacgccttcgctggctatcagg ggatcgtcaccctcgcccgccagctgtgtcagac catcaacagccccatctggcgcaaacccattctc gcgccccgtggcgctaa	ATGAAGGGTAACGAGATTCT TGCTCTGCTGGACGAACCGG CCTGTGAACACAACCATAAA CAGAAATCCGGCTGTAGCGC CCCAAAGCCGGGTGCGACGG CGGCTGGCTGCGCTTTCGATG GTGCGCAGATCACCTGCTCC CGATTGCGGACGTTGCCACC TCGTGCATGGCCCAATCGGTT GCGCAGGTAGCTCTTGGGAC AACCCTGGCAGCGCCTCCAG CGGTCCGACCCTGAATCGTTT GGGCTTTACCACTGACTTGAA TGAACAAGATGTGATCATGG GTCGCGCGGAGCGTCCGCTG TTCCACGCTGTGCGCCATATT GTCACCCGTTACCAACCCAGCG GCAGTATTCTCTACAATACG TGCGTGCCGGCTATGGAAGG CGATGACCTGGAGGCCGTGT GTCAGGCAGCCAGACTGCG ACCGCGCTCCCGGTAATCGC AATTGATGCGGCTGGCTTCTA CGGTTTCGAAGAACCTGGGCA ACCGTCCGGCAGGCGATGTC ATGGTTAAACGCGCTATTGGC CAACGTGAGCCAGCGCCGTG GCCGAGAGCACCTGTGTTG CCCGGAGCAACGTCATGAC ATTGGCTTGATCGGTGAGTTT AACATTGCGGGCGAGTTTTG GCACATTACGCGCTGCTTGA TGAGTGGGTATCCGCGTTTT GGGTTTCGCTCAGCGGCGATG GTCGTTTCGCCGAGATTCAAA CCATGCACCGTGCCAGGCG AACATGCTGGTGTGCAGCCG TGCTCTGATCAATGTTGCGCG TGCTCTGGAACAGCGCTATG GCACCCCGTGGTTTGAAGGCT CGTTCTATGGTATCCGCGCGA CCAGCGACGCCCTGCGCCAG TTAGCGGCGCTGCTGGGCGA TGACGACCTCCGTACGCGCA CCGAGGCGCTGATCGCGCGT GAAGAACAGGCGGCTGAGCT GGCCCTGCAACCGTGGCGTG AACAGCTGCGTGCCGCAAG GCCCTGCTCTACACGGGTGGT GTCAAAAGCTGGTCTGTGGT GTCCGCGCTTCAGGATCTGGG TATGACCGTGGTTGCCACGG GCACGCGTAAGAGCACGGAA GAGGATAAACAGCGCATCCG	80,8	47,6

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			CGAATTGATGGGCGAAGAGG CCGTGATGCTTGAAGAAGGC AACGCACGTACCTTATTGGAT GTAGTTTATCGCTATCAAGCA GACCTGATGATTGCCGGTGG CCGCAACATGTATACCGCCTA CAAAGCGCGCTTGCCGTTCT GGACATCAACCAGGAACGCG AGCACGCGTTTGCGGGCTAC CAAGGCATCGTGACCTTAGC GCGCCAGCTGTGCCAAACGA TTAACAGCCCGATCTGGCCGC AGACTCATTCCCGGCACCGT GGCGCTAA		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
N	El gen N WT es SEQ ID NO:13 El gen N sintético es SEQ ID NO:14	atggcagacattttcgcaccgataagccgctggc ggcagccccatcaaacccggcagccgctcgg cgcaalcctcgcagccctcgggacgaacacagc atccctcggcgcacggcgccaggggtgcagc gccttcgcaaaagctttttttcaacatttcacga ccgggttccctcgcagtcgacggcgatggacccc acgtcgcagattatggcgccggcagcgaataattt taaccgctggataccctctgcagcgcaacaatc cgagggctatcgtactcgcagcagcggtcgtc ggagggccagggcagcgatattcccgctgggt cgccagtttcgcgaagagtatcccgccataagg gggtggcgatattgacgggtaaacacggcgatttt atggctccatggagaacggctcagcgcggtgta gagagcgtcattgagcagtggtgcccggcgcc ccgcgcccggctcagcgcaalcgcgggtcaat ctgctggcagccatctctgttcgcccggcgatc gagtggtcgtgcgcagtcgtcgaagccttggct gcagccgataatctgcggacgtggcgcaatcg atggacggccacctggcgagggcgatttctgc cgctgaccagggcgggacgcccgtgcgcag atagagcagatggggcaaaagcctgtgcagctcg ccattggcttcccttcacgcgcctcatcgtcgc tggccccgcgtgcgcggcgagggtatcgccct gcccgcacctgagaccccgaaacgtgcgcagc ctttattatcaactggcgaaaatttcggacgcgc cgttccggagtggtggaaagccagcgccgcca gctacagatgcaatgacgactccatattggc tccagggccagcgcatggcgatagcgccggaa ggcgatttgcgtgcgcgtggtgatttgcgcaa cagccaggggatgcaagccggcccgctggcg ccctacggctcaccagccctgcgcagctgcc gggtgaacgggtggtgcgggggacatggagga tctgcaaacctgctgtgcgcgcatcccgcgac ctgctgtggcgaaactgcagcggcgacgtg gcggagcagtttgcgtgcgcgtggtgcgcgcg gggttccgctcttgacaagctggcgaaatccgc cggtgcgacaggggtatagcggatgcccgat acgtgtttgagctggcaaacctgatacgagcgc gtcaccacacctcggccactaccgacggcgct gcgcaggaacccgaatcgtcactctccacagga ggcgcttatgcccgatata	ATGGCAGACATTTTCCGCACT GATAAGCCGTTGGCTGTGTCG CCGATCAAGACCGGCCAGCC GCTGGGTGCGATCCTGGCGTC CCTGGGTATCGAGCACTCGAT TCCGCTGGTACATGGCGCGC AGGGCTGTTTCGGCTTTTGCCA AGGTTTTCTTTATCCAGCACT TCCAGGATCCGGTCCCGCTGC AAAGCACGGCAATGGACCCG ACCAGCACCATCATGGGCGC TGATGGTAACATCTTCACCGC GCTGGACACTCTCTGCCAACG CAATAACCCGCAAGCAATTG TGCTGCTGAGCACCGGCTCT CCGAGGCGCAGGGCAGCGAC ATTTCCCGTGTAAGTGCGTCAG TTCCGTGAAGAATATCCGCGT CATAAAGGCGTGCGGATTCT GACTGTTAACACCCCGGACTT TTACGGTAGCATGGAGAACG GCTTTTCCGCTGTCTGGAGT CTGTGATTGAACAGTGGGTTT CGCCAGCCCCACGTCCGCGC CAGCGCAATCGTCGCGTCAA TCTTTTGGTGAGCCATCTCTG TAGCCCAGGCGATATTGAAT GGCTGCGCCGTTCGCTCGAG GCCTTCGGTCTGCAACCGATC ATTCTGCCGATCTGGCTCAG AGCATGGACGGCCACCTTGC TCAGGGTGACTTTTCGCCGCT GACGCAGGGCGGCACGCCGT TGCGCCAAATCGAGCAGATG GGCCAGAGCCTTTGCTCTTT GCGATTGGCGTCAGCCTGCA CCGTGCGAGCAGCCTGCTGG CTCCGCTTGTCTGGCGAAG TCATCGCCTTGCCGACCTCA TGACCTTGGAACGCTGCGAC GCCTTTATCCATCAGTTGGCG AAAATCAGCGGTTCGCGCCGT TCCGGAGTGGCTGGAACGCC AGCGCGGTGAGCTGCAAGAC GCCATGATCGATTGCCACATG TGGCTGCAAGGCCAGCGCAT GGCGATTGCCGCCGAAGCG ACCTGCTGGCAGCGTGGTGC GATTTGCGGAACCTCAAGGT ATGCAGCCGGGTCCACTGGTT GCTCCGACGGGTATCCGAG CCTGCGTCAGTTGCCGGTGGA GCGCGTGGTGCCGGGTGATC TGGAGGATCTTCAGACCTCT	79,7	47

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			TATGCGCACATCCGGCCGACT TACTGGTGGCGAACTCCAC GCCCGTGATTAGCAGAGCA ATTCGCCC TGCCGCTGGTGCG CGCAGGCTTCCCGCTGTTGA CAAACTGGGCGAATTCGTC GTGTTGCCAGGGTTATAGCG GTATGCGTGATACCCTGTTTCG AGTTGGCGAACCTGATCCGT GAACGCCATCATCATCTGGCT CATTATCGCAGCCCGCTGCGC CAGAACCCAGAATCCTCGTT GTCTACGGGTGGCGCGTACG CAGCGGATTAA		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
U	El gen U WT es SEQ ID NO:15 El gen U sintético es SEQ ID NO:16	atgtggaattactccgagaaagtgaagaccatttt tttaacccccgcaatgcgcgcgtggaggacaacg ccaacgcggtaggcgacgtcgggtgtaagctg cggcgacgccctgcgcctgctgctgcgcgtcga cccgcaaaagcgaatcattgaggagcgggcttc cagaccttcggctgcggcagcggcctcctcct cctccgcgtgacggagctgattatcgccataacc ctcgcgaagcggcgagataaccnaccagcag attgccgattatctcgacggactgcgcgggaga aaatgcactgctcgtgtagggccaggaggccct gcgcgcggccatcgccacttgcgcggcgaag cttgaagaggagcagcagcaggcgcaagctgat ctgcaaatgcttcggcgtcgtgaaggcgcatatc gccgcgcggctacagaacaacggcgtgaccaccc ttgccgaggtgatcaactacacaaagcggcg cggtgcacctcttccacgaaaaatcgagctg gccctggcgagatcctcggcagcagccgcag acgacgccagccgtggccagcggcaaatcc gcactggcagagcgtcgtcgataaccatcgagaa ctgcggcccatattcagccgacggcgccgat atggcgctactcagcgtcaccacccaggtga ccgtcagcctctcggcagcgtgtagcggctgat gatgaccgataatgacctggcctggcgtcagcaa aaactgatggaactgaccggctgtatattggaagt ggtagcgccctga	ATGTGGAAGTACAGCGAGAA AGTCAAGGACCATTTCTTCAA TCCGCGCAACGCGCGTGTGT GGATAACGCAAATGCGGTGG GCGACGTCGGCAGCTTATCTT GTGGCGATGTCTCCGCTTGA TGCTGCGCGTGGACCCGAG AGCGAAATCATCGAAGAAGC GGGCTTTCAGACCTTCGGCTG CGGACGCGGATGTGCGTGT CCAGCGCACTGACGGAGCTG ATCATCGGTACACCCCTGGCG GAAGCGGTGATGATCACCAG CCAGCAGATCGCCGACTATCT GGACGGCTTACCGCCGAGAA AGATGCACTGCTCTGTAATGG GCCAGGAAGCTCTTCGTGCG GCCATTGCTAACTTTCGCGGT GAATCGCTGGAAGAGGAGCA TGACGAGGGTAAGCTGATCT GCAAGTGCTTCGGCGTCGAT GAAGGCCATATTCGCCGTGCT GTCCAGAACAACGGTCTTAC GACTCTGGCCGAGGTGATCA ATTACACCAAGGCAGGTGGC GGTTGTACCAAGCTGCCATGA GAAAATCGAGCTGGCCCTGG CCGAGATTCTCGCCCAACAG CCGCAAAACACCCCGGAGT TGCGTCCGGTAAAGATCCGC ACTGGCAGAGCGTCGTGGAT ACCATCGCTGAAGTGCCTCCA CATATCCAAGCGACGGTGG TGACATGGCGCTGTTGTCCGT GACGAACCAACCAAGTGAAGT TTTCGCTGTGCGGCAGCTGTT CTGGCTGCATGATGACCGAC ATGACCCTGGCGTGGCTGCA ACAGAAATTGATGGAGCGTA CCGGCTGCTATATGGAAGTTG TTGCCGCTTAA	82,3	53,5

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
S	El gen S WT es SEQ ID NO:17 El gen S sintético es SEQ ID NO:18	atgaacaggtttatctcgataacaacgccaccac ccgtctggaccgcatggctctggaagcgaatg ccctttttgaccgattttaccggcaacccctcgtcga tacacgatttggcattccggccagggcgctctg gaacgcgcgcacagcaggctcgccgcgtcgtg ggcgccggagatccagcagatcatctttacctc ctgcgccaccgaagccaccggccaccgcatcgc ctcgccgacgcccctgctcctgagcgtcgcgaa atcatcaccagcgtggtcgaacatccggcgacgc tggcgccctcgccagcacatggagcgcgagggtc accggattcatcgcacgcggtagatggcgaggg ggcgctggacatggcgccagttccggcgccgct cagcccgccgctcgttggcagcgtgatgtgg gcgaataacgaacccgggtgcttttccgacgc gcgaatggcgagctggcccatgaacaaggg gcgctgtttcactgcgatgcgggagcgtggcgg gaaataccgacgcggctggccagaccgcgat cgatatgctctcgtcgcgcgataagttccacg ggccaaaaggcgtaaggctgttcttgcggcg gggaacgcgctttcggccgctgctgcgcggcggt caccaggagtagcgtcggcgagccgggacaga aaatatctgcggaatcgtcggcatggcgcgcc tgcgagctggcgaatattcatctgccgggaatgac gcatacggcccaatggcgcaacngctggagcat cgctcgtcggcagcgtgcgcgtcgtcatgttga tggcgccggccagccggcggtgcccggcagc gtgaatciggcctttgagttattgaagggtgaagcc attctgctgtgttaaccaggccgggacgcgcgc ctcagcggcagcgcctgcacctcaggtcgtcgtg gaacccctccacgtgagtcggcgatgaataacc cctacaccgcgccacggcaccatccgctttct ctctcgcgtacacccggagaaagagatcgatt acgtcgtccacgcgtccgccgattatcgaccg gctcgcgcgcgtcgtcgccttactggcagaacgg caagccggcccgccggagcgcgtatcaccgccc ggttacggctaa	ATGAAACAAGTGTACCTGGA CAACAACGCGACCAACCCGCC TGGACCCGATGGTTCTTGAA GCGATGATGCCGTTTCTCACG GATTCTATGGCAATCCGTCC AGCATCCATGACTTCGGCATC CCGGCACAAGCGCGCTGGA ACGTGCGCACCAAGCAAGCTG CGGCACTGCTGGGCGCAGAG TACCCGTCTGAAATCATTTTC ACGAGCTGTGCGACCGAGGC CACTGCAACCGCCATTGCGTC GGCCATCGCGTTATTGCCGGA ACGCCGCGAAATCATCACCT CGGTAGTGGAGCAACCCGGCT ACGCTGGCGCGGTGCGAGCA CATGGAACGCGAAGGCTATC GCATCCATCGCATTTGCGGTGG ATGGCGAAGGTGCGCTGGAC ATGGCCCAGTTCCTGTCAGC GCTCtccCGCGTGTGCGGTTG GTGAGCGTGTATGTTGGCCAA CAACGAAACCCGGCGTGTGT TCCCGATTGGCGAAATGGCC GAGCTTGCCACGAGCAGGG CGCTCTGTTCCACTGCGATGC CGTTACAGGTGTTGGCAAAAT CCCAATTGCTGTTGGCCAGAC GCGCATCGACATGCTGTCTTG CTCCGCGCACAAGTTTCATGG TCCGAAGGGTGTGGTTGCTT GTACTTACGTCGTGGCACGCG CTTTCGTCCGCTGCTTCGCGG TGGCCATCAAGAATATGGTC GCCGTGCCGGCACTGAGAAT ATCTGTGGCATCGTCGGCATG GGCGCTGCGTGCGAACTGGC GAACATCCATCTGCCGGGTAT GACCCATATTGGCCAGTTACG CAATCGCCTGGAGCACCGTCT GCTCGCCAGCGTGCCGTCCGT GATGGTTATGGCGGTGGTC AGCCGGCTGTACCGGGTACT GTCAACCTGGCGTTTCGAGTTT ATCGAAGGTGAAGCGATCCT GCTCTTGCTGAACCAAGGCTGG CATTGCCGCAAGCTCCGGTC CGCGTGTACCTCTGGCAGCTT GGAGCCGAGCCATGTGATGC GCGCATGAACATTCCATAC ACCGCGGCTCACGGACCAT TCGTTTTAGCCTGAGCCGTTA TACGCGCGAGAAAGAGATCG ACTACGTGTTGCGACCCCTCC	79,1	43,6

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			CGCCAATCATTGATCGTCTGC GTGCCCTTGCCCCGTATTGGC AGAATGGTAAGCCGCGTCCG GCAGATGCAGTCTTTACCCCG GTTTACGGTTAA		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
V	El gen V WT es SEQ ID NO:19 El gen V sintético es SEQ ID NO:20	atggaacgcgtgctgattaacgataccaccctgcgcgacggcgagcagagccccggcgctgccttccgacccagcgaagggcgcattgcccaggcgctttacggccgaggaataacggcgatggaggtcggcaccggcgatggcgacgaggagatcgcgcggtaccagctggcgctgcgcagctgcccagacggacccctgatgacctgggtgcggaatgaacgcgcgtggatccggcagagcgccgatctgggcacgactgggtggatctcgtatccggcttcggataagctgcggcagtaacaactgcgcgagccgctggcggtgctgctggagcggcgtggcgatgttatccattggccataccctcggcctgaaggatgcatcggctgcgaggacgctcggggccagcgccagaccctgcgcgctatccggaggtcgcgcagcaatgcgccggcgcccgctcgtatgcccagatcggctggcctgctgacccctttaccacggcgccgaatcctggccctgcgcgacgtctggtccggcgaatcgaatgcatgccataacgatctgggtatggcgaccgcgaatagctggcgagcgtgaaagcggggccacagcggtgaatagcaggtcctcggctcggcgagcggcgggcgaacgcgcgcgtggaaaccgtgcgctgggcttgaacgctgctggcggtggagaccggcggtgcatcttgcggcgtgcccgcgtctgcaagggtcgcggaagccgcagcgcgccatcgaccgcagcagccgctgggtcggcgagctgggtttaccatgagtcaggtgtccacgtggcgggcgctcgcggcacagcgagagctaccagtcacgtcccttccctgatggccgcagctaccggctgggtgtgggcaaacacitccggcgctcagcggtgcaacggcgagattaaccagctcgtgccgccatccggcgcttccggcgaactgggaagcgagccgaaagattacagagctgtggctatcagcagagctgtgcgtgtaatccgctcgtggcgaggggggtaa	ATGGAGCGCGTCTTGATCAA CGATACTACCCTGCGTGATGG CGAACAAATCTCCGGGCGTAG CGTTTCGTACCTCCGAGAAAG TTGCCATCGCGGAGGCACTGT ACGCTGCGGGTATCACCGCG ATGGAAGTCGGCACTCCGGC GATGGGTGATGAAGAGATCG CCCGCATTCAGCTGGTGCGTC GTCAACTGCCGGACGCGACG CTTATGACCTGGTGCCGTATG AACGCTCTGGAATCCGTCA GAGCGCGGATCTGGGTATTG ACTGGGTGGATATCTCGATCC CAGCATCCGACAAGCTGCGT CAGTACAAGCTGCGTGAGCC GCTGGCCGTGCTGCTGGAGC GCCTTGCGATGTTTATCCATC TGGCCACACGTTAGGCCTCA AAGTATGTATTGGTTGCGAG GATGCGAGCCGTGCGTCTGG TCAGACCCTGCGCGCCATTGC CGAGGTGGCCAGCAATGCG CGGCTGCGCGCTTGCGTTACG CTGACACCGTGGGCCTGCTG GACCCGTTACACACCGCAGC CCAGATCAGCGCCCTGCGTG ACGTTTGGTCGGGCGAGATC GAGATGCATGCTCAAAATGA TCTGGGCATGGCTACCGCGA ACACGCTGGCGGCAGTTTCG GCTGGCGCCACGTCGGTGAA CACTACCGTCCTCGGTCTGGG TGAACGTGCAGGCAACGCAG CCCTGGAAACCGTTGCGCTG GGCCTGGAACGCTGCCTGGG CGTGGAACCGCGTCCATT CAGCGCGCTCCAGCGAGCT GTCAGCGCTGCGGAGGCT GCACAGCGCGCAATCGACCC GCAACAGCCGCTGGTGGGTG AATTGGTTTTACCCACGAGT CTGGTGTTCAGTTGCGGCGC TGCTGGCCACAGCGAATCCT ATCAATCTATTGCCCAAGCC TCATGGGCCGTAGCTACCGTC TGGTGTCTCGCAAGCATTTCG GGTCGTACGGCTGTCAACGG TGTTTTCGACCAGATGGGTTA CCACCTGAATGCGGCGCAGA TCAATCAGTTGCTGCCGGCCA TTCGCCGCTTCGCCGAGAATT GGAAACGCTCTCCGAAAGAC TACGAACGTGTTGCGATCTAT	80,8	48,3

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			GACGAATTGTGCGGTGAATC CGCCCTTCGTGCTCGCGGCTA A		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
W	El gen W WT es SEQ ID NO:21 El gen W sintético es SEQ ID NO:22	atggagtggtttataaaatccggcggtggacga acttcgctccgcgaatctttttcagttttcgcgt ccctatcagcccgagctgcttggcgcgtgcagc ctgccggtgctggcaacgtttcgcgcaaacctcg cgccggaggtgccgtgcaaacggcctcgagga taacgaccgcgcgcctggtgctggtgcggaag actgctcgggagagctatcagcaacagtticag gagagcggaacatga	ATGGAGTGGTTTTACCAGATT CCGGGTGTAGACGAATTGCG CAGCGCTGAATCCTTCTTTCA GTTCTTCGCGGTTCCATACCA GCCGGAAGTGTGGGCCGCT GCTCGCTTCCGGTGTTAGCGA CGTTCCACCGTAAACTGCGTG CGGAGGTCCCGCTGCAAAAC CGTCTGGAGGACAATGATCG TGCGCCGTGGCTCTTGGCGCG CCGCCTCCTGGCCGAATCTTA TCAGCAGCAATTCAGGAGA GCGGCACCTAA	77,9	43

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
Z	El gen Z WT es SEQ ID NO:23 El gen Z sintético es SEQ ID NO:24	atgagaccgaaatcacctttagcgaagaggtccg cgtcgtacgcgcgattcgtaacgacggcacccgtg gcgggcttcgcgccggcgctgctggtcagg cgcggcngcacccgctttgtgcgcgactggggc gttttttgcagatcagattatctaccagatccaatt tcggaaaccgatcggatcgcggctgccgcgag caggagctgatcccatcaccacggcgtggcgtgg ccggaatttgaatacagggatagcgtgacctg ccagatggcgtcgcgcgtcaacggcgtatgtgtc gtgagcggcggccagcggggacgcgttgaggct accgatcggggagagctcggcgacagctacacc gtcgactttagcggcgcgtgttcagggtcccggt gcaggccatcgcccttatagaggaaagagaaga atga	ATGCGCCCGAAATTCACCTTC TCTGAAGAGGTCCGCGTAGTT CGCGCGATTCTGTAATGATGG CACCCTGGCGGGTTTTGCGCC AGGTGCGCTGCTGGTTCGTCTG CGGTTCGACGGGCTTTGTGCG TGAATGGGGTGTGTTCTGCA AGACCAGATCATCTATCAAA TCCACTTTCCGGAAACCGACC GCATTATCGGCTGTCGCGAGC AGGAGTTAATCCCGATTACCC AGCCGTGGTTGGCTGGTAAC CTCCAGTATCGTGACAGCGTC ACGTGCCAAATGGCACTGGC TGTCAACGGTGACGTGGTTGT GAGCGCCGGTCAACGTGGCC GTGTGGAGGCCACTGATCGT GGCGAACTTGGCGATTCTAC ACCGTGGAATTCAGCGGCCG TTGGTTCCGCGTTCCGGTCCA GGCCATCGCGCTGATTGAAG AGCGCAAGAATAA	80,1	48,3

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
M	El gen M WT es SEQ ID NO:25 El gen M sintético es SEQ ID NO:26	atgaacccgtggcaacgttttgcggcagcggc tggcgcgcagccgtggaatcgcgatccggcgg ccctggatccggcgcacagccgctttgaaca ggcctggcaacgccagtgcctatggagcagac gatcgtcgcggggtccctgaaggcgatattccg gcggcggttctggagaataatcgtccctccctgc catctggctcgcagagggggaatttgcgcgcgc gagcgcgcgtccatcgtgcgcacacagccgg ctggaactgccttcgcgatcgcgcgcaggg cgccgcagccggatctcgcaggtacagggcatg gtatctgcgcaccagacgcagtttatgcgcgcg gaacagcgtcgaaccgcatctacgtcgcaggt cgataacgaccgcgaagccgtgcacacgcggat cctcggcctglatcggcaaatcaacgcctcgcgg gacgcttctgcgcgcgtggccagcgcctatcc actcccgcgcgcgtgggaaggggtcgtttagg ctggattagccgtggcctgcctatccgcagctcg agaccgcgcgtgtttacgtggcggaaacgcgc aagccttccatcgcagcgaactggcgtggcat cttttatgggtgcgaagcgaatcgcgcgcgcgc catggagccgcagcagcgcgtggagagcgcgc gcgattatcttggcagcagagccagcagcga tcagcgcaggtggcgtggaacagatgattccgctc agccgggactgtgcggtag	ATGAATCCGTGGCAGCGCTTT GCCCCTCAACGCCTTGCTCGC AGCCGCTGGAACCGTGATCC GGCTGCTCTCGACCCAGCCG ATACCCAGCGTTTCGAGCAG GCGTGGCAGCGTCAATGCCA TATGGAACAAACCATCGTAG CGCGTGTCCCGGAAGGCAT ATTCCGGCTGCCTTACTGGAA AACATCGCGGCCAGCCTGGC GATCTGGCTGGACGAGGGTG ACTTCGCTCCGCCGAGCGC GCTGCGATTGTGCGTCATCAT GCACGTCTGGAGCTGGCGTTT GCCGACATTGCCCGCCAGGC ACCGCAACCGGATCTGAGCA CGGTTCAAGCGTGGTATCTGC GTCACCAGACTCAATTCATGC GTCCGGAGCAGCGTCTGACC CGTCACCTGCTCCTGACGGTC GATAATGATCGCGAGGCGGT GCATCAACGCATCCTTGGCCT GTATCGTCAGATCAACGCGA GCCGTGACGCCTTCGCCCCAC TGGCACAGCGCCACTCTCATT GCCCGTCCGCCTTGGAAGAA GGCCGTCTGGGCTGGATCTCC CGTGGTCTGTGTACCCGCAG CTCGAAACCGCGTTGTTTAGC CTGGCGGAAAACGCACTGTC GCTGCCGATTGCGTCGGAATT GGGTTGGCACCTGTTATGGTG CGAGGCCATTCTCCGGCAG CCCCGATGGAGCCGCAACAG GCCCTTGAATCTGCGCGCGAC TACTTGTGGCAGCAGAGCCA GCAGCGCCACCAGCGTCAAT GGCTGGAGCAGATGATTTC CGCCAACCGGCCTGTGTGG TTAA	79,7	45,3

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
F	El gen F WT es SEQ ID NO:27 El gen F sintético es SEQ ID NO:28	atggcgaacatiggtatttcttggcaccgatacc ggtaaaaccgcgaatcgcgaatgatccata agcaactgggcgagctggcggtatgcccggtca acattaacggcagcagctgggaactttaaggcc tatccgggtgctgctgctgggcacgcgcgctcg ggcagggccagctgcggggctggagggcggga tgcgaaagcgagtcattggagcgaattatcagcg gcctcgacgagccagcctgaagggaacacg tggcgtgttggcctcggcgatcagagaggcta tccggacaacttcgtcagcgggataggcccgctg ttcgagccctgagcgcgcggcgccgagatg attggcagctggccaaatgagggttatgaattcag cgcgctcctggcgctggaaggcgaccgcttgggt ggcgtgtgtggtcagcagataaccagttcgacc agaccgaagcgcgcttggcgagctggcttgagg agattaaacgcaccgtgctgtag	ATGGCGAACATCGGCATCTTC TTTGGTACGGATACCGCAA AACCCGCAAGATTGCGAAGA TGATTCACAAACAGCTGGGC GAGCTGGCCGATGCCCGGT TAACATCAATCGTACCACTTT GGATGACTTTATGGCTTACCC AGTCCTGTTGCTCGGCACGCC GACGCTTGGTGATGGTCAACT GCCGGGCTTAGAGGCGGGCT GCGAGAGCGAAAGCTGGTCT GAGTTTATCTCCGGTCTGGAT GACGCTTCCCTGAAGGGCAA AACCGTGGCGCTGTTTGGCCT GGCGGACACGCTGGTTACC CGGACAACTTCGTGTCGGGT ATGCGTCCGCTGTTTCGACGC CTGAGCGCCCGTGGCGCCA GATGATTGGTAGCTGGCCGA ACGAAGGTTATGAGTTTAGC GCATCGTCCGCGCTGGAAGG CGACCGCTTCGTGGCTTGGT GCTGGATCAAGACAATCAGT TCGACCAGACCGAAGCGCGC CTGGCGTCTTGGCTTGAAGAG ATCAAACGCACCGTTCTGTAA	81,5	51,4

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
B	El gen B WT es SEQ ID NO:29 El gen B sintético es SEQ ID NO:30	atgacttccctgctcctcttcttctggcggaagcct gcccgcggcggaatgacagcgatgacggcgc ttgtggcgataangctgcccgcacccctgctac tctcgccatggcgatcacccgttctcgcggaatgca tctgcccgtcgccgcctgcaattgacagtgca actactgtaatcgcaaatcgatgacagcaacgagt cccgcggcgggatcgtaacgctgctgacggc tgaacaggcggctgtaaaagtgcgtcaggtcgcg caggcgatcccgcagcttctggtggcgatcg ccggcgccggcgatccgctcgcaaatatgcgc gcaccttctgacccctggagctgacgcggaaca gctgcccggacgtgaaattatgcttgcgaccaacg gactgggtgctgctgacggcggtggacgcctgct ggatgctggcggtgaccacgicacggtcaccatla acaccctcgacggagatgcccgcgaatcta cgctggctatggctggacggcgaaacgctacag cgggcggaagcgggagagatcctgattgcccg tcagcttgaggcggtacgaggtgacggcgaana ggcggtgctggtaanaataatcggtgctgaccc cggtatcaacgataggcgatggcggtgag ccgcgcgtcgggcgccagcgcggttatacat aatattatgccgtgacggcggcgagcagc gcacgggtgttggcctaacggcgccggagc cgagcgccgagacgctcgccgcccgcgcgagc cggtgcggcggaatgacggcgagatgacccac tgccaccagtgctgcggcgacccatgggagtc tcggcggaagaccgagccagcagtttaccagct tccggcgccagagagctcctggcctggctgccc gatccctcaccagcgcgcgacgtgacgcgag catlccgaccccgcggaatctgaagccgagac gcccgtgctgctcgctgctgcaagccgcggg gacgtcattgattgctacttggcagcgacgg ttctacattacagcctcctggcgccggtatgggt ctggtaacgagcgctttagcccaaatatgta ggggcgcgatgactgagcgccgaggataacg cagcccgttggcggaatcccgaaactgctg ggagctaaagccggtatctgctgctgacggcc atacgccgtggcaacagctggaacaggaaaggca ttgaacccctgcttgcagggcggtggcgccggt ctccgaagtgctgcccgcgtgggtggcaacagcgt cgggggagctggcctgcccgttgcgcataag ggggctgcctga	ATGACCTCTTGTTCGTCGTTT TCTGGCGGTAAAGCGTGCCG TCCGGCCGATGACTCCGCGCT GACTCCGCTGTTGGCCGACA AGGCAGCTGCGCACCCGTGC TATAGCCGCCACGGCCATCA CCGCTTCGCGCGTATGCACCT GCCAGTCGCTCCGGCCTGCA ACTTACAATGCAACTACTGCA ACCGCAAGTTCGATTGCAGC AATGAAAAGCCGTCCGGGCGT GTCCTCTACCTCTGTGACGCC GGAACAGGCTGTGGTGAAGG TGCGCCAGGTCCGCCAAGCT ATCCCGCAGCTGtgcGTGGTCTG GTATTGCTGGTCCGGGCGATC CGCTTGCGAATATCGCCCGCA CCTTCCGTACCTTGGAGCTTA TTCGCGAAGCTTGGCCGAC CTGAAACTGTGCTGAGCAC CAACGGCTTGGTGTGTCAG ATGCGGTGATCGTCTGCTCG ATGTGGGCGTGGATCACGTT ACCGTCAACATTAACACCCCTG GACGCAGAAATCGCAGCGCA AATCTACGCTGGTTGTGGCT GGATGGCGAAGCTACTCGC GTCGCGAAGCCGCGGAAATT CTCATTTGCCCGCAGCTGGAA GGCGTACGTCGCTGACCGC GAAAGGTGTGCTCGTCAAGA TCAACAGCGTATTGATTCCCG GCATCAATGACAGCGGCATG GCGGGTGTTAGCCGTGCGCT GCGCGCGTCTGGTGCCTCAT CCACAACATCATGCCACTGAT TGCGCGTCCGGAGCATGGCA CTGTTTTCGGTCTGAACGGCC AGCCGGAACCGGACGCGGAA ACCTGGCGGCGACGCGCTC CCGCTGCGGCGAGGTTATGC CACAAATGACCACTGCCAC CAGTGCCGTGCCGACGCGAT TGGCATGCTTGGTGAGGATC GCTCGCAACAGTTTACGCAAT TACCGGCTCCGGAGTCCCTCC CGGCTTGCTGCGGATCCTGC ATCAGCGTGCTCAGTTGCATG CGAGCATCGCCACGCGCGGT GAGAGCGAAGCCGATGACGC CTGCTGTGGTGGCGTTGCGTC GAGCCGTGGCGATGTAATTG ACTGCCATTTCCGGCATGCCG ACCGTTTCTATATCTATAGCC	80,2	45,4

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
			TGTCTGCGGCTGGTATGGTTC TGGTTAACGAACGTTTCACCC CGAAATACTGCCAGGGTCGC GATGACTGCGAGCCGCAGGA CAATGCCGCACGCTTTGCTGC CATCCTTGAGTTGCTGGCGGA CGTCAAAGCGGTGTTTGTGT GCGTATCGGCCATACCCGTG GCAACAGCTGGAGCAGGAAG GCATCGAACCGTGCGTGGAT GGCCTGCGGTCCGGTATCC GAGGTCCTGCCGGCATGCTG GCAGCAGCGCCGTGGTAGCT GGCCGGCTGCATTGCCGCAC AAAGGCGTTGCGTAA		

Gen	SEQ ID NO:	Secuencia WT	Secuencia sintética	% Identidad nucleótidos	% Identidad codón
Q	El gen Q WT es SEQ ID NO:31 El gen Q sintético es SEQ ID NO:32	atcccgccgctcgaactggttcggcgcttatggct gctgtaccacgcggggaaggcagctttccgctg cgcatggggcttagcccgcgcatggcagggcg ctgcggcgccgctggcgaggtggaaacgccc gctcgacggcgagacgctcacccgctcgccgct gatggcgagctcaacgccaccgcccgaagagg agcgccagcagctggcgccctggctggcgggct ggatgcagcaggatgccggccgatggcgcnag attatgccgaggtttcgtggcgfttaaccatctet ggcaggatcttggctggcgtcgcgcgcgaatt gcgcctgctgatgagcgaactgctttcacagctgg tgggatgaacgaacacaatatgcgctggaaaaa gttctttatgctcagcgtgtttgctcaacagggg gaagttatctgccgttcgccaagctgcgacgagtg ctgggaacgcagcgctgtttgagtag	ATGCCGCCATTGGACTGGTTG CGTCGTTTGTGGTTACTCTAT CACGCCGGCAAAGGCAGCTT TCCGCTTCGTATGGGCTTGTG GCCGCGTGACTGGCAAGCTC TGCGCCGTGCGCTGGGCGAG GTGGAAACGCCGCTGGATGG CGAAACCCGTACCCGTCGCC GTCTGATGGCGGAGCTGAAT GCCACCCGCGAAGAAGAACG CCAGCAGCTGGGTGCCTGGC TGGCCGTTGGATGCAACAG GATGCCGGTCCGATGGCGCA GATTATCGCAGAGGTGAGCC TGGCGTTCAACCATCTCTGGC AGGACCTTGGCTCGCGAGC CGCGCTGAACGCGTCTGCTG ATGCTGACTGCTTCCCGCAG CTGGTTGTTATGAACGAGCAC AACATGCGCTGGAAGAAATT CTTTTACCGCCAGCGTTGCCT GCTGCAACAGGGCGAAGTCA TCTGTGCGAGCCCGTCTTGCG ATGAATGCTGGGAACGTCTCT GCGTGCTTTGAGTAA	81,3	54,2

Fig. 19. Controlador sintético Ptac-T7

N.º orden 8-19: el esqueleto es pincW (KT_96)

SEQ ID NO:	Orden	Nombre	Secuencia
SEQ ID NO:33	1	lacI	caattcgcgctaacttacatlaattgcgttcgcctactgccgcgttccagtcgggaaacctgctgfgcca gcctgaatlaatgaatcggccaacgcgcggggagagcggttttcgctattggcgccaggggttttttc tttccaccagtggagacggggcaacagctgaltgccctaccgcctggcccgagagaggttgcagcaacg ggctccagctgggtttgccccagcgcgaaatctgtttgatgtgggttcagcggcgcgatataacaatg agctgtcttcggatcgtcgtatcccatlaccgagatalcgcaccaacgcgcagcccgactcggtaaat ggcgcgcattgcgccccagcgcctatctgactgttggcaaccagcaatcgagtggaacgatgccctcat tcagcatttgcattgtttgttgaanaaccggacatggcactcagctccgtccgttcgcgtatccgttga atttgattgcgagtgagatatttatgccaccagccagacgcagacgcgccgagacagaacttaattggg cccgctaacacgcgcgatttgcgttggtagcccaatgcgaccagatgtctcacgccagtcgcgtaccgtct lcatgggaganaaataaiaactgttgaatgggtgtcgttgcagagacatcaagaaalaacgccggaaactatg tgcagcagcgtctccacagcaatggcactctgtgtcatccagcgagatgttaattgataccgccactgcgc gttgcgcgagaaattgtgcaccgccgtttacaggcttcgacgcgcctctgttctaccatcgacaacac cagctggccaccagtgtgatcggcgcgagatttaatcggccgcgaacttgcgcagggcgctgcagggg ccagactggaggttggcaacgccaatcagcaacgactgtttgccgccagttgttggccacgcgggttg ggaaatgtaattcagctccgcctatcccgcttccacttttccgcgttttcgcagaaacgtggctggcctg gttcacacacgcgggaaacggcttgataagagacaccggcctactctgcgacatcgtataacgtttactgg tttccac
SEQ ID NO:34	2	promotor lacI	attcaccaccctgaattgactcttccggggcgctatcatgccataccgcgaagggttttgcaccattcgat gggtgtcaacgtataatgcattgcccttcgccttcgcgcgcgaattggccgcattgccgcgataatggcc tgccttctgccgaaacgttttgggtggcgggacagctgacgaaggcttgagcgaaggcgctgcgaagattcc gaataccgcaagcagacagggccgatcatctcgcgtctcagcgaaagcggtctcgcgcgaaatgacc cagagcgtctgcggcaccgtctctacaggttgcattgataaagaagacagtcataaagctgcgcagcat agtcatgccccgcgccaccgggaaggagctgactgggttgaaagcctcaaggcgatcggcgagct tatcgactgcacggctgcacaaatgtcttggcgtaaggcagccatcgggaagctgtgtgatgtcgtgtga ggctgtaaatcactgcataatctgtgtcgtcaaggcgcaactccggtctggaataatttttttgcgccgac atcataacgggttctggcaataattctgaaatgagctg
SEQ ID NO:35	3	promotor tac	ttgacaattaatatcggcctgtataaattgtggaaattgtgagcggataacaatt
SEQ ID NO:36	4	aislador 1	TGCAGTTTTATTCTCTCGCCAGCACTGTAATAGGCACTAA
SEQ ID NO:37	5	rbs ARNP T7	TATCCAAACCAGTAGCTCAATTggagctCGTCTAT
SEQ ID NO:38	6	etiqueta de degradación umuD	Gtgtgttttatcaagcctgcggatctccgcgaattgtgacttttccgtatttagcgaatctgttcagtgtgg cttttcttaccggcagcagattactgtgaacagcgcatcgtctg

SEQ ID NO:	Orden	Nombre	Secuencia
SEQ ID NO:39	7	Andamio de ARNP T7	gggtggcAACACGATTAACATCGCTAAGAACGACTTCTCTGACATC GAACTGGCTGCTATCCCGTTCAACACTCTGGCTGACATTACG GTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTC TTACGAGATGGGTGAAGCACGCTTCCGCAAGATGTTTGAGCGT CAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAG CCTCTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCA ACGACTGGTTTGAGGAAGTGAAAGCTAAGCGCGGCAAGCGCC CGACAGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGT AGCGTACATCACCATTAAGACCACTCTGGCTTGCCTAACCACT GCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTC GGGCCATTGAGGACGAGGCTCGCTTCGGTCGTATCCGTGACCT TGAAGCTAAGCACTTCAAGAAAAACGTTGAGGAACAACCTAA CAAGCGCGTAGGGCACGTCTACAGAAAGCATTATGCAAGT TGTCGAGGCTGACATGCTCTCTAAGGGTCTACTCGGTGGCGAG GCGTGGTCTTCGTGGCATAAGGAAGACTCTATTCTAGTAGGAG TACGCTGCATCGAGATGCTCATTGAGTCAACCGGAATGGTTAG CTTACACCGCCAAAATGCTGGCGTAGTAGGTCAAGACTCTGAG ACTATCGAACTCGCACCTGAATACGCTGAGGCTATCGCAACCC GTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTG CGTAGTTCCCTCCTAAGCCGTTGGAGTGGCATTACTGGTGGTGGC TATTGGGCTAACGGTCGTCTCCTCTGGCGCTGGTGCGTACTC ACAGTAAGAAAGCACTGATGCGCTACGAAGACGTTTACATGC CTGAGGTGTACAAAGCGATTAACTTGGCGAAAACACCGCAT GGAAAATCAACAAGAAAGTCTAGCGGTGCGCAACGTAATCA CCAAAGTGGAAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGA GCGTGAAGAATCCCGATGAAACCGGAAGACATCGACATGAA TCCTGAGGCTCTACCGCGTGGAAACGTGCTGCCGCTGCTGTG TACCGCAAGGACAAGGCTCGCAAGTCTCGCCGTATCAGCCTTG AGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCCATAAGGC CATCTGGTTCCTTACAACATGGACTGGCGCGGTGCTGTTTAC GCTGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAG GACTGCTTACGCTGGCGAAAGGTAAACCAATCGGTAAGGAAG GTTACTACTGGCTGAAAATCCACGGTGCAAACTGTGCGGGTGT CGACAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAA AACCACGAGAACATCATGGCTTGCCTAAGTCTCCACTGGAGA ACACTGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGC GTTCTGCTTTGAGTACGCTGGGGTACAGCACACCGGCTGAGC TATAACTGCTCCCTTCCGCTGGCGTTTGACGGGTCTTGTCTGG CATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGT CGCGCGGTTAACTTGCTTCTAGTGAAACCGTTCAGGACATCT ACGGGATTGTTGCTAAGAAAGTCAACGAGATTCTACAAGCAG ACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCG ATGAGAACACTGGTGAAATCTCTGAGAAAGTCAAGCTGGGCA CTAAGGCACCTGGCTGGTCAATGGCTGGCTTACGGTGTACTCG CAGTGTGACTAAGAGTTCAGTCATGACGCTGGCTTACGGGTCC AAAGAGTTCGGCTTCCGTCAACAAGTGTGGAAGATACCATTC AGCCAGCTATTGATTCCGGCAAGGGTCTGATGTTCACTCAGCC GAATCAGGCTGCTGGATACATGGCTAAGCTGATTGGGAATCT GTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAAGTGG CTTAAGTCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATA AGAAGACTGGAGAGATTCTTCGCAAGCGTTGCGCTGTGCATTG GGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAG CCTATTGACGCGCTTGAACCTGATGTTCTCGGTCAAGTTCC GCTTACAGCCTACCATTAAACCAACAAGATAGCGAGATTG ATGCACACAAAACAGGAGTCTGGTATCGTCTCAACTTTGTACA CAGCCAAGACGGTAGCCACCTTCGTAAGACTGTAGTGTGGG ACACGAGAAGTACGGAATCGAATCTTTGCACTGATTACGAC TCCTTCGGTACGATTCCGGCTGACGCTGCGAACCTGTTCAAAG CAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGATGT ACTGGCTGATTCTACGACAGTTTCGCTGACCAGTTGACACGAG

SEQ ID NO:	Orden	Nombre	Secuencia
SEQ ID NO:40	8	Sufijo Biobrick (BBa_G00001)	Tactagtagcgccgctgcag
SEQ ID NO:41	9	Parada de traducción (BBa_B0042)	ttagtttagttag
SEQ ID NO:42	10	Terminador (BBa_B0054)	attagcagaaagtcaaaagcctccgaccggaggcttttgactaaaactccctgggggtatcattggg
SEQ ID NO:43	11	Secuencia cebador VR (BBa_G00102)	gctcactcaaggcggtaat
SEQ ID NO:44	12	Terminador (BBa_B0062)	cagataaaaaaaaaatccttagctttcgctaaggatgatttct

SEQ ID NO:	Orden	Nombre	Secuencia
SEQ ID NO:45	13	Origen pSa	gctagagctgtcagaccagagctcacaacacgttttatctataccgctacagggtatttaattctatit atctgcgctagaatgagcatgtttaaccgaatcagctgttccctgcaaatgccaaaatactatgcctat ctccgggtttcgcgtgacggccaagaccggaaaacaaaatacgggttgcctgaatacgcgaacgc caaaggcttgcgcgcctacagatcgaaggaatgtccagcagagcaaaaggactggcgcaagcg caagctcggagcaatcatcgaaaaggccgagcgtgacgactgactgacgccggagattacgcgc attgccgttccgcctcgcgccttggaaattctcaaaaggcgagcgcgcggcctaactgtccat gtgaccaacagaagatcatcatggagggcagcctacaagcgacatcagcgaacggctgtggctt ggctgcacagatcagcggcattcattcaggcagctaccaccaggcgctacaagtcgccagagag cggcgcaagagcgtcggcgaccacaaggcgagcgaatcagcgccttgaagctggagacggcgtatt gatgaagtacaggcatacgtgaaccttggcttgcgcgaagtcgcgcagccggagtgttagcgctcag ccctcacaccttgcgcctgttcatcaaacgcgggaacatcaaacacaaaacacagacacacacac catgccgggggggaacacatgccctagaacacaaagcccccggcctatcgaacagagatcat caagacgagcctcgcgtcgaatggagatgcccggaagcgtgcttagtgcgtacatgtgccgtt gccttgcgaagcaccatgcccacaccgaccccaagaccagctactttagcgcgaacacagtgacac gtcaccttgcgtatcagggcaagcggagcctgcgccttgcctatgctatgcgcgcaccttgcctt cttggatagtcaccgagcggcggaacgaagaccgcgttgaaccttggcggctgcgaatcggaa tttctacaagggctcggnaatgcacacgagggcgttacagggccaccccttgcgaatcagggcgcaacg cctgtttcatcctatgttgccttgcgcgcgagcaagggcaatgacttgcgcatggagaacgtgcattgc caagcgcgttcttctatctggaatcccaagcggccagaagatcggcgctatgggtagcagccctcac cctcacaggcgatttcttgcaggaagtcacccgctacccgttctatcgaatcgactacgtcatgcct tgcggcagctcctcgttgcgagtgacatttacagtggtgctgacctatcgcgtgttcttgcgggcaaa ggggcgcccttgcgtcaaatcccttgggtcggcctcgaagcgcaatcggctatcctatggcgagccg cgcacgcaactcggcgaactggacgataagggcgagagcggcgagagcggcgagcagcactcgcca gttcaaatataacttcaaaaagcgctacgcgaagtgttattgtctatcccgaggcgcaagcagctcat cgaagatgacggcgaatgcctgcgcataaaccacagcctgcatgtcaccgcgcaccggcgcaag ggcgctcgcacggcccccctcgcactgacagggcgaacgctacgcttggcttggcgaagccttcca tcacaagcccgctcgcgcggcgttttlatcccggaagccttgggtagagggtagtattccagct gaaacgctaatgcgcgcgaagccttgcgtacggggcgttccggccgctccaaaactacacagct gaaatcgttaatcagggtacgtgaatcgcatacggagtagctgaatcgtataaggtcacgtgaa atcgttaataaaaaggcagctgagaacgctaatagccctttagatcaacagcttgaanaacccctc gctcggcaagtagttacagcaagtagtatgttcaattagctttcaattatgaatatataatattatgtt cgcccttggcttggacaaatgcgtacgcgcacgggctcggccgggggacaaccgaagcggttgc ccaccgtcgaagcggcgccttgcgcacacccggcgggcgcaacagatcgtttataaatttttttt tgaanaagaaaagcccgaagggcggaacctctcggccttctgatttccgatcaacgcagagatgc ttcggaagtagctgttccagaattatagggcgagagacaccagattccagatggctctgttaatttgtt gtatgtatgtatcatataacatactacagtagcagggcccgcaagaatggcaatcacataaagac atttggcgagcggcgaactggagcggcgaagcgtatcgcccaacttggcgccgtgfcgcaaga aatcgggaagcggtagcttcaaacatttccgatgcaatggcgtgaatgaaaaacggcaagaccgcc acccttccctcatagaccatttccgggttcagctaacgagcatcttgcgagcttggcaatgcgctat gggctatcgccttggcgacgccaacggcggttgcgaagatcggaaacagatcgaggccgaca aaaggccatcagccagcagcttgcgaagcaatcgaactagccgacaccttaccgcggaacgga ccagctcggcggaacgagtgaaicagctgaacctatggaacgcgagcgcgaagcgtggcgacca gcttccggaagtgaacggcgagcggcggaagaaactaaaccgctgcatggaaaagctacccaacg cgataacgaggtatcaggcccgcaaacggcgaagggccatcagagcgcgcgcgcagcttgcga aggctcaggtgggaagccctcaagagcaggtcggcaatctcacagccgttcttgaanaacagggcgaaa caatgaaaagcggccttgcgccttgcgcggctcggccgtctgtccttaactgtcaggttgcgcg cttggccgacttccgggttgaagctgtccgaatcttgcgggtgacactatcagcgttgggtgaacgg tcagacatcgcgtgagattggcgatattgacgcacgggaagcggcgaagccttgggtcctccgtg ctcggcaacggctcggcagcttgaacttgcgaagaggttcaagtgaccgaaaaagaggttgaatcgtt atggcgcaacttgggggtgttacgcgcgttgcgaatcccgccggcgcaaacacacacacacacac atcaatgcatcatgtttcaagaagcgtggttgcgttaccgtttatcagcgaacacacacacacac cagatgtacgagtagtaaaaagggcccgcccaacggcgcggccttgggtcagaccggaatgtctca ggagccttggaaatggctcgcgcctcgaanaatgccagaaatgacacggcgacggccgttgcg acgcgcgcagggcagctgaattaccgcgggagc

SEQ ID NO:	Orden	Nombre	Secuencia
SEQ ID NO:46	14	resistencia a espectinomicina	cccctcgtcccgacacttcagatcgccatagcgcacagcgcctcagcgggtggaacggcgagtg ggcgttttcatgcttggtatgactgttttttgggtacagctctatgcctcggcatccaagcagcaagcg cgttacgccgtgggtcgaatgttgatgttatggagcagcaacgaatgttacgcagcagggcagtcgccct aaaacaaagttaaactatgagggaagcgggtgacgccgaagtatcgaactcaactatcagaggtagti ggctgtatcagcgcctcgaaccgacgttgctggccgtacattgtacggctcgcagtggaaggc ggctgaagccacacagtgatattgattgtgtgtacgggtgacggtaaggcttgatgaacaacgcgg cgagcttgatcaacgacctttggaaactcggcttccctggagagagcggagattctccgcgtgtag aagtcacattgtgtgcacgacacatcctcgtggcgttatccagctaaagcgcgaactgcaatttga gaatggcagcgcgaatgacattcttcaggtatcttcgagccagccacgacgacattgacatctggctatct gctgacaaaagcaagagaacatagcgttgccttggtaggtcagcggcggagggaactcttgatccgg ttctgaacaggtatctatttggcgcctaaatgaaaccttaacgctatggaactcggccgcgactgggc tggcgatgagcgaatgtatgtcttacgttgcctccgcatlttggtacagcgcagtaaccggcaaatcgc gccgaagatgtcgttccgactgggcaatggagcgccttccggccagatcagccgtcatacttg aagctagacaggttattcttggacaagaagaagatcgttggcctcgcgcagatcagttggagaat ttgtccactacgtgaaagcgcgagatccaaggtagtcggcaataatgtctaaacttgcgtcaagccg acgccgttcgcggcgcggttaactcaagcgttagatgcactaagcacataattgctcagcgaac tatcaggtcaagctcgttttattatttttaagcgtgcataataagccctacacaaatggtaccc
SEQ ID NO:47	15	Terminador (BBa_B0053)	tcggcgaacaaacgggcaagggtgcaccaccctgccctttttctttaaaccgaaaagattacttcggtt
SEQ ID NO:48	16	Secuencia cebador VF2 (BBa_G000101)	tggcacctgacgtctaagaa
SEQ ID NO:49	17	Terminador (BBa_B0055)	aagggaattatcagaatttgcctgcccgaagaaaggccaccctggaagggtgagccagtgagttgattgctacgttaa
SEQ ID NO:50	18	Parada de traducción (BBa_B0042)	ttagttagttag
SEQ ID NO:51	19	Prefijo Biobrick x (Bba_G00000)	gaattcgcggccgcttctagag

Fig. 20. Lista de partes del complejo sintético completo

Orden N.º 57-67 tiene el esqueleto pSB4C5

[illegible]

				catggatagacctgaataaccagcgtggaatgaactgaccgcaccgtggctgaatcgcataa
SEQ ID NO:58	7	aislador 2		CAAACACCCCATGTCGATACTGAACGAATCGACGCACACTCC CTTCCTTG
SEQ ID NO:59	8	rbs nifK	50	CAATCTCATACTCTCAAAAATTAGGCGAGGTAAC
SEQ ID NO:60	9	nifK		atgtctcaactatcgataaaataaactctgttacccgcgttcgagcaggacgaatacaggaaactgtt ccgtaacaaacgtcagctggagaagcgacagcgcacagcgcgtgcaggaaggttcgcatggac caccaccgcggaatacgaagctcgaactccgtcgcgaagccctgacgggtgacccggcgaagcg tgccagccctcgggtgcgggttcgtgcagccctgggtttgccaaacccctccgtatgtccacgggtccc agggtcgtgtagcctacttccgtacctatttcaaccgccactttaaagaaccaatcgctgctgctcga cagcatgacggaggacgcggcagtttccgggtgaacaacaacatgaacctgggctgcgaataatgctt ccgcactgtacaacccggaataatcgcagctgtaccactgcatggcagaggttatgggtgatgact gcaagcatttattgcccaacgcaaagaaagacgggttcgttgacagctctatcgggtccgcacgctcat accgcgtcttcatcgggttcacgttaactgggtgggacaacatgttcgaaggtctgcgaataactttac cgacagactatcaaggccaaccgggttaactgcggaagcgaacctgggtgacgggttgaacactacc tgggcaacttcgtgctcgaagcgcatgagtgagcagatggggtccgttctcgtgctgctgaccc gltcagggttctggacactccagcggacggccactatcgcatgtatctgggtggcaccactcagcagga aatgaaaggggcccagacgcgaatgacacctgctgctgcaaccgtggcagctgctgaaagcaag aaaagtgttcaggaaatgtgggaacagccggcaacgggaaggttgaatcccgctgggttctggcagctac tgacgaactcgtgatgaccgtgtccaaactgagcgggcaaccaatcgcggatgctcgaacctgggaac gggtcgccttgggtggacatgagctgacagccacacgtggctgcatggcgaagaaattggcctgtac gggtgacccggacttctaaggccctgacccgttccgtcgtgaactgggtcgtgagccgactgttate ctgtctcacaacgctaaacaacgttggcagaaggccatgaacaaatgcggatgcgagcccatagcgg ccgtgatagcgaagtgttcatcaactgcgacctgtggcatttccgtctcgtatgttaccgctcagccgg attcatgatcggtaactcttacggcaattatcagcgtgacactcggcgaaggcgaagcggttga agtgcccgctgaticgtcgggttccgtgtgacgggtaccactcgaaccgcaagaccacccgggg ttacgaaggcgcatgaacatcgtactactcgtgaacgcagctaggaaaagctggacagcgata cttccacgtgggcaaacggactattcttctgatcgtgtcgttaa
SEQ ID NO:61	10	aislador 3		CCTGATTGTATCCGCATCTGATGCTACCGTGGTTGAGTTA
SEQ ID NO:62	11	rbs nifY	240	CCATACTCACTCCCGGAGGTACTTCT
SEQ ID NO:63	12	nifY		ATGTCGACAAATGATACCTGTTTTGGCGCATGCTGGCGCTGT TTCAGTCGCTGCCGATTGTCAGCCGGCTCAAATCGTCGATTG GCTGGCGCAGGAATCCGGCGAAACCCCTGACGCGGAGCGCCT TGCCACCCTGACCCAACCGAACTCGCGCGCTCGTTCCCATCC GCGACGGCAGTGATGAGCCCGGCTCGCTGGAGCCGCGTTATG GCTTCTCTGCAAGGCGCCCTCCAGCCCACTTGCGCATCGTAC GTCCGGCGCAGCGTACCCCGCAACTGCTCGCCGCGTTTTGCAG CCAAGACGGCCTTGTATCAATGGTCAATTCGGCCAGGGTCGT CTGTTCTTCAATTTACGCCCTTGACGAGCAGGGCGGCTGGCTGT ATGACTTGCGCCGCTATCCGAGCGCACCGCACAGCAGGAAG CGAATGAGGTGCGTGCTCGTCTGATTGAAGATTGCCAGCTGCT GTTCTGCCAGGAGATTGGCGGTCCGGCAGCAGCGCTCCGAT CCGCCACCGCATCCATCCGATGAAGGCGCAGCCGGGTACTAC GATTACGGCGCAGTGTGAAGCTATCAACACCCTGCTGGCCGG TCGCTGCGCCGCTGGCTCGCCAAACGTTTGAACCGTGATAAC CCGCTGGAAGAGCGTGTGTTTTAA
SEQ ID NO:64	13	aislador 4		CATTTTTCCTTGGCAGACAGACCTCCTACTTAGATTGCCAC
SEQ ID NO:65	14	rbs nifE	40	ACTATTCAATACATCACTGGAGGTTATTACAA

SEQ ID NO:66	15	nifE		ATGAAGGGTAACGAGATTCTTGCTCTGCTGGACGAACCGGCC TGTGAACACAACCATAAACAGAAATCCGGCTGTAGCGCCCCA AAGCCGGGTGCGACGGCGGCTGGCTGCGCTTTCGATGGTGCG CAGATCACCTGCTCCCGATTGCGGACGTTGCCACCTCGTGC ATGGCCCAATCGGTTGCGCAGGTAGCTCTTGGGACAACCGTG GCAGCGCTCCAGCGGTCCGACCCTGAATCGTTTGGGCTTTAC CACTGACTTGAATGAACAAGATGTGATCATGGGTGCGGGCGA GCGTCGCCTGTTCCACGCTGTGCGCCATATTGTCACCCGTTAC CACCCAGCGGCAGTATTCTACAATACGTGCGTGCGCGGCTA TGGAAGGCGATGACCTGGAGGCGGTGTGTCAGGCAGCCAGA CTGCGACCGGCTCCCGTAATCGCAATTGATGCGGCTGGCTT CTACGGTTCGAAGAACCCTGGGCAACCGTCCGGCAGGCGATGT CATGGTTAAACGCGTCATTGGCCAACGTGAGCCAGCGCGTG GCCGGAGAGCACCTGTTTGGCCCGGAGCAACGTATGACAT TGGCTTGATCGGTGAGTTCAACATTGCGGGCGAGTTTGGCAC ATTACGCGCTGCTTGATGAGCTGGGTATCCGCGTTTGGGTT CGCTCAGCGGCGATGGTCGTTTGGCCGAGATTCAAACATGCA CCGTGCCAGGCGAATATGCTGGTGTGACGCGGTGCTCTGATC AATGTTGCGGTGCTCTGGAACAGCGCTATGGCACCCCGTGT TTGAAGGCTCGTTCTATGGTATCCGCGCGACAGCGACGCCCT GCGCCAGTTAGCGGCGCTGCTGGGCGATGACGACCTCCGTCA GCGCACCGAGGCGCTGATCGCGCGTGAAGAACAAGCGGCTGA GCTGGCCCTGCAACCGTGGCGTGAACAGCTGCGTGGCCGCAA GGCCCTGCTCTACACGGGTGGTGTCAAAAGCTGGTCTGTGGTG TCCGCGCTTCAGGATCTGGGTATGACCGTGGTTGCCACGGGCA CGCGTAAGAGCACGGAAGAGGATAAACAGCGCATCCGCGAAT TGATGGGCGAAGAGGCGGTGATGCTTGAAGAAGGCAACGCAC GTACCTTATTGGATGTAGTTTATCGCTATCAAGCAGACCTGAT GATTGCGGTTGGCCGCAACATGTATACCGCTACAAGCGCG CTTGCCGTTCTGGACATCAACCAGGAACGCGAGCAGCGGTT GCGGGTACCAAGGCATCGTGACCTTAGCGCGCCAGCTGTGC CAAACGATTAAACAGCCCGATCTGGCCGACAGACTATTCCCGC GCACCGTGGCGCTAA
SEQ ID NO:67	16	aislador 5		TGTCACGCTAGGAGGCAATTCTATAAGAATGCACACTGCA
SEQ ID NO:68	17	rbs infN	40	CCTAAACCTACCACACCTGGAAGAAGTAATT
SEQ ID NO:69	18	nifN		ATGGCAGACATTTCCGCACTGATAAGCCGTTGGCTGTGTGCG CGATCAAGACCGGCCAGCCGCTGGGTGCGATCCTGGCGTCCC TGGGTATCGAGCACTCGATTCCGCTGGTACATGGCGCGCAGG GCTGTTGCGGCTTTTGCCAAGGTTTTCTTTATCCAGCACTCCAC GATCCGGTCCCGCTGCAAAAGCACGGCAATGGACCCGACCAGC ACCATCATGGGCGCTGATGGTAACATCTTACCGCGCTGGACA CTCTCTGCCAACGCAATAACCCGCAAGCAATTGTGCTGCTGAG CACCGGCTCTCCGAGGCGCAGGGCAGCGACATTTCCCGTGT AGTGCGTCAGTTCCGTGAAGAATATCCGCGTCATAAAGGCGT GGCGATTCTGACTGTAAACACCCCGGACTTTTACGGTAGCATG GAGAACGGCTTTTCCGCTGTCTGGAGTCTGTGATTGAACAGT GGGTTCCGCGCAGCCACGTCGCGCGCAGCGCAATCGTCGCG TCAATCTTTTGGTGAGCCATCTCTGTAGCCCAGGCGATATTGA GTGGCTGCGCGCTTTCGCTCGAGGCTTTCGGTCTGCAACCGATC ATTCTGCCGATCTGGCTCAGAGCATGGACGGCCACCTTGCTC AGGGTGACTTTTTCGCGCTGACGCAAGGGCGCACGCGGTTGC GCCAAATCGAGCAGATGGGCCAGAGCCTTTGCTCTTTTGCAT TGGCGTCAGCCTGCACCGTGCGAGCAGCTGCTGGCTCCGCGT TGTCGTGGCGAAGTATCGCCTTGCCGACCTCATGACCTTGG AACGCTGCGACGCTTTATCCATCAGTTGGCGAAAATCAGCG

				GTCGCGCCGTTCCGGAGTGGCTGGAACGCCAGCGCGGTACAGC TGCAAGACGCCATGATCGATTGCCACATGTGGCTGCAAGGCC AGCGCATGGCGATTGCCGCCGAAGGCGACCTGCTGGCAGCGT GGTGCATTTCGCGAACTCTCAAGGTATGCAGCCGGGTCCACT GGTTGCTCCGACGGGTATCCGAGCCTGCGTCAGTTGCCGGTG GAGCGCGTGGTGCCGGGTGATCTGGAGGATCTTCAGACCTCT TATGCGCACATCCGGCCGACTTACTGGTGGCGAACTCCACGC CCGTGATTTAGCAGAGCAATTCGCCCTGCCGCTGCTGCGCGCA GGCTTCCCGCTGTTTGACAACTGGGCGAATTTCTGCTGTTTC GCCAGGGTTATAGCGGTATGCGTGATACCTGTTTCGAGTTGGC GAACCTGATCCGTGAACGCCATCATCATCTGGCTCATTATCGC AGCCCGCTGCGCCAGAACCAGAATCTCTGTTGTCTACGGGTG GCGGTACGCAGCGGATTAA
SEQ ID NO:70	19	rbs nifJ	100	ctagagattaagaggagaaaattaagc
SEQ ID NO:71	20	niJ		ATGAAAACATATGGACGGTAACGCTGCGGCTGCATGGATTAGC TACGCCTTTACCGAAGTGGCTGCGATCTACCCGATTACGCCGA GCACCCGATGGCGGAAAAATGTGGACGAATGGCTGCGCAGG GCAAGAAGAACCCTTTCGCCAGCCGGTGCGCTGATGGAGA TGCAGTCGGAAGCGGGTGCAGCAGGTGCTGTGCATGGCGCCT TGCAAGCTGGCGCACTGACGACCACCTACACCGCGTGCAGG GCCTGTTGCTGATGATCCCAAACATGTACAAAATCGCGGGTG AACTGCTGCCGGGTGTCCTTCATGTTTCGGCACGCGCACTGGC CACCAATAGCCTCAACATCTTTGGCGATCATCAGGATGTAATG GCGGTGCGCCAAACGGGCTGCGCGATGTTGGCCGAGAATAAC GTCCAGCAAGTTATGGATTGTCCGCGGTAGCCCACTTGGCAG CGATCAAAGGTGCGATTCCGTTCTGTGAACCTTCTTCGATGGCTT TCGCACGAGCCACGAAATCCAGAAGATCGAGGTTCTGGAATA TGAACAGCTGGCCACCTTGTGGATCGTCCGGCCCTGGACAGC TTCCGCCGTAACGCCCTTACCCGGACACCCGGTCATCCGTG GCACCGCCAGAAACCCGACATCTACTTCCAGGAACGTGAGG CCGGTAACCGTTTCTATCAGGCGCTCCCGGATATTGTGGAATC TTACATGACCCAGATTCTGCCCTGACTGGTCGCGAGTATCAC CTGTTTAACCTACACTGGTGCTGCGGATGCGGAGCGCGTGATCA TCGCGATGGGCTCTGTCTGTGACACCGTCCAAGAGGTGGTTGA CACGCTGAATGCAGCGGGTGAGAAAGTTGGTCTGCTCTCCGTT CATCTTTTCCGCCCGTTTTCGTTAGCGCACTTCTTCGCCCAACT GCCGAAAACGTGTACAGCGTATCGCAGTATTGGACCGTACGAA AGAGCCAGGTGCTCAAGCAGAGCCGCTGTGCCTCGATGTGAA GAATGCCCTTTACCACCATGACGATGCCCGTTGATTGTGGGT GGTCGTATGCTTGGGCGGTAAGGACGTGTTGCCGAACGAT ATTGCGGCCGTGTTTGATAACCTGAACAAACCGCTGCCGATGG ACGGCTTACGCTGGGTATCGTGGACGATGTTACCTTACCTC TCTCCGCGCAGCGCAGCAGACCTGGCGGTTTCTCACGACGGC ATCACGGCATGTAAGTTTTGGGGCATGGGCTCCGACGGCACG GTTGGTGCGAACAAGTCCGCGATCAAGATTATCGGCGACAAA ACGCCACTGTATGCGCAAGCGTACTTTTCTACGACTCGAAGA AGAGCGGTGGTATTACCGTCAGCCATCTGCGTTTGGTGATCG CCCGATCAACTCCCGTATTTGATCCATCGCGCGGATTTTCATC TCGTGCAGCCAGCAAAGCTATGTTGAACGCTACGATCTGCTGG ATGGCCTTAAACCGGGTGGCACCTTCTGCTGAACTGCTCCTG GAGCGATGCCGAACTGGAGCAACATCTGCCGGTCCGGTTTCAA ACGTTATCTGGCACGCGAGAATATCCACTTCTACACTCTCAAC GCTGTGGACATCGCCCGTGAGCTTGGTTTGGGTGGCCGTTTCA ACATGCTGATGCAGGCTGCCTTCTTCAAACCTGGCCGCGATCAT TGACCCGACAGACTGCTGCGGACTATCTGAAGCAGGCTGTTGA GAAAAGCTATGGCAGCAAAGGTGCGGCGGTATCGAGATGAA CCAGCGTGCCATCGAGCTTGGCATGGCCAGCCTGCACCAAGT

				GACGATCCCGGCACATTGGGCCACCCTGGATGAGCCAGCGGC GCAGGCGTCCGCGATGATGCCGGACTTTATCCGCGACATCTG CAACCGATGAACCGTCAGTGCGGGACCAGCTTCCGGTGTCG GCTTTTGTCCGCATGGAAGATGGCACCTTCCCGTCCGGCACGG CCGCATGGGAGAAACGTGGCATCGCCCTTGAGGTGCCAGTCT GGCAGCCGGAAGGCTGCACGCAGTGC AACCAGTGCGCCTTCA TTTGTCCGCACGCCGCGATTCTGCCGGCGTTGTTGAATGGCGA AGAGCATGATGCTGCCCCGTTGGCCTGCTGAGCAAACCGGC ACAAGGCGCTAAAGAATATCACTATCATCTGGCGATTAGCCC GCTGGACTGCTCCGGCTGTGGCAACTGCGTTGACATTTGTCCA GCTCGTGGCAAAGCGTTGAAGATGCAGTCTCTGGATAGCCAA CGCCAGATGGCTCCGGTGTGGGATTATGCGCTGGCGCTGACCC CGAAGTCTAACCCTTTTCGTAAAACACCGTCAAAGGCTCGC AGTTCGAAACCCCGCTGCTGGAGTTAGCGGTGCGTGCGCTGG TTGTGGCGAAACGCCGATGCGCGCCTCATTACCCAGCTGTTT GGCGACCGCATGCTGATTGCCAATGCCACCGCTGTTCCAGCA TCTGGGCGCATCTGCGCCGAGCATCCCGTATACCAATCA TCGTGGTCAATGGTCCGGCTGGGCGAATAGCCTGTTTGAGGAC AATGCCGAATTTGGTTTAGGTATGATGCTGGGCGGTCAAGCTG TGCGTCAACAGATCGCGGACGATATGACGGCTGCGTTAGCGC TCCCGGTTTCCGATGAGCTGAGCGACGCGATGCGCCAGTGGTT GGCGAAACAGGACGAGGGTGAAAGGCACGCGTGAGCGTGCGG ACCGTCTGAGCGAGCGCTTAGCCGCGGAGAAAGAGGGCGTTC CGCTGTTAGAGCAGCTGTGGCAAAATCGTGATTACTTTGTGCG TCGCAGCCAGTGGATTTTCGGCGGTGACGGCTGGGCCTATGAT ATTGGCTTCGGTGGCTGGACACGTCTCGCCAGCGGTGAGG ATGTGAACATTCTGGTATTTGACACCGAAGTCTACTCGAACAC CGGCGGTCAAAGCAGCAAAATCGACCCCGGTGCGGGCCATCGC CAAGTTCGCGGCTCAGGGCAAGCGCACCCGCAAGAAAGACCT GGGTATGATGGCGATGAGCTACGGCAACGTCTATGTAGCCCA GGTGGCGATGGGTGCGGATAAAGATCAAACTCTGCGCGCCAT TGCGGAAAGCTGAAAGCGTGGCCAGGCCGCTCGCTGGTGATTGC GTATGCGGCCTGCATCAATCATGGCTGAAGGCCGGTATGCGT TGCAGCCAACGTGAGGCGAAGCGCGCTGTTGAGGCGGGCTAC TGGCACCTGTGGCGTTATCACCCGACGCGGAAGCGGAAGGC AAGACGCCGTTTATGTTAGATAGCGAAGAACCGGAAGAGTCG TTCCGTGACTTTCTGTTGGGTGAGGTGCGCTACGCATCCCTGC ACAAGACCACCCGCACCTCGCCGATGCCCTTTTCAGCCGTAC CGAAGAAGATGCGCGTGCGCGCTTTGCGCAATACCGTGCCT GGCTGGCGAAGAGTAATAA
SEQ ID NO:72	21	terminador T7 25	1600	TACTCTAACCCCATCGGCCGTCTTAGGGGTTTTTTGT
SEQ ID NO:73	22	aislador 6		CCGTGTTGAGTCAGCGTCGAGCACGCGGC
SEQ ID NO:74	23	terminador T7 mut 2	1500	TAATACGACTACTAGAGAGAGA
SEQ ID NO:75	24	aislador 7		CGCGACTTCCAGAGAAGAAGACTACTGACTTGAGCGTTCC
SEQ ID NO:76	25	rbs nifB	220	CTCTCTGTAATACATCAAAATCAATCATAGGAGGGCTAAA
SEQ ID NO:77	26	nifB		ATGACCTCTTGTTCGTCGTTTCTGGCGGTAAAGCGTGCCGTC CGGCCGATGACTCCGCGCTGACTCCGCTGGTGGCCGACAAGG CAGCTGCGCACCCGTGCTATAGCCGCCACGGCCATCACCGCTT CGCGCGTATGCACCTGCCAGTCGCTCCGGCCTGCAACTTACAA TGCAACTACTGCAACCGCAAGTTCGATTGCAGCAATGAAAGC

				CGTCCGGGCGTGTCTTACCCTGCTGACGCCGGAACAGGCTG TGGTGAAGGTGCCAGGTCGCCAAGCTATCCCGAGCTGteg GTGGTCGGTATTGCTGGTCCGGGCGATCCGCTTGCGAATATCG CCCGCACCTTCCGTACCTTGGAGCTTATTCGCGAACAGTTGCC GGACCTGAAACTGTGCTGAGCACCAACGGCTTGGTGTGCC AGATGCCGTTGATCGTCTGCTCGATGTGGGCGTGGATCACGTT ACCGTCACCATTAACACCCTGGACGCAGAAATCGCAGCGCAA ATCTACGCGTGGTTGTGGCTGGATGGCGAACGCTACTCCGGTC GCGAAGCCGGCGAAATTCTCATTGCCCGCCAGCTGGAAGGCG TACGTCGCCTGACCGCGAAAGGTGTGCTCGTCAAGATCAACA GCGTATTGATTCCGGGCATCAATGACAGCGGCATGGCGGGTG TTAGCCGTGCGCTGCGCGCGTCTGGTGCCTTCATCCACAACAT CATGCCACTGATTGCGCGTCCGGAGCATGGCACTGTTTTCGGT CTGAACGGCCAGCCGGAACCGGACCGCGAAACCTGGCGGGC ACGCGCTCCCGCTGCGGCGAGGTTATGCCACAAATGACCCAC TGCCACCAAGTCCCGTGGCGACGCGATTGGCATGCTTGGTGAG GATCGCTCGCAACAGTTTACGCAATTACCGGCTCCGGAGTCCC TCCCGGCTGGCTGCCGATCCTGCATCAGCGTGTCTAGTTGCA TGCGAGCATCGCCACGCGCGGTGAGAGCGAAGCCGATGACGC CTGCCCTGGTGGCGTTGCGTCGAGCCGTGGCGATGTAATTGAC TGCCATTTCCGCCATGCCGACCGTTTCTATATCTATAGCCTGTC TGCGGCTGGTATGGTTCTGGTTAACGAACGTTTCACCCCGAAA TACTGCCAGGGTCGCGATGACTGCGAGCCGAGGACAATGCC GCACGCTTTGCTGCCATCCTTGAGTTGCTGGCGGACGTCAAAG CGGTGTTTTGTGTGCGTATCGGCCATACCCCGTGGCAACAGCT GGAGCAGGAAGGCATCGAACCCTGCGTGGATGGCGCCTGGCG TCCGGTATCCGAGGTCTTGGCGGCATGGTGGCAGCAGCGCCG TGGTAGCTGGCCGGCTGCATTGCCGCACAAAGGCGTTGCGTA A
SEQ ID NO:78	27	aislador 8		ACTACGAGATTTGAGGTAACCAATAAGCACGTTAGTGGC
SEQ ID NO:79	28	rbs nifQ	5	TGCGTTTAGCAGTTCCCGTAGGAATATTTCTTT
SEQ ID NO:80	29	nifQ		ATGCCGCCATTGGACTGGTTGCGTCGTTTGTGGTTACTCTATC ACGCCGGCAAAGGCAGCTTTCGCTTCGTATGGGCTTGTGCC GCGTGACTGGCAAGCTCTGCGCCCTGCGCTGGGCGAGGTGGA AACGCCGCTGGATGGCGAAACCTGACCCGTGCGCGTCTGAT GGCGGAGCTGAATGCGACCCGCGAAGAAGAAGCAGCAGAGCT GGGTGCCTGGCTGGCCGTTGGATGCAACAGGATGCCGGTCC GATGGCGCAGATTATCGCAGAGGTGAGCCTGGCGTTCAACCA TCTCTGGCAGGACCTTGGCCTGCGAGCCGCGCTGAACGCGT CTGCTGATGTCTGACTGCTTCCCGCAGCTGGTTGTTATGAACG AGCACAAACATGCGCTGGAAGAAATCTTTTACCGCCAGCGTTG CCTGCTGCAACAGGGCGAAGTCATCTGTCGACGCCGCTCTTGC GATGAATGCTGGGAACGTTCTGCGTGCTTTGAGTAA
SEQ ID NO:81	30	terminador T7 1	1300	TACATATCGGGGGGGTAGGGGTTTTTTGT
SEQ ID NO:82	31	aislador 9		GTCTGTAGCACGTGCATC
SEQ ID NO:83	32	promotor T7 mut 3	2750	TAATACGACTCACTAATGGGAGA
SEQ ID NO:84	33	rbs nifF	45	GACAAGAGTCTCAATTATAAGGAGGCTTTACTAC

SEQ ID NO:85	34	nifF		ATGGCGAACATCGGCATCTTCTTTGGTACGGATACCGGCAAA ACCCGCAAGATTGCGAAGATGATTCACAAACAGCTGGGCGAG CTGGCCGATGCCCCGGTTAACATCAATCGTACCACTTTGGATG ACTTTATGGCTTACCCAGTCTGTTGCTCGGCACGCCGACGCT TGGTGATGGTCAACTGCCGGGCTTAGAGGCGGGCTGCGAGAG CGAAAGCTGGTCTGAGTTTATCTCCGGTCTGGATGACGCTTCC CTGAAGGGCAAAACCGTGGCGCTGTTTGGCTGGGCGACCAG CGTGGTTACCCGGACAACCTTCGTGTCGGGTATGCGTCCGCTGT TCGACGCGCTGAGCGCCGTTGGCGCCAGATGATTGGTAGCT GGCCGAACGAAGGTTATGAGTTTACGCGCATCGTCCGCGCTGG AAGGCGACCGCTTCGTGCGCTTGGTGCTGGATCAAGACAATC AGTTCGACCAGACCGAAGCGCGCTTGGCGTCTTGGCTTGAAG AGATCAACGCACCGTTCTGTAATAA
SEQ ID NO:86	35	terminador T7 1	1300	TACATATCGGGGGGTAGGGGTTTTTTGT
SEQ ID NO:87	36	aislador 10		GGTCATTACAACGGTTAT
SEQ ID NO:88	37	promotor T7 mut 2	1500	TAATACGACTCACTAGAGAGAGA
SEQ ID NO:89	38	aislador 11		AACATAGCGTTCCATGAGGGCTAGAATTACCTACCGGCCT
SEQ ID NO:90	39	rbs nifU	2800	CAGATACTGACAAATAAACCAGCGAAGGAGGTTCTTA
SEQ ID NO:91	40	nifU		ATGTGGAACTACAGCGAGAAAGTCAAGGACCATTCTTCAAT CCGCGCAACGCGCGTGTGTGGATAACGCAATGCGGTGGGC GACGTGCGCAGCTTATCTGTGGCGATGCTCTCCGCTTGATGC TGCGCGTGGACCCGAGAGCGAAATCATCGAAGAAGCGGGCT TTCAGACCTTCGGCTGCGGCAGCGCGATTGCGTCCAGCGC ACTGACGGAGCTGATCATCGGTACACCCCTGGCGGAAGCGGG TCAGATCACCAACCAGCAGATCGCCGACTATCTGGACGGCTT ACCGCGGAAAAGATGCACTGCTCTGTAATGGGCCAGGAAGC TCTTCGTGCGGCCATTGCTAACTTTTCGCGGTGAATCGCTGGAA GAGGAGCATGACGAGGGTAAGCTGATCTGCAAGTGCTTCGGC GTCGATGAAGGCCATATTCGCCGTGCTGTCCAGAACACGGT CTTACGACTCTGGCCGAGGTGATCAATTACACCAAGGCAGGT GGCGGTTGTACCAGCTGCCATGAGAAAATCGAGCTGGCCCTG GCCGAGATTCTCGCCCAACAGCCGCAAAACACCCCGGCAGTT GCGTCCGGTAAAGATCCGCACTGGCAGAGCGTCGTGGATACC ATCGCTGAACCTGCGTCCACATATCCAAGCGGACGGTGGTGAC ATGGCGCTGTTGTCCGTGACGAACCACCAAGTGACTGTTTCGC TGTCGGGCAGCTGTTCTGGCTGCATGATGACCGACATGACCT GGCGTGGCTGCAACAGAAATTGATGGAGCGTACCGGCTGCTA TATGGAAGTTGTTGCCGCTAA
SEQ ID NO:92	41	aislador 12		CATTGTAATAGCCACCAAAAGAGTGATGATAGTCATGGGT
SEQ ID NO:93	42	rbs nifS	175	GATACCCGTAGACCATTCTGAAATCGAAGGAGGTTTTCC
SEQ ID NO:94	43	nifS		ATGAAACAAGTGTACCTGGACAACAACGCGACCACCCGCTG GACCCGATGGTTCTGGAAGCGATGATGCCGTTTCTACGGATT TCTATGGCAATCCGTCCAGCATCCATGACTTCGCGCATCCCGGC ACAAGCGGCGCTGGAACGTGCGCACCAGCAAGCTGCGGCACT GCTGGGCGCAGAGTACCCGTCTGAAATCATTTTACAGAGCTGT GCGACCGAGGCCACTGCAACCGCCATTGCGTCGGCCATCGCG

				TTATTGCCGGAACGCCGCGAAATCATCACCTCGGTAGTGGAG CACCCGGCTACGCTGGCGGCGTGCAGACACCTGGAACGCCAA GGCTATCGCATCCATCGCATTGCGGTGGATAGCGAAGGTGCG CTGGACATGGCCAGTTCCGTGCAGCGCTC _{leg} CCGCGTGTGCG GTTGGTGAGCGTGATGTGGGCCAACAACGAAACCGGCGTGCT GTTCCCGATTGGCGAAATGGCCGAGCTTGCCACGAGCAGGG CGCTCTGTTCCACTGCGATGCGGTTACGGTCGTTGGCAAAATC CCAATTGCTGTTGGCCAGA _g CGCGCATCGACATGCTGTCTTGCT CCGCGCACAAGTTTCATGGTCCGAAGGGTGTGGTTGCTTGTA CTTACGTGCTGGCACGCGCTTTCGTCCGCTGCTTCGCGGTGGC CATCAAGAATATGGTCGCGGTGCCGGCACTGAGAATATCTGT GGCATCGTCGGCATGGGCGCTGCGTGCGAACTGGCGAACATC CATCTGCCGGGTATGACCCATATTGGCCAGTTACGCAATCGCC TGGAGCACCGTCTGCTCGCCAGCGTGCCGTCCGTGATGGTTAT GGGCGGTGGTCAGCC _g GTGTACCGGGTACTGTCAACCTGGCG TTCGAGTTTATCGAAGGTGAAGCGATCCTGCTCTTGCTGAACC AGGCTGGCATTGCCGCAAGCTCCGGCTCCGCGTGTACCTCTGG CAGCTTGGAGCCGAGCCATGTGATGCGCGCCATGAACATTCC ATACACCGCGGCTCACGGCACCATTCGTTT _g AGCCTGAGCCGT TATACGCGCGAGAAAGAGATCGACTACGTGCTTGCGACCCCTC CCGCCAATCATTGATCGTCTGCGTGCCCTGTCCCGTATTGGC AGAATGGTAAGCCGCGTCCGGCAGATGCAGTCTTTACCCCGG TTTACGGTTAA
SEQ ID NO:95	44	aislador 13		GAGTTACTGGCCCTGATTCTCCGCTTCTAATACCGCACA
SEQ ID NO:96	45	rbs nifV	5	GCGACTAGGAGCCTAACTCGCCACAAGGAAACAT
SEQ ID NO:97	46	nifV		ATGGAGCGCGTCTTGATCAACGATACTACCTGCGTGATGGCG AACAATCTCCGGGCGTAGCGTTTCGTACCTCCGAGAAAAGTTGC CATCGCGGAGGCACTGTACGCTGCGGGTATCACCGCGATGGA AGTCGGCACTCCGGCGATGGGTGATGAAGAGATCGCCCGCAT TCAGCTGGTGCGTCTGCAACTGCCGGACGCGACGCTTATGACC TGGTGCCGTATGAACGCTCTGGAATCCGTGAGAGCGCGGAT CTGGGTATTGACTGGGTGGATATCTCGATCCCAGCATCCGACA AGCTGCGTCAGTACAAGCTGCGTGAGCCGCTGGCCCGTGTGCT GGAGCGCCTTGCGATGTTTATCCATCTGGCCACACGTTAGGC CTCAAAGTATGTATTGGTTGCGAGGATGCGAGCCGTGCGTCTG GTCAGACCCCTGCGCGCCATTGCCGAGGTGGCCAGCAATGCG CGGCTGCGCGCTTGCGTTACGCTGACACCGTGGGCTGCTGGA CCCGTTACACCCGAGCCAGATCAGCGCCCTGCGTGACGTT TGGTCGGGCGAGATCGAGATGCATGCTCACAATGATCTGGGC ATGGCTACCGCGAACACGCTGGCGGCAAGTTTCGGCTGGCGCC ACGTGCGGTGAACACTACCGTCTCGGTCTGGGTGAACGTGCA GGCAACGCAGCCCTGGAAACCGTTGCGCTGGGCCTGGAAACGC TGCTGGGCGTGGAACCGGCGTCCATTTACGCGCGCTCCGAG CGAGCTGTACGCGCTGCGGAGGCTGCACAGCGCGCAATCG ACCCGCAACAGCCGCTGGTGGGTGAATTGGTTTTACCCACGA GTCTGGTGTTACGTTGCGGCGCTGCTGCGCCACAGCGAATCC TATCAATCTATTGCCCAAGCCTCATGGGCCGTAGCTACCGTC TGGTGCTCGGCAAGCATTTCGGGTCGTGAGGCTGTCAACGGTGT TTTCGACCAGATGGGTACCACCTGAATGCGGCGCAGATCAAT CAGTTGCTGCGCGCCATTTCGCCGCTTCGCCGAGAATTGGAAAC GCTCTCCGAAAGACTACGAACCTGGTTGCGATCTATGACGAATT GTGCGGTGAATCCGCCCTTCGTGCTCGCGCTAA
SEQ ID NO:98	47	aislador 14		GACTCAACACGCTAGGGACGTGAAGTCGATTCTTCGATG

SEQ ID NO:99	48	rhs nifW	40	CAGAAGGCGAGAACTAGATTAAAGGGCCATTATAG
SEQ ID NO:100	49	nifW		ATGGAGTGGTTTTACCAGATTCCGGGTGTAGACGAATTGCGCA GCGCTGAATCCTTCTTTTCAGTTCTTCGCGGTTCATACCAGCC GGAAGTGTGGGCGCTGCTCGCTTCGGGTGTTAGCGACGTTT CACCGTAAACTGCGTGCAGGAGTCCCGCTGCAAAACCGTCTG GAGGACAATGATCGTGCGCGTGGCTCTTGGCGCGCGCTCC TGGCCGAATCTTATCAGCAGCAATTCAGGAGAGCGGCACCT AA
SEQ ID NO:101	50	aislador 15		TCGAGAAACAAGGCAGTTCGGGGCTGAAAGTAGCGCCGGG
SEQ ID NO:102	51	rhs nifZ	365	ACAAGTCCCGTATTATAACCGCTAGGAGGTGTTGG
SEQ ID NO:103	52	nifZ		ATGCGCCCGAAATTCACCTTCTCTGAAGAGGTCCGCGTAGTTC GCGCGATTTCGTAATGATGGCACCGTGGCGGGTTTTGCGCCAG GTGCGCTGCTGGTTCGTGCGGGTTCGACGGGCTTTGTGCGTGA CTGGGGTGTGTTCTGCAAGACCAGATCATCTATCAAATCCAC TTTCCGGAACCGACCGCATTATCGGCTGTGCGGAGCAGGAG TTAATCCCGATTACCCAGCGTGGTTGGCTGGTAACCTCCAGT ATCGTGACAGCGTCACGTGCCAAATGGCACTGGCTGTCAACG GTGACGTGGTTGTGAGCGCCGTCACAGTGGCCGTGTGGAGG CCACTGATCGTGGCGAACTTGGCGATTCTACACCGTGGACTT CAGCGGCGGTGGTTCCGCGTTCGGTCCAGGCCATCGCGCTG ATTGAAGAGCGCGAAGAATAA
SEQ ID NO:104	53	aislador 16		ACGCCACGCGTAGTGAGACATACACGTTCGTGGGTTTAC
SEQ ID NO:105	54	rhs nifM	750	TCAGAGACTGAAGTTATTACCCAGGAGGTCTATA
SEQ ID NO:106	55	nifM		ATGAATCCGTGGCAGCGCTTTGCCCGTCAACGCCTTGCTCGCA GCCGCTGGAAACCGTGATCCGGCTGCTCTCGACCCAGCCGATAC CCCAGCGTTCGAGCAGGCGTGGCAGCGTCAATGCCATATGGA ACAAACCATCGTAGCGCTGTCCCGGAAGGCGATATCCGGC TGCTTACTGGAAAACATCGCGGCCAGCCTGGCGATCTGGCTG GACGAGGGTGACTTCGCTCCGCCGGAGCGCGCTGCGATTGTG CGTCATCATGCACGTCTGGAGCTGGCGTTTGCCGACATTGCCC GCCAGGCACCGCAACCGGATCTGAGCACGGTTCAAGCGTGGT ATCTGCGTCACCAGACTCAATTCATGCGTCCGGAGCAGCGTCT GACCCGTACCTGCTCTGACGGTCGATAATGATCGCGAGGC GGTGATCAACGCATCCTTGGCCTGTATCGTCAGATCAACGCG AGCGGTGACGCCTTCGCCCCACTGGCACAGCGCCACTCTCATT GCCCGTCCGCTTGGAAGAAGGCCGTCTGGGCTGGATCTCCCG TGGTCTGCTGTACCCGAGCTCGAAACCGCGTTGTTAGCCTG GCGGAAAACGCACTGTGCTGCGGATTGCGTCGGAATTGGGT TGGCACCTGTTATGGTGCGAGGCCATTGCTCCGCGAGCCCCGA TGGAGCCGCAACAGGCCCTTGAATCTGCGCGCGACTACTTGTG GCAGCAGAGCCAGCAGCGCCACCAGCGTCAATGGCTGGAGCA GATGATTTCGCGCAACCGGGCTGTGTGGTTAA
SEQ ID NO:107	56	terminador T7 WT	1400	TAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTT TTTG
SEQ ID NO:108	57	sufijo Biobrick : (BBa_G00001)		Tactagtagcgccgctgcag
SEQ ID	58	Traducción		ttagttagttag

NO:109		parada (BBa_B0042)		
SEQ ID NO:110	59	Terminador (BBa_B0054)		attagcagaaagtcaaaagccctcgaccggaggcttctgactaaaactcccttggggtatcatitggg
SEQ ID NO:111	60	Secuencia cebador VR (BBa_G00102)		gtcactcaaaaggcggaat
SEQ ID NO:112	61	Terminador (BBa_B0062)		cagataaaaaaatccttagcttctgctaaggatgatttct
SEQ ID NO:113	62	origen pSC101 (BBa_I50042)		ctgtcagaccaagtttacgagctcgttggactcctgtttagatagccagtaagacccagaaatccatc tggatttgttcagaacgctcgttggcggcggttttttattgggtgagaatccaagcactaggacaggt aagacgggtaagccgtgttgaatgacgcgtccttactgggtgcallagccagtcgaatgacgtgtcac gggataatccgaagtggcagactggaataacagagggcaggaaactgctgaacagcaaaaagtcag atagcaccacatagcagaccgccataaaaagccctgagaagcccgtagcgggttttctgtattatg ggtagtttcttgcataatccataaaaagcgcctgtatgtccatttacccttactgccaagagccgt gagcgagcgaactgaatgacgaaaaagacagcagcaggtgctgagtgctggagacaaaag gaatattcagcgatttggccgagcttgcgaggggtgctacttaagcctttaggggttlaaggctgtttttag aggagcaaacagcgttggcagatccttttgaatactgctggaactgactaaagtgtggttatacaca gggtggggtatcttcttttattcttttttattcttcttattctataaattataaacacttgaatataacaaaa aaacacacaaagggttcagcggaaattacagaggggtcagcagaattacaagtttccagcaagggtcta gcagaatttacagataccacacacaaaggaaggaagcagtgtaattatcagtagccatctcaatt ggtagtagtgatttaaaatcacctagacaaattgagatgtatgtctgaattagtttcaaaagcaaatgaact agcgaattagtcgtatgactaacggagcatgaacaaagcgaattttatgctgtgtggcactactcaacc ccacgattgaaaacccacaaaggaaggaagcggaggtatcgttacttataaccaatagcgcagatga tgaacatcagtaggggaaaagcgttattgggtgtattagctaaagcaaccagagagcgtgtagcagagaactg tggaaatcaggaatccttttggtaaggcctttagatgttccagtggaacaaactatgccaagtgtcgaagc gaaaaattagaatttagtttttagtgaagagataattgcttatttccagttaaaaaatcataaataatatt ctggaaatgttaagcttttgaaaaacaatactatgagggatttatgagtggttattaaaagaactaacac aaaagaaaactacaaagcgaataatagagattagccttgatgaatttaagttcatgttaattgcttgaataa actaccatgagtttaaaagccttaacaaatgggttttgaacaaataagtaagatttaaacacttacagc aatatgaatttgggtgttgaataagcggagccgccgactgatactgttatttccagttgaactagatag acaaatggtatctcgttaaccgaacttgagaacaaccagataaaaaatgaatgggtgacaaaataaccacaa ccattacatcagatttctacctacgttaacggactaagaaaaacacacacagatgctttaaactgcaaaaatt cagctcaccagttttaggcaaaaatttttagtgatcgtcaaaagtaagcatgactcaatggttctgttca tggctcagcgaacaaacacgaaccacactagagaacatactggctaaatcagggaaggtatcgtgaggttc ttatggctctgtatctatcagtgaaacatcaagactaacaacaaaagtagaacaactgttaccgttag atatcaaaagggaactgtccatagcagatgaaaacgggtgttaaaaagatagatacatcagagctt ttacgagtttttgggtgcattlaaagctgttaccatgaacagatcgacaaatgtaac
SEQ ID NO:114	63	resistencia a cloranfenicol (BBa_P1004)		gttgcacggcagcgttaagaggttccacttaccataatgaataagatcactaccggcggtatttttgg agtatcagagattttaggagcctaagggaagcctaaatggagaaaaaatcacgggatataccacgggtg atatacccattggcagtcgtaaagaacatttttagggcattttagcagtttgcataatgtactataaccaga ccgttcagctggtatattacggcctttttaaagaccgtlaaagaaaaataagcacaagttttatccggccttta ttcacattcttggccgctgatgaacgctcaccggagtttctgtatggccatgaagacgggtgagctgggt gatctgggtagatgttacccttggtaaccgttttcatgagcaaacctgaacgttttctcctctggaggt gaataaccagcagatttccggcagtttctccacatatttccagagatgtggcgtgttaccgggtgaacact ggcctatttccctaaagggtttattgagaataatgtttttgtctcagccaatccctgggtgagtttaccagtt ttgatttaaacgtggcgaatattggacaacttcttgcgccctgttttccagatgggcaaatattatagcgaag gcgacaaggtgtgtatgctggcgtgacaggttcatatgcccgtttgtgtatggttccatgtcgtgcc gcatgcttaattgaattacaacagtactgtgtatgagtgccagggcgccggcgtaataa
SEQ ID NO:115	64	Terminador (BBa_B0053)		tccggcaaaaaaacgggcaagggtgtcaccacctgccccttttctttaaaccgaaaagattacttcggt tt
SEQ ID NO:116	65	Secuencia cebador VF2 (BBa_G00101)		tggcacctgacgtctaagaa

SEQ ID NO:117	66	Terminador (BBa_B0055)		aaggaatattcagcaattgcccgaggcgaagaaggccaccggaaggtagccagtagttgat tgctacgtaa
SEQ ID NO:118	67	Parada de traducción (BBa_B0042)		ttagttagttag

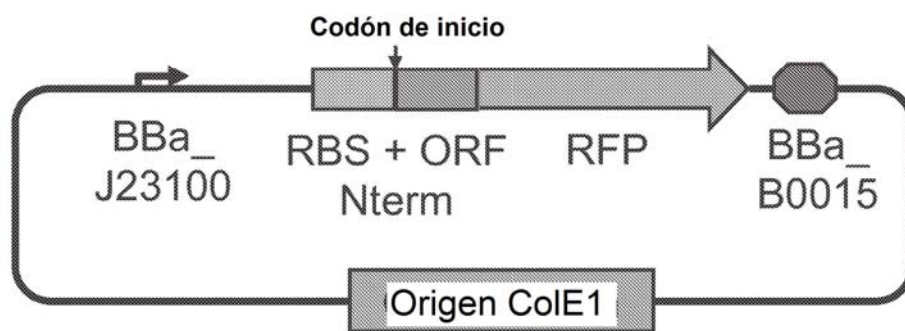


FIG. 21

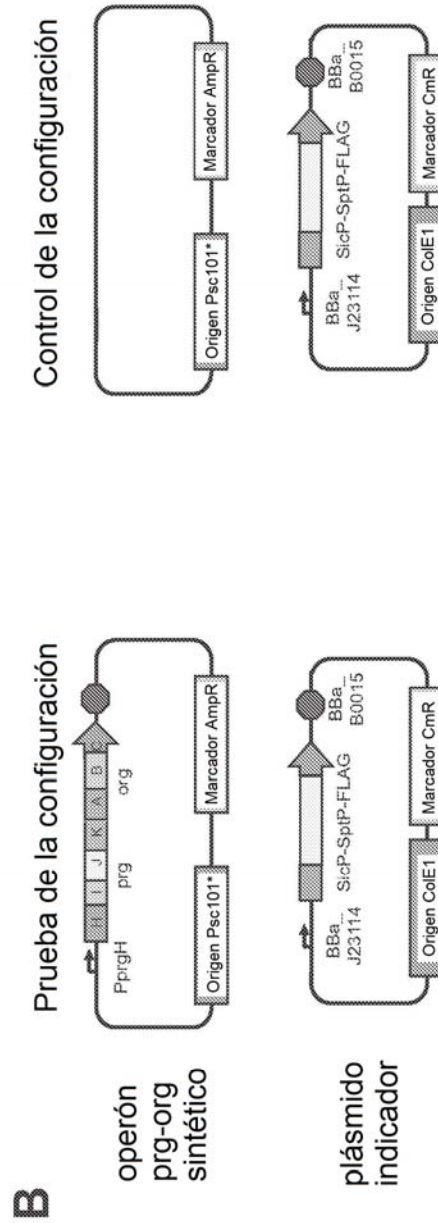
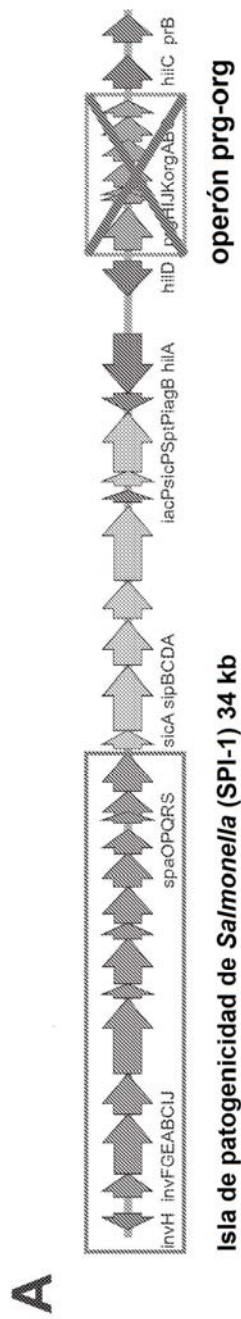


FIG. 22

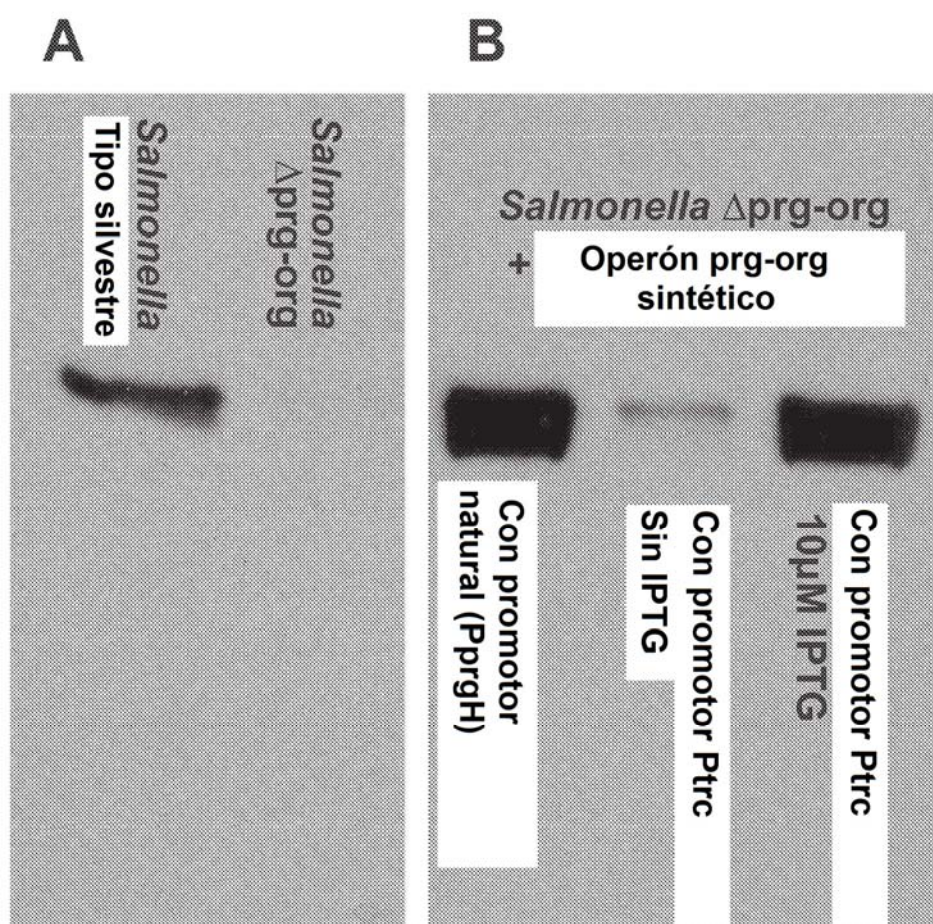


FIG. 23

Fig. 24.

SEQ ID NO:	Nombre	Secuencia
SEQ ID NO:119	SynPrgH_RBS	ATTAAACAGGATAATAAGC
SEQ ID NO:120	SynPrgH	ATGGAACTAGCAAGGAGAAAACGATTACGTCCCTGGTCCGTATATCGTT CGTCTGTTGAACTCGTCCCTGAACGGTTGTGAATTCGGCTGCTGACTGGTC GTACGCTGTTCGTCTGGGTACAGAGCGATGCTCTGACCGCGTCTGGTCAGCT GCCGGACATTCTGCCACTCTTCTTCATCCCGCTGGATCATGGCGGTGTT AATTTGAGATTCAAGTGGACACTGACGCGACGGAAAATCATCTGCACGAA CTGAAAGAGGGCAACTCCGAGAGCGCTCCGTGCAACTGAACACCCGATC CAAGTTGGTGAACGTGCTGATTTTGATTCTGTCGGAGAGCGAGCCGTGGGTG CCGGAACAGCCGGAGAAGTTGGAAACTTCTGCGAAAAAGAACGAACCGCG CTTTAAAAACGGCATCGTCGCGGCACTGGCGGGTTTCTTTATCCTGGGTATC GGCACGGTTGGCACCCCTGTGGATTCTGAACCTGCGCGCAACGTCGAAGCAGCC GAATTTGGACAGCCTGTTGGGTACAGGAGAAGGAGCGTTTTCAGGTGCTGCCG GGTCGCGACAAGATGCTGTATGTGCGCGCGCAAAACGAACCGGACACCCCTG TGGGCACGTCAAGTCTTGGCACGCGCGATTACGATAAAAAACGCACGCGTT ATTAATGAAAATGAGGAGAATAAACGTATCAGCATCTGGCTGGACACGTAT TATCCACAACCTGGCATATTACCGTATCCATTTTGATGAACCACGTAAGCCGG TGTTTTGGCTGTCCCGTCAACGCAACACGATGAGCAAGAAAGAGCTGGAGG TGCTGTCCAGAAAATTGCGTGCCTGATGCCGTACGCCGACAGCGTCAATAT TACTCTGATGGATGACGTGACCGCAGCAGGCCAAGCCGAGGCAGGTCTGAA ACAACAGGCGCTGCCATACAGCCGCCGTAACCACAAAGGTGGTGTACGTT CGTTATTCAGGGCGCCTTGGACGACGTTGAGATTCTGCGTGCGCCCAGTTT GTCGACTCCTATTATCGTACCTGGGGTGGTCTGTTACGTTCAATTCGCAATTG AATTGAAAGACGATTGGCTGAAAGGCCGCTCGTTCCAATACGGTGCAGGAAAG GCTACATTAAGATGAGCCAGGTCATTGGTATTTCCGTCTCCTCTGTAATA G
SEQ ID NO:121	SynPrgI_RBS	ATTAAAGAGGAGAAATTAAGC
SEQ ID NO:122	SynPrgJ_RBS	ATTAAGGAGGATAAATTAAGC
SEQ ID NO:123	SynPrgJ	ATGTCTATCGCGACTATCGTGCCTGAAAATGCCGTTATTGGTCAAGCGGTGA ATATTCGCAGCATGGAAACGGACATCGTCAGCTTGGACGACCGTTTGCTGC AAGCATTTTCGGGACAGCGCCATCGCTACCGCCGTCGATAAGCAGACCATTA CCAATCGCATTGAAGACCCTAATCTGGTTACCGATCCGAAGGAGCTGGCGA TTAGCCAGGAAATGATTTCCGACTACAATCTGTACGTCAGCATGGTTAGCAC CCTGACGCGTAAGGGCGTTGGCGCTGTTGAGACTTTGCTGCGTTCTGATAG
SEQ ID NO:124	SynPrgK_RBS	ATTAAACAGGATAATAAGC
SEQ ID NO:125	SynPrgK	ATGATCCGCCGTTACCTGTATACCTTCTTGCTGGTTATGACTTTGGCCGGCTG TAAAGATAAGGATCTGCTGAAAGGCTTGGACCAAGAGCAAGCGAATGAGG TCATTGCGGTTCTGCAAAATGCAACAATTGAGGCTAACAAGATTGATAGCG GCAAGCTGGGTTACAGCATTACCGTCGCGGAACCGGATTTACCGCGGCAG TGTATTGGATTAAGACCTACCAGTTGCCGCTCGCCCGCGTGTGAAAATCGC CCAAATGTTTCCGGCAGACAGCCTGGTTAGCTCTCCGCTGCGGAGAAAAGC ACGCTCTGTAACGCGGATTGAACAGCGCCGAGGAGCAGTCGCTGCAAAACGAT GGAAGGTGTTCTGTGCGCCCGTGTCCACATCAGCTATGATATTGATGCGGGC GAGAACGGTCTCGCCTAAGCCGGTGCACCTGTGCGCTTTGGCGGTGTAT GAACGCGGTTCCCTCTGGCCCATCAGATTTCGGATATTAAGCGCTTCCTGA AAAACAGCTTCGCGGATGTTGACTATGATAACATCAGCGTGGTTCTGTCCGA

SEQ ID NO:	Nombre	Secuencia
		GCGTAGCGACGCACAGTTGCAGGCGCCGGGACGCCGGTCAAGCGCAATAG CTTCGCTACCTCCTGGATTGTGCTGATTATCCTGTTGTCTGTTATGAGCGCGG GTTTCGGTGTCTGGTACTACAAAAATCACTATGCGCGTAATAAGAAAGGCA TTACTGCCGATGACAAGGCAAAGTCCAGCAACGAGTAATAA
SEQ ID NO:126	SynOrgA_RBS	GAAAGAAGGGACAGACTAG
SEQ ID NO:127	SynOrgA	ATGATTCTGTCGTAACCGTCAAATGAACCGTCAACCACTGCCAATTATCTGGC AACGCATTATTTTCGACCCACTGTCCTATATTCACCCACAACGTCGCAAT CGCGCCGGAGATGATCGTGCCTCCGGCAGCGCGTGGGCAAGCAATGAGCT GATTTTGGCGGCGTGGCGTTTGAAGAACGGTGAGAAGGAGTGCATTAGAA TAGCCTGACGCAGCTGTGGTTGCGTCAATGGCGCCGTCTGCCGAGGTTGCT TACCTGCTGGGTTGCCACAAGTTGCGTGCTGACCTGGCCCGTCAAGGTGCTT TATTGGGCTGCCGACTGGGCGCAGGCATTCTTGGCGATGCACAGGGTA CGTCTTGTGCGTTTGTAAATAAGGCGCCGAACACCGTTTCTGCTGTCCGT TGGTTACGCCCAACTGAACGCGCTGAATGAGTTCCTGCCGGAGAGCTTGGC CCAACGCTTTCCTCTGCTGTTTCCACCGTTCATCGAGGAGGCACTGAAACAA GATGCAGTGGAGATGAGCATCCTGCTGCTGGCCCTGCAATACGCGCAAAAG TATCCTAACACCGTCCCGGCGTTTGGCTGCTAATAA
SEQ ID NO:128	SynOrgB_RBS	GAAAGAAGGGACAGACTAG
SEQ ID NO:129	SynOrgB	ATGTTGAAAAATATCCCGATTCCATCCCCTCTGTCTCCGGTTGAAGGTATCC TGATTAACGCAAGACGTTGGAGCGTTACTTCTCGATTGAGCGCCTGGAAC AACAGGCGCATCAGCGTGCAAAAGCGCATTTTGGCTGAGGCAGAAGAAGAA GCCAAGACCCTGCGCATGTATGCGTACCAGGAGGGCTACGAGCAGGGCATG ATTGACGCACTGCAACAGGTGGCCGCCTATTTGACGACAACAGACGATG GCTTGGAAATGGATGGAGAAAAATCAAAATCTATGCGCGTGAGTTGTTGAGC GCGGCTGTGATCACC CGGAAACGTTGTTGACGGTGCTGGACGAGTGGCTG CGTGATTTGATAAGCCGGAAGGTCAGCTGTTTTGACCCTGCCGGTGAACG CAAAGAAAGATCATCAGAAACTGATGGTGCTGCTGATGGAAAAATGGCCGG GCACCTTCAATCTGAAGTATCATCAGGAGCAACGTTTTATCATGTCTGTGG CGATCAGATTGCCGAGTTTTCGCCGGAACATTTGTTGAAACGCGGTTGGT GTTATCAAGCACCATCTGGATGAGCTGCCTCAGGACTGTGCGCACCATTTCGG ACAATGCGATTAACGCGCTGATTGATGAATGGAAGACGAAACCCAAGCTG AAGTTATTCGCTGATAA
SEQ ID NO:130	SynInvA_RBS	CTTGGGCACGCGTCCATTAACAGGAGTAATTAAGC
SEQ ID NO:131	SynInvA	ATGCTGCTGTCCCTGCTGAATAGCGCGCTCTGCGTCTGAGCTGCTGATTC TGGTTCTGATGGTTATGATCATCAGCATGTTTCGTTATCCCGTTGCCGACCTAT TTGGTTGACTTCTTGATCGCTTTGAACATTGTCCTGGCAATTCTGGTGTTCAT GGGCTCCTTCTACATCGACCGCATTCTGAGCTTCAGCACCTTTCGGCGGTT CTGCTGATCAGACTCTGTTCCGTTTGGCACTGAGCATCAGCACCAGCCGCC TGATCCTGATTGAAGCAGATGCGGGTGAGATCATCGCGACCTTTGGTCAGTT TGTGATCGGTGACAGCCTGGCGGTTGGTTTCGTCGTATTCTCCATCGTGACG GTGGTGCAGTTTATCGTTATTACCAAGGGCAGCGAACGTTGTGGCGGAGGTC GCCGCTCGCTTCAGCCTGGACGGCATGCCGGGTAACAGATGTCTATTGAT GCAGACCTGAAAGCCGGCATTATTGATGCTGATGCAGCGCGGAGCGCCGC AGCGTCTGGAGCGTGAAGCCAACTGTACGGTTTCCTTCGACGGTGCCATG AAGTTCATTAAAGGTGATGCGATTGCGGGCATCATTATCATCTTCGTTAACT TCATTGGCGGTATCAGCGTCGGTATGACCCGTCATGGTATGGATCTGAGCAG CGCCTGAGCACCTACACCATGCTGACGATTGGTGATGGTCTGGTTGCCCAA ATTCCGGCGTTGCTGATCGCGATTCTGCGGGCTTCATCGTTACCCGCGTCA

SEQ ID NO:	Nombre	Secuencia
		ACGGTGATAGCGATAACATGGGTTCGTAACATTATGACCCAGCTGCTGAATA ATCCGTTTGTCTCGTTGTAAACGGCGATTTTGACCATCAGCATGGGCACGCT GCCGGGCTTTCCGTTGCCGCTTTTCGTTATTCTGTCTGTTGTGCTGTCCGTCC TGTTTTACTTTAAGTTCCGCGAGGCGAAACGTAGCGCTGCGAAACCAAAAA CGAGCAAGGGCGAGCAACCGTTGTCCATCGAGGAGAAGGAAGGTAGCAGC CTGGGCTGATTGGCGACCTGGATAAGGTTAGCACGGAACCGTCCCGCTG ATTTTGCTGGTGCCGAAATCGCGTCGTGAGGATCTGGAGAAAAGCGCAGCTG GCGGAACGTCTGCGCAGCCAATTCTTTATTGATTATGGTGTGCGTCTGCCAG AAGTACTGCTGCGTGACGGTGAGGGTCTGGATGACAACCTCTATCGTCCTGCT GATTAATGAGATTCCGTTGAACAGTTTACTGTCTATTTGACCTGATGCGT GTGGTTAACTACAGCGACGAGGTGGTGAGCTTTGGCATCAACCCGACCATT CACCAGCAAGGTTCCAGCCAGTACTTTTGGGTGACCCATGAGGAAGGCGAA AAGCTGCGCGAGCTGGGCTACGTCCTGCGTAATGCACTGGACGAACTGTAC CACTGTCTGGCGGTGACGCTGGCACGCAATGTGAACGAGTATTTCCGGTATCC AAGAAACGAAACACATGCTGGACCAACTGGAAGCAAAGTTTCTGACCTGC TGAAGGAGGTTTTCGCCACGCCACCGTGCAGCGCATTTCCGAAAGTGCTGC AACGTCTGCTGTCCGAACGCGTGAGCGTCCGTAACATGAAGCTGATCATGG AAGCCCTGGCACTGTGGGCTCCGCGTGAGAAAAGATGTGATCAATCTGGTGG AGCACATCCGTGGTGCGATGGCGCGTTATATCTGCCACAAGTTCCGAAATG GTGGTGAACCTGCGTGCCGTTATGGTTTCCGCCGAAGTTGAGGATGTCATTCTG TAAAGGCATTCTGCAAACTTCTGGCTCCACCTTTTGGAGCTGGACCCGGAG GCTTCGGCAAACTCTGATGGACCTGATCACGCTGAAGCTGGACGACCTGTTG ATTGCGCATAAGGACCTGGTCTCTGTTGACCAGCGTTGACGTGCGTCTGTTTA TCAAGAAAATGATTGAAGGTCGTTTCCGGATCTGGAGGTCCTGTCCTTCGG TGAGATTGCAGATAGCAAAAGCGTGAATGTCATCAAAACCATCTGA
SEQ ID NO:132	SynInvC_RBS	CTTGGGCACGCGTCCATTAAGAGGACCAATTAAGC
SEQ ID NO:133	SynInvC	ATGAAAACCCACGCTCTGCTGCAATACCTGGCCTACCCGCAGAAAATCACT GGCCCTATCATTGAAGCAGAACTGCGTGATGTTGCAATTGGTGAATTGTGCG AGATCCGTCGCGGTGGCACCAGAAAGCAGGTTGTGGCCCGTGCGCAAGTGG TTGGTTTGCAGCGCGAACGTACCGTCTGAGCCTGATCGGCAATGCCCAAG GCCTGAGCCGTGATGTGGTCTTGTACCCGACCGGCCGTGCTCTGAGCGCGTG GGTTGGTTACAGCGTCTGGGCGCAGTACTGGACCCGACGGGTAATAATCGT TGAACGTTTACCCCGGAAGTCGCACCGATTCCGAGGAGCGCGTTATCGA CGTGGCACCGCCGAGCTACGCAAGCCGTGTCGGTGTGCGCAACCGCTGAT CACGGGTGTCCGCGCAATTGATGGTCTGCTGACGTGTGGTGTGGGCCAGCG TATGGGTATTTTCGAAGCGCGGGTGTGGTAAGACCATGTTGATGCACATG CTGATCGAGCAAAACCGAAGCGGATGTCTTTGTTATTGGCCTGATTGGCGAGC GTGGTCTGTAAGTTACCGAATTTGTAGACATGTTGCGTGCATCTCATAAGAA AGAAAAATGTGTCTGGTTTTTGCCACGTCCGACTTCCCAAGCGTTGACCGC TGCAACGCTGCCAGCTGGCAACGACGGTCGCGAGTATTTCCGCGACCAG GGTAAACGTGTTGTCTGTTTATCGACAGCATGACCCGTTATGCACGCGCGT TGGTGTGATGTCGCGCTGGCGAGCGCGAGCGTCCGGCTCGCCGTGGTTATC CGGCGTCTGTGTTTCGACAATCTGCCGCGTTTGTCTGGAGCGTCCGGGTGCGAC GAGCGAGGGTAGCATTACCGCCTTCTATACCGTCTGCTGGAGTCGGAAGA AGAAGCGGACCCGATGGCGGACGAGATCCGTTCTATTCTGGACGGTCACCT GTACCTGTCCCGCAAACTGGCGGGTCAGGGTCATTACCCGGCTATCGATGTG CTGAAGAGCGTGAGCCGTGTGTTTGGTCAAGTGACCACCCGACTCACCGG GAGCAAGCGAGCGCGGTCCGTAAGCTGATGACCCGTCTGGAAGAGCTGCAA CTGTTTATTGACCTGGGCGAGTATCGTCCGGGCGAGAACATTGACAATGAT CGTGCGATGCAAAATGCGCGATAGCCTGAAGCGTGGCTGTGTACGCTGTT GCGCAATACAGCAGCTTCGATGATACGCTGTCCGGCATGAACCGCTTTGCG GATCAGAACTGA

SEQ ID NO:	Nombre	Secuencia
SEQ ID NO:134	SynInvE_RBS	TACTTGGGCACGCGTCCATTAATTAGGATCAATAGC
SEQ ID NO:135	SynInvE	ATGATTCCGGGCAGCACCTCCGGTATTTCTTTAGCCGTATCCTGAGCCGTCAGACCTCCCACCAGGATGCAACCCAGCATAACGACGCACAAACAGCGGAAATTCAACAAGCGGCGGAAGATAGCTCGCCGGGTGCGGAGGTTCAGAAATTCGTCCAGAGCACGGACGAGATGTCTGCTGCGTTGGCGCAGTTCCGCAATCGCCGTGACTATGAGAAAAAGAGCAGCAATTTGTCTAACTCCTTCGAGCGCGTTCTGGAGGACGAGGCACTGCCGAAAGCGAAGCAGATTCTGAAACTGATCAGCGTGCATGGCGGTGCGTTGGAGGATTTCTGCGTCAGGCGCGCAGCCTGTTCCCGGACCCAAGCGATCTGGTGCTGGTTCTGCGCGAGCTGTTGCGTCGTAAGGACCTGGAGGAGATCGTGCCTAAGAAGCTGGAGAGCCTGCTGAAGCACGTGGAGGAACAACCGACCCGAAACCCCTGAAGGCCGGTATTAAGTGCCTGTAAGGCGCTGTGTTTGCAAGACGCTGTCTCTGAAACCTGGTCTGCTGCGTGCAGCTACCGCCAGTTCATCCAAAGCGAAAGCCACGAAGTCGAGATTACACGATTGGATCGCCAGCTACGGTTATCAGCGTCGCGTGGTTGTTCTGGATTTCATTGAAGGCAGCCTGCTGACTGACATCGATGCTAACGATGCAAGCTGCTCCCGTCTGGAGTTTGGCCAACCTGCTGCGCCGCTGACCCAGCTGAAAAATGTTGCGTAGCGCCGACCTGCTGTTTGTCTCGACGTTGCTGTCTTACAGCTTCACGAAAGCATTTAACGCTGAGGAGAGCAGCTGGCTGTTGCTGATGCTGTCTTTGCTGCAACAGCCGACGAAGTGGATAGCCTGCTGGCGGACATTATCGGTCTGAATGCGCTGCTGTTGTCCACAAAGAGCAGCCAGCTTCCTGCAAACTCTTCTATCAGGTCTGTAAGGCAATCCCGTCTAGCCTGTTTATGAAGAGTACTGGCAAGAAGAACTGCTGATGGCACTGCGCTCCATGACGGACATTGCTTACAAACACGAAATGGCCGAACAACGTCGTACCATCGAAAAGCTGTCTAA
SEQ ID NO:136	SynInvF_RBS	CTTGGGCACGCGTCCATTAATAAGGAGTAATTAAGC
SEQ ID NO:137	SynInvF	ATGAGCTTCAGCGAGAGCGCCACAATGAAAACGTCTGATTCAAGAAGGCGCACTGCTGTTTTGTGAGCAAGCAGTCGTGGCGCCTGTGACGGGTGATCTGGTTTTTCGTCCGCTGAAAATCGAGGTCCCTGAGCAAGCTGCTGGCGTTCATCGACGGCGAGGTCTGGTGGATACGACCTACGCGGAGTCGGACAAATGGGTTCTGTGTCTCCGGAGTTCCGTGCTATTTGCAAGACCGTAACGTTGCGAATATTGGTTTTTGACGAGATTATCACCCCATCTCCGGCGTTCAACAAGTTCTGGCACTGTTGCGTAAGAGCGAAAAGCTATTGGTTGGTTCGGCTACTTGTGGCCCAAGCACAGCGGCAATACTATGCGTATGTTGGGTGAGGATTACGGTGTTAGCTACACGCATTTCCGCCGTCTGTGCAGCCGCGCTCTGGGCGGTAAGGCGAAAGCGAGCTGCGCAATTGGCGCATGGCCAGTCCCTGCTGAATAGCGTGGAAGGTCATGAAAACATCACCCAGCTGGCGGTCAACCACGGTTATAGCAGCCGTCCCACTTTAGCTCTGAAATCAAGGAGCTGATTGGTGTTTCCCGCGTAAGCTGTCTAACATCATTACAGCTGGCCGACAAATGA
SEQ ID NO:138	SynInvG_RBS	TGGGCACGCGTCCATTAATGAGGAAAAATTATTAGC
SEQ ID NO:139	SynInvG	ATGAAAACGCACATTCTGTTGGCCCGTGTGCTGGCTTGCGCAGCTCTGGTGTGGTTCACCCAGGTTATAGCTCCGAGAAGATCCCGGTTACGGGACGCGGCTTCGTTGCAAAAGGACGATTCTCTGCGCACCTTTTTCGATGCGATGGCACTGCAATTGAAGGAGCCGGTGATTGTGACGAAGATGGCGGCTCGCAAAAAGATTACCGCAATTTTCGAGTTCCACGATCCAAACGCGCTGCTGGAGAAACTGTCCCTGCAACTGGGTCTGATCTGGTACTTTGATGGTCAGGCGATCTACATCTACGACGCGAGCGAAATGCGTAATGCGGTTGTGAGCCTGCGTAATGTGAGCCTGAACGAGTTCAACAATTTCTGAAGCGCAGCGGCTGTACAATAAGAACTACCCTCTGCGTGGTGATAATCGTAAAGGCACCTTCTATGTGACGGGTCCGCCGGTGTATGTTGATATGGTTGTAAATGCGGCCACCATGATGGACAAACAGAATGATG

SEQ ID NO:	Nombre	Secuencia
		GCATCGAGCTGGGTGCGCAAAAGATCGGTGTTATGCGCCTGAACAACACTT TTGTGGGCGACCGCACCTACAACCTGCGTGATCAAAAGATGGTCAATCCGG GTATTGCTACGGCAATTGAACGCTGTTGCAAGGCGAAGAAACAACCGCTGG GTAACATTGTAAGCTCCGAGCCTCCGGCCATGCGCGCCTTTAGCGCAAAACG GCGAGAAAGGTAAGGCAGCGAATTACGCGGGTGGTATGAGCCTGCAAGAA GCGCTGAAACAGAACGCGAGCGGAGGCAACATCAAAATTGTGCGCTATCCG GACACCAACAGCCTGCTGGTGAAAGGTACGGCGGAGCAGGTGCATTTTCATC GAGATGCTGGTTAAAGCCCTGGACGTGGCGAAACGTCACGTTGAATTGAGC CTGTGGATTGTGGATTGTAATAAGAGCGACCTGGAACGCTTGGGCACCAGC TGGAGCGGTAGCATCACCATCGGCGACAAGCTGGGTGTGAGCCTGAACCAG AGCAGCATCTCCACGCTGGACGGTAGCCGCTTTATTGCGGCGGTCAACGCTC TGGAGGAAAAGAAACAGGCCACTGTTGTGACCGCTCCGTTCTGCTGACCC AGGAGAAATGTCCGCGGATTTTGGACAATAATCGTACTTTCTATACCAAAC GATCGGTGAACGTAATGTTGCATTGGAACACGTGACCTATGGCACCATGAT CCGTGTCTGCGCGTTTCAGCGCGGACGGCCAGATTGAGATGAGCCTGGA CATCGAAGATGGTAATGACAAAACCCCGCAGTCTGATACGACCACCTCCGT TGATGCGCTGCCAGAAGTGGGTGCTACCCTGATCTCGACGATTGCACGTGTC CCGATGGTAAATCTCTGCTGGTTGGTGGCTACACGCGTGATGCAAAACAG GACACGGTCCAAGCATCCGTTCTGGGTAAGCTGCCGCTGATTGGCTCGT TGTTTCGCTACAGCAGCAAAAACAGTCTAATGTCGTCGTTCTTTATGAT TGAGCCGAAAGAAATCGTTGACCCGCTGACCCGGATGCCAGCGAGAGCGT TAACAACATTCTGAAACAGTCCGGTGGTGGAGCGGTGACGATAAGCTGCA AAAGTGGGTTCTGTGTATTGACCCGTTGGTCAAGGAGCCATTAAGTAA
SEQ ID NO:140	SynInvI_RBS	CTTGGGACGCGTCCATTAAGGACCAATTAAGC
SEQ ID NO:141	SynInvI	ATGCACTCTCTGACTCGTATCAAGGTCCTGCAACGTCGTTGTACCGTGTTC ATTCTCAGTGGGAGTCCATTCTGTTGCGTTATCAGGACGAAGATCGCGGCTT GCAGGCGGAGGAGGAGGCCATCTGGAACAGATCGCAGGTCTGAAACTGCT GTTGGACACCCTGCGTGCTGAAATCGTCAACTGAGCCGTGAAGAAATCTA TACGCTGTGCGCAACAGAGCATTGTTCTGCGCCAGATTAAAGGATCTGGA GCTGCAAAATCATTAGATTCAAGAAAAGCGTAGCGAGCTGGAAAAGAAAC GTGAGGAATTTCAAAAGAAAAGCAAAATACTGGCTGCGTAAAGAAGGTAAC ACCAGCGCTGGATTATCCGTCAAAACGCTTCTACATTCAACGTGAGATCCA GCAAGAGGAGGCGGAGAGCGAAGAGATCATTAA
SEQ ID NO:142	SynInvJ_RBS	AAAAAAAAAGCTCTAAAAGATTAAGAGGGGGTAACAT
SEQ ID NO:143	SynInvJ	ATGGGTGACGTGAGCGCGGTGAGCAGCAGCGGTAACTTCTGCTGCCGAG CAGGACGAGGTTGGTGGCTGTCCGAAGCGCTGAAGAAAGCGGTTGAAAA ACAAAAACCGAATACAGCGGTGACAAGAAAGATCGTGATTATGGTGACGC CTTTGTTATGCACAAGGAAACCGCGCTGCCGTTGTTGCTGGCAGCTTGGCGC CACGGCGCACCGGCGAAAAAGCGAGCACCATAACGGTAACGTAAGCGGTCT GCATCACAACGTAAGAGCGAGCTGCGTATTGCTGAGAAAAGTGTGAAGGT GACGGCGGAGAAGAGCGTTGGTCTGATTAGCGCTGAAGCGAAGGTGGATA AATCTGCCGCGCTGCTGTCTAGCAAGAATCGTCCGCTGGAATCGGTGAGCG GCAAAAAGTTGTCCGCCGATCTGAAAAGCAGTGGAGTCCGTTGCCGAGGTCA CGGATAACGCCACCGGCATTTCCGATGACAACATCAAAGCATTGCCGGGTG ACAATAAGGCCATCGCCGTTGAGGGTGTGCGTAAAGAAGGTGCGCCGCTGG CGCGTGACGTGGCTCCGGCACGCATGGCGGCAGCAAAACGCGGAAGCCGG AGGATAAAGACCACAAGAAGGTCAAGGACGTTAGCCAGCTGCCGCTGCAA CCGACTACCATCGCCGATCTGTCTCAACTGACGGGTGGCGATGAAAAAGATG CCGCTGGCAGCGCAGTCCAAACCGATGATGACCATTTTCCCAACCGCCGAC GGCGTTAAAGGTGAGGACAGCTCTCTGACCTATCGTTCCAGCGCTGGGGC

SEQ ID NO:	Nombre	Secuencia
		AACGATTACTCCGTCAATATCCAGGCACGCCAAGCGGGCGAATTTAGCCTG ATTCTTAGCAATACCCAGGTTGAACATCGTCTGCACGACCAATGGCAAAAT GGCAATCCACAACGCTGGCATTGACGCGTGATGACCAGCAAAACCCGCAA CAGCAACAGCATCGTCAGCAGTCCGGTGAAGAGGACGACGCGTAA
SEQ ID NO:144	SynSpaO_RBS	CTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG
SEQ ID NO:145	SynSpaO	ATGAGCTTGC CGGTACGCCAAATTGACCGCGTGAATGGCTGCTGGCCCAG ACTGCGACCGAATGTCAGCGTCACGGTCTGTGAGGCCACCTGGAGTATCCG ACCCGTCAGGGTATGTGGGTCCGCCTGTCTGATGCCGAAAAACGCTGGTCT GCGTGGATCAAACCGGGTATTGGTTGGAACACGTTAGCCAGCACTGGCG GGTGCCGCGGTACGCGCAGGCGCGGAGCACCTGGTGGTTCCGTGGCTGGCG GCAACCGAACGCCCGTTTCGAGCTGCCGGTCCCGCACTTGAGCTGCCGTGCTC TGTGCGTGGAGAACCCGGTTCGGGTTCCGCACTGCCTGAGGGCAAGCTGC TGACATCATGTGCGATCGTGGTGGCCTGTGGTTGAGCATCTGCCGGAAGCT GCCGGCTGTTGGCGGTGGCCGTCCGAAGATGCTGCGTTGGCCGCTGCGTTTC GTTATTGGCAGCAGCGACACCCAGCGCAGCCTGCTGGGTGCTATCGGTATC GGTGATGTTCTGCTGATCCGCACCGCGTGTGAGGTTTACTGTTACGCGA AGAAACTGGGCCACTTTAATCGTGTGGAAGGTGGCATTATTGTCGAAACGC TGGACATTCAACATATCGAGGAGGAGAAACAACACGACGGAACCGCGGAA ACCCTGCCGGGTCTGAATCAACTGCCGGTGAAGCTGGAGTTCGTTCTGTATC GTAAAAACGTGACGTTGGCCGAACGGAAGCAATGGGTACGCAACAACCTGC TGTCCTTGCCAACCAATGCGGAACCTGAACGTGCAAAATCATGGCGAATGGCG TGCTGCTGGGTAAACGGCGAACTGGTGCAGATGAATGACACCTGGGTGTCG AGATTTCATGAGTGGTTGTCCGAGAGCGGTAAATGGCGAGTAG
SEQ ID NO:146	SynSpaP_RBS	CTTGGGCACGCGTCCATGAAAGAAACGACATACTAG
SEQ ID NO:147	SynSpaP	ATGGGTAAACGATATTAGCTTGATTGCATTGCTGGCGTTTCCACCCTGCTGC CGTTCATCATCGCTCTGGTACTTGCTTCGTCAAATTCAGCATCGTCTTTGTG ATGGTGCACAACGCGCTGGGTCTGCAACAAATTCGAAGCAATATGACCTG AATGGCGTCGCACTGCTGCTGTCGATGTTGTTATGTGGCCGATTATGCACG ACGCGTATGTGATTTCGAGGATGAAGATGTGACCTTTAACGACATCTCCAG CCTGAGCAAGCATGTTGATGAGGGCTGGACGTTATCGCGACTACCTGAT CAAGTATTCGACCGTGAGCTGGTGCAGTTCTTTGAGAATGCCAGTTGAAA CGTCAGTACGGTGAAGAAACGGAACCGTTAAACGTGACAAGGACGAGAT TGAAAAGCCGAGCATTTTCGCACTGTTGCCTGCTTACGCCCTGAGCGAGATT AAGAGCGCATTCAAAATTGGTTTTTACCTGTACCTGCCGTTCTGTTGTGGTCG ATCTGGTTGTCTCCAGCGTTCTGCTGGCCCTGGGCATGATGATGATGTCCTC GGTTACCATCAGCAGCGCGATCAAACTGGTCCTGTTTGTGGCCCTGGATGGT TGGACGCTGCTGTCTAAAGGCCTGATCCTGCAATACATGGACATCGCGACCT AA
SEQ ID NO:148	SynSpaQ_RBS	CTTGGGCACGCGTCCATTAAGAAGGAGGAATTAAGC
SEQ ID NO:149	SynSpaQ	ATGGACGATCTGGTTTTTCGCCGGCAACAAAGCCCTGTACTTGGTGTGATTCT TGTCGGTTGGCCGACGATTGTCGCAACCAATTATCGGTCTGCTGGTTGGTCT GTTTCAAACCGTGACGCAAGTTGCAGGAGCAAAACCTGCCGTTCCGTATCAA GCTGCTGGGTGTGTGCTGTGCCTGTTTTGTCTGTCTGGCTGGTATGGCGAA GTTCTGCTGTCTACGGCCGTCAAGTTCATCTTCTGGCTCTGGCGAAAGGTT AA
SEQ ID NO:150	SynSpaR_RBS	CTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG

SEQ ID NO:	Nombre	Secuencia
SEQ ID NO:151	SynSpaR	ATGTTCTATGCATTGTATTTGAGATCCACCATCTGGTGGCGTCCGCGGCTC TGGGTTTTGCGCGTGTTCCTCCGATCTTTTTCTTTCTGCCGTTCCTGAATAGC GGTGTCTGAGCGGTGTCTCCGCGCAACGCGATCATCATCTGGTGGCGCTGG GTGTATGGCCGACGCCCTGAATGAGGCGCCACCGTTTCTGTCTGTGGCAAT GATTCGCTGGTCTTGCAGAGAGGAGCGCGTGGGTGTTATGCTGGGTTCCTG TTGTCTTGGCCGTTTTGGGTTATGCACGCGTTGGGCTGTATCATTTGATAACC AACGCGGTGCAACCCTGTCCAGCAGCATTGATCCTGCGAATGGCATCGACA CCAGCGAGATGGCCAATTTCTGAACATGTTCCGCGCTGTCTGTATTTGCA GAACGGTGGCTTGGTCACGATGGTGGACGTGCTGAATAAGCTTACCAGCT GTGTGATCCGATGAACGAGTGTACGCCGAGCCTGCCTCCGTTGCTGACCTTC ATTAATCAAGTGGCCGAGAAGCGACTGGTGTGGCGTCCCCGGTCTGTCTGG TTCTGCTGCTGTCGGAAGTTTTCTGGGCTGCTGTCTGTTTTGCACCGCAA ATGAACGCGTTCGCCATTAGCCTGACTGTTAAAAAGCGGTATTGCGGTTTTGA TCATGCTGCTGTATTTACGCCCCTGCGGACAATGTTCTGCTGCTGAG CTTTCAGCGACCGCCTGAGCAGCTGGTTCTACGAACGTGGCGCAACGCA TGTGCTGGAATAA
SEQ ID NO:152	SynSpaS_RB S	CTTGGGCACGCTCCATGAAAGACAGGACCCACTAG
SEQ ID NO:153	SynSpaS	ATGTCCAGCAACAAAACCGAAAAACCGACTAAGAAACGTCTGGAGGATAG CGCAAAGAAAGGTGAGAGCTTCAAGAGCAAGGACCTGATTATCGCGTGCCT GACCTGGGTGGTATCGCTTATTTGGTGAAGTACGCGAGCTTCAATGAGTTT ATGGGTATCATTAAGATTATCATCGCTGATAACTTTGATCAGTCGATGGCAG ATTATAGCCTGGCCGTGTTTGGTATTGGCCTGAAATACCTGATTCCGTTTAT GCTGCTGTGTTTGGTTTGTCCGCACTGCCGCACTGCTGCAAGCGGGCTTC GTTCTGGCAACCGAGGCCCTGAAGCCGAATCTGTCCGCCCTGAACCCGGTT GAAGGCGCAAGAACTGTTTCCATGCGCACCGTCAAAGACACGGTCAAG ACGCTGCTGTATCTGTGAGCTTTGTGGTTGCGGCAATCATTTGCTGGAAAA AGTATAAAGTCGAGATTTTCAGCCAACGTAACGTAATATCGTGGGTATTG CGGTTATCTGGCGTGAATTGCTGCTGGCGTTGGTTCTGACCTGTCTGGCGTG CGCGCTGATCGTGTGTTGCTGGATGCTATTGCCGAGTACTTTCTGACCATG AAAGATATGAAGATGGACAAAAGAAGAAGTTAAACGCGAGATGAAAGAGCA GGAGGGTAACCCGGAGGTGAAGTCTAAACGTCGTGAAGTCCACATGGAAAT CCTGAGCGAACAAGTCAAGTCTGACATTGAAAATAGCCGTCTGATTGTGGC AAACCCTACGCATATTACCATCGGCATCTACTTCAAACCGGAACTGATGCCG ATTCCAATGATTAGCGTCTATGAAACCAATCAACGCGCGCTGGCGGTCCGT GCGTACGCCGAGAAAGTGGGTGTTCCGTTATTGTAGACATCAAGCTGGCG CGCAGCCTGTTCAAAACGCACCGTCTGTTACGACCTGGTGAGCCTGGAGGAG ATCGACGAGTTTTGCGCTGCTGGTTTGGTTGGAAGAGTCAAAAACGCA GGCAAGGATGTGATCCAACCGCAGGAGAATGAAGTGCGTCATTAA
SEQ ID NO:154	Syn_PrgOrg Operon	GCATGCATTAAAGAGGAGAAATTAAGCATGGCTACTCCGTGGTCTGGTTAT CTGGATGATGTTTCTGCAAAATTTGACACGGGTGTTGACAACCTTGCAACCC AAGTTACCGAAGCCCTGGACAAGCTGGCTGCGAAGCCGTCGATCCGGCGC TGCTGGCGCGTATCAATCCAACTGTCTGAGTATAACTTGTACCGTAATGC GCAGTCGAACACGGTCAAGGTCTTCAAAGATATTGATGCAGCGATTATTCA GAACTTTCGTTAGTAACCTAGGATTAAGGAGGATAAATTAAGCATGTCTATC GCGACTATCGTGCTGAAAAATGCCGTTATTGGTCAAGCGGTGAATATTCGCA GCATGGAACCGACATCGTCAGCTTGGACGACCGTTTGCTGCAAGCATTTTC GGGCAGCGCCATCGCTACCGCCGTCGATAAGCAGACCATTACCAATCGCAT TGAAGACCCTAATCTGGTTACCGATCCGAAGGAGCTGGCGATTAGCCAGGA AATGATTTCCGACTACAATCTGTACGTCAGCATGGTTAGCACCTGACGCGT AAGGGCGTTGGCGCTGTTGAGACTTTGCTGCGTTCTGATAGGCTAGCATTA AACAGGATAATAAGCATGGAACTAGCAAGGAGAAAAACGATTACGTCCCT GGTCCGTATATCGTTCGTCTGTTGAACTCGTCCCTGAACGGTTGTGAATTC

SEQ ID NO:	Nombre	Secuencia
		CGCTGCTGACTGGTCGTACGCTGTTTCGTCTGGGTCAGAGCGATGCTCTGAC CGCGTCTGGTCAGCTGCCGGACATTCCTGCCGACTCCTTCTTCATCCCGCTG GATCATGGCGGTGTTAATTTTCGAGATTCAAGTGGACACTGACGCGACGGAA ATCATCTGCACGAACTGAAAGAGGGCAACTCCGAGAGCCGCTCCGTGCAA CTGAACACCCCGATCCAAGTTGGTGAACGTGCTGATTTTGATTCTCCGGAGA GCGAGCCGTGGGTGCCGGAAACAGCCGGAGAAAGTTGGAAACTTCTGCGAAA AAGAACGAACCCGCTTTAAAAACGGCATCGTCGCGGCACGTGGCGGGTTTC TTTATCCTGGGTATCGGCACGGTTGGCACCTGTGGATTCTGAACTCGCGC AACGTCAAGCAGCCGAATTGGACAGCCTGTTGGGTGAGGAGAAGGAGCGTT TTCAGGTGCTGCCGGGTGCGGACAAGATGCTGTATGTCGCCCGCGCAAAACG AACGCGACACCTGTGGGCACGTCAAGTCCTGGCACGCGGCGATTACGATA AAAACGCACGCGTTATTAATGAAAATGAGGAGAATAAACGTATCAGCATCT GGCTGGACACGTATTATCCACAACCTGGCATATTACCGTATCCATTTTGATGA ACCACGTAAGCCGGTGTGTTGGCTGTCCCGTCAACGCAACACGATGAGCAA GAAAGAGCTGGAGGTGCTGTCCAGAAATTGCGTGCCTGATGCCGTACGC CGACAGCGTCAATAATTACTCTGATGGATGACGTGACCGCAGCAGGCCAAGC CGAGGCAGGTCTGAACAACAGGGCGCTGCCATACAGCCCGCGTAACCAAA AGGTGGTGTACGTTCTGTTATTCAGGGCGCCTTGGACGACGTTGAGATTCTG CGTGCGCGCCAGTTTGTGCGACTCCTATTATCGTACCTGGGGTGGTCTACG TTCAATTGCAATTGAATTGAAAGACGATTGGCTGAAAGGCCGCTCGTTCCA ATACGGTGCGGAAGGCTACATTAAGATGAGCCAGGTCATTGGTATTTTCC GTCTCCTGTGAATAGAAGCTTATTAACAGGATAATAAGCATGATCCGCCG TTACCTGTATACCTTCTGCTGGTTATGACTTGGCCGGCTGTAAAGATAAG GATCTGTGTAAGGCTTGGACCAAGAGCAAGCGAATGAGGTCAATTGCGGTT CTGCAAAATGCACAACATTGAGGCTAACAAGATTGATAGCGGCAAGCTGGGT TACAGCATTACCGTCGCGGAACCGGATTTACCGCGGCAGTGTATTGGATTA AGACCTACCAGTTGCCGCCTCGCCCGCTGTGAAAATCGCCCAAAATGTTTCC GGCAGACAGCCTGGTTAGCTCTCCGCGTGGGAGAAAGACGCTGTGACTC GGCGATTGAACAGCGCCTGGAGCAGTCGCTGCAAAACGATGGAAGGTGTTCT GTGCGCCCGTGTCCACATCAGCTATGATATTGATGCGGGCGGAGAACGGTCG TCCGCCTAAGCCCGGTGCACCTGTCCGGCTTGGCGGTGTATGAACGCGGTTCC CCTCTGGCCCATCAGATTTCCGATATTAAGCGCTTCTGAAAAACAGCTTCG CGGATGTTGACTATGATAACATCAGCGTGGTTCTGTCCGAGCGTAGCGAGC CACAGTTGCAGGCGCCGGGCACGCCGTTCAAGCGCAATAGCTTCGTACCT CCTGGATTGTGTGATTATCCTGTTGTCTGTTATGAGCGCGGGTTTCGGTGTG TGGTACTACAAAAATCACTATGCGCGTAATAAGAAAGGCATTACTGCCGAT GACAAGGCAAAAGTCCAGCAACGAGTAATAAGGTACCGAAAGAAGGGACAG ACTAGATGATTTCGTGTAACCGTCAAAATGAACCGTCAACCACTGCCAATTAT CTGGCAACGCATTATTTTCGACCCACTGTCCTATATTCACCCACAACGTCGT CAAATCGCGCCGGAGATGATCGTGCCTCCGGCAGCGCGTGCAGGAGCGAAT GAGCTGATTTTGGCGCGGTGGCGTTTGAAGAACGGTGAGAAGGAGTGCAATT CAGAATAGCCTGACGCACTGTGGTTGCGTCAATGGCGCCGTCTGCCGCAG GTTGCTTACCTGCTGGGTTGCCACAAGTTGCGTGCTGACCTGGCCCGTCAAG GTGCTTTATTGGGCTGCCGGACTGGGCGCAGGCATTCTTGGCGATGCACCA GGGTACGTCCTGTGCGGTTTGAATAAGGCGCCGAACACCGTTTCTGCTG TCCGTTGGTTACGCCCAACTGAACGCGCTGAATGAGTTCCTGCCGAGAGCT TGGCCCAACGCTTTCTCTGCTGTTTCCACCGTTTCATCGAGGAGGCACTGAA ACAAGATGCACTGGAGATGAGCATCTGCTGTGGCCCTGCAATACGCGCA AAAGTATCCTAACACCGTCCCGGCGTTTGCCTGCTAATAAAGATCTGAAAG AAGGGACAGACTAGATGTTGAAAAATATCCCGATTCCATCCCTCTGTCTCC GGTTGAAGGTATCCTGATTAAACGCAAGACGTTGGAGCGTTACTTCTCGATT GAGCGCTTGAACAACAGGCGCATCAGCGTGCAAGCGCATTTTGCCTGAG GCAGAAGAAGAAGCCAAGACCCTGCGCATGTATGCGTACCAAGGAGGGCTA CGAGCAGGGCATGATTGACGCACTGCAACAGGTGGCCGCCTATTTGACCGA CAACCAGACGATGGCTTGGAAATGGATGGAGAAAAATTCAAATCTATGCGCG TGAGTTGTTACGCGCGGCTGTGATCACCCGGAACGTTGTTGACGGTGCTG

SEQ ID NO:	Nombre	Secuencia
		GACGAGTGGCTGCGTGATTTCGATAAGCCGGAAGGTCAGCTGTTTTTGACCC TGCCGGTGAAACGCAAGAAAGATCATCAGAAACTGATGGTGTCTGCTGATGG AAAATTGGCCGGGCACCTTCAATCTGAAGTATCATCAGGAGCAACGTTTTAT CATGTCTCTGTGGCGATCAGATTGCCGAGTTTTCGCCGGAACAATTTGTTGAA ACGGCGGTTGGTGTATCAAGCACCATCTGGATGAGCTGCCTCAGGACTGTC GCACCATTTCCGACAAATGCGATTAAACGCGCTGATTGATGAATGGAAAGACGA AAACCCAAGCTGAAGTTATTTCGCTGATAAGGATCCGGCATCAATAAAACG AAAGGCTCAGTCGAAAGACTGGGCCCTTTCGTTTTATCTGTTGTTTTCGCGTG AACGCTCTCTGAGTAGGACAAATCCGCCGCCCTAGAACTAGTccataaagaatGC GGCCGC
SEQ ID NO:155	Syn_InvSpa_Operon	CTCGAGAAAAAAAGCTCTAAAAAGATTAAGAGGGGGTAACATATGGGTGA CGTGAGCGCGGTGAGCAGCAGCGGTAAACATTCTGCTGCCGACAGCAGGACGA GGTTGGTGGCTGTCCGAAGCGCTGAAGAAAGCGGTTGAAAAACACAAAAAC CGAATACAGCGGTGACAAGAAAGATCGTGATTATGGTGACGCTTTGTTAT GCACAAGGAAACCGCGCTGCCGTTGTTGCTGGCAGCTTGGCGCCACGGCGC ACCGGCGAAAAGCGAGCACCATAACGGTAACGTAAGCGGTCTGCATCACAA CGGTAAGAGCGAGCTGCGTATTGCTGAGAACTGCTGAAGGTGACGGCGGA GAAGAGCGTTGGTCTGATTAGCGCTGAAGCGAAGGTGGATAAATCTGCGGC GCTGCTGTCTAGCAAGAATCGTCCGCTGGAATCGGTACGCGGCAAAAAGTT GTCCGCCGATCTGAAAGCAGTGGAGTCCGTGTCCGAGGTACGGGATAACGC CACCGGCATTTCGGATGACAACATCAAAGCATTGCCGGGTGACAATAAGGC CATCGCCGGTGAGGGTGTGCGTAAAGAAAGGTGCGCCGCTGGCGCGTGACGT GGCTCCGGCAGCGATGGCGGCAGCAAAATACGGGCAAGCCGGAGGATAAAG ACCACAAGAAGGTCAAGGACGTTAGCCAGCTGCCGCTGCAACCGACTACCA TCGCCGATCTGTCTCAACTGACGGGTGGCGATGAAAAGATGCCGCTGGCAG CGCAGTCCAAACCGATGATGACCATTTTCCCAACCCCGACCGCGTTAAAG GTGAGGACAGCTCTCTGACCTATCGTTTCCAGCGCTGGGGCAACGATTACTC CGTCAATATCCAGGCACGCCAAGCGGGCGAATTAGCCTGATTCTTAGCAA TACCCAGGTTGAACATCGTCTGCACGACCAATGGCAAAATGGCAATCCACA ACGCTGGCATTTGACGCGTGATGACCAGCAAAACCCGCAACAGCAACAGCA TCGTACGAGTCCGGTGAAGAGGACGACGCGTAACCTGGGCAACGCGTCCAT TAAGAAGGAGGAATTAAGCATGGACGATCTGGTTTTCGCCGGCAACAAAGC CCTGTAATTTGGTGTGATTCTGTCCGTTGGCCGACGATTGTGCGAACCAATT ATCGGTCTGCTGGTTGGTCTGTTTCAAACCGTGACGAGTTGACAGGAGCAAA CCCTGCCGTTCCGGTATCAAGCTGCTGGGTGTGTGCTGTGCTGTTTTGCTG TCTGGCTGGTATGGCGAAGTTCTGCTGTCTACGGCCGTCAGGTATCTTCC TGGCTCTGGCGAAAGGTTAAGACGTCCTTGGGCACGCGTCCATTAACAGG AGTAATTAAGCATGCTGCTGTCCCTGCTGAATAGCGCGCTGTGCGTCTGA GCTGCTGATTCTGGTTCTGATGGTTATGATCATCAGCATGTTGTTATCCCGT TGCCGACCTAATTTGGTTGACTTCTTGATCGCTTTGAACATTGTCTGGCAATT CTGGTGTTCATGGGCTCCTTCTACATCGACCGCATTCTGAGCTTCAGCACCT TTCCGGCGGTTCTGCTGATCAGACTCTGTTCCGTTTGGCACTGAGCATCAG CACCAGCCGCTGATCCTGATTGAAGCAGATGCGGGTGAGATCATCGCGAC CTTTGGTCAGTTTGTGATCGGTGACAGCCTGGCGGTTGGTTTCGTCGTATTCT CCATCGTGACGGTGGTGCAGTTTATCGTTATTACCAAGGGCAGCGAACGCTG GGCGGAGGTGCGCGCTCGCTTCAGCCTGGACGGCATGCCGGGTAAACAGAT GTCTATTGATGACAGACCTGAAAGCCGGCATTATTGATGCTGATGCAGCGCG CGAGCGCCGACGCTCTGGAGCGTGAAAGCCAACCTGTACGGTTCCCTCGA CGGTGCCATGAAGTTCAATTAAGGTGATGCGATTGCGGCGCATCTATCATC TTCGTTAACTTCATTGGCGGTATCAGCGTCGGTATGACCCGTCATGGTATGG ATCTGAGCAGCGCCCTGAGCACCTACACCATGCTGACGATTGGTGATGGTCT GGTTGCCCAAAATCCGGCGTTGCTGATCGCGATTCTGCGGCTTCATCGTT ACCCGCGTCAACGGTGATAGCGATAACATGGGTGTAACATTATGACCCAG CTGCTGAATAATCCGTTTGTCTGGTTGTAACGCGGATTTTGACCATCAGCA TGGGCACGCTGCCGGGCTTCCGTTGCCGGTTTTCGTTATCTGCTGTGTG

SEQ ID NO:	Nombre	Secuencia
		CTGTCCGTCCTGTTTTACTTTAAGTTCCGCGAGGCGAAACGTAGCGCTGCGA AACCAAAAACGAGCAAGGGCGAGCAACCGTTGTCCATCGAGGAGAAGGAA GGTAGCAGCCTGGGCCCTGATTGGCGACCTGGATAAGGTTAGCACGGAACCC GTCCCGCTGATTTTGTCTGGTGCCGAAATCGCGCTCGTGAGGATCTGGAGAAA GCGCAGCTGGCGGAACGTCTGCGCAGCCAATTCTTTATTGATTATGGTGTGC GTCTGCCAGAAGTACTGCTGCGTGACGGTGAGGGTCTGGATGACAACCTCTA TCGTCCTGCTGATTAATGAGATTCGCGTTGAACAGTTTACTGTCTATTTTGAC CTGATGCGTGTGGTTAACTACAGCGACGAGGTGGTGAGCTTTGGCATCAAC CCGACCATTACAGCAAGGTTCCAGCCAGTACTTTTGGGTGACCCATGAG GAAGGCGAAAAGCTGCGCGAGCTGGGCTACGTCTGCGTAATGCACTGGAC GAACTGTACCACTGTCTGGCGGTGACGCTGGCAGCAATGTGAACGAGTAT TTCGGTATCCAAGAAAACGAAACACATGCTGGACCAACTGGAAGCAAAAGTTT CCTGACCTGCTGAAGGAGGTTTTTGGCGCCACGCCACCGTGCAGCGCATTTTCGG AAGTGCTGCACGCTCTGCTGTCCGAACGCGTGAGCGTCCGTAACATGAAGC TGATCATGGAAGCCCTGGCACTGTGGGCTCCGCGTGAGAAAAGATGTGATCA ATCTGGTGGAGCACATCCGTGGTGCGATGGCGCGTTATATCTGCCACAAGTT CGCAAATGGTGGTGAACGCGTGCCGTTATGGTTTCCGCGCAAGTTGAGGA TGTCAATTCGTAAAGGCATTCTGCAAACTTCTGGCTCCACCTTTTGTAGCTTG GACCCGGAGGCTTCGGCAAACTCTGATGGACCTGATCAGCTGAAGCTGGAC GACCTGTTGATTGCGCATAAGGACCTGGTCTGTTGACCAGCGTTGACGTGC GTCGTTTTATCAAGAAAATGATTGAAGGTCGTTTTCCGGATCTGGAGGTCCT GTCCTTCGGTGAGATTGCAGATAGCAAAAGCGTGAATGTCATCAAAACCAT CTGAATTTAAATCTTGGGCACGCGTCCATTAAGGAGTAATTAAGCATG AGCTTCAGCGAGAGCCGCCACAATGAAAACCTGCTGATTCAAGAAGGCGCA CTGCTGTTTTGTGAGCAAGCAGTCTGGCGCCTGTGACGCGTGATCTGGTTT TTCGTCCGCTGAAAATCGAGGTCCTGAGCAAGCTGCTGGCGTTTCATCGAGC GCGCAGGTCTGGTGGATACGACCTACGCGGAGTCGGACAATGGGTTCTGC TGCTCTCCGAGGTTCCGTGCTATTGGGCAAGACCGTAAACGTTGCGAATATTG GTTTTTGACAGAGATTATCACCCATCTCCGGCGTTCAACAAGGTTCTGGCA CTGTTGCGTAAGAGCGAAAAGCTATTGGTTGGTGGGCTACTTGTGGCCAAA GCACGACCGCAATATATGCGTATGTTGGGTGAGGATTACGGTGTAGCT ACACGCATTTCCGCCGTCTGTGACGCCGCGCTCTGGGCGGTAAGGCGAAAA GCGAGCTGCGCAATTGGCGCATGGCCAGTCCCTGCTGAATAGCGTTGGAAG GTCATGAAAACATCACCCAGCTGGCGGTCAACCACGGTTATAGCAGCCCGT CCCACTTTAGCTCTGAAAATCAAGGAGCTGATTGGTGTTCCTCCGCGTAAGCT GTCTAACATCATTACGCTGGCCGACAAATGAATTTAAATGCTAGCCTTGGGC ACGCGTCCATTAAAGAGGACCAATTAAGCATGAAAACCCACGTCGTCTGC AATACCTGGCCTACCCGAGAAAATCACTGGCCCTATCATTGAAGCAGAAC TGCGTGATGTTGCAATTGGTGAATTGTGCGAGATCCGTGCGGCTGGCACCA GAAGCAGGTTGTGGCCCGTGCGCAAGTGGTTGGTTTGACAGCGCGAACGTAC CGTCCTGAGCCTGATCGGCAATGCCCAAGGCCTGAGCCGTGATGTGGTCTTG TACCCGACCGGCCGTGCTCTGAGCGCGTGGGTTGGTTACAGCGTTCTGGGCG CAGTACTGGACCCGACGGGTAAAATCGTTGAACGTTTACCCCGGAAGTCG CACCGATTTCCGAGGAGCGGTTATCGACGTGGCACCGCCGAGCTACGCAA GCCGTGTGCGGTGTGCGCAACCGTGTATCACGGGTGTCCGCGCAATTGATG GTCTGTGACGTGTGGTGTGGGCCAGCGTATGGGTATTTTCGCAAGCGCGG GTTGTGGTAAGACCATGTTGATGCACATGCTGATCGAGCAAAACGGAAGCGG ATGTCTTTGTTATTGGCCTGATTGGCGAGCGTGGTCTGTAAGTTACCGAATT TGTAAGACATGTTGCTGTCATCTCATAAGAAAAGAAAAATGTGTGCTGGTTTTT GCCACGTCCGACTTCCCAAGCGTTGACCGCTGCAACGCTGCCAGCTGGCA ACGACGGTCGCCGAGTATTTCCGCGACCAAGGTAACGTTGTTGCTCTGTTTA TCGACAGCATGACCCGTTATGACGCGCGTGTGCGTGTGTCGCGCTGGCGA GCGGCGAGCGTCCGGCTCGCCGTGGTTATCGGCGCTGTGTGTTGACAATCT GCCGCGTTTGTGGAGCGTCCGGGTGCGACGAGCGAGGGTAGCATTACCGC CTTCTATACCGTCTGCTGGAGTCGGAAGAAGAAGCGGACCCGATGGCGGA CGAGATCCGTTCTATTCTGGACGGTCACTGTACCTGTCCCGCAAACTGGCG

SEQ ID NO:	Nombre	Secuencia
		GGTCAGGGTCATTACCCGGCTATCGATGTGCTGAAGAGCGTGAGCCGTGTG TTTGGTCAAGTGACCACCCGACTCACGCGGAGCAAGCGAGCGCGGTCCGT AAGCTGATGACCCGTCTGGAAGAGCTGCAACTGTTTCATTGACCTGGGCGAG TATCGTCCGGGCGAGAACATTGACAATGATCGTGCGATGCAAAATGCGCGAT AGCCTGAAGGCGTGGCTGTGTCAGCCTGTTGCGCAATACAGCAGCTTCGAT GATACGCTGTCCGGCATGAACGCCCTTTGCGGATCAGAACTGACTTGGGCAC GCGTCCATGAAAGAAACGACATACTAGATGGGTAACGATATTAGCTTGATT GCATTGCTGGCGTTTTCACCCTGCTGCCGTTTCATCATCGCGTCTGGTACTTG CTTCGTCAAATTCAGCATCGTCTTTGTGATGGTGCGCAACGCGCTGGGTCTG CAACAAATTCCAAGCAATATGACCCTGAATGGCGTCGCACTGCTGCTGTCG ATGTTTGTATGTGGCCGATTATGCACGACGCGTATGTGTATTTTCGAGGATG AAGATGTGACCTTTAACGACATCTCCAGCCTGAGCAAGCATGTTGATGAGG GCCTGGACGGTTATCGCGACTACCTGATCAAGTATTCGACCGTGAGCTGGT GCAGTTCTTTGAGAAATGCCAGTTGAAACGTCAGTACGGTGAAGAAACGGA AACCGTTAAACGTGACAAGGACGAGATTGAAAAGCCGAGCATTTCGCACT GTTGCTGCTTACGCCCTTGAGCGAGATTAAAGAGCGCATTCAAAATTGGTTTT TACCTGTACTGCGCTTCGTTGTGGTCGATCTGGTTGTCTCCAGCGTCTGCT GGCCCTGGGCATGATGATGATGTCCCGGTTACCATCAGCACGCCGATCAA ACTGGTCTGTTTGTGGCCCTGGATGGTTGGACGCTGCTGTCTAAAGGCTG ATCCTGCAATACATGGACATCGCGACCTAAGAGCTCTGGGCACGCGTCCAT TAATGAGGAAAAATTATTAGCATGAAAACGCACATTCTGTTGGCCCGTGTG CTGGCTTGCGCAGCTCTGGTGTGGTCAACCCAGGTTATAGCTCCGAGAAGA TCCCGGTTACGGGCAGCGGCTTCGTTGCAAAAGGACGATTCTCTGCGCACCTT TTTCGATGCGATGGCACTGCAATTGAAGGAGCCGGTGATTGTACAGCAAGAT GGCGGCTCGCAAAAAGATTACCGCAATTTTCGAGTTCCACGATCCAAACGC GCTGCTGGAGAACTGTCCCTGCAACTGGGTCTGATCTGGTACTTTGATGGT CAGGCGATCTACATCTACGACGCGAGCGAAATGCGTAATGCGGTTGTGAGC CTGCGTAATGTGAGCCTGAACGAGTTCAACAATTTCTGAAGCGCAGCGGC CTGTACAATAAGAACTACCCCTGCGTGGTGATAATCGTAAAGGCACCTTCT ATGTCAGCGGTCCGCCGGTGTATGTTGATATGGTTGTAATGCGGCCACCAT GATGGACAAACAGAATGATGGCATCGAGCTGGGTGCGCAAAAGATCGGTGT TATGCGCTGAACAACACTTTTGTGGGCGACCGCACCTACAACCTGCGTGAT CAAAAGATGGTCATTCCGGGTATTGCTACGGCAATTGAACGCCGTGTTGCAA GGCGAAGAACAACCGCTGGGTAACATTGTAAGCTCCGAGCCTCCGGCCATG CCGGCCTTTAGCGCAAAACGGCGAGAAAGGTAAGGCAGCGAATTACGCGGGT GGTATGAGCCTGCAAGAAGCGCTGAAACAGAACGCAAGCGGCAAGCAACAT CAAAATTGTGGCCTATCCGGACACCAACAGCCTGCTGGTGAAAGGTACGGC GGAGCAGGTGCATTTATCGAGATGCTGGTTAAAGCCCTGGACGTGGCGAA ACGTCACGTTGAATTGAGCCTGTGGATTGTGGATTGTAATAAGAGCGACCT GGAACGCTTGGGCACCAAGCTGGAGCGGTAGCATCACCATCGGCGACAAGCT GGGTGTGAGCCTGAACAGAGCAGCATCTCCACGCTGGACGGTAGCGCTT TATTGGCGGGTCAACGCTCTGGAGGAAAAGAAACAGGCCACTGTTGTGAG CCGTCCGGTTCTGCTGACCCAGGAGAATGTCCCGGCGATTTTGTACAATAAT CGTACTTTCTATACAAACTGATCGGTGAACGTAATGTTGCAATTGGAACAG TGACCTATGACACCATGATCCGTGTCCTGCCGCGTTTCAGCGCGGACGGCCA GATTGAGATGAGCCTGGACATCGAAGATGGTAATGACAAAACCCCGCAGTC TGATACGACCACTCCGTTGATGCGCTGCCAGAAGTGGGTGCTACCTGATC TCGACGATTGCACGTGTCCCGCATGGTAAATCTCTGCTGGTTGGTGGCTACA CGCGTGATGCAAAACACGGACACGGTCCAAAGCATCCCGTTCCTGGGTAAGC TGCCGCTGATTGGCTCGTTGTTTCGCTACAGCAGCAAAAACAAGTCTAATGT CGTCCGTGCTTTATGATTGAGCCGAAAGAAATCGTTGACCCGCTGACCCCG GATGCCAGCGAGAGCGTTAAACAACATTCTGAAACAGTCCGGTGGCGGAGC GGTGACGATAAGCTGCAAAAAGTGGGTTCTGTGTGATTGGAACCGTGGTCAG GAGGCCATTAAAGTAACTTGGGCACGCGTCCATGAAAGACAGGACCCACTAG ATGTCCAGCAACAAAACCGAAAAACCGACTAAGAAACGTCTGGAGGATAG CGCAAAGAAAGGTCAGAGCTTCAAGAGCAAGGACCTGATTATCGCGTGCCT

SEQ ID NO:	Nombre	Secuencia
		GACCCTGGGTGGTATCGCTTATTTGGTGAGCTACGGCAGCTTCAATGAGTTT ATGGGTATCATTAAGATTATCATCGCTGATAAATTTGATCAGTCGATGGCAG ATTATAGCCTGGCCGTGTTTGGTATTGGCCTGAAATACCTGATTCCGTTTCAT GCTGCTGTGTTTGGTTTGTTCGCACTGCCGCACTGCTGCAAGCGGGCTTC GTTCTGGCAACCGAGGCCCTGAAGCCGAATCTGTCCGCCCTGAACCCGGTT GAAGGCGCGAAGAACTGTTTCCATGCGCACCGTCAAAGACACGGTCAAG ACGCTGCTGTATCTGTCGAGCTTTGTGGTTGCGGCAATCATTTGCTGGAAAA AGTATAAAGTCGAGATTTTCAGCCAACGGAACGGAATATCGTGGGTATTG CGGTTATCTGGCGTGAATTGCTGCTGGCGTTGGTTCTGACCTGCTGGCGTG CGCGCTGATCGTGCTTGTGCGGATGCTATTGCCGAGTACTTCTGACCATG AAAGATATGAAGATGGACAAAGAAGAAGTTAAACGCGAGATGAAAGAGCA GGAGGGTAACCCGGAGGTGAAGTCTAAACGTCGTGAAGTCCACATGGAAAT CCTGAGCGAACAAGTCAAGTCTGACATTGAAAAATAGCCGCTGATTGTGGC AAACCTACGCATATTACCATCGGCATCTACTTCAAACCGGAACTGATGCCG ATTCCAATGATTAGCGTCTATGAAACCAATCAACGCGCGCTGGCGGTCCGT GCGTACGCCGAGAAAGTGGGTGTTCCGGTTATTGTAGACATCAAGCTGGCG CGCAGCCTGTCAAAACGCACCGTCGTTACGACCTGGTGAGCCTGGAGGAG ATCGACGAGGTTTGGCGCTGCTGGTTTGGTTGGAAGAGGTCGAAAACGCA GGCAAGGATGTGATCCAAACGCGAGGAGAATGAAGTGCCTATTAACGTCAG GTTTAACTACTTGGGCACGCGTCCATTAATTAGGATCAATAGCATGATTCC GGGCAGCACCCTCCGGTATTTCCTTTAGCCGTATCCTGAGCCGTCAGACCTCC CACCAGGATGCAACCCAGCATACCGACGCACAACAAGCGGAAATTCACA AGCGCGGAAGATAGCTCGCCGGGTGCGGAGGTTACAGAAATTCGTCCAGAG CACGGACGAGATGTCTGCTGCGTTGGCGCAGTTCCGCAATCGCCGTGACTAT GAGAAAAAGAGCAGCAATTTGTCTAACTCCTTCGAGCGCGTTCTGGAGGAC GAGGCACTGCCGAAAGCGAAGCAGATTCTGAAACTGATCAGCGTGCATGGC GGTGCCTGGAGGATTTCCTGCGTCAGGCGCGCAGCCTGTTCCCGGACCCA AGCGATCTGGTGCTGGTTCTGCGCGAGCTGTTGCGTCGTAAGGACCTGGAG GAGATCGTGCCTAAGAAGCTGGAGAGCCTGCTGAAGCACGTGGAGGAACA AACCGACCCGAAAACCTGAAGGCCGTTATTAAGTGCAGCTGAAGGCGCG TCTGTTTGGCAAGACGCTGTCTCTGAAACCTGGTCTGCTGCGTGCCAGCTAC CGCCAGTTTATCCAAAGCGAAAGCCACGAAGTCGAGATTTACAGCGATTGG ATCGCCAGCTACGGTTATCAGCGTCGCCGTTGGTTGTTCTGGATTTCAATTGA GCAGCCTGCTGACTGACATCGATGCTAACGATGCAAGTGTCTCCGCTCTGG AGTTTGGCCAACTGCTGCGCCGCTGACCCAGCTGAAAATGTTGCGTAGCGC CGACCTGCTGTTTGTCTCGACGTTGCTGTCTTACAGCTTACGAAAAGCATTT AACGCTGAGGAGAGCAGCTGGCTGTTGCTGATGCTGTCTTTGCTGCAACAG CCGCACGAAGTGGATAGCTGCTGGCGGACATTATCGGCTCTGAATGCGCTG CTGTTGTCCACAAAAGAGCAGCCAGCTTCTGCAAAATCTTCTATCAGGTCT GTAAGGCAATCCCGTCTAGCCTGTTTATGAAGAGTACTGGCAAGAAGAAC TGCTGATGGCACTGCGCTCCATGACGGACATTGCTTACAAACACGAAATGG CCGAACAACGTCGTACCATCGAAAAGCTGTCTTAAGTTTAAACCTTGGGCA CGCGTCCATGAAAAGACAGGACCACTAGATGAGCTTGCGCGTACGCCAAAT TGACCGCCGTGAATGGCTGCTGGCCAGACTGCGACCGAATGTCAGCGTCA CGGTCGTGAGGCCACCTGGAGTATCCGACCCGTCAGGGTATGTGGGTCCG CCTGTCTGATGCCGAAAAACGCTGGTCTGCGTGGATCAAACCGGGTGATTG GTTGGAACACGTTAGCCAGCACTGGCGGGTGCCGCGTACGCGACGCGC GGAGCACCTGGTGGTTCCGTGGCTGGCGGCAACCGAACCGCCGTTGAGCT GCCGGTCCCGCACTTGAAGTGGCGTCTGTGCGTGGAGAACCCGGTTCCG GGTTCCGCACTGCTGAGGGCAAGCTGCTGCACATCATGTCCGATCGTGGT GGCCTGTGGTTTGAACATCTGCCGAGCTGCCGGCTGTTGGCGGTGGCCGTC CGAAGATGCTGCGTTGGCCGCTGCGTTTCTGTTATTGGCAGCAGCGACACCA GCGCAGCCTGCTGGGTGCTATCGGTATCGGTGATGTTCTGCTGATCCGCACC AGCCGTGCTGAGGTTTACTGTTACGCGAAGAACTGGGCCACTTTAATCGTG TGGAAGGTGGCATTATTGTCGAAACGCTGGACATTCAACATATCGAGGAGG AGAACAACACGACGGAAACGGCGGAAACCTGCCGGGTCTGAATCAACTG

SEQ ID NO:	Nombre	Secuencia
		CCGGTGAAGCTGGAGTTCGTTCTGTATCGTAAAAACGTGACGTTGGCCGAA CTGGAAGCAATGGGTCAGCAACAACCTGCTGTCCTTGCCAACCAATGCGGAA CTGAACGTCGAAATCATGGCGAATGGCGTGCTGGGTAAACGGCGAACTG GTGCAGATGAATGACACCCTGGGTGTCGAGATTATGAGTGGTTGTCCGAG AGCGGTAATGGCGAGTAGAGATCTCTTGGGCACGCGTCCATTAAGGAC CAATTAAGCATGCACTCTCTGACTCGTATCAAGGTCCTGCAACGTCGTTGTA CCGTGTTCCATTCTCAGTGCAGTCCATTCTGTGCGTTATCAGGACGAAGA TCGCGGCTTGCAGGCGGAGGAGGAGGCCATCCTGGAACAGATCGCAGGTCT GAAACTGCTGTGGACACCCTGCGTGCTGAAAAATCGTCAACTGAGCCGTA AGAAATCTATACGCTGCTGCGCAAAACAGAGCATTGTTGCTGCCAGATTAA GGATCTGGAGCTGCAAAATCATTAGATTCAAGAAAAGCGTAGCGAGCTGGA AAAGAAACGTGAGGAATTTCAAAAAGAAAAGCAAATACTGGCTGCGTAAAG AAGGTAACTACCAGCGCTGGATTATCCGTCAAAAACGCTTCTACATTCAAC GTGAGATCCAGCAAGAGGAGGCGGAGAGCGAAGAGATCATTTAACTGGG CACGCGTCCATGAAAGACAGGACCCACTAGATGTTCTATGCATTGTAATTTG AGATCCACCATCTGGTGGCGTCCGCGGCTCTGGGTTTTGCGCGTGTTGCTCC GATCTTTTTCTTCTGCGGTTCTGAATAGCGGTGCTCTGAGCGGTGCTCCGC GCAACGCGATCATATTCTGGTGGCGCTGGGTGTATGGCCGCACGCCCTGA ATGAGGCGCCACCGTTTCTGTCTGTGGCAATGATTCCGCTGGTCTGCAAGA GGCAGCCGTGGGTGTTATGCTGGGTGCTGTTGTCCTGGCCGTTTTGGGTT ATGCACGCGTTGGGCTGTATCATTGATAACCAACGCGGTGCAACCCTGTCCA GCAGCATTGATCCTGCGAATGGCATCGACACCAGCGAGATGGCCAATTTCC TGAACATGTTGCGGCTGTCGTGTAATTTGCAGAACGGTGGCTTGGTCAAGAT GGTGGACGTGCTGAATAAGTCTTACCAGCTGTGTGATCCGATGAACGAGTG TACGCCGAGCCTGCCCTCCGTTGCTGACCTTCATTAATCAAGTGGCCAGAAC GCACTGGTGTGGCGTCCCGGTCGTTCTGGTTCTGCTGTGTCGGAAGTTT TCCTGGGCCTGCTGTCTCGTTTTCACCGCAAAATGAACGCGTTTCGCCATTAG CCTGACTGTAAAAAGCGGTATTGCGGTTTGTATCATGCTGCTGTATTTCAGC CCGGTCTGCCGACAATGTTCTGCGTCTGAGCTTTCAGGCGACCGGCCTGA GCAGCTGGTTCTACGAACGTGGCGCAACGCATGTGCTGGAATAAGGATCCG GCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTAT CTGTTGTTTGTGCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGCCCTAG AACTAGTccataagaatGCGGCCGC

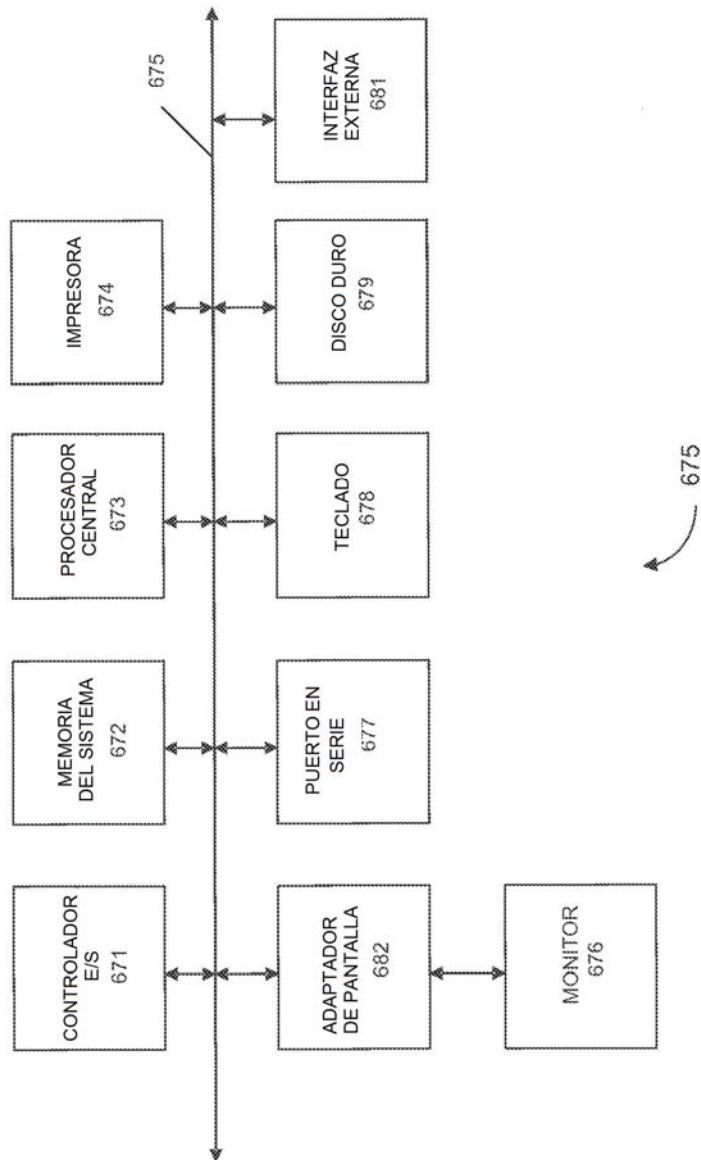


FIG. 25

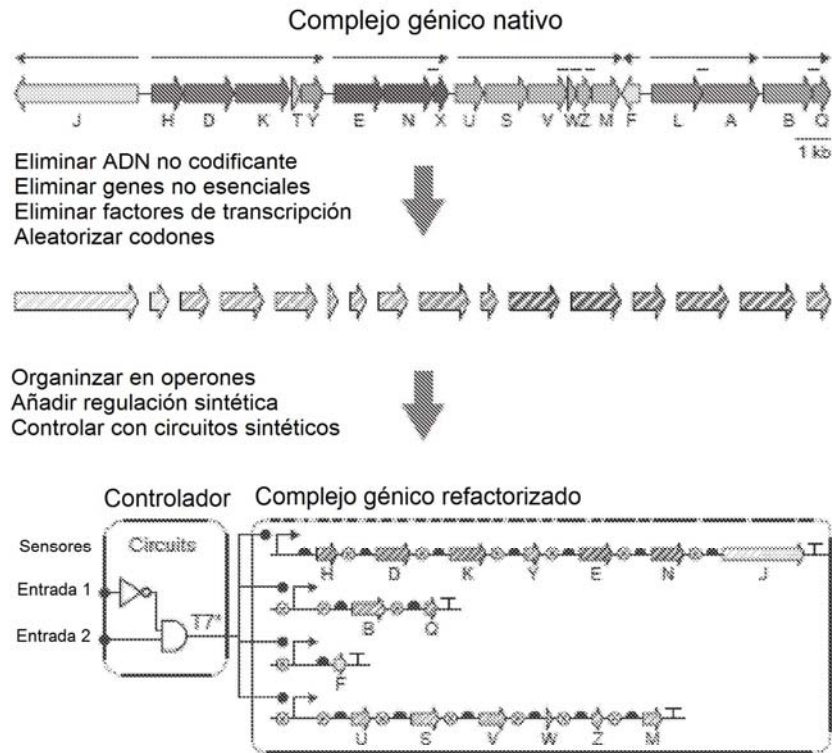


FIG. 26

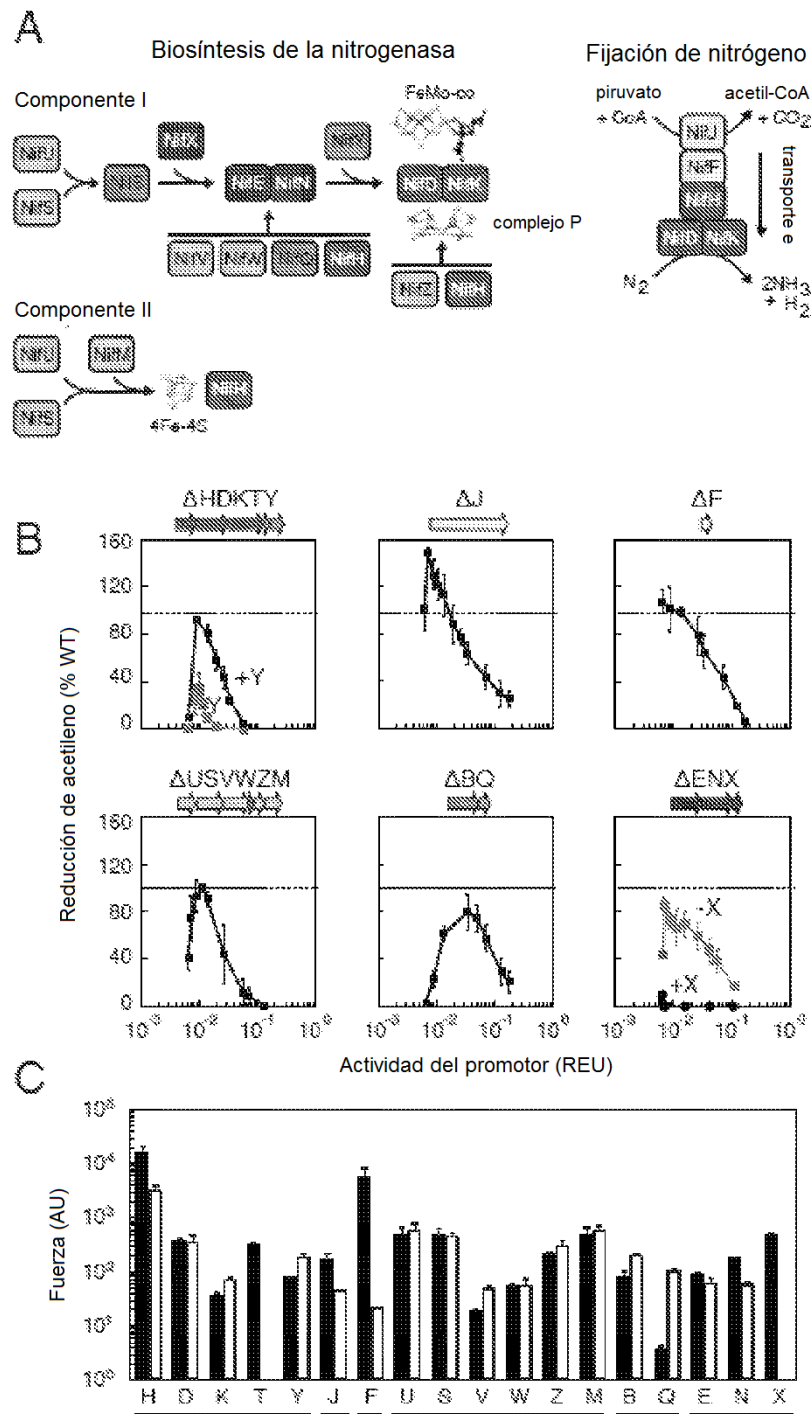


FIG. 27

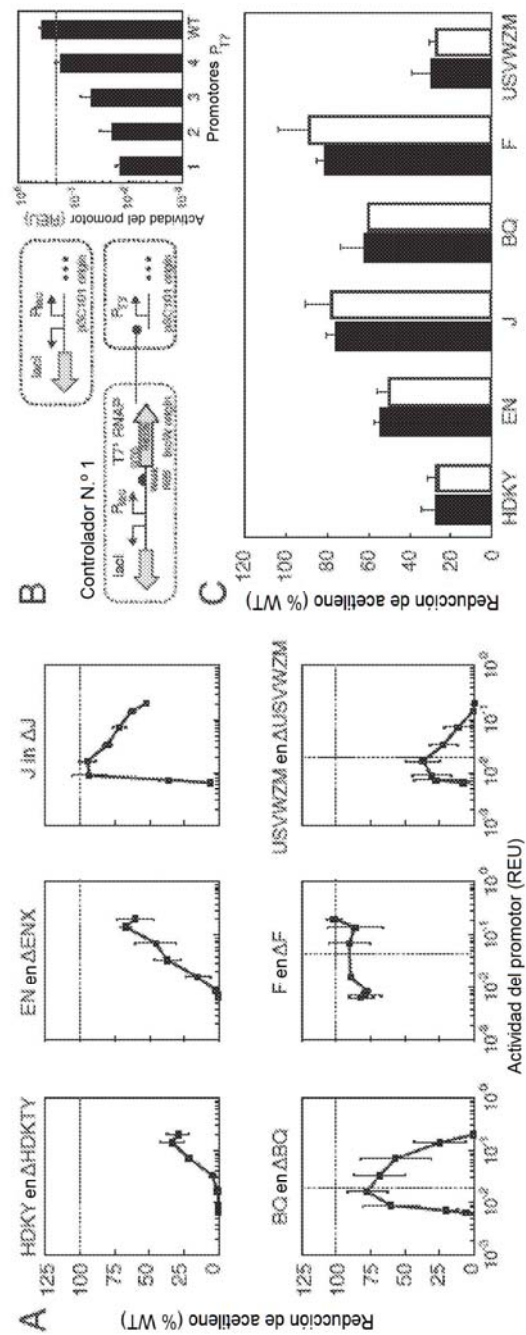


FIG. 28

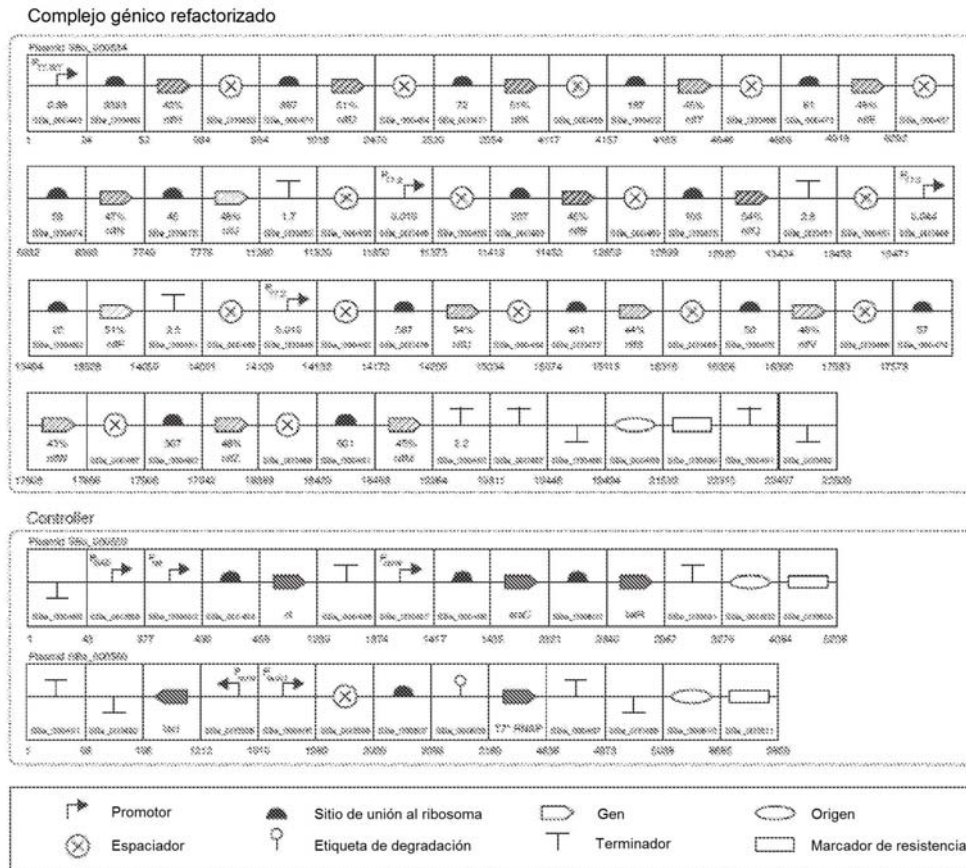


FIG. 29

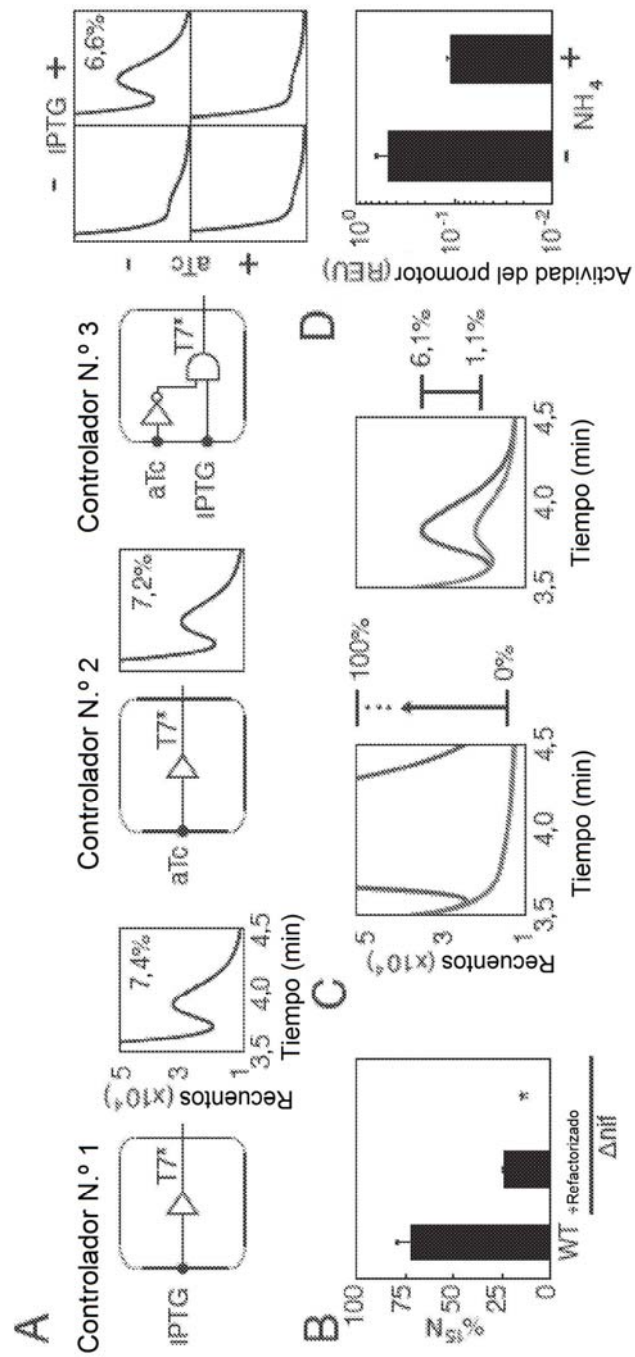


FIG. 30

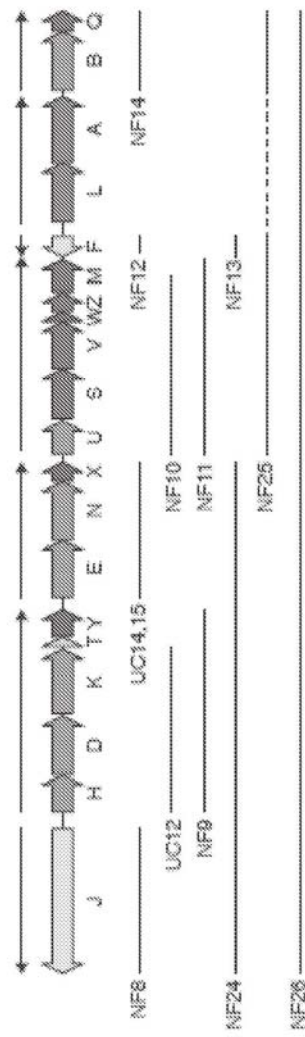


FIG. 31

FIG. 32

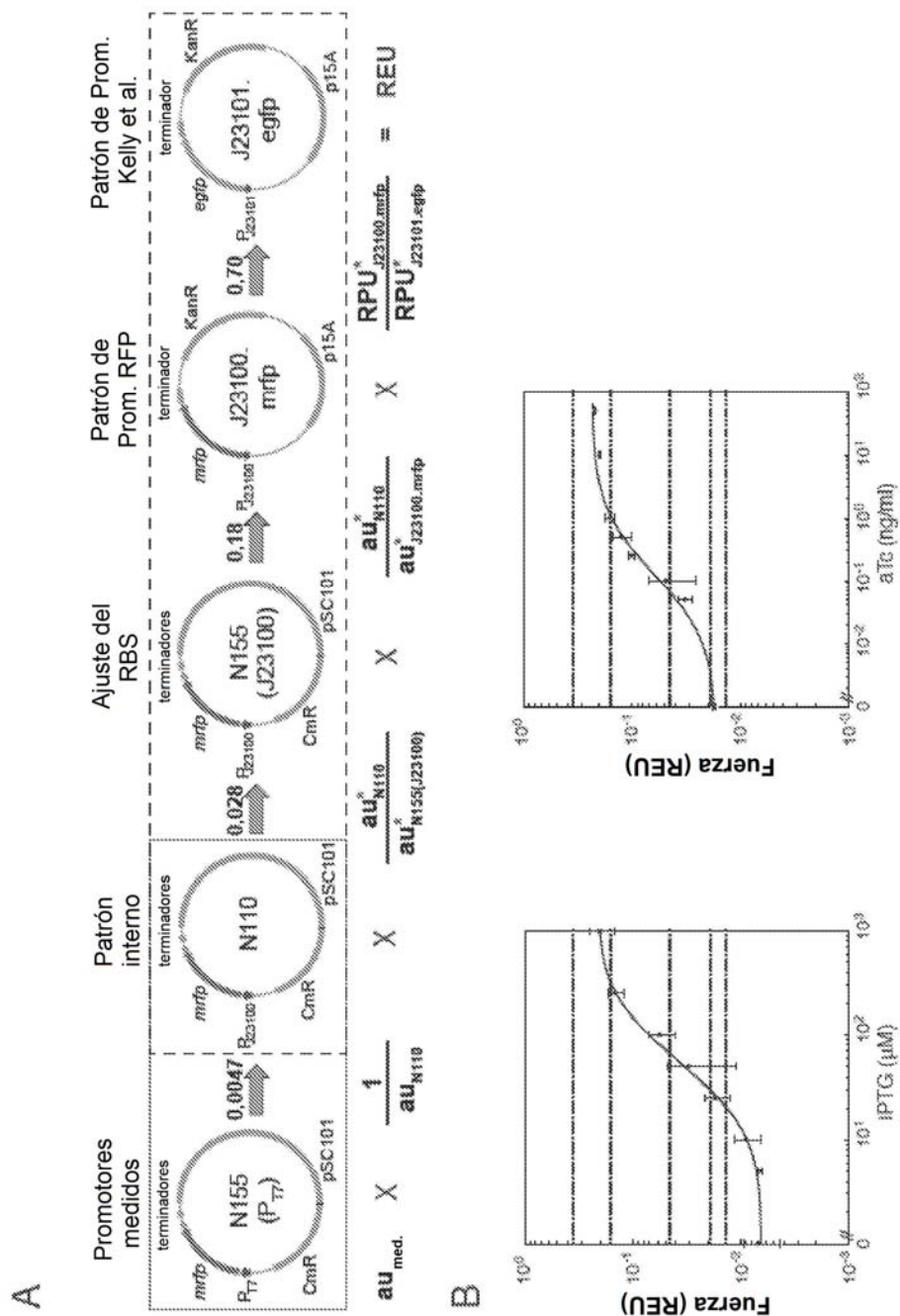
Tabla. Construcción y verificación de todos los mutantes de delección del gen <i>nif</i> de <i>K. oxytoca</i>					
Nombre	Genotipo	Nota	Cepa parental	Plásmido de intercambio alélico ⁴	Secuencia de ADN de la región de delección ⁵
NF8	$\Delta nifJ, Km$	delección de operón	M5aL	pNif16	gcaggagaactaaaggcccgGCATGC (Km cassette)
UC12	$\Delta nifHDK$	delección en marco	M5aL	pRHB296	cataaagggcagcgctggTGGTACccatcaggtgcccgcgtca
NF9	$\Delta nifHDKTY$	delección de operón	M5aL	pNif9	atggcccggcaggcgcaatGCATGCgacgctcttcccacgttac
UC14	$\Delta nifENX$	delección de operón	M5aL	pRHB294	gacgctcttcccacgttacGCATGCgacggaacccgcgcgtta
UC15	$\Delta nifENX$	delección de operón	M5aL	pRHB295	ctcttcccacgttacgctcGCATGCgacggaacccgcgcgtta
NF10	$\Delta nifUSVWZM$	delección de operón ¹	M5aL	pNif10	cggaacccgcgcgcgttagccgGCATGCatctttggcagcagacccag
NF11	$\Delta nifUSVWZM$	delección de operón	M5aL	pNif11	cggaacccgcgcgcgttagccgGCATGCagcctcggcggtaccggtt
NF12	$\Delta nifF$	delección de operón ²	M5aL	pNif12	actgggtatcctgatccagcGCATGCctttgcaactacccggccca
NF13	$\Delta nifF$	delección de operón	M5aL	pNif13	ccgttaacgctacagcagcGCATGCctttgcaactacccggccca
NF14	$\Delta nifBQ$	delección de operón	M5aL	pNif14	attcagggaacgcgggttgc-----atgtgattatgcgacgtctt
NF24	$\Delta nifJ, HDKTY, ENX$	delección de operón ³	M5aL	pNif32	gtcaaacaccagaatattgaGAGCTCgtgtgtttctgtgacaaagc
NF25	$\Delta nifJ, SVWZM, F, BQ$	delección de operón	NF14	pNif51	cggaacccgcgcgcgttagccgGCATGCctttgcaactacccggccca
NF26	$\Delta nifJ, HDKTY, ENX, USVWZM, F, LA, BQ$	delección completa de <i>nif</i> ³	NF24	pNif54	gtcaaacaccagaatattgaCTCGAGatgtgattatgcgacgtctt

1) 83bp de *nifM* 3' retenido2) 82bp de *nifF* 3' retenido3) 500bp de *nifJ* 3' retenido

4) El esqueleto de todos los plásmidos de intercambio alélico en pDS132(1).

5) La región en negrita muestra la secuencia del sitio de restricción que sustituye a la región de delección. Se muestran los 20 nucleótidos de la región de delección.

flanking each site



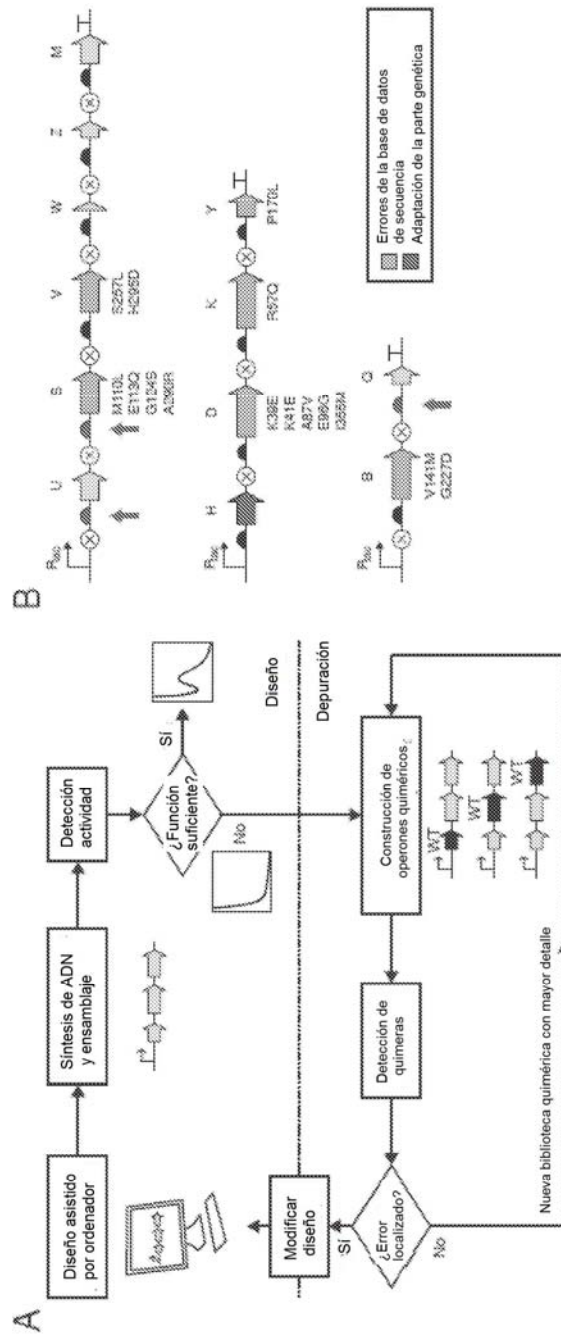


FIG. 34

FIG. 35

Tabla: Errores de secuencia del ADN en la secuencia del complejo <i>nif</i> X13303.1			
Localización	X13303.1	Este estudio	Impacto
5187	a	g	NiD K39E
5193	a	g	NiD K41E
5332	ct	tc	NiD A87V
5359	a	g	NiD E96G
6137	c	g	NiD B55M
6140	a	c	silente
6749	ga	ag	NiK R57Q
7371	t	c	silente
7371	t	c	silente
8168	cgg	ggc	<i>nifT</i> 5' UTR
8919	c	t	NiY P170L
9341	-	inserción t	<i>nifE</i> 5' UTR
9498	cg	gc	NiE A35G
9697	ac	ct	silente
9891	c	t	NiE P166L
12541	ca	ac	NiX Q131T
12770	t	c	<i>nifU</i> 5' UTR
13646	gc	cg	<i>nifS</i> 5' UTR
13999	a	c	NiS M110L
14008	g	c	NiS E113Q
14037	a	g	silente
14041	g	a	NiS G124S
14539	gc	cg	NiS A290R
15657	tc	ct	NiV S257L
15770	gc	cg	NiV H295D
16739	g	a	silente
16814	c	t	silente
22050	g	a	NiB V141M
22309	g	a	NiB G227D

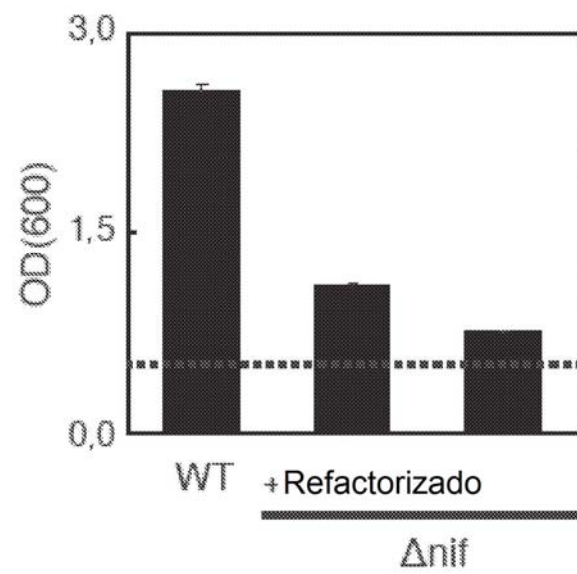


FIG. 36

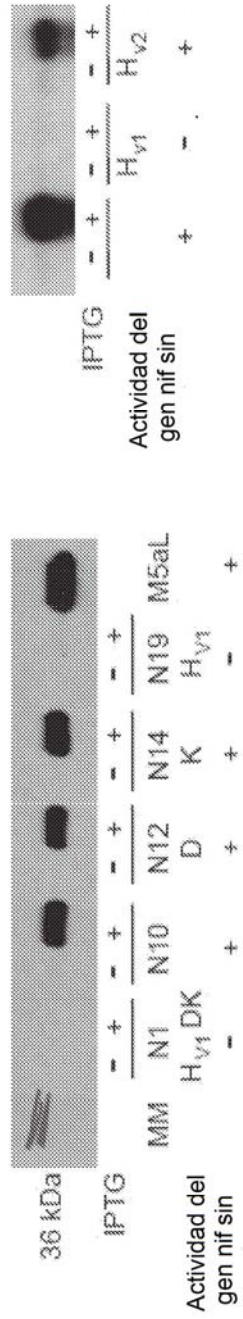


FIG. 37

FIG. 38

Tabla. Secuencias de ADN de partes sintéticas		
Número de parte	Tipo de parte	Secuencia
SBa_000443	Promotor de T7 WT	taatacgactcactatagggaga
SBa_000444	Promotor de T7 1	taatacgactcactacaggcaga
SBa_000445	Promotor de T7 2	taatacgactcactagagagaga
SBa_000446	Promotor de T7 3	taatacgactcactataggagaga
SBa_000447	Promotor de T7 4	taatacgactcactataggagaga
SBa_000450	Terminador de T7 WT	tagcataaaccccttggggcctctaaacgggtctttgaggggttttttgt
SBa_000451	Terminador de T7	tacatatcggggggttaggggttttttgt
SBa_000452	Terminador de T7	taacttaaaccccttggggcctcttaggggttttttgt
SBa_000453	Aislador	gggtggttacaccttaatacgcgcttcatgctaaaggtcctggctgcattgc
SBa_000454	Aislador	caaacaccccatgtcgatactgaacgaatcgaagcacaactcccttccttg
SBa_000455	Aislador	cctgattgtatccgcatctgctgctacgggtggttgagttta
SBa_000456	Aislador	caattttgccttgcgcacagacctctaacttaggttgccac
SBa_000457	Aislador	tgtcagctagggagcaattctataagaatgcacactgca
SBa_000458	Aislador	cctggttgagtgagctcagcagcagcggc
SBa_000459	Aislador	cgcgacttcacagagaagaagcactactgacttgagcgttcc
SBa_000460	Aislador	actacgagatttgaggttaaaccaataaagcagctagtggc
SBa_000461	Aislador	gtctgtagcagctgcattc
SBa_000462	Aislador	ggctattacaacggttat
SBa_000463	Aislador	aacatagcgttccatgagggctagaattacctacggcct
SBa_000464	Aislador	caattgtaataggcacaacaaagagtgatgatagctatgggt
SBa_000465	Aislador	gagttactggcctgatcttccgctttctaataccgcaca
SBa_000466	Aislador	gactcaacacgctagggacgtgaagtcgattccttcgatg
SBa_000467	Aislador	tcgagaaacaaaggcagttccgggctgaaagtagcgcggg
SBa_000468	Aislador	acgcacgcgttagtgagacataacagttcgttgggttcc
SBa_000508	Aislador	tgcagtttatctctgcgcagcactgttaagggactaa
SBa_000475	RBS sintético	ttaaagcggagaaattcagcATGaaactatggacggtaacgctgcggctgcattggattagctacgctttacgcgaagtggtctgcgctacacggattacgcgcagacacc
SBa_000469	RBS sintético	tactagagaaattcaactcaactcaaggagataaactATGacactgctcagtgccgagatttatg
SBa_000470	RBS sintético	gcacaaaggtggtattggcacaagcagcagcaccgaacttggggcgccctggccgagatg
SBa_000471	RBS sintético	gcgtgcgtacaccttaatacgcgcttcaatgctaaagtcctggctgcattgcacaaactccatccctactagggagcggcATGatgactaatgctactggcgaaagtaacctggcactgattca
SBa_000472	RBS sintético	agaagtagtggaaggtgtcccggaacccgcgcgaagaagagccgct
SBa_000473	RBS sintético	caaacaccccatgtcgatactgaacgaatcgaagcacaactcccttccttcgaaactcactactc
SBa_000474	RBS sintético	caaaaattagggagggaacATGctcacaactatcgataaaatcaactcttgtaacccgctgtt
SBa_000475	RBS sintético	cgagcaggacgaatcaggaactgttccgttaacaaacgtcagctg
SBa_000476	RBS sintético	cctgattgtatccgcattcgtatgctacgggtggttgagttaactactcactccggagggtaact
SBa_000477	RBS sintético	ctATGctcgaacataataacccgttttggcgcatgctggcgctgttccagtcgctgcgggattt
SBa_000478	RBS sintético	gcagcgcgctcaaatcgtcgtattggctg
SBa_000479	RBS sintético	caattttgccttgcgcacagacctctaacttagattgcacacacttccaatacaactccctgggggt
SBa_000480	RBS sintético	tattcaaaATGaaagggttaacagagattcttgcctgctgctggaacacggcctgtgaacacaaaca
SBa_000481	RBS sintético	taaacacgaataccggctgtagcgcacaaagcgc
SBa_000482	RBS sintético	tgtcagctagggagcaattctataagaatgcacactgcacactaaacacacacacacagggag
SBa_000483	RBS sintético	agtaattATGgcagacattttccgcactgataaagcgttggtgtgtgcgcgatcaagaacggc
SBa_000484	RBS sintético	cagcgcgtgggtgcgatcctggcgtccctgggt
SBa_000485	RBS sintético	aacatagcgttccatgagggctagaattacctacggcctcagatcctgcacaaatcaaacagcgc
SBa_000486	RBS sintético	aaagaggttccctATGtggaactacagcggagaaagtcaaggacattttctcaatccgcgcgaac
SBa_000487	RBS sintético	gcgcgtgttgtggataaacgcaaatgcggtggcgagctc
SBa_000488	RBS sintético	caattgtaatagcacacaaagagtgatgatagtcaggtgtatcccgtagacaaattctgaat
SBa_000489	RBS sintético	cgaaggaggttttccATGaaacaaagtgtacctggacaaacacgcgacacccgctggacccga
SBa_000490	RBS sintético	tgggttcggaagcagatgatgcgctttctcagcgtttctat
SBa_000491	RBS sintético	gagttactggcctgatcttccgcttctaataccgcacacgcgactaggagcctacactgcgcac
SBa_000492	RBS sintético	aaagaaactATGagcgcgtcttgatcaacgatactacccctggtgatgggacacactcctcg
SBa_000493	RBS sintético	ggcgtagcgtttcgtacctccgagaagttgcccato
SBa_000494	RBS sintético	gaactcaacacgctagggagctgaagtcgattccttcgatgtaggaagggcagaaactagcttaag
SBa_000495	RBS sintético	ggcctattatagATGagtggttttaccagattccgggttagacgaattgcgcagcgtgaatc
SBa_000496	RBS sintético	cttctttcagttcttgcgggttccataccagcggaa
SBa_000497	RBS sintético	tcgagaaacaaaggcagttccgggctgaaagttagcgcgggacaaagtcctgtatataacacgct
SBa_000498	RBS sintético	agggaggtgttgcATGgcgcgaataatcaactctctgaagaggtccgcgtagtttcgcgcgattc
SBa_000499	RBS sintético	gtaatgatggcacgcgtggcggttttgcgcaggtgcg
SBa_000500	RBS sintético	acgcacgcgttagtgagacatacagcttcttgggttcaactcagagactcaggttatttccag
SBa_000501	RBS sintético	gaaggtctataATGaatccgtggcagcgttttgcgcgtcaacgccttgcctcgcagcgttggaac
SBa_000502	RBS sintético	cgtgatccgcgtgctctgcacccagcgcgataccca
SBa_000503	RBS sintético	gaacaggtctcactatataaggggcttttccATGgcgaactcggcactccttcttggtacg
SBa_000504	RBS sintético	gataccgcgaacacccgcgaagattgcgaagatgattcaacaaacagctgggcagcgtgcgc
SBa_000505	RBS sintético	cgcgacttccgcgaagagagcactgacttgagcgttccctctcgttaactcaactcaactcaat
SBa_000506	RBS sintético	catagggagggcaaaATGactcttctgtcgtgctttctggcggtaaagcgtgcgcgtccggcgcg
SBa_000507	RBS sintético	atgactccgcgctgactccgctggtggcgcacaaagcagct
SBa_000508	RBS sintético	aaactacagatttgaggttaaaccaataaagcagctagtggtcatataagggagaaatttaagcAT
SBa_000509	RBS sintético	Gcgcgcaattggcaggttgcgtggtttgtgtgtactctatcagccgcgcaaaagcagcttccgc
SBa_000510	RBS sintético	cttcgtatggccttgcgcgcgct
SBa_000561	Promotor P _{rec} Y	caatttcgcgctcaacttacattaatfgcgttgcgctcactgcgcgcttccagtcgggaacctg

SBa_000562	Promotor P _{tr} y casete tetr	casete lacI
SBa_000484	Indicador fluorescente mRFP	
SBa_000514	ARNP de T7*	

[illegible]

nifK sintético

nifY sintético

nifE sintético

nifN sintético

		tgggtgcgaatcctggcgctccctgggtatcgagcaatcgattcccgctggtaaatggcgccaggg ctgttcggcttttgcgaaggttttctttatccagcaatcccgatccggctcccgctgcaaaagc acggcaatggacccgacccagcccatcattggcgctgattggttaacatctcccgccgtggaca ctctctgccaaacgaataaacccgcaagcaattgtgctgctgagcaacccgctctccgagcgca ggcgagcgacattcccgctgtagtgcgtcagttccgtgaagaatcccgctcataaagcgctg ggcattctgaatgttaacaccccggaattttacggtagcaaggagaacggctttcccgctgctc tggagctgtgattgaacagtggttcccgccagcccaacgtccggcgccagcgaatcgtccggt caatcttttggtagccatctctgtagcccgagcgaattgagtggtgctggcgctgtgctgag gccttcggtctgcacccgacattctgcccgtatcggtcagagcaatggacggccacctgctc agggtagctttcccgctgagcgaggcgccagccgcttggcccaaatccgagcagatgggcca gagcctttgctcttttgcgattggcgtagcctgcaacgtcgagcagcctgctggtcccggt tgtcgtggcgagctcatcgcttcccgcaacctcatgacattggaacgctcgagcagccttatcc atcaattggcgcaaatccagcgctcccgccgtccggagtggtggaaacccagcagcgtcagct gcaagcgcacatgatcgattgcaacatgtggtgcaagccagcagcagcagcagcagcagcagc ggcagcctgctggcagcgtggtgctgatttccgcaacctccaggtatgcaagccggctccatgg ttgctccgagcggctcatccgagcctgctgctgctgctgctgctgctgctgctgctgctgctg ggaggtatctccagcctcttatgcccacatccggcgccacttactggtgggaactccagcc cgtgatttagcagagcaattccgctgctgctgctgctgctgctgctgctgctgctgctgctg tggcgcaatttccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg gaaactgctccgtagcagccatcatcatctggtcatctccagcccgctgctgctgctgctgctg gaactctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg atgtggaactacagcgagaaagtcaaggacatttctcaatccgcaacccgctgctgctgctg ataacgcaaatccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg cgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg gagatcccaacccagcagatcccgactatctggagcgttaccgcccgaagatgactgctc tgtaatggcgccaggaagctctctgctgcccacttgttaacttccgctggaatcgtcggaag gagcaatgagcagggtaagctgctgctgcaagtgcttccgctgctgctgcaagcgaatctccg ctgctccagacaaacggctcttaccgctctgctgcccaggtgctgcaatcaacaaaggcaggtg ttgtaccagctgcaatgagaaatcgagctggcctggcggagattctccgcccacagccgcaa accaccccgccagltgctgctgggaagatccgcaatggcgagcgtcgtggaacacatcgtg aantcgctcccaatactccaaaggcagcgtggtgcaatggcgtgctgctgctgctgctgctg agtgactgcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ctgcaacagaaattgctgagcgtgctgctgctgctgctgctgctgctgctgctgctgctgctg atgaaacgaagtacatccgcaacccagcagcagcagcagcagcagcagcagcagcagcagc tgatccgcttctccagcagcttcttggcaatccgctccagcagcagcagcagcagcagcagc accagcggcgtgctggaacgtgctgcaacccagcagcagcagcagcagcagcagcagcagc gaaactcaatttccagcagctgctgcaacccagcagcagcagcagcagcagcagcagcagc tattgcccgaaccccgcaaatccatccatccgctgctgctgctgctgctgctgctgctgctg cgagc gacatggcccagctccgctgcaacgctctccgcccgtgctgctgctgctgctgctgctgctg accagcaaacccgctgctgcttccgctgctgctgctgctgctgctgctgctgctgctgctg tctgttccacccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg atcgcaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg taoetgctgcaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg cggcaatgagaaatctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ctgcccgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg tgccgtccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg cgagttatccgaagtgagcagctcctgctgctgctgctgctgctgctgctgctgctgctg ggctccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg acaacggcgtcctcagcagcagcttctgctgctgctgctgctgctgctgctgctgctgctg ctaagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg aatggttaagccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg atggagcggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg tctgctcctccgcaaatgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg agtcggcaatccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ccggaacgcaacgcttatacctgctgctgctgctgctgctgctgctgctgctgctgctgctg tgggtattgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg tgagcggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ctcaagctatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ccgaggtggcccaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ccgcttcaacacccgagccagctgctgctgctgctgctgctgctgctgctgctgctgctgctg cagcttcaacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg cgtcgtgcaacatccgctcctcgtgctgctgctgctgctgctgctgctgctgctgctgctg tgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg aattgcttccacccagctcgtgctgctgctgctgctgctgctgctgctgctgctgctgctg tcaatctatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg cgtcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg agctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg ggttgcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg atggagtggttttccagattccgggtgtagagcaattggcagcagcagcagcagcagcagc
	<i>nifU</i> sintético	
	<i>nifS</i> sintético	
	<i>nifV</i> sintético	
	<i>nifW</i> sintético	
	<i>nifZ</i> sintético	

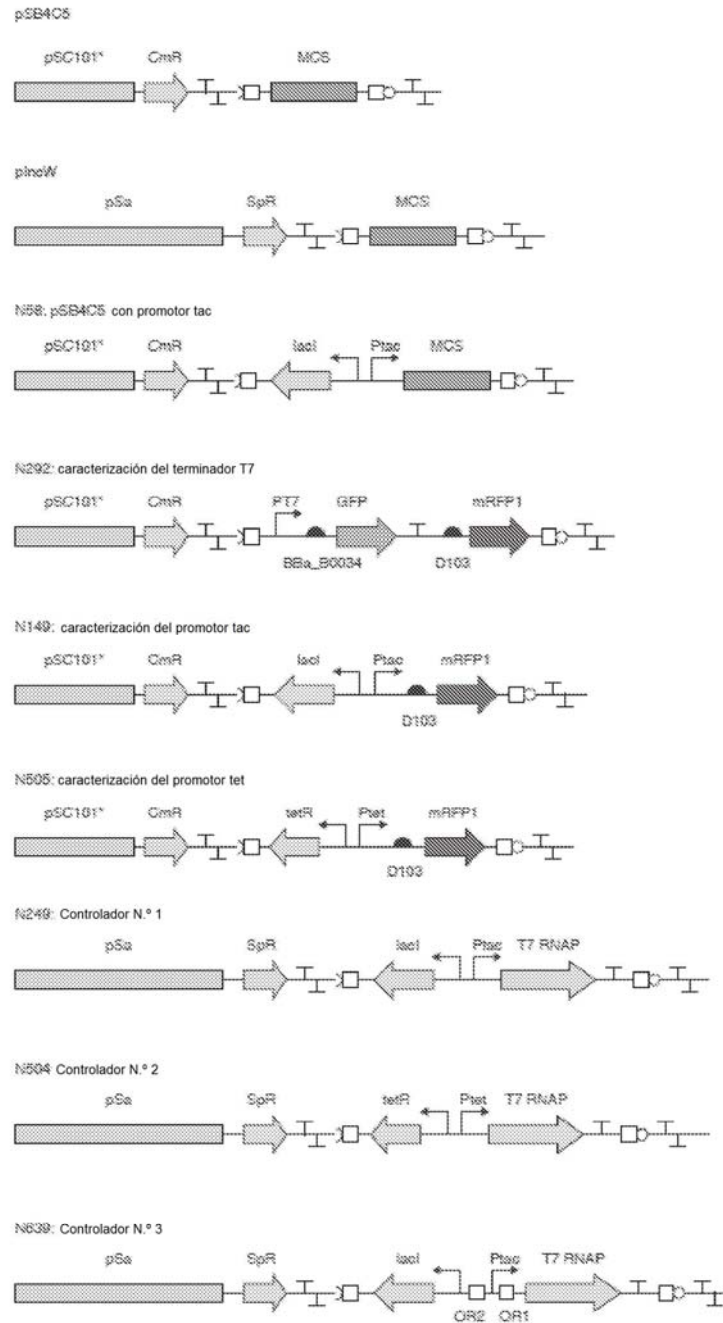


FIG. 39