

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 875 579**

51 Int. Cl.:

C12N 15/10 (2006.01)

C12N 15/73 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **06.06.2018 PCT/US2018/036333**

87 Fecha y número de publicación internacional: **13.12.2018 WO18226880**

96 Fecha de presentación y número de la solicitud europea: **06.06.2018 E 18734715 (8)**

97 Fecha y número de publicación de la concesión europea: **24.03.2021 EP 3485013**

54 Título: **Plataforma de ingeniería genómica de HTP para mejorar Escherichia coli**

30 Prioridad:

06.06.2017 US 201762515870 P

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

10.11.2021

73 Titular/es:

**ZYMERGEN, INC. (100.0%)
5980 Horton Street Suite 105
Emeryville, California 94608, US**

72 Inventor/es:

**DAVIS, MATTHEW;
WISNEWSKI, CHRISTY;
WESTFALL, PATRICK;
SERBER, ZACH;
DEAN, ERIK JEDEDIAH;
MANCHESTER, SHAWN;
GORA, KATHERINE;
SHELLMAN, ERIN;
KIMBALL, AARON;
SZYJKA, SHAWN;
FREWEN, BARBARA;
TREYNOR, THOMAS;
FLASHMAN, MICHAEL;
HAUSHALTER, ROBERT;
MORGAN, STACY-ANNE;
BLAISSE, MICHAEL;
RAMAKRISHNAN, PRABHA;
ROTHSCHILD-MANCINELLI, KYLE y
KIM, YOUNGNYUN**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 875 579 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

Plataforma de ingeniería genómica de HTP para mejorar *Escherichia coli*

REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

5 Esta solicitud reivindica el beneficio de prioridad a la Solicitud Provisional de EE.UU. Número de serie 62/515.870, presentada el 6 de junio de 2017.

CAMPO

10 La presente descripción se dirige a la ingeniería genómica microbiana de alta capacidad (HTP) para *Escherichia coli*. La plataforma de ingeniería genómica de HTP descrita está dirigida computacionalmente e integra biología molecular, automatización y protocolos avanzados de aprendizaje automático. Esta plataforma integradora usa una serie de conjuntos de herramientas moleculares de HTP para crear bibliotecas de diseño genético de HTP, que se obtienen, entre otras cosas, a partir de conocimientos científicos profundos y reconocimiento de patrones iterativos.

DECLARACIÓN RESPECTO A LA LISTA DE SECUENCIAS

15 La lista de secuencias asociada con esta solicitud se proporciona en formato de texto en lugar de una copia en papel. El nombre del archivo de texto que contiene la lista de secuencias es ZYMR_012_01WO_SeqList_ST25.txt. El archivo de texto es de ≈ 127 KB, se creó el 6 de junio de 2018 y se envía electrónicamente a través de EFS-Web.

ANTECEDENTES

20 Los seres humanos han estado aprovechando las posibilidades que ofrecen las rutas biosintéticas celulares microbianas durante milenios para producir productos de interés, cuyos ejemplos más antiguos incluyen alcohol, vinagre, queso y yogur. Estos productos todavía tienen una gran demanda en la actualidad y también han ido acompañados de un repertorio cada vez mayor de productos producibles por microbios. La llegada de la tecnología de la ingeniería genética ha permitido a los científicos diseñar y programar nuevas rutas biosintéticas en una variedad de organismos para producir una amplia variedad de productos industriales, médicos y de consumo. De hecho, los cultivos celulares microbianos se utilizan ahora para producir productos que van desde moléculas pequeñas, antibióticos, vacunas, insecticidas, enzimas, combustibles y productos químicos industriales.

25 Dada la gran cantidad de productos producidos por microbios industriales modernos, no es de extrañar que los ingenieros estén bajo una tremenda presión para mejorar la velocidad y la eficiencia con la que un microorganismo dado es capaz de producir un producto objetivo. Se ha utilizado una variedad de enfoques para mejorar la economía de los procedimientos industriales de base biológica "mejorando" el microorganismo implicado. Por ejemplo, muchas industrias farmacéuticas y químicas se basan en programas de mejora de cepas microbianas en los que las cepas parentales de un cultivo microbiano mutan continuamente mediante la exposición a sustancias químicas o radiación UV y, posteriormente se criban buscando aumentos de rendimiento, tales como la productividad, el rendimiento y el título. Este procedimiento de mutagénesis se repite extensamente hasta que una cepa demuestra un aumento adecuado en el rendimiento del producto. La siguiente cepa "mejorada" se usa después en la producción comercial. La identificación de cepas microbianas industriales mejoradas por mutagénesis requiere mucho tiempo y es ineficaz. El procedimiento, por su propia naturaleza, es asistemático y se basa en tropezar con una mutación que tiene un resultado deseable en la producción del producto. Los programas tradicionales de mejora de cepas microbianas no solo son ineficaces, sino que el procedimiento también puede dar lugar a cepas industriales con un alto grado de carga mutagénica perjudicial. La acumulación de mutaciones en cepas industriales sometidas a este tipo de programas puede llegar a ser significativa y puede conducir a un eventual estancamiento en la mejora de la tasa de rendimiento.

40 Quizás no hay mejor ejemplo del estancamiento que resulta de los programas de mejora de cepas tradicionales que con *E. coli*, que es uno de los sistemas hospedantes microbianos más modificados que existen. El microbio se ha sometido a los métodos tradicionales mencionados antes de mejora de la cepa microbiana durante décadas. A pesar de la gran cantidad de esfuerzo que se ha dedicado a la modificación de *E. coli*, el microbio todavía tiene una enorme cantidad de potencial sin explotar. Esto se debe a que *E. coli* presenta desafíos únicos para los investigadores que intentan mejorar el microbio con fines de producción. Estos desafíos han obstaculizado el campo de la ingeniería genómica en *E. coli* y han impedido a los investigadores aprovechar todo el potencial de este sistema microbiano.

45 En particular, la industria todavía no ha desarrollado un sistema de alta capacidad para la modificación genómica en *E. coli*. Está claro que los métodos tradicionales de mejora de cepas han alcanzado una meseta con respecto a este sistema de organismos, pero los investigadores aún no tienen las herramientas de modificación genómica que se necesitan para atravesar esta meseta.

50 Lee Michael E et al., *ASC Synthetic Biology*, 18 de septiembre de 2015, vol. 4, nº 9, páginas 975 a 986, describen una plataforma para la modificación de microbios (p. ej., levaduras) basada en un ensamblaje modular de múltiples plásmidos separados (p. ej., seleccionados de bibliotecas de elementos tales como plásmidos de parte del promotor o terminador) para formar una unidad transcripcional que permite la expresión de genes individuales en microbios.

Kincade JM et al., *Gene*, vol. 97, nº 1, 1 de enero de 1991, páginas 7 a 12, describen los promotores pL y pR del bacteriófago lambda determinantes de la secuencia de la actividad in vivo y la sensibilidad al inhibidor de la ADN girasa, cumermicina.

5 Rainer Knaus et al., *The EMBO Journal*, vol. 7, nº 9, 25 de agosto de 1988, páginas 2919-2923, describen diferentes secuencias de promotores químicas que comprenden secuencias de promotores pL y pR variantes, de aproximadamente 43 nucleótidos de longitud y que consisten en una parte distal del promotor pR del fago lambda, regiones variables -35 y -10 de los promotores pL y pR del fago lambda que tienen cada uno seis nucleótidos de longitud y porciones centrales de pL y pR del fago lambda.

10 El documento EP 2 657 250 A2 describe un método para mejorar la actividad de la acetil-coA sintetasa mediante la mejora de la expresión génica mediante modificación de secuencias de nucleótidos de la región promotora y la región 5'-UTR del gen asc.

El documento US 2006/141578 A1 describe un método de producción de una proteína deseada por recombinación génica. Los vectores usados comprenden la parte 5'-UTR/sitio de unión al ribosoma (RBS) del promotor del gen acs de *E. coli*.

15 William R McCleray, *Applied Microbiology and Biotechnology*, vol. 84, nº 4, 25 Julio de 2009, páginas 641-648, es un artículo de revisión que analiza la aplicación de técnicas de intercambio de promotores para controlar la expresión de genes cromosómicos.

Pengfei GU et al., *Microbial Cell Factories*, vo. 11, nº 1, 2 de marzo de 2012, página 30, describen el uso del intercambio de promotores para mejorar la producción de L-triptófano en *Escherichia coli*.

20 Rice Christopher D et al., *Applied and Enviromental Microbiology*, vol. 75, nº 3, febrero de 2009, páginas 573-583, describen que el uso de una técnica de intercambio de promotores muestra que PhoU modula la actividad del transportador de ABC PstSCAB₂ en *Escherichia coli*. Gueguen Erwan et al., *Applied and Environmental Microbiology*, vol. 79, nº 1, enero de 2013, páginas 32-38, describen el uso de intercambio de promotores para revelar la función del sistema de secreción de *Citrobacter rodentium* CTS1 tipo VI en la competencia interbacteriana.

25 Por lo tanto, existe una gran necesidad en la técnica de nuevos métodos de modificación de *E. coli* con fines de producción, que no tengan los inconvenientes antes mencionados inherentes a los programas tradicionales de mejora de cepas. Específicamente, un sistema de alta capacidad para descubrir y consolidar mutaciones beneficiosas en *E. coli* revolucionaría el campo y permitiría a los investigadores aprovechar todo el potencial de este organismo.

RESUMEN DE LA DESCRICCIÓN

30 La presente descripción proporciona una plataforma de ingeniería genómica de alta capacidad (HTP) para *E. coli* que no tiene la miríada de problemas asociados con los programas tradicionales de mejora de cepas microbianas.

Además, la plataforma de HTP enseñada en el presente documento es capaz de rehabilitar cepas de *E. coli* que han acumulado mutaciones no beneficiosas durante décadas de programas de mejora de cepas basados en mutagénesis aleatoria.

35 La descripción también proporciona conjuntos de herramientas y procedimientos de ingeniería genómica únicos, que respaldan la funcionalidad de la plataforma de HTP en un sistema de *E. coli*.

40 La plataforma de ingeniería genómica de HTP descrita se dirige computacionalmente e integra biología molecular, automatización y protocolos avanzados de aprendizaje automático. Esta plataforma integradora usa una serie de conjuntos de herramientas moleculares de HTP para crear bibliotecas de diseño genético de HTP, que se obtienen, entre otros, de conocimientos científicos profundos y reconocimiento de patrones iterativos.

45 Las bibliotecas de diseño genético de HTP enseñadas funcionan como conductoras del procedimiento de ingeniería genómica, proporcionando bibliotecas de alteraciones genómicas particulares para ensayar en *E. coli*. Los microbios modificados usando una biblioteca particular, o combinación de bibliotecas, se seleccionan de manera eficiente de una manera HTP para un efecto de resultado, p. ej. la producción de un producto de interés. Este procedimiento de utilizar las bibliotecas de diseño genético de HTP para definir alteraciones genómicas particulares para ensayar en un microbio y después posteriormente cribar genomas microbianos del hospedante que albergan las alteraciones se implementa de una manera eficiente e iterativa. En algunos ejemplos, el ciclo iterativo o "rondas" de campañas de ingeniería genómica puede ser al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más iteraciones/ciclos/rondas.

50 Por lo tanto, en algunos ejemplos, la presente descripción enseña métodos para llevar a cabo al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650,

675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000 o más "rondas" de ingeniería genética de HTP (p. ej., rondas de intercambio de SNP, intercambio de PRO, intercambio de STOP o combinaciones de los mismos) en un sistema hospedante de *E. coli*.

5 En algunos ejemplos, la presente descripción enseña un enfoque lineal, en el que cada ronda subsiguiente de ingeniería genética de HTP se basa en la variación genética identificada en la ronda previa de ingeniería genética. En otros ejemplos, la presente descripción enseña un enfoque no lineal, en el que cada ronda subsiguiente de ingeniería genética de HTP se basa en la variación genética identificada en cualquier ronda previa de ingeniería genética, incluyendo análisis realizados previamente y ramas de ingeniería genética de HTP separadas.

10 Los datos de estos ciclos iterativos permiten el análisis de datos a gran escala y el reconocimiento de patrones, que utiliza la plataforma integradora para informar a rondas posteriores de implementación de la biblioteca de diseño genético de HTP. En consecuencia, las bibliotecas de diseño genético de HTP utilizadas en la plataforma enseñada son herramientas altamente dinámicas que se benefician de algoritmos de reconocimiento de patrones de datos a gran escala y se vuelven más informativas por medio de cada ronda iterativa de modificación microbiana. Dicho sistema no se ha desarrollado nunca para *E. coli* y se necesita desesperadamente en la técnica.

15 En algunos ejemplos, las bibliotecas de diseño genético de la presente descripción comprenden al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975, 1000 o más cambios genéticos individuales (p. ej., al menos un número X de combinaciones de promotor:gen en la biblioteca de intercambio de PRO).

20 En algunos ejemplos, la presente descripción enseña un método de ingeniería genómica de alta capacidad (HTP) para transformar una cepa de *E. coli* para que adquiriera un fenotipo deseado, que comprende: a) perturbar los genomas de una pluralidad inicial de cepas de *E. coli* que tienen el mismo contexto de cepa, para crear así una biblioteca de cepas de *E. coli* de diseño genético de HTP inicial que comprende cepas individuales con variaciones genéticas únicas; b) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de diseño genético de HTP inicial según el fenotipo deseado; c) proporcionar una pluralidad subsiguiente de microbios de *E. coli* que comprende cada uno una combinación única de variación genética, seleccionada dicha variación genética de la variación genética presente en al menos dos cepas individuales de *E. coli* cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de diseño genético de HTP subsiguiente; d) cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de diseño genético de HTP subsiguiente según el fenotipo deseado; e) repetir las etapas c)-d) una o más veces, de forma lineal o no lineal, hasta que una cepa de *E. coli* haya adquirido el fenotipo deseado, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas de *E. coli* de diseño genético de HTP que comprende cepas de *E. coli* individuales que albergan variaciones genéticas únicas que son una combinación de variación genética seleccionada de entre al menos dos cepas de *E. coli* individuales de una biblioteca de cepas de *E. coli* de diseño genético de HTP precedente.

25 En algunos ejemplos, la presente descripción enseña que la biblioteca de cepas de *E. coli* de diseño genético de HTP inicial es al menos una seleccionada del grupo que consiste en una biblioteca de cepas microbianas de intercambio de promotor, biblioteca de cepas microbianas de intercambio de SNP, biblioteca de cepas microbianas de codones de inicio/parada, biblioteca de cepas microbianas de secuencia optimizada, una biblioteca de cepas microbianas de intercambio de terminadores, una biblioteca de cepas microbianas de marcadores de solubilidad de proteínas, una biblioteca de cepas microbianas de marcadores de degradación de proteínas o cualquier combinación de las mismas.

30 En algunos ejemplos, la presente descripción enseña métodos para hacer una pluralidad subsiguiente de cepas de *E. coli* que cada una comprende una combinación única de variaciones genéticas, en donde cada una de las variaciones genéticas combinadas se obtiene de la biblioteca de cepas de *E. coli* del diseño genético HTP inicial o la biblioteca de cepas de *E. coli* de diseño genético HTP de la etapa precedente.

35 En algunos ejemplos, la combinación de variaciones genéticas en la pluralidad subsiguiente de cepas de *E. coli* comprenderá un subconjunto de todas las posibles combinaciones de las variaciones genéticas en la biblioteca de cepas de *E. coli* de diseño genético de HTP inicial o la biblioteca de cepas de *E. coli* de diseño genético de HTP de la etapa precedente.

40 En algunos ejemplos, la presente descripción enseña que la biblioteca de cepas de *E. coli* de diseño genético de HTP subsiguiente es una biblioteca de cepas combinatoria completa derivada de las variaciones genéticas en la biblioteca de cepas de *E. coli* de diseño genético de HTP inicial o la biblioteca de cepas de *E. coli* de diseño genético de HTP de la etapa precedente.

45 Por ejemplo, si la biblioteca de cepas de *E. coli* de diseño genético de HTP previa solo tenía variaciones genéticas A, B, C y D, entonces una combinación parcial de dichas variaciones podría incluir una biblioteca de cepas de *E. coli* de diseño genético de HTP subsiguiente que comprende tres cepas comprendiendo cada una de las combinaciones únicas AB, AC o AD de variaciones genéticas (el orden en el que se representan las mutaciones no es importante).

Una biblioteca de cepas de *E. coli* combinatoria completa derivada de las variaciones genéticas de la biblioteca de diseño genético de HTP de la etapa precedente incluiría seis microbios, comprendiendo cada uno combinaciones únicas de variaciones genéticas AB, AC, AD, BC, BD o CD.

5 En algunos ejemplos, los métodos de la presente descripción enseñan la perturbación del genoma de *E. coli* utilizando al menos un método seleccionado del grupo que consiste en: mutagénesis aleatoria, inserciones de secuencias dirigidas, deleciones de secuencias dirigidas, reemplazos de secuencias dirigidas o cualquier combinación de los mismos.

En algunos ejemplos de los métodos descritos actualmente, la pluralidad inicial de *E. coli* comprende variaciones genéticas únicas derivadas de una cepa de *E. coli* de producción industrial.

10 En algunos ejemplos de los métodos descritos actualmente, la pluralidad inicial de *E. coli* comprende cepas de *E. coli* de producción industrial denominadas S₁Gen₁ y cualquier número de generaciones microbianas subsiguientes derivadas de las mismas denominadas S_nGen_n.

15 En algunos ejemplos, la presente descripción enseña un método para generar una biblioteca de cepas de *E. coli* de intercambio de SNP, que comprende las etapas de: a) proporcionar una cepa de *E. coli* de referencia y una segunda cepa de *E. coli*, en donde la segunda cepa de *E. coli* comprende una pluralidad de variaciones genéticas identificadas seleccionadas de polimorfismos de un solo nucleótido, inserciones de ADN y deleciones de ADN, que no están presentes en la cepa de referencia; b) perturbar el genoma de la cepa de referencia, o de la segunda cepa, para crear así una biblioteca de cepas de *E. coli* de intercambio de SNP inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas
20 individuales, en donde cada una de dichas variaciones genéticas únicas corresponde a una única variación genética seleccionada de la pluralidad de variaciones genéticas identificadas entre la cepa de referencia y la segunda cepa.

En algunos ejemplos de una biblioteca de intercambio de SNP, el genoma de la cepa de *E. coli* de referencia se perturba para añadir uno o más de los polimorfismos de un solo nucleótido, inserciones de ADN o deleciones de ADN identificados, que se encuentran en la segunda cepa de *E. coli*.

25 En algunos ejemplos de una biblioteca de intercambio de SNP, el genoma de la segunda cepa de *E. coli* se perturba para eliminar uno o más de los polimorfismos de un solo nucleótido, inserciones de ADN o deleciones de ADN identificados, que no se encuentran en la cepa de *E. coli* de referencia.

En algunos ejemplos, las variaciones genéticas de la biblioteca de intercambio de SNP comprenderán un subconjunto de todas las variaciones genéticas identificadas entre la cepa de *E. coli* de referencia y la segunda cepa de *E. coli*.

30 En algunos ejemplos, las variaciones genéticas de la biblioteca de intercambio de SNP comprenderán todas las variaciones genéticas identificadas, identificadas entre la cepa de *E. coli* de referencia y la segunda cepa de *E. coli*.

35 En algunos ejemplos, la presente descripción enseña un método para rehabilitar y mejorar el rendimiento fenotípico de una cepa de *E. coli* industrial, que comprende las etapas de: a) proporcionar una cepa de *E. coli* de linaje parental y una cepa de *E. coli* industrial derivada de la misma, en donde la cepa industrial comprende una pluralidad de variaciones genéticas identificadas seleccionadas de polimorfismos de un solo nucleótido, inserciones de ADN y deleciones de ADN, no presentes en la cepa de linaje parental; b) perturbar el genoma de la cepa de linaje parental o de la cepa industrial, para crear así una biblioteca de cepas de *E. coli* de intercambio de SNP inicial que comprende una pluralidad de cepas individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas individuales, en donde cada una de dichas variaciones genéticas únicas corresponde a una única
40 variación genética seleccionada de la pluralidad de variaciones genéticas identificadas entre la cepa de linaje parental y la cepa industrial; c) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de intercambio de SNP inicial para mejorar el rendimiento fenotípico frente a una cepa de *E. coli* de referencia, identificando así variaciones genéticas únicas que confieren a dichas cepas de *E. coli* mejoras en el rendimiento del fenotipo; d) proporcionar una pluralidad subsiguiente de cepas de *E. coli* que cada una comprende una combinación única de variación genética, seleccionada dicha variación genética de la variación genética presente en al menos dos cepas individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de SNP subsiguiente; e) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de intercambio de SNP subsiguiente para mejorar el rendimiento del fenotipo frente a la cepa de referencia, identificando así combinaciones
45 únicas de variación genética que confieren a dichas cepas de *E. coli* mejoras adicionales del rendimiento del fenotipo; y f) repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que una cepa presenta un nivel deseado de rendimiento del fenotipo mejorado en comparación con el rendimiento del fenotipo de la cepa industrial de *E. coli*, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas de *E. coli* de intercambio de SNP que comprende cepas microbianas individuales que albergan variaciones genéticas únicas que son una combinación de variación genética seleccionada de entre al menos dos cepas microbianas individuales de una biblioteca de cepas
50 de *E. coli* de intercambio de SNP precedente.

En algunos ejemplos, la presente descripción enseña métodos para rehabilitar y mejorar el rendimiento fenotípico de una cepa industrial de *E. coli*, en donde el genoma de la cepa de *E. coli* de linaje parental se perturba para añadir uno

o más de los polimorfismos de un solo nucleótido, ADN inserciones o deleciones de ADN identificados, que se encuentran en la cepa industrial de *E. coli*.

5 En algunos ejemplos, la presente descripción enseña métodos para rehabilitar y mejorar el rendimiento fenotípico de una cepa industrial de *E. coli*, en donde el genoma de la cepa industrial de *E. coli* se perturba para eliminar uno o más de los polimorfismos de un solo nucleótido, inserciones de ADN, o deleciones de ADN identificados, que no se encuentran en la cepa de *E. coli* del linaje parental.

10 En algunos ejemplos, la presente descripción enseña un método para generar una biblioteca de cepas de *E. coli* de intercambio de promotores, comprendiendo dicho método las etapas de: a) proporcionar una pluralidad de genes diana endógenos para una cepa de *E. coli* base, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa de *E. coli* base; b) modificar el genoma de la cepa de *E. coli* base, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprenda una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno de los promotores de la escalera de promotores unido operativamente a uno de los genes diana endógenos para la cepa de *E. coli* base.

20 En algunos ejemplos, la presente descripción enseña un método de intercambio de promotores de ingeniería genómica para transformar una cepa de *E. coli* para que adquiera un fenotipo deseado, comprendiendo dicho método las etapas de: a) proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*; b) modificar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprenda una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno de los promotores de la escalera de promotores operativamente unido a uno de los genes diana endógenos para la cepa base de *E. coli*; c) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores inicial para el fenotipo deseado; d) proporcionar una pluralidad subsiguiente de cepas de *E. coli* que cada una comprende una combinación única de variación genética, dicha variación genética seleccionada de la variación genética presente en al menos dos cepas individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente; e) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente para el fenotipo deseado; f) repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que un microbio haya adquirido el fenotipo deseado, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas de *E. coli* de intercambio de promotores que comprende cepas individuales que albergan variaciones genéticas únicas que son una combinación de variación genética seleccionada de entre al menos dos cepas individuales de una biblioteca de cepas de *E. coli* de intercambio de promotores precedente.

40 En algunos ejemplos, la presente descripción enseña un método para generar una biblioteca de cepas de *E. coli* de intercambio de terminadores, comprendiendo dicho método las etapas de: a) proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de terminadores, en donde dicha escalera de terminadores comprende una pluralidad de terminadores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*; b) modificar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de terminadores inicial que comprenda una pluralidad de cepas individuales con variaciones genéticas únicas que se encuentran dentro de cada cepa de dicha pluralidad de cepas individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno de los genes diana endógenos para la cepa base de *E. coli* operativamente unida a uno o más de los terminadores de la escalera de terminadores.

45 En algunos ejemplos, la presente descripción enseña un método de intercambio de terminadores de ingeniería genómica para transformar una cepa de *E. coli* para que adquiera un fenotipo deseado, comprendiendo dicho método las etapas de: a) proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de terminadores, en donde dicha escalera de terminadores comprende una pluralidad de terminadores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*; b) modificar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de terminadores inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno de los genes diana endógenos para la cepa base de *E. coli* operativamente unida a uno o más de los terminadores de la escalera de terminadores; c) cribar y seleccionar cepas microbianas individuales de la biblioteca de cepas de *E. coli* de intercambio de terminadores inicial para el fenotipo deseado; d) proporcionar una pluralidad subsiguiente de cepas de *E. coli* que cada una comprende una combinación única de variación genética, dicha variación genética seleccionada de la variación genética presente en al menos dos cepas individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de terminadores subsiguiente; e) cribar y seleccionar cepas individuales de la biblioteca de cepas de *E. coli* de intercambio de terminadores subsiguiente para el fenotipo deseado; f) repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que un microbio haya adquirido el fenotipo deseado, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas de *E. coli* de intercambio de terminadores que comprende cepas individuales que albergan variaciones genéticas únicas que son una combinación

de variación genética seleccionada de entre al menos dos cepas individuales de una biblioteca de cepas de *E. coli* de intercambio de terminadores precedente.

En algunos ejemplos, la presente descripción enseña la mejora iterativa del diseño de cepas candidatas de *E. coli* (a) accediendo a un modelo predictivo equipado con un conjunto de entrenamiento que comprende (1) entradas que representan cambios genéticos para una o más cepas de *E. coli* originales y (2) medidas de rendimiento correspondientes; (b) aplicar entradas de ensayo al modelo predictivo que representan cambios genéticos, correspondiendo las entradas de ensayo a cepas candidatas de *E. coli* que incorporan esos cambios genéticos; (c) predecir el rendimiento fenotípico de las cepas candidatas de *E. coli* basándose, al menos en parte, en el modelo predictivo; (d) seleccionar un primer subconjunto de las cepas candidatas de *E. coli* basándose, al menos en parte, en su comportamiento previsto; (e) obtener el rendimiento fenotípico medido del primer subconjunto de las cepas candidatas de *E. coli*; (f) obtener una selección de un segundo subconjunto de las cepas candidatas de *E. coli* basándose, al menos en parte, en su rendimiento fenotípico medido; (g) añadir al conjunto de entrenamiento del modelo predictivo (1) entradas correspondientes al segundo subconjunto seleccionado de cepas de *E. coli* candidatas, junto con (2) el rendimiento medido correspondiente del segundo subconjunto seleccionado de cepas candidatas de *E. coli*; y (h) repetir (b)-(g) hasta que el rendimiento fenotípico medido de al menos una cepa candidata de *E. coli* satisfaga un indicador de rendimiento. En algunos casos, durante una primera aplicación de entradas de ensayo al modelo predictivo, los cambios genéticos representados por las entradas de ensayo comprenden cambios genéticos en una o más cepas de *E. coli* originales; y durante aplicaciones posteriores de entradas de ensayo, los cambios genéticos representados por las entradas de ensayo comprenden cambios genéticos en cepas de *E. coli* candidatas dentro de un segundo subconjunto previamente seleccionado de cepas candidatas de *E. coli*.

En algunos ejemplos, la selección del primer subconjunto se puede basar en efectos epistáticos. Esto se puede lograr: durante una primera selección del primer subconjunto: determinando los grados de diferencia entre las medidas de rendimiento de una o más cepas de *E. coli* originales en respuesta a la aplicación de una pluralidad de entradas respectivas que representan cambios genéticos en una o más cepas de *E. coli* originales; y seleccionando para su inclusión en el primer subconjunto al menos dos cepas candidatas de *E. coli* basándose, al menos en parte, en los grados de diferencia en las medidas de rendimiento de una o más cepas de *E. coli* originales en respuesta a la aplicación de cambios genéticos incorporados en las al menos dos cepas candidatas de *E. coli*.

En algunos ejemplos, la presente descripción enseña la aplicación de efectos epistáticos en la mejora iterativa de cepas de *E. coli* candidatas, comprendiendo el método: obtener datos que representan el rendimiento medido en respuesta a los cambios genéticos correspondientes hechos en al menos una cepa de *E. coli* original; obtener una selección de al menos dos cambios genéticos basada, al menos en parte, en un grado de diferencia entre las correspondientes medidas de rendimiento sensibles de los al menos dos cambios genéticos, en donde el grado de diferencia se relaciona con el grado en que los al menos dos cambios genéticos afectan a sus correspondientes medidas de rendimiento sensibles a través de diferentes rutas biológicas; y diseñar cambios genéticos en una cepa de *E. coli* original que incluye los cambios genéticos seleccionados. En algunos casos, la cepa original de *E. coli* para la que se diseñan los al menos dos cambios genéticos seleccionados es la misma que la al menos una cepa de *E. coli* original para la que se obtuvieron datos que representan el rendimiento sensible medido.

En algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de la cepa de *E. coli* que utilizan solo un tipo único de biblioteca genética. Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que utilizan únicamente bibliotecas de intercambio de SNP. En otros ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que utilizan únicamente bibliotecas de intercambio de PRO. En algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* de que utilizan únicamente bibliotecas de intercambio de STOP. En algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que utilizan únicamente bibliotecas de intercambio de codones de inicio/parada.

En otros ejemplos, la presente descripción enseña métodos de mejora de HTP de la cepa de *E. coli* que utilizan dos o más tipos de bibliotecas genéticas. Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos de mejora HTP de la cepa de *E. coli* que combinan bibliotecas de intercambio de SNP y de intercambio de PRO. En algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que combinan bibliotecas de intercambio de SNP y de intercambio de STOP. En algunos ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que combinan bibliotecas de intercambio de PRO y de intercambio de STOP.

En otros ejemplos, la presente descripción enseña métodos de mejora de HTP de cepas de *E. coli* que utilizan múltiples tipos de bibliotecas genéticas (véase, por ejemplo, la Figura 5). En algunos ejemplos, las bibliotecas genéticas se combinan para producir mutaciones de combinación (p. ej., escaleras de combinación de promotores/terminadores aplicadas a uno o más genes). En otros ejemplos más, los métodos de mejora HTP de cepas de *E. coli* de la presente descripción se pueden combinar con uno o más métodos tradicionales de mejora de cepas.

En algunos ejemplos, los métodos de mejora de HTP de cepas de *E. coli* de la presente descripción dan como resultado una célula hospedante de *E. coli* mejorada. Es decir, la presente descripción enseña métodos para mejorar una o más propiedades de las células hospedantes de *E. coli*. En algunos ejemplos, la propiedad mejorada de la célula

hospedante de *E. coli* se selecciona del grupo que consiste en: productividad volumétrica, productividad específica, rendimiento o título, de un producto de interés producido por la célula hospedante de *E. coli*. En algunos ejemplos, la propiedad mejorada de la célula hospedante de *E. coli* es la productividad volumétrica. En algunos ejemplos, la propiedad mejorada de la célula hospedante de *E. coli* es la productividad específica. En algunos ejemplos, la propiedad mejorada de la célula hospedante de *E. coli* es el rendimiento.

En algunos ejemplos, los métodos de mejora HTP de cepas de *E. coli* de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de mejora de HTP de la cepa (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos). En algunos ejemplos, los métodos de mejora de HTP de cepas de *E. coli* de la presente descripción se seleccionan del grupo que consiste en intercambio de SNP, intercambio de PRO, intercambio de STOP, intercambio de MARCADORES DE SOLUBILIDAD, intercambio de MARCADORES DE DEGRADACIÓN y combinaciones de los mismos.

Por lo tanto, en algunos ejemplos, los métodos de intercambio de SNP de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de intercambio de SNP (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos).

Por lo tanto, en algunos ejemplos, los métodos de intercambio de PRO de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de intercambio de PRO (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos).

Por lo tanto, en algunos ejemplos, los métodos de intercambio de TERMINADORES de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de intercambio de TERMINADORES (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos).

Por lo tanto, en algunos ejemplos, los métodos de intercambio de MARCADORES DE SOLUBILIDAD de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de intercambio de MARCADORES DE SOLUBILIDAD (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos).

Por lo tanto, en algunos ejemplos, los métodos de intercambio de MARCADORES DE DEGRADACIÓN de la presente descripción dan como resultado una célula hospedante de *E. coli* que presenta un 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%,

- 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 31%, 32%, 33%, 34%, 35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100%, 150%, 200%, 250%, 300% o más de una mejora en al menos una propiedad de la célula hospedante de *E. coli* frente a una célula hospedante de *E. coli* de control que no se somete a los métodos de intercambio de MARCADORES DE DEGRADACIÓN (p. ej., una mejora de X% en el rendimiento o la productividad de una biomolécula de interés, incorporando cualquier intervalo y subintervalo entre ellos).
- 5
- 10 En algunos ejemplos, la presente descripción enseña un método para generar una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad de proteínas, que comprende las etapas de: a. proporcionar una pluralidad de genes diana endógenos para una cepa de *E. coli* base, y una escalera de marcadores de solubilidad, en donde dicha escalera de marcadores de solubilidad comprende una pluralidad de marcadores de solubilidad que presentan diferentes perfiles de solubilidad en la cepa base de *E. coli*; y b. modificar el genoma de la cepa base de *E. coli*, para
- 15 crear así una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los marcadores de solubilidad de la escalera de marcadores de solubilidad operativamente unidos a uno de los genes diana endógenos para la cepa base de *E. coli*.
- 20 En algunos ejemplos, la presente descripción enseña un método de intercambio de marcadores de solubilidad de proteínas para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción, que comprende las etapas de: proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de marcadores de solubilidad, en donde dicha escalera de marcadores de solubilidad comprende una pluralidad de marcadores de solubilidad que presentan diferentes perfiles de expresión en la cepa base de *E. coli*; modificar el genoma de la cepa
- 25 base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de las marcadores de solubilidad de la escalera de marcadores de solubilidad operativamente unidos a uno de los genes diana endógenos para la cepa base de *E. coli*; cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad inicial para mejorar el rendimiento fenotípico frente a una cepa de *E. coli* de referencia, identificando así variaciones genéticas únicas que confieren mejoras del rendimiento fenotípico; proporcionar una pluralidad subsiguiente de microbios de *E. coli* que comprenden cada uno una combinación de variaciones genéticas únicas de las variaciones genéticas presentes en al menos dos cepas de *E. coli* individuales cribadas en la etapa precedente, para crear así una biblioteca
- 30 de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente; cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente para mejorar el rendimiento fenotípico frente a la cepa de *E. coli* de referencia, identificando así combinaciones únicas de variación genética que confieren mejoras de rendimiento fenotípico adicionales; y repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que una cepa de *E. coli* presente un nivel deseado de rendimiento fenotípico mejorado en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas microbianas de cepas de *E. coli* de intercambio de marcadores de solubilidad, donde cada cepa de la nueva biblioteca comprende variaciones genéticas que son una combinación de variaciones genéticas seleccionadas de entre al menos dos cepas individuales de *E. coli* de una biblioteca precedente.
- 35
- 40 En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente es una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad inicial.
- 45
- En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente es un subconjunto de una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad inicial.
- 50 En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente es una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad precedente.
- 55 En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente es un subconjunto de una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad precedente.
- En algunos ejemplos, las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente presenta al menos un aumento de 10% en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.

En algunos ejemplos, las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de marcadores de solubilidad subsiguiente presenta al menos un aumento de una vez en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.

- 5 En algunos ejemplos, el rendimiento fenotípico mejorado de la etapa f) se selecciona del grupo que consiste en: productividad volumétrica de un producto de interés, productividad específica de un producto de interés, rendimiento de un producto de interés, título de un producto de interés, y combinaciones de los mismos.

- 10 En algunos ejemplos, el rendimiento fenotípico mejorado de la etapa f) es: producción mayor o más eficiente de un producto de interés, dicho producto de interés seleccionado del grupo que consiste en: una molécula pequeña, enzima, péptido, aminoácido, ácido orgánico, compuesto sintético, combustible, alcohol, metabolito extracelular primario, metabolito extracelular secundario, molécula de componente intracelular y combinaciones de los mismos.

- 15 En algunos ejemplos, la presente descripción enseña un método para generar una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación de proteínas, que comprende las etapas de: a. proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de marcadores de degradación, en donde dicha escalera de marcadores de degradación comprende una pluralidad de marcadores de degradación que presentan diferentes perfiles de solubilidad en la cepa base de *E. coli*; y b. modificar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas que se encuentran dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas
20 únicas comprende uno o más de los marcadores de degradación de la escalera de marcadores de degradación unidos operativamente a uno de los genes diana endógenos para la cepa de *E. coli* base.

- 25 En algunos ejemplos, la presente descripción enseña un método de intercambio de marcadores de degradación de proteínas para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción, que comprende las etapas de: proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de marcadores de degradación, en donde dicha escalera de marcadores de degradación comprende una pluralidad de marcadores de degradación que presentan diferentes perfiles de expresión en la cepa base de *E. coli*; modificar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas que se encuentran dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los marcadores de degradación de la escalera de marcadores de degradación unidos operativamente a uno de los genes diana endógenos para la cepa base de *E. coli*; cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación inicial para mejorar el rendimiento fenotípico frente a una cepa de *E. coli* de referencia, identificando así variaciones genéticas únicas que confieren mejoras del rendimiento fenotípico; proporcionar una pluralidad subsiguiente de microbios de *E. coli* que comprende cada uno una combinación de variaciones genéticas únicas de las variaciones genéticas presentes en al menos dos cepas de *E. coli* individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente; cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente para mejorar el rendimiento fenotípico frente a la cepa de *E. coli* de referencia, identificando así combinaciones únicas de variación genética que confieren mejoras adicionales del rendimiento fenotípico; y repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que una cepa de *E. coli* presenta un nivel deseado de rendimiento fenotípico mejorado en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas microbianas de cepas *E. coli* de intercambio de marcadores de degradación, donde cada cepa de la nueva biblioteca comprende variaciones genéticas que son una combinación de variaciones genéticas seleccionadas de entre al menos dos cepas individuales de *E. coli* de una biblioteca precedente.
45

En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente es una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación inicial.

- 50 En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente es un subconjunto de una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación inicial.

- 55 En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente es una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación precedente.

En algunos ejemplos, la biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente es un subconjunto de una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación precedente.

- En algunos ejemplos, las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente presenta al menos un aumento de 10% en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.
- 5 En algunos ejemplos, las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de marcadores de degradación subsiguiente presenta al menos un aumento de una vez en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.
- 10 En algunos ejemplos, el rendimiento fenotípico mejorado de la etapa f) se selecciona del grupo que consiste en: productividad volumétrica de un producto de interés, productividad específica de un producto de interés, rendimiento de un producto de interés, título de un producto de interés, y combinaciones de los mismos.
- 15 En algunos ejemplos, el rendimiento fenotípico mejorado de la etapa f) es: producción mayor o más eficiente de un producto de interés, dicho producto de interés seleccionado del grupo que consiste en: una molécula pequeña, enzima, péptido, aminoácido, ácido orgánico, compuesto sintético, combustible, alcohol, metabolito extracelular primario, metabolito extracelular secundario, molécula de componente intracelular y combinaciones de los mismos.
- 20 En algunos ejemplos, la presente descripción enseña un promotor quimérico sintético operativamente unido a un gen heterólogo para la expresión en una célula hospedante microbiana, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que tienen seis nucleótidos de longitud cada uno, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/Sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda.
- 25 En algunos ejemplos, las secuencias de ácido nucleico de la parte distal del promotor p_R del fago lambda, las regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda, las partes centrales de los promotores p_L y p_R del fago lambda y la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda se seleccionan de las secuencias de ácidos nucleicos que se encuentran en la Tabla 1.5.
- 30 En algunos ejemplos, la presente descripción enseña un promotor quimérico sintético operativamente unido a un gen heterólogo para la expresión en una célula hospedante microbiana, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que es cada uno de seis nucleótidos de longitud, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli*.
- 35 En algunos ejemplos, las secuencias de ácido nucleico de la parte distal del promotor p_R del fago lambda, las regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda, las partes centrales de los promotores p_L y p_R del fago lambda y la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli* se seleccionan de las secuencias de ácidos nucleicos que se encuentran en la Tabla 1.5.
- 40 En algunos ejemplos, el promotor quimérico sintético consiste en una secuencia de ácido nucleico seleccionada de las SEQ ID NOs. 132-152, 159-160, 162, 165, 174-175, 188, 190, 199-201 o 207.
- En algunos ejemplos, el promotor quimérico sintético consiste en una secuencia de ácido nucleico seleccionada de las SEQ ID NOs. 153-158, 161, 163-164, 166-173, 176-187, 189, 191-198 o 202-206.
- 40 En algunos ejemplos, la célula hospedante microbiana es *E. coli*.
- En algunos ejemplos, el gen heterólogo que codifica un producto proteína de interés que se encuentra en la Tabla 2.
- En algunos ejemplos, el gen heterólogo es un gen que es parte de una ruta biosintética de la lisina.
- 45 En algunos ejemplos, el gen heterólogo se selecciona del gen *asd*, el gen *ask*, el gen *hom*, el gen *dapA*, el gen *dapB*, el gen *dapD*, el gen *ddh*, el gen *argD*, el gen *dapE*, el gen *dapF*, el gen *lysA*, el gen *lysE*, el gen *zwf*, el gen *pgi*, el gen *ktk*, el gen *fbp*, el gen *ppc*, el gen *pck*, el gen *ddx*, el gen *pyc* o el gen *icd*.
- En algunos ejemplos, el gen heterólogo es un gen que es parte de una ruta biosintética del licopeno.
- En algunos ejemplos, el gen heterólogo se selecciona del gen *dxs*, el gen *ispC*, el gen *ispE*, el gen *ispD*, el gen *ispF*, el gen *ispG*, el gen *ispH*, el gen *idi*, el gen *ispA*, el gen *ispB*, el gen *crtE*, el gen *crtB*, el gen *crtI*, el gen *crtY*, el gen *ymgA*, el gen *dxr*, el gen *elbA*, el gen *gdhA*, el gen *appY*, el gen *elbB* o el gen *ymgB*.
- 50 En algunos ejemplos, el gen heterólogo codifica un producto biofarmacéutico o es un gen en una ruta para generar un producto biofarmacéutico.

- En algunos ejemplos, el producto biofarmacéutico se selecciona de humulina (rh insulina), intrónA (interferón alfa2b), roferón (interferón alfa2a), humatrope (hormona del crecimiento rh somatropina), neupogen (filgrastim), detaferón (interferón beta-1b), lispro (insulina de acción rápida), rapilysin (reteplasa), infergen (interferón alfacon-1), glucagón, beromun (tasonermina), ontak (denileuquina diftotox), lantus (insulina glargina de acción prolongada), kineret (anakinra), natrecor (nesiritida), somavert (pegvisomant), calcitonina (calcitonina recombinante de salmón), lucentis (ranibizumab), preotact (hormona paratiroidea humana), kyrstexal (rh urato oxidasa, PEGilada), nivestim (filgrastim, rhGCSF), voraxaze (glucarpidasa), o preos (hormona paratiroidea).
- En algunos ejemplos, la presente descripción enseña un gen heterólogo operativamente unido a un promotor quimérico sintético con una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 132-207.
- En algunos ejemplos, el gen heterólogo codifica un producto proteína de interés encontrado en la Tabla 2.
- En algunos ejemplos, el gen heterólogo es un gen que es parte de una ruta biosintética de la lisina.
- En algunos ejemplos, el gen heterólogo se selecciona del gen *asd*, el gen *ask*, el gen *hom*, el gen *dapA*, el gen *dapB*, el gen *dapD*, el gen *ddh*, el gen *argD*, el gen *dapE*, el gen *dapF*, el gen *lysA*, el gen *lysE*, el gen *zwf*, el gen *pgi*, el gen *ktk*, el gen *fbp*, el gen *ppc*, el gen *pck*, el gen *ddx*, el gen *pyc* o el gen *icd*.
- En algunos ejemplos, el gen heterólogo es un gen que es parte de la ruta biosintética del licopeno.
- En algunos ejemplos, el gen heterólogo se selecciona del gen *dxs*, el gen *ispC*, el gen *ispE*, el gen *ispD*, el gen *ispF*, el gen *ispG*, el gen *ispH*, el gen *idi*, el gen *ispA*, el gen *ispB*, el gen *crtE*, el gen *crtB*, el gen *crtI*, el gen *crtY*, el gen *ymgA*, el gen *dxr*, el gen *elbA*, el gen *gdhA*, el gen *appY*, el gen *elbB* o el gen *ymgB*.
- En algunos ejemplos, el gen heterólogo codifica un producto biofarmacéutico o es un gen en una ruta para generar un producto biofarmacéutico.
- En algunos ejemplos, el producto biofarmacéutico se selecciona de humulina (rh insulina), intrónA (interferón alfa2b), roferón (interferón alfa2a), humatrope (hormona del crecimiento rh somatropina), neupogen (filgrastim), detaferón (interferón beta-1b), lispro (insulina de acción rápida), rapilysin (reteplasa), infergen (interferón alfacon-1), glucagón, beromun (tasonermina), ontak (denileuquina diftotox), lantus (insulina glargina de acción prolongada), kineret (anakinra), natrecor (nesiritida), somavert (pegvisomant), calcitonina (calcitonina recombinante de salmón), lucentis (ranibizumab), preotact (hormona paratiroidea humana), kyrstexal (rh urato oxidasa, PEGilada), nivestim (filgrastim, rhGCSF), voraxaze (glucarpidasa), o preos (hormona paratiroidea).
- En un primer aspecto de la invención se proporciona un método para generar una biblioteca de cepas de *E. coli* de intercambio de promotores, que comprende las etapas de:
- a. proporcionar una pluralidad de genes diana endógenos para una cepa de *E. coli* base, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa base *E. coli*, en donde al menos uno de la pluralidad de promotores es un promotor quimérico sintético, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que son cada uno de seis nucleótidos de longitud, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) que es la parte 5' UTR/RBS del promotor p_R del fago lambda o la parte de 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli*; y
 - b. transformar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los promotores de la escalera de promotores operativamente unidos a uno de los genes diana endógenos para la cepa base de *E. coli*.
- En un segundo aspecto de la invención, se proporciona un método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción, que comprende las etapas de:
- c. proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*, en donde al menos uno de la pluralidad de promotores es un promotor quimérico sintético, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que son cada uno de seis nucleótidos de longitud, partes centrales de los promotores p_L y p_R del fago lambda y una parte de 5' UTR/sitio de unión al ribosoma (RBS) que es la parte 5' UTR/RBS del promotor p_R del fago lambda o la parte de 5' UTR/sitio de unión al ribosoma del promotor del gen *acs* de *E. coli*;
 - d. transformar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones

genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los promotores de la escalera de promotores unidos operativamente a uno de los genes diana endógenos para la cepa base de *E. coli*;

5 e. cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores inicial para mejorar el rendimiento fenotípico frente a una cepa de *E. coli* de referencia, identificando así variaciones genéticas únicas que confieren mejoras del rendimiento fenotípico;

10 f. proporcionar una pluralidad subsiguiente de microbios de *E. coli* que comprende cada uno una combinación de variaciones genéticas únicas de las variaciones genéticas presentes en al menos dos cepas de *E. coli* individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente;

g. cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente para mejorar el rendimiento fenotípico frente a la cepa de *E. coli* de referencia, identificando así combinaciones únicas de variación genética que confieren mejoras adicionales del rendimiento fenotípico; y

15 h. repetir las etapas d)-e) una o más veces, de una forma lineal o no lineal, hasta que una cepa de *E. coli* presenta un nivel deseado de rendimiento fenotípico mejorado en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción, donde cada iteración subsiguiente crea una nueva biblioteca de cepas microbianas de cepas de *E. coli* de intercambio de promotores, donde cada cepa de la nueva biblioteca comprende variaciones genéticas que son una combinación de variaciones genéticas seleccionadas de entre al menos dos cepas de *E. coli* individuales de una biblioteca precedente.

20 Las realizaciones preferidas de la invención en cualquiera de sus diversos aspectos son las descritas a continuación o como se definen en las reivindicaciones.

Breve descripción de las figuras

25 La **figura 1** representa un método de recombinación de ADN de la presente descripción para aumentar la variación en grupos de diversidad. Las secciones de ADN, tal como las regiones de genoma de especies relacionadas, se pueden cortar por medios físicos o enzimáticos/químicos. Las regiones de ADN cortadas se funden y se deja que se reasocien, de modo que las regiones genéticas superpuestas ceban las reacciones de extensión de la polimerasa. Las reacciones de fusión/extensión subsiguientes se llevan a cabo hasta que los productos se vuelven a ensamblar en ADN quimérico, que comprende elementos de una o más secuencias de partida.

30 La **figura 2** describe métodos de la presente descripción para generar nuevas cepas de *E. coli* hospedantes con modificaciones de secuencia seleccionadas (p. ej., 100 SNP para intercambiar). Brevemente, el método comprende (1) los insertos de ADN deseados se diseñan y generan combinando uno o más oligos sintetizados en una reacción de ensamblaje, (2) los insertos de ADN se clonan en plásmidos de transformación, (3) los plásmidos completos se transfieren a las cepas de producción deseadas, donde se integran en el genoma de la cepa hospedante, y (4) los marcadores de selección y otros elementos de ADN no deseados se sacan en bucle de la cepa hospedante. Cada etapa de ensamblaje de ADN puede implicar etapas de control de calidad (CC) adicionales, tales como la clonación de plásmidos en bacterias *E. coli* para amplificación y secuenciación.

40 La **figura 3** representa el ensamblaje de los plásmidos de transformación de la presente descripción y su integración en un genoma de *E. coli* hospedante. El inserto de ADN se genera combinando uno o más oligos sintetizados en una reacción de ensamblaje. Los insertos de ADN que contienen la secuencia deseada están flanqueados por regiones de ADN homólogas a la región diana del genoma. Estas regiones homólogas facilitan la integración genómica y, una vez integradas, forman regiones de repetición directas diseñadas para sacar un bucle de ADN de la cadena principal del vector en etapas subsiguientes. Los plásmidos ensamblados contienen el inserto de ADN y, opcionalmente, uno o más marcadores de selección.

45 La **figura 4** representa un procedimiento para la salida bucles de regiones seleccionadas de ADN de cepas de *E. coli* hospedantes. Las regiones de repetición directa del ADN insertado y el genoma del hospedante pueden "salir en un bucle" en un suceso de recombinación. El contador de células seleccionado para el marcador de selección contiene delecciones del ADN del bucle flanqueado por las regiones de repetición directa.

50 La **figura 5** representa un ejemplo del procedimiento de mejora de cepas de *E. coli* de la presente descripción. Las secuencias de la cepa hospedante que contienen modificaciones genéticas (diseño genético) se ensayan para mejorar el rendimiento de la cepa en varios contextos de cepas (construcción de cepas). Las cepas que presentan mutaciones beneficiosas se analizan (ID de aciertos y análisis) y los datos se almacenan en bibliotecas para su análisis adicional (p. ej., bibliotecas de intercambio de SNP, bibliotecas de intercambio de PRO y combinaciones de las mismas, entre otras). Las reglas de selección de la presente descripción generan nuevas secuencias de cepas hospedantes de *E. coli* propuestas basadas en el efecto predicho de combinar elementos de una o más bibliotecas para un análisis iterativo adicional.

- La **figura 6A-B** representa las etapas de ensamblaje de ADN, transformación y cribado de cepas de *E. coli* de uno de los ejemplos de la presente descripción. La **figura 6A** representa las etapas para construir fragmentos de ADN, clonar dichos fragmentos de ADN en vectores, transformar dichos vectores en cepas de *E. coli* hospedantes y salida de bucle de secuencias de selección mediante contraselección. La **figura 6B** representa las etapas para el cultivo, cribado y evaluación de alta capacidad de cepas hospedantes de *E. coli* seleccionadas. Esta figura también describe las etapas opcionales de cultivo, cribado y evaluación de cepas de *E. coli* seleccionadas en tanques de cultivo.
- La **figura 7** representa un ejemplo del sistema automatizado de la presente descripción. La presente descripción enseña el uso de sistemas robóticos automatizados con varios módulos capaces de clonar, transformar, cultivar, cribar y/o secuenciar *E. coli* hospedantes.
- La **figura 8** representa una visión general de un ejemplo del programa de mejora de cepas de *E. coli* de la presente descripción.
- La **figura 9** es una representación del genoma de *Corynebacterium glutamicum*, que comprende aproximadamente 3,2 millones de pares de bases.
- La **figura 10** representa los resultados de un experimento de transformación de la presente descripción. Se seleccionaron insertos de ADN en el intervalo de 0,5 kb a 5,0 kb para insertar en varias regiones (mostradas como posiciones relativas 1-24) del genoma de *Corynebacterium glutamicum*. El color claro indica una integración con éxito, mientras que el color más oscuro indica fallo de inserción.
- La **figura 11** representa los resultados de un programa de intercambio de PRO de ingeniería de HTP de segunda ronda. Las combinaciones de promotor: :gen superiores identificadas durante la primera ronda de intercambio de PRO se analizaron de acuerdo con los métodos de la presente descripción para identificar combinaciones de dichas mutaciones que sería probable que presentaran efectos beneficiosos aditivos o combinatorios en el rendimiento del hospedante. Por lo tanto, los mutantes de intercambio de PRO de segunda ronda comprendían combinaciones de pares de varias mutaciones de promotor: : gen. Los mutantes de la segunda ronda resultantes se cribaron según las diferencias en el rendimiento de la célula hospedante de una biomolécula seleccionada. Un par de mutaciones de combinaciones que se había predicho que presentarían efectos beneficiosos se enfatiza con un círculo.
- La **figura 12** representa los resultados de un experimento que ensaya el ensamblaje de plásmido con éxito para plásmidos transformados en *E. coli*. Escoger cuatro colonias es suficiente para lograr una tasa de fallo de 13% para plásmidos que contienen secuencias de inserción de 1 y 2 kb. Las inserciones más grandes pueden requerir un cribado de colonias adicional para lograr resultados consistentes.
- La **figura 13** representa los resultados de un experimento que ensaya la transformación con éxito de *Corynebacterium glutamicum* con vectores de inserción. Los tamaños de insertos de ADN de 2 y 5 kb presentaron tasas altas de transformación con tasas bajas de fallo del ensamblaje.
- La **figura 14** representa los resultados de las selecciones de salidas de bucle en *Corynebacterium glutamicum*. La resistencia a la sacarosa de las bacterias transformadas indica salida de bucle del marcador de selección sacB. El tamaño del inserto de ADN no parece tener impacto en la eficiencia de la salida de bucle.
- La **figura 15** es una matriz de similitud calculada usando la medida de correlación. La matriz es una representación de la similitud funcional entre variantes de SNP. Se espera que la consolidación de SNP con baja similitud funcional tenga una mayor probabilidad de mejorar el rendimiento de las cepas, en contraposición a la consolidación de SNP con mayor similitud funcional.
- La **figura 16A-B** representa los resultados de un experimento de mapeo de epistasia. La combinación de intercambios de SNP y de PRO con bajas similitudes funcionales da un mejor rendimiento de las cepas. La **figura 16A** representa un dendrograma agrupado por similitud funcional de todos los intercambios de SNP/PRO. La **figura 16B** representa el rendimiento de la cepa hospedante de SNP consolidados medidos por el rendimiento del producto. Una mayor distancia de agrupamientos se correlaciona con un mejor rendimiento de consolidación de la cepa hospedante.
- La **figura 17A-B** representa diferencias de SNP entre variantes de cepas en el grupo de diversidad. La **figura 17A** representa la relación entre las cepas de este experimento. La cepa A es la cepa hospedante de tipo natural. La cepa B es una cepa modificada intermedia. La cepa C es la cepa de producción industrial. La **figura 17B** es un gráfico que identifica el número de SNP únicos y compartidos en cada cepa.
- La **figura 18** representa un experimento de intercambio de SNP de primera ronda de acuerdo con los métodos de la presente descripción. (1) todos los SNP de C se clonarán de forma individual y/o combinatoria en la cepa A base ("onda ascendente" de A a C). (2) todos los SNP de C se eliminarán de forma individual y/o combinatoria de la cepa comercial C ("onda descendente" de C a A). (3) todos los SNP de B se clonarán de forma individual y/o combinatoria en la cepa base A (onda ascendente de A a B). (4) todos los SNP de B se eliminarán de forma individual y/o combinatoria de la cepa comercial B (onda descendente de B a A). (5) todos los SNP únicos de C se clonarán de forma individual y/o combinatoria en la cepa B comercial (onda ascendente de B a C). (6) todos los SNP únicos de C se eliminarán de forma individual y/o combinatoria de la cepa comercial C (onda descendente C a B).

La **figura 19** ilustra dianas génicas de ejemplo que se usarán en un procedimiento de intercambio de promotores. Los 4 subrayados son genes de desvío que se pueden dirigir para regulación por disminución, mientras que los 19 genes restantes en la ruta se pueden dirigir para sobreexpresión.

5 La **figura 20** ilustra una biblioteca de promotores de ejemplo que se está utilizando para llevar a cabo un procedimiento de intercambio de promotores para las dianas de genes identificadas. Los promotores utilizados en el procedimiento de intercambio de PRO (es decir, intercambio de promotores) son P₁-P₈, cuyas secuencias e identidad se pueden encontrar en la Tabla 1.

La **figura 21** ilustra los diferentes enfoques disponibles para el intercambio de promotores dependiendo de si el gen diana comprende su propio promotor o es parte de un operón.

10 La **figura 22** representa datos de intercambio de promotores de HTP de ejemplo que muestran modificaciones que afectan significativamente al funcionamiento en el rendimiento de lisina. El eje X representa diferentes cepas dentro de la biblioteca de cepas microbianas de diseño genético de intercambio de promotores, y el eje Y incluye valores relativos de rendimiento de lisina para cada cepa. Cada letra del gráfico representa un gen diana de intercambio de PRO. Cada punto de datos representa una repetición. Los datos demuestran que una herramienta molecular adaptada para aplicaciones de HTP, como se describe en el presente documento (es decir, intercambio de PRO), es capaz de crear y optimizar de manera eficiente el rendimiento de cepas microbianas para la producción de un compuesto o molécula de interés. En este caso, el compuesto de interés era lisina; sin embargo, la herramienta molecular de intercambio de PRO enseñada se puede utilizar para optimizar y/o aumentar la producción de cualquier compuesto de interés. Un experto en la técnica comprendería cómo elegir genes diana, que codifican la producción de un compuesto deseado, y después utilizar el procedimiento de intercambio de PRO enseñado. Un experto en la técnica apreciaría fácilmente que los datos demostrados que ilustran los aumentos de rendimiento de lisina enseñados en el presente documento, junto con la descripción detallada presentada en la solicitud, permiten que la herramienta molecular de intercambio de PRO sea un avance ampliamente aplicable en la ingeniería genómica de HTP.

15 La **figura 23** ilustra la distribución de los rendimientos de cepas relativos para los datos de entrada considerados. Un rendimiento relativo de cero indica que la cepa modificada funcionaba igualmente bien que la cepa base en la placa. Los procedimientos descritos en el presente documento están diseñados para identificar las cepas que es probable que funcionen significativamente por encima de cero.

20 La **figura 24** ilustra los valores del coeficiente de regresión lineal, que representan el cambio promedio (aumento o disminución) en el rendimiento relativo de las cepas asociado con cada cambio genético incorporado en las cepas representadas.

25 La **figura 25** ilustra la composición de los cambios para los 100 diseños superiores de cepas pronosticadas. El eje x da el conjunto de cambios genéticos potenciales (las mutaciones dss son intercambios de SNP y las mutaciones Pcg son intercambios de PRO), y el eje y muestra el orden de clasificación. Las celdas negras indican la presencia de un cambio particular en el diseño candidato, mientras que las celdas blancas indican la ausencia de ese cambio. En este ejemplo particular, todos los 100 diseños superiores contienen los cambios pcg3121_pgi, pcg1860_pyc, dss_339 y pcg0007_39_lysa. Además, el diseño del candidato superior contiene los cambios dss_034, dss_009.

30 La **figura 26** representa las etapas de ensamblaje y transformación de ADN de uno de los ejemplos de la presente descripción. El diagrama de flujo representa las etapas para construir fragmentos de ADN, clonar dichos fragmentos de ADN en vectores, transformar dichos vectores en cepas de *E. coli* hospedantes y la salida de bucle de secuencias de selección mediante contraselección.

35 La **figura 27** representa las etapas para el cultivo, cribado y evaluación de alta capacidad de cepas de *E. coli* hospedantes seleccionadas. Esta figura también representa las etapas opcionales de cultivo, cribado y evaluación de cepas de *E. coli* seleccionadas en tanques de cultivo.

40 La **figura 28** representa perfiles de expresión de promotores ilustrativos que presentan un intervalo de expresión reguladora, de acuerdo con las escaleras de promotores de la presente descripción. La expresión del promotor A alcanza su máximo en la fase de retardo de los cultivos bacterianos, mientras que el promotor B y C alcanzan su máximo en la fase exponencial y estacionaria, respectivamente.

45 La **figura 29** representa perfiles de expresión de promotores ilustrativos que presentan un intervalo de expresión reguladora, de acuerdo con las escaleras de promotores de la presente descripción. La expresión del promotor A alcanza su máximo inmediatamente después de la adición de un sustrato seleccionado, pero vuelve rápidamente a niveles indetectables a medida que se reduce la concentración del sustrato. La expresión del promotor B alcanza su máximo inmediatamente después de la adición del sustrato seleccionado y baja lentamente hasta niveles indetectables junto con la correspondiente reducción en el sustrato. La expresión del promotor C alcanza su máximo tras la adición del sustrato seleccionado y permanece con expresión alta a lo largo de todo el cultivo, incluso después de que el sustrato se haya disipado.

50

55

La **figura 30** representa perfiles de expresión de promotores ilustrativos que presentan un intervalo de niveles de expresión constitutivos, de acuerdo con las escaleras de promotores de la presente descripción. El promotor A presenta la expresión más baja, seguido de niveles de expresión crecientes del promotor B y C, respectivamente.

5 La **figura 31** representa gráficamente un ejemplo del sistema LIMS de la presente descripción para la mejora de cepas de *E. coli*.

La **figura 32** representa gráficamente una implementación computacional en la nube de ejemplos del sistema LIMS de la presente descripción.

La **figura 33** representa un ejemplo del flujo de trabajo de diseño predictivo de cepas iterativo de la presente descripción.

10 La **figura 34** representa gráficamente un ejemplo de un sistema informático, de acuerdo con ejemplos de la presente descripción.

La **figura 35** representa el flujo de trabajo asociado con el ensamblaje de ADN según un ejemplo de la presente descripción. Este procedimiento se divide en 4 etapas: generación de partes, ensamblaje del plásmido, CC del plásmido y preparación del plásmido para la transformación. Durante la generación de las partes, los oligonucleótidos diseñados por el sistema de gestión de información de laboratorio (LIMS) se encargan a un proveedor de secuenciación de oligonucleótidos y se utilizan para amplificar las secuencias diana del organismo hospedante por PCR. Estas partes de la PCR se limpian para eliminar contaminantes y se evalúa su éxito por análisis de fragmentos, comparación de control de calidad *in silico* de tamaños de fragmentos observados con teóricos y cuantificación de ADN. Las partes se transforman en levaduras junto con un vector de ensamblaje y se ensamblan en plásmidos por recombinación homóloga. Los plásmidos ensamblados se aíslan de la levadura y se transforman en *E. coli* para el subsiguiente control de calidad del ensamblaje y amplificación. Durante el control de calidad del ensamblaje del plásmido, se aíslan varias copias de cada plásmido, se amplifican por amplificación en círculo rodante (RCA) y se evalúa el ensamblaje correcto por digestión enzimática y análisis de fragmentos. Los plásmidos correctamente ensamblados identificados durante el procedimiento de CC se seleccionan como aciertos para generar reservas permanentes y el ADN del plásmido se extrae y cuantifica antes de la transformación en el organismo hospedante diana.

15
20
25

La **figura 36** representa los resultados de un experimento que caracteriza los efectos de terminadores T1-T8 en dos medios en dos puntos de tiempo. Las condiciones A y C representan los dos puntos de tiempo para los medios BHI, mientras que los puntos B y D representan los dos puntos de tiempo para los medios de ensayo HTP.

30 La **figura 37** representa los resultados de un experimento que compara la eficacia de enfoques tradicionales de mejora de cepas tales como mutagénesis por UV frente a las metodologías de ingeniería de HTP de la presente descripción. La gran mayoría de las mutaciones por UV no produjeron un aumento notable en el rendimiento de las células hospedantes. Por el contrario, las metodologías de intercambio de PRO de la presente descripción produjeron una alta proporción de mutantes que presentaban aumentos de 1,2 a 2 veces en el rendimiento de las células hospedantes.

35 La **figura 38** representa los resultados de la primera ronda de un programa de intercambio de SNP de ingeniería de HTP. Se identificaron 186 mutaciones de SNP individuales y se clonaron individualmente en una cepa base. Los mutantes resultantes se cribaron según las diferencias en el rendimiento de la célula hospedante de una biomolécula seleccionada.

40 La **figura 39** representa los resultados de la segunda ronda de un programa de intercambio de SNP de ingeniería de HTP. Se clonaron individualmente 176 mutaciones de SNP individuales de una primera ronda del programa de intercambio de SNP en una cepa de células hospedantes de la segunda ronda que contenía un SNP beneficioso identificado durante un programa de SNP de la primera ronda. Por lo tanto, los mutantes resultantes representan el efecto de dos pares de combinaciones de mutaciones. Se muestran los resultados del cribado según las diferencias en el rendimiento de la célula hospedante (eje Y) y la productividad (eje X) para la biomolécula seleccionada.

45 La **figura 40** representa los resultados de un experimento de validación de fermentación en tanque. Los pares de mutaciones superiores de la segunda ronda de intercambio de SNP de HTP se cultivaron en tanques de fermentación. Se muestran los resultados del rendimiento y la productividad de las células hospedantes para la biomolécula seleccionada (es decir, lisina). Como puede verse, en una ronda de ingeniería genómica, los autores de la invención utilizaron el procedimiento de intercambio de PRO para determinar que un mutante de intercambio de PRO particular (zwf) presentaba mayor rendimiento de una biomolécula seleccionada en comparación con la cepa base (es decir, se compara la cepa base con la cepa base + zwf). Después, los autores de la invención llevaron a cabo otra ronda de ingeniería genómica, en donde se utilizó un procedimiento de intercambio de SNP para determinar mutaciones de SNP beneficiosas que podrían afectar al rendimiento de la biomolécula, cuando se combinan con dicho mutante de intercambio de PRO. La combinación del procedimiento de intercambio de PRO y el procedimiento de intercambio de SNP creaba mutantes con rendimientos incluso más altos que los mutantes de intercambio de PRO solo anteriores (es decir, se compara la cepa base + zwf + SNP121 con la cepa base + zwf previamente descrita). Esta figura ilustra las mejoras notables en el rendimiento que se pueden lograr al combinar los procedimientos de intercambio de PRO e intercambio de SNP de la descripción. En ejemplos, la combinación de una campaña de ingeniería genómica de

50
55

intercambio de PRO con una campaña de ingeniería genómica de intercambio de SNP puede conducir a un mayor rendimiento y/o productividad de una biomolécula/producto de interés en un factor de 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 40%, 45%, 50%, o más, en relación con una cepa base.

5 La **figura 41** representa los resultados de la primera ronda de un programa de intercambio de PRO de ingeniería de HTP. Los genes seleccionados que se cree que están asociados con el rendimiento del hospedante se combinaron con una escalera de promotores para crear una biblioteca de intercambio de PRO de la primera ronda, de acuerdo con los métodos de la presente descripción. Los mutantes resultantes se cribaron según las diferencias en el rendimiento de la célula hospedante de una biomolécula seleccionada (es decir, lisina).

10 La **figura 42** es un diagrama de flujo que ilustra la consideración de los efectos epistáticos en la selección de mutaciones para el diseño de una cepa microbiana, de acuerdo con los ejemplos de la descripción.

15 La **figura 43** representa una secuencia reguladora de diseño bicistrónico (BCD), de acuerdo con la presente descripción. En algunos ejemplos, la presente descripción enseña que se pueden usar los BCD en lugar de los promotores tradicionales con el fin de mejorar la consistencia de la expresión entre diferentes combinaciones de promotores: genes diana en intercambios de PRO. En algunos ejemplos, los BCD comprenden un promotor, un primer sitio de unión al ribosoma (SD1), una primera secuencia cistrónica (Cis1), un segundo sitio de unión al ribosoma (SD2), operativamente unido a un gen diana de interés (Cis2). En algunos ejemplos, la presente descripción enseña que Cis1 puede ser cualquier secuencia codificante de péptidos. Se proporciona información adicional sobre el diseño y uso de BCD en secciones posteriores de la memoria descriptiva.

20 La **figura 44** es una ilustración de la colocalización de enzimas de rutas mediante dominios de unión a ADN recombinantes. Una célula modificada codifica las enzimas de ruta Enz1-3 con dominios de unión al ADN. Cuando se expresan, estas enzimas se unen a un ADN almacén u otra ubicación diana, que comprende motivos de ADN que son reconocidos por los dominios de unión de ADN recombinantes fusionados con las enzimas de ruta. Cuando los dominios de unión de ADN fusionados se unen a sus motivos de ADN cognado en el plásmido de almacén, las enzimas están constreñidas cerca unas de otras en el espacio, lo que puede mejorar la productividad de la ruta.

25 La **figura 45** es un diagrama esquemático para la incorporación de secuencias de nucleótidos que codifican dominios de unión de ADN en enzimas de ruta. El GOI codifica una enzima de ruta. Las células de *E. coli* se transforman con un plásmido que codifica una versión mutante de GOI que incluye una secuencia de nucleótidos que codifica un dominio de unión de ADN (indicado con una estrella). El plásmido también codifica un marcador de resistencia a antibióticos (Ab) que permite la selección de células con "entrada de bucle" ("*Loop-in*") y un marcador de contraselección (Contador) que permite la contraselección posterior de las células con "salida de bucle" ("*Loop-out*"). En la etapa de "entrada de bucle", el plásmido completo, incluido el GOI mutante, se incorpora en el genoma por recombinación homóloga (HR). Durante la etapa de "salida de bucle", algunas de las células volverán a la secuencia nativa del GOI a través de HR, mientras que otras sufrirán un suceso de HR que dejará al GOI mutante en el genoma.

30 La **figura 46** es un diagrama de puntos para el rendimiento previsto frente al rendimiento medido de los datos de entrenamiento para un modelo de rendimiento de la presente descripción. El modelo subyacente es un modelo de regresión de Kernel Ridge (con kernel polinomial de orden 4). El modelo se entrena en 1864 construcciones genéticas únicas y el rendimiento fenotípico asociado. El modelo ajustado tiene un valor r^2 de 0,52.

35 La **figura 47** representa la composición genética de los diseños de candidatos generados por los algoritmos de predicción de la presente descripción. Estos diseños de candidatos se enviaron para la construcción y análisis de HTP. Aquí, el diseño de candidato se define como la combinación de la id de la cepa parental y la mutación o mutaciones introducidas.

40 La **Figura 48** es un diagrama de puntos del rendimiento previsto frente al rendimiento medido de los diseños de candidatos generados por los algoritmos de predicción de la presente descripción, y construido de acuerdo con los métodos de construcción de HTP de la presente descripción. Esta figura demuestra que el modelo puede predecir el rendimiento de la cepa candidata con un grado aceptable de precisión.

45 La **figura 49** es un diagrama de caja y bigotes que representa el porcentaje de cambio de rendimiento de las cepas candidatas con respecto a las cepas parentales. En el eje y, un valor de 0,01 corresponde al 1%. Esta figura demuestra que las cepas diseñadas por un modelo computacional (gris claro) logran una mejora cuantificable frente a sus cepas parentales correspondientes. Además, la figura demuestra que estas mejoras de la cepa base del modelo son comparables en magnitud a las mejoras logradas por cepas diseñadas por expertos humanos.

50 La **figura 50** ilustra la distribución del comportamiento del rendimiento para las cepas diseñadas por el modelo computacional (gris oscuro) y por un experto humano (gris claro). Las cepas diseñadas por ordenador presentaban distribuciones más ajustadas con ganancias medianas más altas.

55 La **figura 51** es un diagrama de caja y bigotes que representa el rendimiento absoluto de cepas candidatas generadas por ordenador (gris claro) o por un experto humano (gris oscuro). Los resultados se agregan por cepa parental.

La **figura 52** es una representación del genoma de *Escherichia coli*, que comprende aproximadamente 4,6 millones de pares de bases.

La **figura 53** ilustra el efecto de las partes aislantes y terminadoras en las cadenas principales del vector en la eficiencia de la transformación y la integración del plásmido.

5 La **figura 54** ilustra el diseño combinatorio de secuencias de promotor-5'UTR sintéticas de la Tabla 1.4.

La **figura 55** representa el mapa de plásmido que ilustra los componentes de la cadena principal del vector 1.

La **figura 56** representa el mapa de plásmido que ilustra los componentes de la cadena principal del vector 2.

La **figura 57** representa el mapa de plásmido que ilustra los componentes de la cadena principal del vector 3.

La **figura 58** representa el mapa de plásmido que ilustra los componentes de la cadena principal del vector 4.

10 La **figura 59** representa la ruta biosintética del licopeno de *E. coli*.

La **figura 60** representa ediciones del terminador en las dianas de la ruta del licopeno *idi* e *ymgA*. El terminador TyjbE demuestra una disminución del rendimiento de la cepa en relación con el control, destacando así la utilidad de estos tipos de bibliotecas para identificar las dianas de las rutas críticas.

La **figura 61** representa ediciones del terminador en múltiples dianas de la ruta del licopeno.

15 La **figura 62** representa los intercambios de promotores (para comparación), marcadores de degradación y terminadores en la diana *dxs* de la ruta del licopeno. El marcador de degradación *ssrA_LAA* demuestra un rendimiento de cepa mejorado con respecto al control. Esto es inesperado ya que esta cepa es una combinación de un PROSWP con un marcador de degradación en una única diana de ruta. Se espera que el PROSWP inicial aumente la abundancia de proteínas, y se espera que el marcador de degradación disminuya la abundancia de proteínas, demostrando así la utilidad de las combinaciones de tipos de bibliotecas para ajustar el rendimiento óptimo de la cepa.

20 La **figura 63** representa intercambios de marcadores de solubilidad, promotores y terminadores en la diana *gdhA* de la ruta del licopeno. El marcador de solubilidad FH8 demuestra un rendimiento de cepa mejorado con respecto al control, pero el marcador de solubilidad GB1 no, lo que demuestra la necesidad de evaluar bibliotecas de cada tipo de modificación.

25 Descripción detallada

Definiciones

Aunque se cree que los siguientes términos son bien entendidos por un experto en la técnica, las siguientes definiciones se exponen para facilitar la explicación de la materia objeto actualmente descrita.

30 El término "un" o "una" se refiere a uno o más de esa entidad, es decir, se puede referir a varios referentes. Así pues, los términos "un" o "una", "uno o más" y "al menos uno" se usan indistintamente en el presente documento. Además, la referencia a "un elemento" por el artículo indefinido "un" o "una" no excluye la posibilidad de que esté presente más de uno de los elementos, a menos que el contexto requiera claramente que haya uno y solo uno de los elementos.

35 Como se usa en el presente documento, los términos "organismo celular" "microorganismo" o "microbio" deben tomarse de manera amplia. Estos términos se usan indistintamente e incluyen, pero no se limitan a los dos dominios procariotas, Bacteria y Archaea, así como a ciertos protistas y hongos eucariotas. En algunos ejemplos, la descripción se refiere a los "microorganismos" u "organismos celulares" o "microbios" de listas/tablas y figuras presentes en la descripción. Esta caracterización se puede referir no solo a los géneros taxonómicos identificados de las tablas y figuras, sino también a las especies taxonómicas identificadas, así como a las diversas cepas nuevas y recién identificadas o diseñadas de cualquier organismo en dichas tablas o figuras. La misma caracterización es válida para
40 la mención de estos términos en otras partes de la memoria descriptiva, tal como en los ejemplos.

El término "procariotas" es reconocido en la técnica y se refiere a células que no contienen núcleo u otros orgánulos celulares. Los procariotas se clasifican en general en uno de dos dominios, las Bacteria y las Archea. La diferencia definitiva entre organismos de los dominios Archaea y Bacteria se basa en diferencias fundamentales en la secuencia de bases de nucleótidos en el ARN ribosómico 16S.

45 El término "Archaea" se refiere a una categorización de organismos de la división Mendosicutes, que se encuentran típicamente en ambientes inusuales y se distinguen del resto de los procariotas por varios criterios, que incluyen el número de proteínas ribosómicas y la falta de ácido murámico en las paredes celulares. Basado en el análisis de *ssrARN*, las Archaea consisten en dos grupos filogenéticamente distintos: Crenarchaeota y Euryarchaeota. Basado en su fisiología, las Archaea se pueden organizar en tres tipos: metanógenos (procariotas que producen metano);
50 halófilos extremos (procariotas que viven en concentraciones muy altas de sal (NaCl)); y (hiper) termófilos extremos (procariotas que viven a temperaturas muy altas). Además de las características unificadoras de las arqueas que las

distinguen de las bacterias (es decir, sin mureína en la pared celular, lípidos de membrana unidos por éster, etc.), estos procariotas presentan atributos estructurales o bioquímicos únicos que los adaptan a sus hábitats particulares. La Crenarchaeota consiste principalmente en procariotas hipertermófilos dependientes de azufre y la Euryarchaeota contiene los metanógenos y halófilos extremos.

5 "Bacteria" o "eubacteria" se refiere a un dominio de organismos procariotas. Las bacterias incluyen al menos 11 grupos distintos como sigue: (1) bacterias Gram positivas (Gram+), de las cuales hay dos subdivisiones principales: (1) grupo de alto contenido G+C (*Actinomycetes*, *Mycobacteria*, *Micrococcus*, otros) (2) grupo de bajo contenido G+C (*Bacillus*, *Clostridia*, *Lactobacillus*, *Staphylococci*, *Streptococci*, *Mycoplasmas*); (2) Proteobacterias, p. ej., bacterias Gram negativas fotosintéticas púrpura+no fotosintéticas (incluye las bacterias Gram negativas más "comunes"); (3) 10 Cianobacterias, p. ej., fotótrofos oxigénicos; (4) Espiroquetas y especies relacionadas; (5) Planctomicetos; (6) *Bacteroides*, *Flavobacteria*; (7) *Chlamydia*; (8) Bacterias verdes del azufre; (9) Bacterias verdes no del azufre (también fotótrofos anaeróbicos); (10) Micrococos radiorresistentes y relacionados; (11) *Thermotoga* y *Thermosiphon thermophiles*.

15 Un "eucariota" es cualquier organismo cuyas células contienen un núcleo y otros orgánulos encerrados dentro de membranas. Los eucariotas pertenecen al taxón Eukarya o Eukaryota. La característica definitoria que diferencia a las células eucariotas de las células procariotas (las bacterias y arqueas mencionadas anteriormente) es que tienen orgánulos unidos a la membrana, especialmente el núcleo, que contiene el material genético, y está encerrado por la envoltura nuclear.

20 Las expresiones "célula hospedante genéticamente modificada", "célula hospedante recombinante" y "cepa recombinante" se utilizan indistintamente en el presente documento y se refieren a células hospedantes que han sido modificadas genéticamente mediante los métodos de clonación y transformación de la presente descripción. Por lo tanto, las expresiones incluyen una célula hospedante (p. ej., bacteria, célula de levadura, célula fúngica, CHO, célula humana, etc.) que ha sido alterada, modificada o manipulada genéticamente, de modo que presenta un genotipo y/o fenotipo (p. ej., cuando la modificación genética afecta a las secuencias de ácido nucleico codificantes del microorganismo) alterado, modificado o diferente en comparación con el organismo natural del que se ha obtenido. 25 Se entiende que en algunos ejemplos, los términos se refieren no solo a la célula hospedante recombinante particular en cuestión, sino también a la progenie o la potencial progenie de dicha célula hospedante.

La expresión "microorganismo de tipo natural" o "célula hospedante de tipo natural" describe una célula que se encuentra en la naturaleza, es decir, una célula que no ha sido modificada genéticamente.

30 La expresión "genéticamente transformado" se puede referir a cualquier manipulación del genoma de una célula hospedante (p. ej., por inserción, delección, mutación o reemplazo de ácidos nucleicos).

35 El término "control" o "célula hospedante de control" se refiere a una célula hospedante de comparación adecuada para determinar el efecto de una modificación genética o tratamiento experimental. En algunos ejemplos, la célula hospedante de control es una célula de tipo natural. En otros ejemplos, una célula hospedante de control es genéticamente idéntica a la célula hospedante genéticamente modificada, salvo por la o las modificaciones genéticas que diferencian a la célula hospedante de tratamiento. En algunos ejemplos, la presente descripción enseña el uso de cepas parentales como células hospedantes de control (p. ej., la cepa Si que se usó como base para el programa de mejora de cepas). En otros ejemplos, una célula hospedante puede ser una célula genéticamente idéntica que carece de un promotor o SNP específico que se esté ensayando en la célula hospedante de tratamiento.

40 Como se usa en el presente documento, el término "alelo(s)" significa cualquiera de una o más formas alternativas de un gen, todos cuyos alelos están relacionados con al menos un rasgo o característica. En una célula diploide, los dos alelos de un gen determinado ocupan locus correspondientes en un par de cromosomas homólogos.

Como se usa en el presente documento, el término "locus" (plural locus) significa un lugar o lugares específicos o un sitio en un cromosoma donde, por ejemplo, se encuentra un gen o marcador genético.

45 Como se usa en el presente documento, la expresión "genéticamente ligado" se refiere a dos o más rasgos que se co-heredan en una alta tasa durante la reproducción, de modo que son difíciles de separar por cruzamiento.

Una "recombinación" o "suceso de recombinación" como se usa en el presente documento, se refiere a un cruce cromosómico o una distribución independiente.

50 Como se usa en el presente documento, el término "fenotipo" se refiere a las características observables de una célula individual, cultivo celular, organismo o grupo de organismos que resultan de la interacción entre la composición genética de ese individuo (es decir, genotipo) y el medio ambiente.

55 Como se usa en el presente documento, el término "quimérico" o "recombinante" cuando se describe una secuencia de ácido nucleico o una secuencia de proteína se refiere a un ácido nucleico, o una secuencia de proteína, que une al menos dos polinucleótidos heterólogos, o dos polipéptidos heterólogos, en un solo macromolécula, o que reordena uno o más elementos de al menos una secuencia de proteína o ácido nucleico natural. Por ejemplo, el término "recombinante" se puede referir a una combinación artificial de dos segmentos de secuencia separados de otro modo,

p. ej., por síntesis química o por la manipulación de segmentos aislados de ácidos nucleicos por técnicas de ingeniería genética.

5 Como se usa en el presente documento, una "secuencia de nucleótidos sintética" o "secuencia de polinucleótidos sintética" es una secuencia de nucleótidos que no se conoce que se produzca en la naturaleza o que no se produce de forma natural. Generalmente, dicha secuencia de nucleótidos sintética comprenderá al menos una diferencia de nucleótidos en comparación con cualquier otra secuencia de nucleótidos que se encuentra de forma natural.

10 Como se usa en el presente documento, el término "ácido nucleico" se refiere a una forma polimérica de nucleótidos de cualquier longitud, sean ribonucleótidos o desoxirribonucleótidos, o análogos de los mismos. Este término se refiere a la estructura primaria de la molécula y, por tanto, incluye ADN bicatenario y monocatenario, así como ARN bicatenario y monocatenario. También incluye ácidos nucleicos modificados tales como ácidos nucleicos metilados y/o con caperuza, ácidos nucleicos que contienen bases modificadas, modificaciones de la cadena principal y similares. Los términos "ácido nucleico" y "secuencia de nucleótidos" se usan indistintamente.

15 Como se usa en el presente documento, la expresión "armazón de ADN" o "armazón de ácido nucleico" se refiere a un armazón de ácido nucleico que se produce de forma artificial o a una secuencia de origen natural que se readapta como armazón. En un ejemplo de la presente descripción, el armazón de ácido nucleico es un armazón de ácido desoxirribonucleico sintético. Los desoxirribonucleótidos del armazón sintético pueden comprender bases de purina y pirimidina u otras bases desoxirribonucleotídicas naturales, modificadas química o bioquímicamente, no naturales o derivatizadas. Como se describe con más detalle en el presente documento, el armazón de ácido nucleico de la presente descripción se utiliza para ensamblar e inmovilizar espacial y temporalmente dos o más proteínas implicadas en una ruta biológica, es decir, enzimas biosintéticas, para crear un complejo funcional. El ensamblaje e inmovilización de cada proteína de la ruta biológica en el armazón se produce a través de la interacción de unión entre una de las secuencias de unión a proteínas, es decir, sitios de acoplamiento de proteínas, del armazón y una parte de unión al ADN correspondiente de una enzima biosintética quimérica. Por consiguiente, el armazón de ácido nucleico comprende una o más subunidades, comprendiendo cada subunidad dos o más secuencias de unión a proteínas para acomodar la unión de dos o más proteínas quiméricas diferentes de la ruta biológica.

20 Como se usa en el presente documento, una "secuencia de unión de ADN" o "sitio de unión de ADN" se refiere a una secuencia de ácido nucleico específica que es reconocida y unida por una parte del dominio de unión de ADN de un gen biosintético quimérico (p. ej., enzima biosintética quimérica) codificado por genes modificados de la presente descripción. Muchos dominios de unión de ADN y sus sitios de reconocimiento de ADN de la pareja de unión análoga (es decir, sitios de unión de ADN) son bien conocidos en la técnica. Por ejemplo, se conocen en la técnica numerosos dominios de unión de dedos de zinc y sus correspondientes sitios diana de unión de ADN y son adecuados para usar en la presente descripción. Otros dominios de unión de ADN incluyen, sin limitación, dominios de unión de cremallera de leucina y sus correspondientes sitios de unión de ADN, dominios de unión de ADN de hélice alada y sus correspondientes sitios de unión de ADN, dominios de unión de ADN de hélice alada-giro-hélice y sus correspondientes sitios de unión de ADN, dominios de unión de ADN HMG-caja y sus correspondientes secuencias de unión de ADN, dominios de unión de ADN de hélice-bucle-hélice y sus correspondientes secuencias de unión de ADN, y dominios de unión de ADN de hélice-giro-hélice y sus correspondientes secuencias de unión de ADN. Otros dominios de unión de ADN conocidos con secuencias de unión de ADN conocidas incluyen el dominio de ADN de inmunoglobulina, dominio de unión de ADN B3 y dominios de unión de ADN efector TAL. Las subunidades de armazón de ácido nucleico de la presente descripción pueden comprender dos o más de los sitios de unión de ADN mencionados antes.

25 Como se usa en el presente documento, el término "gen" se refiere a cualquier segmento de ADN asociado con una función biológica. Por lo tanto, los genes incluyen, pero no se limitan a, secuencias codificantes y/o las secuencias reguladoras necesarias para su expresión. Los genes también pueden incluir segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes se pueden obtener de una variedad de fuentes, que incluyen la clonación de una fuente de interés o la síntesis a partir de información de secuencia conocida o prevista, y pueden incluir secuencias diseñadas para tener los parámetros deseados.

30 Como se usa en el presente documento, el término "homólogo" u "ortólogo" se conoce en la técnica y se refiere a secuencias relacionadas que comparten un ancestro o miembro de familia común y se determinan basándose en el grado de identidad de secuencia. Los términos "homología", "homólogo", "sustancialmente similar" y "que se corresponde sustancialmente" se usan indistintamente en el presente documento. Se refieren a fragmentos de ácido nucleico en donde los cambios en una o más bases de nucleótidos no afectan a la capacidad del fragmento de ácido nucleico para mediar en la expresión génica o producir un determinado fenotipo. Estos términos también se refieren a modificaciones de los fragmentos de ácido nucleico de la presente descripción tales como delección o inserción de uno o más nucleótidos que no alteran sustancialmente las propiedades funcionales del fragmento de ácido nucleico resultante con respecto al fragmento inicial no modificado. Por lo tanto, se entiende, como apreciarán los expertos en la técnica, que la descripción abarca más que las secuencias de ejemplo específicas. Estos términos describen la relación entre un gen que se encuentra en una especie, subespecie, variedad, variedad cultivada o cepa y el gen correspondiente o equivalente en otra especie, subespecie, variedad, cultivo o cepa. Para los propósitos de esta descripción, se comparan secuencias homólogas. Se piensa, se cree o se sabe que las "secuencias homólogas" u "homólogos" u "ortólogos" están funcionalmente relacionados. Una relación funcional se puede indicar en cualquiera

de varias formas, que incluyen, pero no se limitan a: (a) grado de identidad de secuencia y/o (b) la misma o similar función biológica. Preferiblemente, se indican tanto (a) como (b). La homología se puede determinar usando programas de software fácilmente disponibles en la técnica, tales como los descritos en *Current Protocols in Molecular Biology* (F.M. Ausubel et al., Eds., 1987) Suplemento 30, sección 7.718, Tabla 7.71. Algunos programas de alineamiento son MacVector (Oxford Molecular Ltd, Oxford, Reino Unido), ALIGN Plus (Scientific and Educational Software, Pensilvania) y AlignX (Vector NTI, Invitrogen, Carlsbad, CA). Otro programa de alineamiento es Sequencher (Gene Codes, Ann Arbor, Michigan), que utiliza parámetros predeterminados.

Como se usa en el presente documento, el término "endógeno" o "gen endógeno" se refiere al gen de origen natural, en la ubicación en la que se encuentra de forma natural dentro del genoma de la célula hospedante. En el contexto de la presente descripción, unir operativamente un promotor heterólogo a un gen endógeno significa insertar genéticamente una secuencia promotora heteróloga delante de un gen existente, en la ubicación donde ese gen está presente de forma natural. Un gen endógeno como se describe en el presente documento puede incluir alelos de genes naturales que se han mutado de acuerdo con cualquiera de los métodos de la presente descripción.

Como se usa en el presente documento, el término "exógeno" se usa indistintamente con el término "heterólogo" y se refiere a una sustancia que proviene de alguna fuente distinta a su fuente nativa. Por ejemplo, las expresiones "proteína exógena" o "gen exógeno" se refieren a una proteína o gen de una fuente o ubicación no nativa, y que se han suministrado artificialmente a un sistema biológico.

Como se usa en el presente documento, la expresión "cambio de nucleótidos" se refiere, p. ej., a la sustitución, delección y/o inserción de nucleótidos, como se entiende bien en la técnica. Por ejemplo, las mutaciones contienen alteraciones que producen sustituciones, adiciones o delecciones silenciosas, pero no alteran las propiedades o actividades de la proteína codificada o cómo se producen las proteínas.

Como se usa en el presente documento, la expresión "modificación de proteínas" se refiere, p. ej., a la sustitución de aminoácidos, modificación, delección y/o inserción de aminoácidos, como se entiende bien en la técnica.

Como se usa en el presente documento, la expresión "al menos una parte" o "fragmento" de un ácido nucleico o polipéptido significa una parte que tiene las características de tamaño mínimo de dichas secuencias, o cualquier fragmento más grande de la molécula de longitud completa, hasta e incluyendo la molécula de longitud completa. Un fragmento de un polinucleótido de la descripción puede codificar una parte biológicamente activa de un elemento regulador genético. Se puede preparar una parte biológicamente activa de un elemento regulador genético aislando una parte de uno de los polinucleótidos de la descripción que comprende el elemento regulador genético y evaluando la actividad como se describe en el presente documento. De manera similar, una parte de un polipéptido puede ser de 4 aminoácidos, 5 aminoácidos, 6 aminoácidos, 7 aminoácidos, etc., subiendo hasta el polipéptido de longitud completa. La longitud de la parte que se va a usar dependerá de la aplicación particular. Una parte de un ácido nucleico útil como sonda de hibridación puede ser tan corta como 12 nucleótidos; en algunos ejemplos, es de 20 nucleótidos. Una parte de un polipéptido útil como un epítipo puede ser tan corta como 4 aminoácidos. Una parte de un polipéptido que realiza la función del polipéptido de longitud completa generalmente sería más larga que 4 aminoácidos.

Los polinucleótidos variantes también abarcan secuencias derivadas de un procedimiento mutagénico y recombinogénico, tal como el barajado de ADN. Las estrategias para dicho barajado de ADN son conocidas en la técnica. Véase, por ejemplo, Stemmer (1994) *PNAS* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Cramer et al. (1997) *Nature Biotech.* 15:436-438; Moore et al. (1997) *J. Mol. Biol.* 272:336-347; Zhang et al. (1997) *PNAS* 94:4504-4509; Cramer et al. (1998) *Nature* 391:288-291; y patentes de EE.UU. nº 5.605.793 y 5.837.458.

Para las amplificaciones por PCR de los polinucleótidos descritos en el presente documento, se pueden diseñar cebadores oligonucleótidos para usar en reacciones de PCR para amplificar las secuencias de ADN correspondientes a partir de ADNc o ADN genómico extraído de cualquier organismo de interés. Los métodos para diseñar cebadores de PCR y la clonación de PCR se conocen en general en la técnica y se describen en Sambrook et al. (2001) *Molecular Cloning: A Laboratory Manual* (3ª ed., Cold Spring Harbor Laboratory Press, Plainview, New York). Véase también Innis et al., eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis y Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); e Innis y Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Los métodos conocidos de PCR incluyen, pero no se limitan a métodos que usan cebadores emparejados, cebadores anidados, cebadores específicos únicos, cebadores degenerados, cebadores específicos de genes, cebadores específicos de vector, cebadores parcialmente mal emparejados y similares.

El término "cebador", como se usa en el presente documento, se refiere a un oligonucleótido que es capaz de hibridar con la diana de amplificación permitiendo que una ADN polimerasa se una, sirviendo así como un punto de inicio de la síntesis de ADN cuando se pone en condiciones en las que se induce la síntesis del producto de extensión del cebador, es decir, en presencia de nucleótidos y un agente para la polimerización tal como ADN polimerasa y a una temperatura y pH adecuados. El cebador (de amplificación) es preferiblemente monocatenario para una máxima eficacia en la amplificación. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente para la polimerización. Las longitudes exactas de los cebadores dependerán de muchos factores, que incluyen la temperatura y la composición (contenido de A/T frente a G/C) del cebador. Un par de cebadores bidireccionales consisten en un

cebador directo y uno inverso como se usa comúnmente en la técnica de la amplificación de ADN, tal como en la amplificación por PCR.

Como se usa en el presente documento, "promotor" se refiere a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En algunos ejemplos, la secuencia de promotor consiste en elementos secuencia arriba proximales y más distales, estos últimos elementos a menudo denominados potenciadores. Por consiguiente, un "potenciador" es una secuencia de ADN que puede estimular la actividad del promotor y puede ser un elemento innato del promotor o un elemento heterólogo insertado para mejorar el nivel o la especificidad tisular de un promotor. Los promotores se pueden obtener en su totalidad de un gen nativo, o estar compuestos de diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintéticos. Los expertos en la técnica entenderán que diferentes promotores pueden dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes etapas de desarrollo, o en respuesta a diferentes condiciones ambientales. Se reconoce además que, puesto que en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido completamente, los fragmentos de ADN de alguna variación pueden tener actividad promotora idéntica.

Como se usa en el presente documento, las frases "construcción recombinante", "construcción de expresión", "construcción quimérica", "construcción" y "construcción de ADN recombinante" se usan indistintamente en el presente documento. Una construcción recombinante comprende una combinación artificial de fragmentos de ácido nucleico, p. ej., secuencias reguladoras y codificantes que no se encuentran juntas en la naturaleza. Por ejemplo, una construcción quimérica puede comprender secuencias reguladoras y secuencias codificantes que derivan de diferentes fuentes, o secuencias reguladoras y secuencias codificantes derivadas de la misma fuente, pero dispuestas de una manera diferente a la que se encuentra en la naturaleza. Dicha construcción se puede usarse por sí misma o se puede usar junto con un vector. Si se usa un vector, entonces la elección del vector depende del método que se usará para transformar las células hospedantes, como es bien conocido por los expertos en la técnica. Por ejemplo, se puede utilizar un vector plasmídico. El experto en la técnica conoce bien los elementos genéticos que deben estar presentes en el vector con el fin de transformar, seleccionar y propagar con éxito las células hospedantes que comprenden cualquiera de los fragmentos de ácido nucleico aislados de la descripción. El experto en la técnica también reconocerá que diferentes sucesos de transformación independientes darán como resultado diferentes niveles y patrones de expresión (Jones et al., (1985) *EMBO J.* 4: 2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218: 78-86) y, por lo tanto, se deben cribar múltiples sucesos para obtener líneas que presenten el nivel y patrón de expresión deseados. Dicho cribado se puede llevar a cabo por análisis Southern de ADN, análisis Northern de expresión de ARNm, análisis de inmunotransferencia de expresión de proteínas o análisis fenotípico, entre otros. Los vectores pueden ser plásmidos, virus, bacteriófagos, pro-virus, fagémidos, transposones, cromosomas artificiales y similares, que se replican de forma autónoma o se pueden integrar en un cromosoma de una célula hospedante. Un vector también puede ser un polinucleótido de ARN desnudo, un polinucleótido de ADN desnudo, un polinucleótido compuesto de ADN y ARN dentro de la misma cadena, un ADN o ARN conjugado con polilisina, un ADN o ARN conjugado con péptido, un ADN conjugado con liposoma, o similar, que no se replica de forma autónoma. Como se usa en el presente documento, el término "expresión" se refiere a la producción de un producto final funcional, p. ej., un ARNm o una proteína (precursora o madura).

"Operativamente unido" significa en este contexto la disposición secuencial del polinucleótido promotor de acuerdo con la descripción con un oligo o polinucleótido adicional, dando como resultado la transcripción de dicho polinucleótido adicional.

La expresión "producto de interés" o "biomolécula" como se usa en el presente documento se refiere a cualquier producto producido por microbios a partir de materia prima. En algunos casos, el producto de interés puede ser una molécula pequeña, enzima, péptido, aminoácido, ácido orgánico, compuesto sintético, combustible, alcohol, etc. Por ejemplo, el producto de interés o biomolécula puede ser cualquier metabolito extracelular primario o secundario. El metabolito primario puede ser, entre otros, etanol, ácido cítrico, ácido láctico, ácido glutámico, glutamato, lisina, treonina, triptófano y otros aminoácidos, vitaminas, polisacáridos, etc. El metabolito secundario puede ser, entre otros, un compuesto antibiótico tal como la penicilina o un inmunosupresor tal como la ciclosporina A, una hormona vegetal tal como la giberelina, un fármaco estatina tal como la lovastatina, un fungicida tal como la griseofulvina, etc. El producto de interés o biomolécula también puede ser cualquier componente intracelular producido por un microbio, tal como: una enzima microbiana, que incluye: catalasa, amilasa, proteasa, pectinasa, glucosa isomerasa, celulasa, hemicelulasa, lipasa, lactasa, estreptoquinasa y muchas otras. El componente intracelular también puede incluir proteínas recombinantes, tales como: insulina, vacuna contra la hepatitis B, interferón, factor estimulante de colonias de granulocitos, estreptoquinasa y otros.

La expresión "fuente de carbono" se refiere en general a una sustancia adecuada para ser usada como fuente de carbono para el crecimiento celular. Las fuentes de carbono incluyen, pero no se limitan a hidrolizados de biomasa, almidón, sacarosa, celulosa, hemicelulosa, xilosa y lignina, así como componentes monoméricos de estos sustratos. Las fuentes de carbono pueden comprender diferentes compuestos orgánicos en diversas formas, que incluyen, pero no se limitan a polímeros, carbohidratos, ácidos, alcoholes, aldehídos, cetonas, aminoácidos, péptidos, etc. Estos incluyen, por ejemplo, varios monosacáridos tales como glucosa, dextrosa (D-glucosa), maltosa, oligosacáridos, polisacáridos, ácidos grasos saturados o insaturados, succinato, lactato, acetato, etanol, etc., o mezclas de los mismos. Los organismos fotosintéticos pueden además producir una fuente de carbono como producto de la

fotosíntesis. En algunos ejemplos, las fuentes de carbono se pueden seleccionar de hidrolizados de biomasa y glucosa.

5 El término "materia prima" se define como una materia prima o mezcla de materias primas suministradas a un microorganismo o proceso de fermentación a partir del cual se pueden producir otros productos. Por ejemplo, una fuente de carbono, tal como la biomasa o los compuestos de carbono derivados de la biomasa son una materia prima para un microorganismo que produce un producto de interés (p. ej., molécula pequeña, péptido, compuesto sintético, combustible, alcohol, etc.) en un proceso de fermentación. Sin embargo, una materia prima puede contener otros nutrientes además de una fuente de carbono.

10 El término "productividad volumétrica" o "tasa de producción" se define como la cantidad de producto formado por volumen de medio por unidad de tiempo. La productividad volumétrica se puede dar en gramos por litro por hora (g/l/h).

15 La expresión "productividad específica" se define como la tasa de formación del producto. La productividad específica se define en el presente documento además como la productividad específica en gramos de producto por gramo de peso seco de células (CDW) por hora (g/g CDW/h). Usando la relación de CDW a DO_{600} para el microorganismo dado, la productividad específica también se puede expresar como gramos de producto por litro de medio de cultivo por densidad óptica del caldo de cultivo a 600 nm (DO) por hora (g/L/h/DO).

20 El término "rendimiento" se define como la cantidad de producto obtenido por unidad de peso de materia prima y se puede expresar como g de producto por g de sustrato (g/g). El rendimiento se puede expresar como porcentaje del rendimiento teórico. El "rendimiento teórico" se define como la cantidad máxima de producto que se puede generar por una cantidad dada de sustrato según lo dictado por la estequiometría de la ruta metabólica usada para producir el producto.

El término "título" se define como la fuerza de una solución o la concentración de una sustancia en solución. Por ejemplo, el título de un producto de interés (p. ej., molécula pequeña, péptido, compuesto sintético, combustible, alcohol, etc.) en un caldo de fermentación se describe como g de producto de interés en solución por litro de caldo de fermentación (g/l).

25 La expresión "título total" se define como la suma de todo el producto de interés producido en un proceso, que incluye, pero no se limita al producto de interés en solución, el producto de interés en fase gaseosa, si se aplica, y cualquier producto de interés retirado del proceso y recuperado con respecto al volumen inicial en el proceso o el volumen operativo en el proceso. Como se usa en el presente documento, la expresión "biblioteca de diseño genético HTP" o "biblioteca" se refiere a colecciones de perturbaciones genéticas según la presente descripción. En algunos ejemplos, las bibliotecas de la presente descripción se pueden poner de manifiesto como i) una colección de información de secuencias en una base de datos u otro archivo informático, ii) una colección de construcciones genéticas que codifican la serie de elementos genéticos mencionada antes, o iii) cepas de células hospedantes que comprenden dichos elementos genéticos. En algunos ejemplos, las bibliotecas de la presente descripción pueden referirse a colecciones de elementos individuales (p. ej., colecciones de promotores para bibliotecas de intercambio de PRO, colecciones de terminadores para bibliotecas de intercambio de STOP, colecciones de marcadores de solubilidad de proteínas para bibliotecas de intercambio de marcadores de solubilidad o colecciones de marcadores de degradación de proteínas para bibliotecas de intercambio de marcadores de degradación). En otros ejemplos, las bibliotecas de la presente descripción también pueden referirse a combinaciones de elementos genéticos, tales como combinaciones de promotores:genes, genes:terminadores o incluso promotores:genes:terminadores. En algunos ejemplos, las bibliotecas de la presente descripción también pueden referirse a combinaciones de promotores, terminadores, marcadores de solubilidad de proteínas y/o marcadores de degradación de proteínas. En algunos ejemplos, las bibliotecas de la presente descripción comprenden además metadatos asociados con los efectos de aplicar cada miembro de la biblioteca en organismos hospedantes. Por ejemplo, una biblioteca como se usa en el presente documento puede incluir una colección de combinaciones de secuencias de promotor::gen, junto con el efecto resultante de esas combinaciones en uno o más fenotipos en una especie en particular, mejorando así el valor predictivo futuro de usar dicha combinación en futuros intercambios de promotores.

50 Como se usa en el presente documento, el término "SNP" se refiere a polimorfismo(s) nuclear(es) pequeño(s). En algunos ejemplos, los SNP de la presente descripción deben interpretarse de manera amplia e incluyen polimorfismos de un solo nucleótido, inserciones, deleciones, inversiones de secuencia y otros reemplazos de secuencia. Como se usa en el presente documento, la expresión "no sinónimos" o "SNP no sinónimos" se refiere a mutaciones que conducen a cambios en la codificación de las proteínas de la célula hospedante. Un método de "alta capacidad (HTP)" de ingeniería genómica puede implicar la utilización de al menos una pieza de equipo automatizado (p. ej., un manipulador de líquidos o una máquina manipuladora de placas) para llevar a cabo al menos un paso de dicho método.

Métodos tradicionales de mejora de cepas

55 Los enfoques tradicionales para la mejora de cepas se pueden clasificar ampliamente en dos tipos de enfoques: transformación de cepas dirigida y mutagénesis aleatoria.

Los métodos de transformación dirigidos para mejorar las cepas implican la perturbación planificada de unos cuantos elementos genéticos de un organismo específico. Estos enfoques se centran típicamente en modular programas

biosintéticos o de desarrollo específicos y se basan en el conocimiento previo de los factores genéticos y metabólicos que afectan a dichas rutas. En sus ejemplos más simples, la ingeniería dirigida implica la transferencia de un rasgo caracterizado (p. ej., gen, promotor u otro elemento genético capaz de producir un fenotipo cuantificable) de un organismo a otro organismo de la misma o diferente especie.

5 Los enfoques aleatorios para la transformación de cepas implican la mutagénesis aleatoria de las cepas parentales, acoplada con un cribado exhaustivo diseñado para identificar mejoras en el rendimiento. Los enfoques para generar estas mutaciones aleatorias incluyen la exposición a radiación ultravioleta o productos químicos mutagénicos tales como el metanosulfonato de etilo. Aunque aleatorio y en gran parte impredecible, este enfoque tradicional para mejorar cepas tenía varias ventajas en comparación con las manipulaciones genéticas más dirigidas. Primero, muchos
10 organismos industriales estaban (y siguen estando) mal caracterizados en términos de sus repertorios genéticos y metabólicos, haciendo difíciles, sino imposibles, los enfoques alternativos de mejora dirigida.

Segundo, incluso en sistemas relativamente bien caracterizados, los cambios genotípicos que dan como resultado mejoras en el rendimiento industrial son difíciles de predecir y, a veces, solo se manifiestan como fenotipos epistáticos que requieren mutaciones acumulativas en muchos genes de función conocida y desconocida.

15 Adicionalmente, durante muchos años, las herramientas genéticas necesarias para realizar mutaciones genómicas dirigidas en un organismo industrial dado no han estado disponibles o eran muy lentas y/o difíciles de usar.

La aplicación extendida de los programas tradicionales de mejora cepas, sin embargo, da ganancias progresivamente reducidas en un linaje de cepa dado y, finalmente conduce a agotar las posibilidades de obtener mayores eficiencias de las cepas. Las mutaciones aleatorias beneficiosas son eventos relativamente raros y requieren grandes grupos de
20 detección y altas tasas de mutación. Esto inevitablemente da como resultado la acumulación accidental de muchas mutaciones neutrales y/o perjudiciales (o parcialmente perjudiciales) en cepas "mejoradas", que finalmente crean un lastre para las futuras ganancias de eficiencia.

Otra limitación de los enfoques tradicionales de mejora acumulada es que se conoce poca o ninguna información sobre cualquier efecto de una mutación particular en cualquier medida de la cepa. Esto limita fundamentalmente la
25 capacidad de un investigador para combinar y consolidar mutaciones beneficiosas, o para eliminar el "bagaje" mutagénico neutro o perjudicial.

Existen otros enfoques y tecnologías para recombinar aleatoriamente mutaciones entre cepas dentro de un linaje mutagénico. Por ejemplo, algunos formatos y ejemplos para la recombinación de secuencias iterativas, a veces denominados barajado de ADN, evolución o reproducción molecular, se han descrito en las solicitudes de patente de
30 EE.UU. N° de serie 08/198.431, presentada el 17 de Feb., 1994, N° de serie PCT/US95/02126, presentada el 17 de Feb., 1995, N° de serie 08/425.684, presentada el 18 de abril, 1995, N° de serie 08/537.874, presentada el 30 de Oct., 1995, N° de serie 08/564.955, presentada el 30 de Nov., 1995, N° de serie 08/621.859, presentada el 25 de Marzo, 1996, N° de serie 08/621.430, presentada el 25 de Marzo, 1996, N° de serie PCT/US96/05480, presentada el 18 de abril, 1996, N° de serie 08/650.400, presentada el 20 de Mayo, 1996, N° de serie 08/675.502, presentada el 3 de Jul.,
35 1996, N° de serie 08/721.824, presentada el 27 de Sep., 1996, y N° de serie 08/722.660 presentada el 27 de Sep., 1996; Stemmer, *Science* 270:1510 (1995); Stemmer et al., *Gene* 164:49-53 (1995); Stemmer, *Bio/Technology* 13:549-553 (1995); Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91:10747-10751 (1994); Stemmer, *Nature* 370:389-391 (1994); Cramer et al., *Nature Medicine* 2(1):1-3 (1996); Cramer et al., *Nature Biotechnology* 14:315-319 (1996).

Estos incluyen técnicas como la fusión de protoplastos y barajado del genoma completo que facilitan la recombinación genómica a través de cepas mutadas. Para algunos microorganismos industriales tales como las levaduras y los hongos filamentosos, también se pueden aprovechar los ciclos de apareamiento naturales para la recombinación genómica por pares. De esta manera, las mutaciones perjudiciales se pueden eliminar por el "retrocruzamiento" de
40 mutantes con cepas parentales y consolidar las mutaciones beneficiosas. Además, las mutaciones beneficiosas de dos linajes de cepas diferentes se pueden combinar potencialmente, lo que crea posibilidades de mejora adicionales de lo que podría estar disponible de la mutación de un solo linaje de cepas por sí solo. Sin embargo, estos enfoques están sujetos a muchas limitaciones que se eluden usando los métodos de la presente descripción.

Por ejemplo, los enfoques recombinantes tradicionales descritos antes son lentos y se basan en un número relativamente pequeño de sucesos cruzados de recombinación aleatoria para intercambiar mutaciones y, por lo tanto, están limitados en el número de combinaciones que se pueden intentar en cualquier ciclo o periodo de tiempo dado.
50 Además, aunque los sucesos de recombinación natural en la técnica anterior son esencialmente aleatorios, también están sujetos al sesgo posicional del genoma.

Lo que es más importante, los enfoques tradicionales también proporcionan poca información sobre la influencia de mutaciones individuales y, debido a la distribución aleatoria de mutaciones recombinadas, muchas combinaciones específicas no se pueden generar ni evaluar.

55 Para superar muchos de los problemas mencionados anteriormente asociados con los programas tradicionales de mejora de cepas, la presente descripción establece una plataforma de ingeniería genómica de HTP única que se dirige computacionalmente e integra biología molecular, automatización, análisis de datos y protocolos de aprendizaje automático. Esta plataforma integradora utiliza una serie de conjuntos de herramientas moleculares de HTP que se

usan para construir bibliotecas de diseño genético de HTP. Estas bibliotecas de diseño genético se elaborarán a continuación. La Figura 8 representa una visión general de un ejemplo del programa de mejora de cepas de *E. coli* de la presente descripción.

5 La plataforma de HTP actualmente descrita y sus bibliotecas de diseño genético microbiano únicas cambian fundamentalmente el paradigma del desarrollo y evolución de cepas microbianas. Por ejemplo, los métodos tradicionales basados en mutagénesis de desarrollo de una cepa microbiana industrial finalmente conducirán a microbios cargados con una carga mutagénica pesada que se ha acumulado a lo largo de años de mutagénesis aleatoria.

10 La capacidad para resolver este problema (es decir, eliminar el bagaje genético acumulado por estos microbios) ha eludido a los investigadores microbianos durante décadas. Sin embargo, utilizando la plataforma de HTP descrita en el presente documento, estas cepas industriales se pueden "rehabilitar" y las mutaciones genéticas que son perjudiciales se pueden identificar y eliminar. De manera congruente, las mutaciones genéticas que se identifican como beneficiosas se pueden mantener y, en algunos casos, mejorar. Las cepas microbianas resultantes demuestran rasgos fenotípicos superiores (p. ej., producción mejorada de un compuesto de interés), en comparación con sus cepas parentales.

15 Además, la plataforma de HTP enseñada en el presente documento es capaz de identificar, caracterizar y cuantificar el efecto que tienen las mutaciones individuales en el rendimiento de la cepa microbiana. Esta información, es decir, qué efecto tiene un cambio genético dado x en el fenotipo de la célula hospedante y (p. ej., producción de un compuesto o producto de interés), se puede generar y después almacenar en las bibliotecas de diseño genético de HTP microbiano descritas a continuación. Es decir, la información de secuencia para cada permutación genética y su efecto en el fenotipo de la célula hospedante se almacenan en una o más bases de datos y están disponibles para análisis posteriores (por ejemplo, mapeo de epistasias, como se describe a continuación). La presente descripción también enseña métodos para guardar/almacenar físicamente permutaciones genéticas valiosas en forma de construcciones de inserción genética, o en forma de uno o más organismos de células hospedantes que contienen dicha permutación genética (p. ej., véanse las bibliotecas descritas a continuación).

20 Cuando se combinan estas bibliotecas de diseño genético de HTP en un procedimiento iterativo que está integrado con un análisis de datos sofisticado y un proceso de aprendizaje automático, surge una metodología radicalmente diferente para mejorar las células hospedantes. Por lo tanto, la plataforma enseñada es fundamentalmente diferente de los métodos tradicionales descritos antes de desarrollo de cepas de células hospedantes. La plataforma de HTP enseñada no adolece de muchos de los inconvenientes asociados con los métodos anteriores. Estas y otras ventajas resultarán evidentes con referencia a los conjuntos de herramientas moleculares de HTP y las bibliotecas de diseño genético derivadas que se analizan a continuación.

Diseño genético e ingeniería microbiana: un enfoque combinatorio sistemático para mejorar cepas utilizando una serie de herramientas moleculares de HTP y bibliotecas de diseño genético de HTP

35 Como se ha mencionado antes, la presente descripción proporciona una plataforma de HTP y una estrategia de diseño genético novedosos para transformar organismos microbianos mediante la introducción y eliminación sistemáticas iterativas de cambios genéticos entre cepas. La plataforma está respaldada por una serie de herramientas moleculares, que permiten la creación de bibliotecas de diseño genético de HTP y permiten la implementación eficiente de alteraciones genéticas en una cepa hospedante dada.

40 Las bibliotecas de diseño genético de HTP de la descripción sirven como fuentes de posibles alteraciones genéticas que se pueden introducir en un contexto de cepa microbiana particular. De esta manera, las bibliotecas de diseño genético de HTP son repositorios de diversidad genética, o colecciones de perturbaciones genéticas, que se pueden aplicar a la ingeniería inicial o posterior de una cepa microbiana dada. Las técnicas para programar diseños genéticos para su implementación en cepas hospedantes se describen en la solicitud de patente de EE.UU. en tramitación, n^o de serie 15/140.296, y solicitud de patente internacional en tramitación de n^o de serie PCT/US17/29725, titulada "Sistema y métodos de diseño de cepas microbianas para mejorar la producción a gran escala de secuencias de nucleótidos modificadas".

45 Los conjuntos de herramientas moleculares de HTP utilizados en esta plataforma pueden incluir, entre otros: (1) Intercambios de promotores (Intercambio de PRO), (2) Intercambios de SNP, (3) Intercambios de codones de inicio/detención, (4) Intercambios de STOP, (5) optimización de secuencia, (6) intercambios de MARCADORES DE SOLUBILIDAD y (7) intercambios de MARCADORES DE DEGRADACIÓN. Los métodos de HTP de la presente descripción también enseñan métodos para dirigir la consolidación/uso combinatorio de conjuntos de herramientas de HTP, que incluyen (8) protocolos de mapeo de epistasias. Como se ha mencionado antes, esta serie de herramientas moleculares, ya sea de forma aislada o en combinación, permite la creación de bibliotecas de células hospedantes de diseño genético de HTP.

50 Como se demostrará, la utilización de las bibliotecas de diseño genético de HTP mencionadas antes en el contexto de la plataforma de ingeniería microbiana de HTP enseñada permite la identificación y consolidación de mutaciones o secciones de genes "causativas" beneficiosas y también la identificación y eliminación de mutaciones o secciones de

genes pasivos o perjudiciales. Este nuevo enfoque permite mejoras rápidas en el rendimiento de la cepa que no podrían lograrse por la mutagénesis aleatoria tradicional o la ingeniería genética dirigida. La eliminación de la carga genética o la consolidación de cambios beneficiosos en una cepa sin carga genética también proporciona un punto de partida nuevo y sólido para mutagénesis aleatoria adicional que puede permitir mejoras adicionales.

5 En algunos ejemplos, la presente descripción enseña que a medida que se identifican cambios ortogonales beneficiosos a través de varias ramas discretas de un linaje de cepas mutagénicas, también se pueden consolidar rápidamente en cepas de mejor rendimiento. Estas mutaciones también se pueden consolidar en cepas que no forman parte de linajes mutagénicos, tales como cepas con mejoras obtenidas por ingeniería genética dirigida.

10 En algunos ejemplos, la presente descripción difiere de los enfoques de mejora de cepas conocidos en cuanto que analiza el efecto combinatorio de mutaciones en todo el genoma a lo largo de múltiples regiones genómicas dispares, incluyendo elementos genéticos expresados y no expresados, y usa la información recopilada (p. ej., resultados experimentales) para predecir las combinaciones de mutaciones que se espera produzcan mejoras en las cepas.

15 En algunos ejemplos, la presente descripción enseña: i) microorganismos industriales y otras células hospedantes susceptibles de mejora a través de las divulgaciones descritas, ii) la generación de grupos de diversidad para análisis secuencial abajo, iii) métodos y hardware para cribado y secuenciación de alta capacidad de grupos de variantes grandes, iv) métodos y hardware para el análisis computacional de aprendizaje automático y la predicción de efectos sinérgicos de mutaciones en todo el genoma, y v) métodos para la ingeniería de cepas de alta capacidad.

20 Las siguientes herramientas y bibliotecas moleculares se describen en términos de ejemplos microbianos ilustrativos. Los expertos en la técnica reconocerán que las herramientas moleculares de HTP de la presente descripción son compatibles con cualquier célula hospedante, incluyendo formas celulares eucariotas y de vida superiores. Además, muchos de los ejemplos ilustrados se llevan a cabo en *Corynebacterium*; sin embargo, se pueden implementar los mismos principios y procedimientos en *Escherichia coli*.

Ahora se describirá cada uno de los conjuntos de herramientas moleculares HTP identificados, que permiten la creación de las diversas bibliotecas de diseño genético de HTP utilizadas en la plataforma de ingeniería microbiana.

25 **1. Intercambios de promotores: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de intercambios de promotores**

En algunos ejemplos, la presente descripción enseña métodos para seleccionar promotores con propiedades de expresión óptimas para producir efectos beneficiosos en el fenotipo general de la cepa del hospedante (p. ej., rendimiento o productividad).

30 Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos para identificar uno o más promotores y/o generar variantes de uno o más promotores dentro de una célula hospedante, que presentan un intervalo de fuerzas de expresión (p. ej., escaleras de promotores descritas más adelante), o propiedades reguladoras superiores (p. ej., control regulador más estrecho para genes seleccionados). Una combinación particular de estos promotores identificados y/o generados se puede agrupar como una escalera de promotores, que se explica con más detalle a continuación.

35 La escalera del promotor en cuestión después se asocia con un gen de interés dado. Por lo tanto, si se tienen los promotores P₁-P₈ (que representan ocho promotores que se han identificado y/o generado para presentar un intervalo de fuerzas de expresión) y se asocia la escalera de promotores con un solo gen de interés en un microbio (es decir, transformar genéticamente un microbio con un promotor dado operativamente unido a un gen diana dado), entonces el efecto de cada combinación de los ocho promotores se puede determinar caracterizando cada una de las cepas transformadas que resultan de cada esfuerzo combinatorio, dado que los microbios transformados tienen un contexto genético por lo demás idéntico excepto el o los promotores particulares asociados con el gen diana.

Los microbios resultantes que se transforman por este procedimiento forman bibliotecas de diseño genético de HTP.

40 La biblioteca de diseño genético de HTP se puede referir a la colección de cepas microbianas físicas reales que se forma por este procedimiento, siendo cada cepa miembro representativa de un promotor dado operativamente unido a un gen diana particular, en un contexto genético por lo demás idéntico, denominándose dicha biblioteca una "biblioteca de cepas microbianas de intercambio de promotores". En el contexto específico de *E. coli*, la biblioteca se puede denominar una "biblioteca de cepas de *E. coli* de intercambio de promotores", pero los términos se pueden usar como sinónimos, ya que *E. coli* es un ejemplo específico de un microbio.

45 Además, la biblioteca de diseño genético de HTP se puede referir a la colección de perturbaciones genéticas, en este caso un promotor dado *x* operativamente unido a un gen *y* dado, dicha colección se denomina "biblioteca de intercambio de promotores".

50 Además, se puede utilizar la misma escalera de promotores que comprende los promotores P₁-P₈ para transformar microbios, en donde cada uno de los 8 promotores está operativamente unido a 10 genes diana diferentes. El resultado de este procedimiento serían 80 microbios que se supone que por lo demás son genéticamente idénticos, excepto por

los promotores particulares operativamente unidos a un gen diana de interés. Estos 80 microbios se podrían cribar y caracterizar adecuadamente y dar lugar a otra biblioteca de diseño genético de HTP. La caracterización de las cepas microbianas en la biblioteca de diseño genético de HTP produce información y datos que se pueden almacenar en cualquier construcción de almacenamiento de datos, incluyendo una base de datos relacional, una base de datos orientada a objetos o una base de datos NoSQL altamente distribuida. Estos datos/información podrían ser, por ejemplo, un efecto del promotor dado (p. ej. P₁-P₈) cuando está operativamente unido a un gen diana dado. Estos datos/información también pueden ser el conjunto más amplio de efectos combinatorios que resultan de unir operativamente dos o más de los promotores P₁-P₈ a un gen diana dado.

Los ejemplos mencionados antes de ocho promotores y 10 genes diana son meramente ilustrativos, ya que el concepto se puede aplicar con cualquier número dado de promotores que se hayan agrupado entre sí basándose en la presentación de un intervalo de fuerzas de expresión y cualquier número dado de genes diana. Las personas con experiencia en la técnica también reconocerán la capacidad de unir operativamente dos o más promotores delante de cualquier gen diana. Por lo tanto, en algunos ejemplos, la presente descripción enseña bibliotecas de intercambio de promotores en las que 1, 2, 3 o más promotores de una escalera de promotores están operativamente unidos a uno o más genes.

En resumen, la utilización de varios promotores para dirigir la expresión de varios genes en un organismo es una herramienta poderosa para optimizar un rasgo de interés. La herramienta molecular de intercambio de promotores, desarrollada por los autores de la invención, usa una escalera de secuencias de promotores que se ha demostrado que varían la expresión de al menos un locus en al menos una condición. Después, esta escalera se aplica sistemáticamente a un grupo de genes en el organismo usando la ingeniería de genoma de alta capacidad. Se determina que este grupo de genes tiene una alta probabilidad de tener impacto en el rasgo de interés basándose en uno cualquiera de una serie de métodos. Estos podrían incluir la selección basada en la función conocida o el impacto en el rasgo de interés, o la selección algorítmica basada en la diversidad genética beneficiosa previamente determinada. En algunos ejemplos, la selección de genes puede incluir todos los genes de un hospedante dado. En otros ejemplos, la selección de genes puede ser un subconjunto de todos los genes en un hospedante dado, elegidos de forma aleatoria.

Después se evalúa en la biblioteca de cepas microbianas de diseño genético HTP resultante de organismos que contienen una secuencia de promotor ligada a un gen su rendimiento en un modelo de cribado de alta capacidad, y se determinan las uniones promotor-gen que conducen a un mayor rendimiento y la información se almacena en una base de datos. La colección de perturbaciones genéticas (es decir, un promotor *x* dado operativamente unido a un gen *y* dado) forma una "biblioteca de intercambio de promotores", que se puede utilizar como una fuente de alteraciones genéticas potenciales para utilizar en el procesamiento de ingeniería microbiana. Con el tiempo, a medida que se implementa un mayor conjunto de perturbaciones genéticas contra una mayor diversidad de contextos genéticos de células hospedantes, cada biblioteca se vuelve más potente como un corpus de datos confirmados experimentalmente que se pueden usar para diseñar de manera más precisa y predecible cambios dirigidos contra cualquier contexto de interés.

Los niveles de transcripción de genes en un organismo son un punto clave de control para afectar al comportamiento del organismo. La transcripción está estrechamente ligada a la traducción (expresión de proteínas), y qué proteínas se expresan en qué cantidades determina el comportamiento del organismo. Las células expresan miles de tipos diferentes de proteínas, y estas proteínas interactúan de numerosas formas complejas para crear funciones. Al variar los niveles de expresión de un conjunto de proteínas de forma sistemática, la función se puede alterar de formas que, debido a la complejidad, son difíciles de predecir. Algunas alteraciones pueden aumentar el rendimiento, por lo que, junto con un mecanismo para evaluar el rendimiento, esta técnica permite la generación de organismos con función mejorada.

En el contexto de una ruta de síntesis de moléculas pequeñas, las enzimas interactúan a través de sus sustratos y productos de moléculas pequeñas en una cadena lineal o ramificada, comenzando con un sustrato y terminando con una molécula pequeña de interés. Debido a que estas interacciones están unidas secuencialmente, este sistema presenta el control distribuido, y el aumento de la expresión de una enzima solo puede aumentar el flujo de la ruta hasta que otra enzima se vuelva limitante de la velocidad.

El análisis de control metabólico (MCA) es un método para determinar, a partir de datos experimentales y primeros principios, qué enzima o enzimas son limitantes de la velocidad. Sin embargo, el MCA está limitado porque requiere una experimentación extensa después de cada cambio de nivel de expresión para determinar la nueva enzima limitante de la velocidad. El intercambio de promotores es ventajoso en este contexto, porque a través de la aplicación de una escalera de promotores a cada enzima en una ruta, se encuentra la enzima limitante, y se puede hacer lo mismo en rondas posteriores para encontrar nuevas enzimas que se conviertan en limitantes de la velocidad. Además, debido a que la función de lectura es una mejor producción de la molécula pequeña de interés, el experimento para determinar qué enzima es limitante es el mismo que la ingeniería para aumentar la producción, acortando así el tiempo de desarrollo. En algunos ejemplos, la presente descripción enseña la aplicación de intercambio de PRO a genes que codifican subunidades individuales de enzimas de múltiples unidades. En otros ejemplos más, la presente descripción enseña métodos para aplicar técnicas de intercambio de PRO a genes responsables de regular enzimas individuales o rutas biosintéticas completas.

En algunos ejemplos, la herramienta de intercambio de promotores de la presente descripción se usa para identificar la expresión óptima de un gen diana seleccionado. En algunos ejemplos, el objetivo del intercambio de promotores puede ser aumentar la expresión de un gen objetivo para reducir los cuellos de botella en una ruta metabólica o genética. En otros ejemplos, el objetivo o el intercambio de promotores puede ser para reducir la expresión del gen diana para evitar gastos de energía innecesarios en la célula hospedante, cuando no se requiere la expresión de dicho gen diana.

En el contexto de otros sistemas celulares como la transcripción, transporte o señalización, se pueden usar varios métodos racionales para probar y descubrir, a priori, qué proteínas son dianas para el cambio de expresión y cuál debería ser ese cambio. Estos métodos racionales reducen el número de perturbaciones que se deben ensayar para encontrar una que mejore el rendimiento, pero lo hacen con un coste significativo. Los estudios de delección de genes identifican proteínas cuya presencia es crítica para una función particular, y después los genes importantes se pueden sobreexpresar. Debido a la complejidad de las interacciones de las proteínas, esto a menudo es ineficaz para aumentar el rendimiento. Se han desarrollado diferentes tipos de modelos que intentan describir, desde los primeros principios, el comportamiento de transcripción o señalización en función de los niveles de proteínas en la célula. Estos modelos a menudo sugieren dianas donde los cambios de expresión pueden conducir a una función diferente o mejorada. Las suposiciones que subyacen a estos modelos son simplistas y los parámetros difíciles de medir, por lo que las predicciones que hacen a menudo son incorrectas, especialmente para organismos que no son modelo. Tanto con la delección de genes como con el modelado, los experimentos necesarios para determinar cómo afectar a un determinado gen son diferentes del trabajo posterior para hacer el cambio que mejora el rendimiento. El intercambio de promotores evita estos desafíos, porque la cepa construida que resalta la importancia de una perturbación particular también es, ya, la cepa mejorada.

Por lo tanto, en ejemplos particulares, el intercambio de promotores es un procedimiento de múltiples etapas que comprende:

1. Seleccionar un conjunto de promotores "x" para que actúen como una "escalera". Idealmente, se ha mostrado que estos promotores conducen a una expresión muy variable a lo largo de múltiples locus genómicos, pero el único requisito es que perturben la expresión génica de alguna manera.

2. Seleccionar un conjunto de genes "n" para direccionamiento. Este conjunto puede ser cada marco de lectura abierto (ORF) en un genoma, o un subconjunto de ORF. El subconjunto se puede elegir usando anotaciones en los ORF relacionados con la función, por relación con perturbaciones beneficiosas previamente demostradas (intercambios de promotores previos o intercambios de SNP previos), por selección algorítmica basada en interacciones epistáticas entre perturbaciones generadas previamente, otros criterios de selección basados en hipótesis relacionadas con ORF beneficiosas para la diana, o mediante selección aleatoria. En otros ejemplos, los genes dirigidos "n" pueden comprender genes que no codifican proteínas, incluyendo ARN no codificantes.

3. Ingeniería de cepas de alta capacidad para llevar a cabo rápidamente, y en algunos ejemplos, en paralelo, las siguientes modificaciones genéticas: Cuando existe un promotor nativo delante del gen diana n y se conoce su secuencia, reemplazar el promotor nativo con cada uno de los promotores x de la escalera. Cuando el promotor nativo no existe, o su secuencia es desconocida, insertar cada uno de los promotores x de la escalera delante del gen n (véase, p. ej., la Figura 21). De esta manera se construye una "biblioteca" (también denominada biblioteca de diseño genético HTP) de cepas, en donde cada miembro de la biblioteca es un caso de promotor x operativamente unido a la diana n, en un contexto genético por lo demás idéntico. Como se describió anteriormente, se pueden insertar combinaciones de promotores, ampliando el intervalo de posibilidades combinatorias sobre las que se construye la biblioteca.

4. Cribado de alta capacidad de la biblioteca de cepas en un contexto en donde su rendimiento frente a una o más medidas es indicativo de que el rendimiento se está optimizando.

Este procedimiento fundamental se puede extender para proporcionar mejoras adicionales en el rendimiento de las cepas, entre otras cosas: (1) Consolidando múltiples perturbaciones beneficiosas en un solo contexto genético de la cepa, ya sea una a la vez en un proceso interactivo, o como múltiples cambios en una sola etapa. Las múltiples perturbaciones pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatoria de cambios. Por ejemplo, si el conjunto de dianas son todos los genes de una ruta, la regeneración secuencial de la biblioteca de perturbaciones en un miembro o miembros mejorados de la biblioteca de cepas previa puede optimizar el nivel de expresión de todos los genes en una ruta, independientemente de qué genes son limitantes de velocidad en cualquier iteración dada; (2) Introduciendo los datos de rendimiento que resultan de la generación individual y combinatoria de la biblioteca en un algoritmo que usa esos datos para predecir un conjunto óptimo de perturbaciones basado en la interacción de cada perturbación; y (3) Implementando una combinación de los dos enfoques anteriores (véase la Figura 20).

La herramienta o técnica molecular descrita antes se caracteriza como intercambio de promotores, pero no se limita a los promotores y puede incluir otros cambios de secuencias que varían sistemáticamente el nivel de expresión de un conjunto de dianas. Otros métodos para variar el nivel de expresión de un conjunto de genes podrían incluir: a) una escalera de sitios de unión a ribosomas (o secuencias de Kozak en eucariotas); b) reemplazar el codón de inicio de

5 cada diana por cada uno de los otros codones de inicio (es decir, intercambios de codones de inicio/parada descritos más adelante); c) unión de varias secuencias estabilizadoras o desestabilizadoras de ARNm al extremo 5' o 3', o en cualquier otra ubicación, de una transcripción, d) unión de varias secuencias estabilizadoras o desestabilizadoras de proteínas en cualquier ubicación de la proteína (es decir, degradación o intercambios de marcadores de solubilización descritos más adelante).

El enfoque se ilustra en la presente descripción con microorganismos industriales, pero es aplicable a cualquier organismo donde se puedan identificar los rasgos deseados en una población de mutantes genéticos. Por ejemplo, esto se podría usar para mejorar el rendimiento de células CHO, levaduras, células de insectos, algas, así como organismos multicelulares, tales como plantas.

10 **2. Intercambios de SNP: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de intercambio de SNP**

En ciertos ejemplos, el intercambio de SNP no es un enfoque mutagénico aleatorio para mejorar una cepa microbiana, sino que implica más bien la introducción o eliminación sistemática de mutaciones de nucleótidos individuales de polimorfismo nuclear pequeño (p. ej., SNP) (de ahí el nombre "intercambio de SNP") a lo largo de las cepas.

15 Los microbios resultantes que se transforman mediante este procedimiento forman bibliotecas de diseño genético de HTP.

20 La biblioteca de diseño genético de HTP se puede referir a la colección de cepas microbianas físicas reales que se forma a través de este procedimiento, siendo cada cepa miembro representativa de la presencia o ausencia de un SNP dado, en un contexto genético por lo demás idéntico, denominándose dicha biblioteca "biblioteca de cepas microbianas de intercambio de SNP". En el contexto específico de *E. coli*, la biblioteca se puede denominar "biblioteca de cepas de *E. coli* de intercambio de SNP", pero los términos se pueden usar como sinónimos, ya que *E. coli* es un ejemplo específico de un microbio.

25 Además, la biblioteca de diseño genético de HTP se puede referir a la colección de perturbaciones genéticas, en este caso estando presente un SNP dado o estando ausente un SNP dado, denominándose dicha colección "biblioteca de intercambio de SNP".

30 En algunos ejemplos, el intercambio de SNP implica la reconstrucción de organismos hospedantes con combinaciones óptimas de "bloques de construcción" de SNP diana con efectos de rendimiento beneficiosos identificados. Por lo tanto, en algunos ejemplos, el intercambio de SNP implica la consolidación de múltiples mutaciones beneficiosas en un solo contexto genético de cepa, ya sea una a la vez en un proceso iterativo o como múltiples cambios en una sola etapa. Los múltiples cambios pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatorizada de mutaciones.

35 En otros ejemplos, el intercambio de SNP también implica la eliminación de múltiples mutaciones identificadas como perjudiciales de una cepa, ya sea una a la vez en un proceso iterativo o como múltiples cambios en una sola etapa. Los cambios múltiples pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatorizada de mutaciones. En algunos ejemplos, los métodos de intercambio de SNP de la presente descripción incluyen tanto la adición de SNP beneficiosos como la eliminación de mutaciones neutras y/o perjudiciales.

40 El intercambio de SNP es una herramienta poderosa para identificar y explotar mutaciones tanto beneficiosas como perjudiciales en un linaje de cepas sometidas a mutagénesis y selección para un rasgo de interés mejorado. El intercambio de SNP utiliza técnicas de ingeniería genómica de alta capacidad para determinar sistemáticamente la influencia de mutaciones individuales en un linaje mutagénico. Las secuencias del genoma se determinan para cepas a lo largo de una o más generaciones de un linaje mutagénico con mejoras de rendimiento conocidas. La ingeniería genómica de alta capacidad se usa luego sistemáticamente para recapitular mutaciones de cepas mejoradas en cepas de linaje anteriores y/o revertir mutaciones en cepas posteriores a secuencias de cepas anteriores. Después, se evalúa el rendimiento de estas cepas y se puede determinar la contribución de cada mutación individual en el fenotipo mejorado de interés. Como se ha mencionado antes, las cepas microbianas que resultan de este procedimiento se analizan/caracterizan y forman la base para las bibliotecas de diseño genético de intercambio de SNP que pueden informar de la mejora de las cepas microbianas entre las cepas hospedantes.

50 La eliminación de mutaciones perjudiciales puede proporcionar mejoras inmediatas en el rendimiento, y la consolidación de mutaciones beneficiosas en el contexto genético de una cepa no sujeta a carga mutagénica puede mejorar rápida y enormemente el rendimiento de la cepa. Las diversas cepas microbianas producidas por el procedimiento de intercambio de SNP forman las bibliotecas de intercambio de SNP de diseño genético HTP, que son cepas microbianas que comprenden los diversos SNP añadidos/eliminados/o consolidados, pero por lo demás contextos genéticos idénticos.

55 Como se ha descrito previamente, la mutagénesis aleatoria y el cribado posterior para mejorar el rendimiento es una técnica comúnmente usada para mejorar cepas industriales, y muchas cepas que se usan actualmente para la producción a gran escala se han desarrollado usando este procedimiento de manera iterativa a lo largo de un periodo de muchos años, a veces décadas. Los enfoques aleatorios para generar mutaciones genómicas, tales como la

5 exposición a la radiación UV o mutágenos químicos tales como el metanosulfonato de etilo, eran un método preferido para mejorar cepas industriales porque: 1) los organismos industriales pueden estar mal caracterizados genética o metabólicamente, haciendo difícil sino imposible la selección de dianas para los enfoques de mejora dirigida; 2) incluso en sistemas relativamente bien caracterizados, los cambios que dan como resultado mejoras en el rendimiento industrial son difíciles de predecir y pueden requerir la perturbación de genes que no tienen una función conocida, y 3) las herramientas genéticas para producir mutaciones genómicas dirigidas en un organismo industrial dado pueden no estar disponible o ser muy lentas y/o difíciles de usar.

10 Sin embargo, a pesar de los beneficios mencionados anteriormente de este procedimiento, también existen una serie de desventajas conocidas. Las mutaciones beneficiosas son sucesos relativamente raros y, con el fin de encontrar estas mutaciones con una capacidad de cribado fija, las tasas de mutaciones deben ser lo suficientemente altas. Esto a menudo da como resultado la incorporación de mutaciones neutras y parcialmente perjudiciales no deseadas en las cepas junto con cambios beneficiosos. Con el tiempo, esta "carga mutagénica" se acumula, dando como resultado cepas con deficiencias en la robustez general y rasgos clave tales como las tasas de crecimiento. Finalmente, la "carga mutagénica" hace que las mejoras adicionales en el rendimiento a través de la mutagénesis aleatoria sean cada vez más difíciles o imposibles de obtener. Sin herramientas adecuadas, es imposible consolidar mutaciones beneficiosas que se encuentran en ramas discretas y paralelas de linajes de cepas.

15 El intercambio de SNP es un enfoque para superar estas limitaciones recapitulando o revertiendo sistemáticamente algunas o todas las mutaciones observadas cuando se comparan cepas dentro de un linaje mutagénico. De esta manera, se puede tanto identificar y consolidar mutaciones beneficiosas ("causativas") como se puede identificar y eliminar mutaciones perjudiciales. Esto permite mejoras rápidas en el rendimiento de la cepa que no podrían lograrse por mutagénesis aleatoria adicional o ingeniería genética dirigida.

20 La eliminación de la carga genética o la consolidación de cambios beneficiosos en una cepa sin carga genética también proporciona un punto de partida nuevo y sólido para la mutagénesis aleatoria adicional que puede permitir mejoras adicionales.

25 Además, a medida que se identifican cambios beneficiosos ortogonales en ramas discretas diversas de un linaje de cepas mutagénicas, se pueden consolidar rápidamente en cepas de mejor rendimiento. Estas mutaciones también se pueden consolidar en cepas que no forman parte de linajes mutagénicos, tales como cepas con mejoras obtenidas por ingeniería genética dirigida.

30 Existen otros enfoques y tecnologías para recombinar mutaciones aleatoriamente entre cepas dentro de un linaje mutagénico. Estos incluyen técnicas como fusión de protoplastos y el barajado del genoma completo que facilitan la recombinación genómica entre cepas mutadas. Para algunos microorganismos industriales, tales como levaduras y hongos filamentosos, también se pueden aprovechar los ciclos de apareamiento naturales para la recombinación genómica por pares. De esta manera, las mutaciones perjudiciales se pueden eliminar por "retrocruzamiento" de mutantes con cepas parentales y consolidar mutaciones beneficiosas. Sin embargo, estos enfoques están sujetos a muchas limitaciones que se eluden usando los métodos de intercambio de SNP de la presente descripción.

35 Por ejemplo, puesto que estos enfoques se basan en un número relativamente pequeño de sucesos cruzados de recombinación aleatoria para intercambiar mutaciones, pueden ser necesarios muchos ciclos de recombinación y cribado para optimizar el rendimiento de la cepa. Además, aunque los sucesos de recombinación natural son esencialmente aleatorios, también están sujetos al sesgo posicional del genoma y puede ser difícil dirigirse a algunas mutaciones. Estos enfoques también proporcionan poca información sobre la influencia de mutaciones individuales sin secuenciación y análisis adicionales del genoma. El intercambio de SNP supera estas limitaciones fundamentales, ya que no es un enfoque aleatorio, sino más bien la introducción o eliminación sistemática de mutaciones individuales entre cepas.

40 En algunos ejemplos, la presente descripción enseña métodos para identificar la diversidad de secuencias de SNP presente entre los organismos de un grupo de diversidad. Un grupo de diversidad puede ser un número n dado de microbios utilizados para el análisis, y representando los genomas de dichos microbios el "grupo de diversidad".

45 En ejemplos particulares, un grupo de diversidad puede ser una cepa parental original (S_1) con una secuencia genética "base" o de "referencia" en un punto de tiempo particular ($S_1\text{Gen}_1$) y después cualquier número de cepas descendientes subsiguientes (S_{2-n}) que derivaron/se desarrollaron a partir de dicha cepa S_1 y que tienen un genoma diferente ($S_{2-n}\text{Gen}_{2-n}$), en relación con el genoma base de S_1 .

50 Por ejemplo, en algunos ejemplos, la presente descripción enseña la secuenciación de los genomas microbianos en un grupo de diversidad para identificar los SNP presentes en cada cepa. En un ejemplo, las cepas del grupo de diversidad son cepas históricas de producción microbiana. Por lo tanto, un grupo de diversidad de la presente descripción puede incluir, por ejemplo, una cepa industrial de referencia y una o más cepas industriales mutadas producidas por programas tradicionales de mejora de cepas.

55 En algunos ejemplos, los SNP dentro de un grupo de diversidad se determinan con referencia a una "cepa de referencia". En algunos ejemplos, la cepa de referencia es una cepa de tipo natural. En otros ejemplos, la cepa de referencia es una cepa industrial original antes de ser sometida a cualquier mutagénesis. La cepa de referencia puede

ser definida por el médico y no tiene que ser una cepa de tipo natural original o una cepa industrial original. La cepa base es simplemente representativa de lo que se considerará la "base", "referencia" o el contexto genético original, por el cual se van a comparar cepas posteriores que se derivaron o se desarrollaron a partir de dicha cepa de referencia.

- 5 Una vez que se identifican todos los SNPS en el grupo de diversidad, la presente descripción enseña métodos de intercambio de SNP y métodos de cribado para delinear (es decir, cuantificar y caracterizar) los efectos (p. ej., creación de un fenotipo de interés) de los SNP individualmente y/o en grupos.

En algunos ejemplos, los métodos de intercambio de SNP de la presente descripción comprenden la etapa de introducir uno o más SNP identificados en una cepa mutada (p. ej., una cepa de entre $S_{2-n}Gen_{2-n}$) en una cepa de referencia (S_1Gen_1) o cepa de tipo natural ("onda ascendente").

10 En otros ejemplos, los métodos de intercambio de SNP de la presente descripción comprenden la etapa de eliminar uno o más SNP identificados en una cepa mutada (p. ej., una cepa de entre $S_{2-n}Gen_{2-n}$) ("onda descendente").

En algunos ejemplos, cada cepa generada que comprende uno o más cambios de SNP (que introduce o elimina) se cultiva y se analiza bajo uno o más criterios de la presente descripción (p. ej., producción de un producto químico o producto de interés). Los datos de cada una de las cepas hospedantes analizadas se asocian o correlacionan con el SNP particular o grupo de SNP presentes en la cepa hospedante y se registran para uso futuro. Por lo tanto, la presente descripción permite la creación de bibliotecas de cepas microbianas de diseño genético HTP grandes y muy registradas que son capaces de identificar el efecto de un SNP en cualquier serie de rasgos genéticos o fenotípicos microbianos de interés. La información almacenada en estas bibliotecas de diseño genético HTP informa a los algoritmos de aprendizaje automático de la plataforma de ingeniería genómica HTP y dirige las iteraciones futuras del procedimiento, que finalmente conduce a organismos microbianos evolucionados que poseen propiedades/rasgos altamente deseables.

3. Intercambios de codones de inicio/parada: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de codones de inicio/parada

25 En algunos ejemplos, la presente descripción enseña métodos para intercambiar variantes de codones de inicio y parada. Por ejemplo, los codones de parada típicos para *S. cerevisiae* y mamíferos son TAA (UAA) y TGA (UGA), respectivamente. El codón de parada típico para plantas monocotiledóneas es TGA (UGA), mientras que los insectos y *E. coli* habitualmente usan TAA (UAA) como codón de parada (Dalphin et al. (1996) *Nucl. Acids Res.* 24: 216-218). En otros ejemplos, la presente descripción enseña el uso de los codones de parada TAG (UAG).

30 La presente descripción enseña de manera similar el intercambio de codones de inicio. En algunos ejemplos, la presente descripción enseña el uso del codón de inicio ATG (AUG) utilizado por la mayoría de los organismos (especialmente eucariotas). En algunos ejemplos, la presente descripción enseña que los procariontes lo que más usan es ATG (AUG), seguido de GTG (GUG) y TTG (UUG).

En otros ejemplos, la presente descripción enseña la sustitución de los codones de inicio ATG por TTG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de inicio ATG por GTG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de inicio GTG por ATG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de inicio GTG por TTG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de inicio TTG por ATG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de inicio TTG por GTG.

40 En otros ejemplos, la presente descripción enseña la sustitución de los codones de parada de TAA por TAG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de parada de TAA por TGA. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de parada de TGA por TAA. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de parada de TGA por TAG. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de parada de TAG por TAA. En algunos ejemplos, la presente descripción enseña la sustitución de los codones de parada de TAG por TGA.

4. Intercambio de Stop: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de intercambio de STOP

En algunos ejemplos, la presente descripción enseña métodos para mejorar la productividad de las células hospedantes mediante la optimización de la transcripción de genes celulares. La transcripción de genes es el resultado de varios fenómenos biológicos distintos, que incluyen el inicio de la transcripción (reclutamiento de RNAP y formación de complejos transcripcionales), elongación (síntesis/extensión de la hebra) y la terminación de la transcripción (separación y terminación de RNAP). Aunque se ha prestado mucha atención al control de la expresión génica a través de la modulación transcripcional de genes (p. ej., cambiando promotores o induciendo factores de transcripción reguladores), se han hecho comparativamente pocos esfuerzos dirigidos a la modulación de la transcripción mediante la modulación de secuencias terminadoras de genes.

La forma más obvia en que la transcripción impacta en los niveles de expresión génica es a través de la velocidad de inicio de la Pol II, que puede ser modulada por combinaciones de promotor o fuerza del potenciador y factores transactivadores (Kadonaga, JT. 2004 "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors" *Cell*. 23 de enero de 2004; 116 (2): 247-57). En eucariotas, la tasa de elongación también puede determinar los patrones de expresión génica al influir en el empalme alternativo (Cramer P. et al., 1997 "Functional association between promoter structure and transcript alternative splicing". *Proc Natl Acad Sci U S A*. 14 de octubre de 1997; 94 (21): 11456-60). La terminación fallida en un gen puede alterar la expresión de genes secuencia abajo al reducir la accesibilidad del promotor a Pol II (Greger IH. Et al., 2000 "Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*". *Proc Natl Acad Sci USA*. 2000 18 de julio; 97 (15): 8415-20). Este proceso, conocido como interferencia transcripcional, es particularmente relevante en eucariotas inferiores, ya que a menudo tienen genes poco espaciados.

Las secuencias de terminación también pueden afectar a la expresión de los genes a los que pertenecen las secuencias. Por ejemplo, los estudios muestran que la terminación transcripcional ineficaz en eucariotas da como resultado una acumulación de pre-ARNm sin empalmar (véase West, S. y Proudfoot, N.J., 2009 "Transcriptional Termination Enhances Protein Expression in Human Cells" *Mol Cell*. 13 de febrero de 2009; 33 (3-9): 354-364). Otros estudios también han mostrado que el procesamiento del extremo 3' se puede retrasar por la terminación ineficaz (West, S et al., 2008 "Molecular dissection of mammalian RNA polymerase II transcriptional termination" *Mol Cell*. 14 de marzo de 2008; 29 (5): 600-10.). La terminación de la transcripción también puede afectar la estabilidad del ARNm al liberar transcripciones de los sitios de síntesis. Además, las secuencias de terminación fuertes pueden aumentar la estabilidad del ARNm, aumentando así la abundancia de proteínas y la actividad general de la ruta.

Terminación del mecanismo de transcripción en eucariotas

La terminación transcripcional en eucariotas opera a través de señales de terminadores que son reconocidas por factores proteicos asociados con la ARN polimerasa II. En algunos ejemplos, el factor de especificidad de escisión y poliadenilación (CPSF) y el factor de estimulación de escisión (CstF) se transfieren desde el dominio carboxilo terminal de la ARN polimerasa II a la señal poli-A. En algunos ejemplos, los factores CPSF y CstF también reclutan otras proteínas en el sitio de terminación, que luego escinden el transcrito y liberan el ARNm del complejo de transcripción. La terminación también desencadena la poliadenilación de los transcritos de ARNm. Los ejemplos ilustrativos de factores de terminación eucariotas validados y sus estructuras conservadas se discuten en partes posteriores de este documento.

Terminación de la transcripción en procariontes.

En los procariontes, dos mecanismos principales, denominados terminación independiente de Rho y dependiente de Rho, median la terminación de la transcripción. Las señales de terminación independientes de Rho no requieren un factor de terminación de la transcripción extrínseco, ya que la formación de una estructura de tallo-bucle en el ARN transcrito a partir de estas secuencias junto con una serie de restos de uridina (U) promueve la liberación de la cadena de ARN del complejo de transcripción. La terminación dependiente de Rho, por otro lado, requiere un factor de terminación de la transcripción llamado Rho y elementos que actúan en cis en el ARNm. El sitio de unión inicial para Rho, el sitio de utilización de Rho (rut), es una región monocatenaria extendida (~70 nucleótidos, a veces 80-100 nucleótidos) caracterizada por un alto contenido de citidina/bajo contenido de guanosina y relativamente poca estructura secundaria en el ARN que está siendo sintetizado, secuencia arriba de la secuencia de terminador real. Cuando se encuentra un sitio de pausa de la polimerasa, se produce la terminación y el transcrito es liberado por la actividad helicasa de Rho.

Intercambio de terminadores (intercambio de STOP)

En algunos ejemplos, la presente descripción enseña métodos para seleccionar secuencias de terminación ("terminadores") con propiedades de expresión óptimas para producir efectos beneficiosos en la productividad general de la cepa hospedante.

Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos de identificación de uno o más terminadores y/o de generación de variantes de uno o más terminadores dentro de una célula hospedante, que presentan un intervalo de fuerzas de expresión (p. ej., escaleras de terminadores descritas más adelante). Una combinación particular de estos terminadores identificados y/o generados se puede agrupar entre sí como una escalera de terminadores, que se explica con más detalle a continuación.

La escalera de terminadores en cuestión después se asocia con un gen de interés dado. Por lo tanto, si se tienen los terminadores T₁-T₈ (que representan ocho terminadores que han sido identificados y/o generados para presentar un intervalo de fuerzas de expresión cuando se combinan con uno o más promotores) y se asocia la escalera de terminadores con un solo gen de interés en una célula hospedante (es decir, transformar genéticamente una célula hospedante con un terminador dado operativamente ligado al extremo 3' de un gen diana dado), entonces el efecto de cada combinación de los terminadores se puede determinar caracterizando cada una de las cepas transformadas que resultan de cada esfuerzo combinatorio, dado que las células hospedantes transformadas tienen un contexto

genético por lo demás idéntico, excepto el o los terminadores particulares asociados con el gen diana. Las células hospedantes resultantes que se transforman por este procedimiento forman bibliotecas de diseño genético HTP.

La biblioteca de diseño genético de HTP se puede referir a la colección de cepas microbianas físicas reales que se forma por este procedimiento, siendo cada cepa miembro representativa de un terminador dado operativamente unido a un gen diana particular, en un contexto genético por lo demás idéntico, denominándose dicha biblioteca una "biblioteca de cepas microbianas de intercambio de terminadores" o "biblioteca de cepas microbianas de intercambio de STOP". En el contexto específico de *E. coli*, la biblioteca puede denominarse "biblioteca de cepas de *E. coli* de intercambio de terminadores" o "biblioteca de cepas de *E. coli* de intercambio de STOP", pero las expresiones se pueden usar como sinónimos, ya que *E. coli* es un ejemplo específico de microbio.

Además, la biblioteca de diseño genético de HTP se puede referir a la colección de perturbaciones genéticas, en este caso un terminador x dado operativamente unido a un gen y dado, denominándose dicha colección "biblioteca de intercambio de terminadores" o "biblioteca de intercambio de STOP".

Además, se puede utilizar la misma escalera de terminadores que comprende los terminadores T_1 - T_8 para transformar microbios, en donde cada uno de los ocho terminadores se une operativamente a 10 genes diana diferentes. El resultado de este procedimiento serían 80 cepas de células hospedantes que por lo demás se suponen genéticamente idénticas, excepto por los terminadores particulares ligados operativamente a un gen diana de interés. Estas 80 cepas de células hospedantes se podrían cribar y caracterizar adecuadamente y dar lugar a otra biblioteca de diseño genético de HTP. La caracterización de las cepas microbianas en la biblioteca de diseño genético HTP produce información y datos que se pueden almacenar en cualquier base de datos, incluyendo, sin limitación, una base de datos relacional, una base de datos orientada a objetos o una base de datos NoSQL altamente distribuida. Estos datos/información podrían incluir, por ejemplo, un efecto de "terminadores" dado (p. ej., T_1 - T_8) cuando se unen operativamente a un gen diana dado. Estos datos/información también pueden ser el conjunto más amplio de efectos combinatorios que resultan de unir operativamente dos o más promotores (p. ej., T_1 - T_8) a un gen diana dado.

Los ejemplos mencionados antes de ocho terminadores y 10 genes diana son meramente ilustrativos, ya que el concepto se puede aplicar con cualquier número dado de terminadores que se hayan agrupado entre sí basándose en la presentación de un intervalo de fuerzas de expresión y cualquier número dado de genes diana. Por ejemplo, otro conjunto de terminadores que se puede usar en los métodos proporcionados en el presente documento (p. ej., intercambio de STOP) es el conjunto de terminadores que se encuentran en la Tabla 1.2 con las SEQ ID NO 225, 226, 227, 228, 229 o 230 de ácido nucleico.

En resumen, la utilización de varios terminadores para modular la expresión de varios genes en un organismo es una herramienta potente para optimizar un rasgo de interés. La herramienta molecular de intercambio de terminadores, desarrollada por los autores de la invención, usa una escalera de secuencias de terminadores que se ha demostrado que varían la expresión de al menos un locus en al menos una condición. Después, esta escalera se aplica sistemáticamente a un grupo de genes en el organismo mediante la ingeniería genómica de alta capacidad. Se determina que este grupo de genes tiene una alta probabilidad de tener impacto en el rasgo de interés basándose en uno cualquiera de varios métodos. Estos podrían incluir la selección basada en la función conocida o el impacto en el rasgo de interés, o la selección algorítmica basada en la diversidad genética beneficiosa previamente determinada.

Después, se evalúa el rendimiento de la biblioteca microbiana de diseño genético de HTP resultante de organismos que contienen una secuencia de terminador unida a un gen en un modelo de cribado de alta capacidad, y se determinan las uniones terminador-gen que conducen a un mayor rendimiento y la información se almacena en una base de datos. La colección de perturbaciones genéticas (es decir, un terminador x dado unido a un gen y dado) forma una "biblioteca de intercambio de terminadores", que se puede utilizar como una fuente de alteraciones genéticas potenciales que se utilizarán en el procesamiento de ingeniería microbiana. Con el tiempo, a medida que se implementa un mayor conjunto de perturbaciones genéticas contra una mayor diversidad de contextos genéticos microbianos, cada biblioteca se vuelve más potente como un corpus de datos experimentalmente confirmados que se pueden utilizar para diseñar de manera más precisa y predecible cambios dirigidos contra cualquier contexto genético de interés. Es decir, en algunos ejemplos, las presentes descripciones enseñan la introducción de uno o más cambios genéticos en una célula hospedante basándose en resultados experimentales previos insertados dentro de los metadatos asociados con cualquiera de las bibliotecas de diseño genético de la descripción.

Por lo tanto, en ejemplos particulares, el intercambio de terminadores es un procedimiento de múltiples etapas que comprende:

1. Seleccionar un conjunto de terminadores " x " para que actúen como una "escalera". Idealmente, se ha mostrado que estos terminadores conducen a una expresión muy variable en múltiples locus genómicos, pero el único requisito es que perturben la expresión génica de alguna manera.

2. Seleccionar un conjunto de genes " n " para dirigir. Este conjunto puede ser todos los ORF de un genoma o un subconjunto de ORF. El subconjunto se puede elegir usando anotaciones en ORF relacionados con la función, por relación con perturbaciones beneficiosas demostradas previamente (intercambios de promotores previos, intercambios de STOP, intercambios de marcadores de solubilidad, intercambios de marcadores de degradación

o intercambios de SNP), por selección algorítmica basada en interacciones epistáticas entre perturbaciones previamente generadas, otros criterios de selección basados en hipótesis relacionadas con el ORF beneficioso para direccionamiento, o mediante selección aleatoria. En otros ejemplos, los genes dirigidos "n" pueden comprender genes que no codifican proteínas, que incluyen los ARN no codificantes.

5 3. Ingeniería de cepas de alta capacidad para llevar a cabo de forma rápida y en paralelo las siguientes modificaciones genéticas: Cuando existe un terminador nativo en el extremo 3' del gen diana n y se conoce su secuencia, se sustituye el terminador nativo por cada uno de los terminadores x de la escalera. Cuando el terminador nativo no existe, o se desconoce su secuencia, se inserta cada uno de los terminadores x en la escalera después del codón de parada del gen. De esta manera, se construye una "biblioteca" (también denominada biblioteca de diseño genético HTP) de cepas, en donde cada miembro de la biblioteca es un caso de terminador x unido a diana n, en un contexto genético por lo demás idéntico. Como se ha descrito previamente, se pueden insertar combinaciones de terminadores, ampliando la variedad de posibilidades combinatorias sobre las que se construye la biblioteca.

10 4. Cribado de alta capacidad de la biblioteca de cepas en un contexto donde su rendimiento frente a uno o más indicadores es indicativo del rendimiento que se está optimizando.

Este procedimiento fundamental se puede extender para proporcionar mejoras adicionales en el rendimiento de la cepa, entre otras cosas: (1) Consolidando múltiples perturbaciones beneficiosas en un solo contexto de cepa, ya sea una a la vez en un proceso interactivo, o como múltiples cambios en una sola etapa. Múltiples perturbaciones pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatorizada de cambios. Por ejemplo, si el conjunto de dianas son todos los genes de una ruta, la regeneración secuencial de la biblioteca de perturbaciones en un miembro o miembros mejorados de la biblioteca de cepas anterior puede optimizar el nivel de expresión de cada gen en una ruta, independientemente de qué genes son limitantes de la velocidad en cualquier iteración dada; (2) Introducir los datos de rendimiento resultantes de la generación individual y combinatoria de la biblioteca en un algoritmo que usa esos datos para predecir un conjunto óptimo de perturbaciones basado en la interacción de cada perturbación; y (3) Implementar una combinación de los dos enfoques anteriores.

El enfoque se ilustra en la presente descripción con microorganismos industriales, pero es aplicable a cualquier organismo donde se puedan identificar los rasgos deseados en una población de mutantes genéticos. Por ejemplo, esto se podría usar para mejorar el rendimiento de células CHO, levaduras, células de insectos, algas, así como organismos multicelulares, tales como plantas.

30 **5. Optimización de secuencias: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de secuencias optimizadas**

En un ejemplo, los métodos de la descripción comprenden codones que optimizan uno o más genes expresados por el organismo hospedante. Los métodos de optimización de codones para mejorar la expresión en varios hospedantes se conocen en la técnica y se describen en la bibliografía (véase la Pub. de solicitud de patente de EE.UU. N° 2007/0292918). Las secuencias de codificación optimizadas que contienen codones preferidos por un hospedante procarionta o eucariota particular (véase también, Murray et al. (1989) *Nucl. Acids Res.* 17: 477-508) se pueden preparar, por ejemplo, para aumentar la velocidad de traducción o para producir transcritos de ARN recombinante que tengan propiedades deseables, tales como una semivida más larga, en comparación con los transcritos producidos a partir de una secuencia no optimizada.

40 La expresión de proteínas se rige por una serie de factores, que incluyen los que afectan a la transcripción, procesamiento del ARNm y la estabilidad y el inicio de la traducción. Por lo tanto, la optimización puede abordar cualquiera de una serie de características de secuencia de cualquier gen en particular. Como ejemplo específico, una pausa de traducción inducida por codones raros puede dar como resultado una expresión de proteína reducida. Una pausa de traducción inducida por codones raros incluye la presencia de codones en el polinucleótido de interés que rara vez se usan en el organismo hospedante y puede tener un efecto negativo en la traducción de proteínas debido a su escasez en el grupo de ARNt disponible.

El inicio de traducción alternativo también puede dar como resultado la expresión reducida de proteína heteróloga. El inicio de traducción alternativo puede incluir una secuencia de polinucleótidos sintética que contiene accidentalmente motivos capaces de funcionar como un sitio de unión al ribosoma (RBS). Estos sitios pueden dar como resultado el inicio de la traducción de una proteína truncada desde un sitio interno del gen. Un método para reducir la posibilidad de producir una proteína truncada, que puede ser difícil de eliminar durante la purificación, incluye eliminar las supuestas secuencias de RBS internas de una secuencia de polinucleótido optimizada.

El deslizamiento de la polimerasa inducido por repetición puede dar como resultado la expresión de proteína heteróloga reducida. El deslizamiento de la polimerasa inducido por repetición implica repeticiones de secuencias de nucleótidos que se ha mostrado que provocan deslizamiento o tartamudeo de la ADN polimerasa, lo que puede dar como resultado mutaciones de cambio de marco. Dichas repeticiones también pueden producir el deslizamiento de la ARN polimerasa. En un organismo con un alto sesgo de contenido de G+C, puede haber un mayor grado de repeticiones compuestas por repeticiones de nucleótidos G o C. Por lo tanto, un método de reducción de la posibilidad

de inducir el deslizamiento de la ARN polimerasa incluye la alteración de repeticiones extendidas de nucleótidos G o C.

La interferencia de estructuras secundarias también puede dar como resultado una expresión de proteína heteróloga reducida. Las estructuras secundarias pueden secuestrar la secuencia de RBS o el codón de inicio y se han correlacionado con una reducción en la expresión de proteínas. Las estructuras de tallo-bucle también pueden estar implicadas en la pausa y atenuación de la transcripción. Una secuencia de polinucleótidos optimizada puede contener estructuras secundarias mínimas en el RBS y las regiones codificantes de genes de la secuencia de nucleótidos para permitir una mejor transcripción y traducción.

Por ejemplo, el procedimiento de optimización puede comenzar identificando la secuencia de aminoácidos deseada para ser expresada por el hospedante. A partir de la secuencia de aminoácidos se puede diseñar un polinucleótido o secuencia de ADN candidata. Durante el diseño de la secuencia de ADN sintético, la frecuencia de uso de codones se puede comparar con el uso de codones del organismo de expresión hospedante y los codones del hospedante raros se pueden eliminar de la secuencia sintética. Además, la secuencia de ADN candidata sintética se puede modificar con el fin de eliminar los sitios de restricción enzimáticos indeseables y añadir o eliminar cualesquiera secuencias señal, conectores o regiones no traducidas deseadas. Se puede analizar en la secuencia de ADN sintético la presencia de una estructura secundaria que pueda interferir con el proceso de traducción, tal como repeticiones G/C y estructuras de tallo-bucle.

6. Intercambio de MARCADORES DE SOLUBILIDAD: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de intercambio de MARCADORES DE SOLUBILIDAD

En algunos ejemplos, la presente descripción enseña métodos para mejorar la productividad de la célula hospedante mediante la optimización de los mecanismos postraduccionales. La mejora tradicional de cepas a menudo se puede lograr mediante la sobreexpresión de genes de la ruta que producen alguna molécula de interés. Típicamente, se pueden duplicar genes de la ruta conocidos o se pueden insertar promotores fuertes para dirigir la expresión de estos genes y, por lo tanto, aumentar los niveles de transcrito de ARNm con el objetivo de aumentar la abundancia de proteínas de la ruta para lograr una velocidad, título o rendimiento mejorados de una ruta determinada. Este enfoque se puede aplicar sistemáticamente a escala del genoma completo para identificar todos los genes que pueden mejorar el rendimiento de la cepa. Otro enfoque aplicado con frecuencia puede ser la delección de genes de la ruta potencialmente competidores con el objetivo de eliminar por completo los productos proteicos que pueden desviar el carbono de la ruta deseada. Sin embargo, estos enfoques de mejora de cepas de sobreexpresión y/o delección conocidos en la técnica pueden tener varias limitaciones.

Comenzando con la duplicación de la ruta o la inserción de un promotor fuerte, el efecto esperado de niveles mayores de transcrito de ARNm puede no dar como resultado necesariamente mayor abundancia de proteínas. Varios productos proteicos pueden tener varias etapas limitantes de la velocidad en su producción, y esta etapa limitante de la velocidad pueden no ser los niveles de transcritos de ARNm. En escenarios donde la transcripción de ARNm no es la etapa limitante de la velocidad, es posible que los mecanismos postraduccionales puedan estar teniendo impacto en la abundancia de proteínas en general. Por ejemplo, la presencia de marcadores de solubilidad de proteínas se puede usar para aumentar la abundancia de proteína activa correctamente plegada que puede contribuir a la producción de una molécula diana, mientras que el simple aumento de los niveles de transcritos de ARNm puede conducir solo a un aumento de proteína inactiva mal plegada. Los efectos de los marcadores de solubilidad de proteínas también se hacen que sean ajustables dependiendo de la secuencia del marcador de solubilidad que se use, lo que permite una optimización precisa del fenotipo diana.

Intercambio de marcadores de solubilidad de proteínas (intercambio de MARCADORES DE SOLUBILIDAD)

En algunos ejemplos, la presente descripción enseña métodos de selección de secuencias de marcadores de solubilidad de proteínas ("marcadores de solubilidad") con propiedades óptimas de solubilidad de proteínas para producir efectos beneficiosos en la productividad general de la cepa hospedante.

Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos de identificación de uno o más marcadores de solubilidad de proteínas y/o generación de variantes de uno o más marcadores de solubilidad de proteínas dentro de una célula hospedante, que presentan un intervalo de fuerzas de solubilidad (p. ej., marcadores de solubilidad de proteínas descritas más adelante). Una combinación particular de estos marcadores de solubilidad de proteínas identificados y/o generados se pueden agrupar entre sí como una escalera de marcadores de solubilidad de proteínas, que se explica con más detalle a continuación.

La escalera de marcadores de solubilidad de proteínas en cuestión después se asocia con un gen de interés dado. Por lo tanto, si se tienen los marcadores de solubilidad de proteínas, PST₁-PST₄ (véase la Tabla 17) que representan un subconjunto de marcadores de solubilidad de proteínas que se han identificado en Costa et al., *Front Microbiol.* 2014; 5: 63 para mejorar la solubilidad de proteínas y también ser de menos de 100 aminoácidos, y asocia la escalera de marcadores de solubilidad de proteínas con un solo gen de interés en una célula hospedante (es decir, transformar genéticamente una célula hospedante con un marcador de solubilidad de proteínas dado operativamente unido a un determinado gen diana para generar una proteína diana marcada en el extremo N o en el extremo C). El efecto de

5 cada combinación del marcador de solubilidad de proteínas se puede determinar caracterizando cada una de las cepas transformadas que resultan de cada esfuerzo combinatorio, dado que las células hospedantes transformadas tienen un contexto genético por lo demás idéntico excepto el o los marcadores de solubilidad particulares asociados con el gen diana. Las células hospedantes resultantes que se transforman por este procedimiento forman bibliotecas de diseño genético de HTP.

10 La biblioteca de diseño genético de HTP se puede referir a la colección de cepas microbianas físicas reales que se forma a través de este procedimiento, siendo cada cepa miembro representativa de una marcador de solubilidad de proteínas dado operativamente unido a una proteína diana particular, en un contexto genético por lo demás idéntico, denominándose dicha biblioteca una "biblioteca de cepas microbianas de intercambio de marcadores de solubilidad" o "biblioteca de cepas microbianas de intercambio de MARCADORES DE SOLUBILIDAD". En el contexto específico de *E. coli*, la biblioteca se puede denominar una "biblioteca de cepas de *E. coli* de intercambio de MARCADORES DE SOLUBILIDAD" o "biblioteca de cepas de *E. coli* de intercambio de MARCADORES DE SOLUBILIDAD", pero los términos se pueden usar como sinónimos, ya que *E. coli* es un ejemplo específico de un microbio.

15 Además, la biblioteca de diseño genético HTP se puede referir a la colección de perturbaciones genéticas, en este caso una marcador de solubilidad de proteína *x* dado operativamente unido a un gen *y* dado, denominándose dicha colección una "biblioteca de intercambio de marcadores de solubilidad de proteínas" o "biblioteca de intercambio de MARCADOR DE SOLUBILIDAD".

20 Además, se puede utilizar la misma escalera de marcadores de solubilidad de proteínas que comprende el marcador de solubilidad de proteínas PST₁-PST₄ para transformar microbios, en donde cada uno de los cuatro marcadores de solubilidad de proteínas está operativamente unido a 10 genes diana diferentes. El resultado de este procedimiento serían 40 cepas de células hospedantes que por lo demás se suponen genéticamente idénticas, excepto por los marcadores de solubilidad de proteínas particulares unidos operativamente a un gen diana de interés. Estas 40 cepas de células hospedantes se podrían cribar y caracterizar adecuadamente y dar lugar a otra biblioteca de diseño genético HTP. La caracterización de las cepas microbianas en la biblioteca de diseño genético HTP produce información y
25 datos que se pueden almacenar en cualquier base de datos, incluyendo, sin limitación, una base de datos relacional, una base de datos orientada a objetos o una base de datos NoSQL altamente distribuida. Estos datos/información podrían incluir, por ejemplo, un efecto del marcador de solubilidad de proteínas dado (p. ej., PST₁-PST₄) cuando está operativamente unido a un gen diana dado. Estos datos/información también pueden ser el conjunto más amplio de efectos combinatorios que resultan de unir operativamente de dos o más marcadores de solubilidad (p. ej., PST₁-PST₄) a un gen diana dado.
30

Los ejemplos mencionados anteriormente de cuatro marcadores de solubilidad de proteínas y 10 genes diana son meramente ilustrativos, ya que el concepto se puede aplicar con cualquier número dado de marcadores de solubilidad de proteínas que se han agrupado entre sí basado en la presentación de un intervalo de fuerzas de solubilidad y cualquier número dado de genes diana.

35 En resumen, la utilización de diversos marcadores de solubilidad de proteínas para modular la solubilidad de diversas proteínas en un organismo es una herramienta poderosa para optimizar un rasgo de interés. La herramienta molecular de intercambio de marcadores de solubilidad de proteínas, desarrollada por los autores de la invención, usa una escalera de secuencias de marcadores de solubilidad de proteínas que han demostrado que varían la solubilidad (p. ej., mejoran) de al menos una proteína en al menos una condición. Después, esta escalera se aplica sistemáticamente a un grupo de genes en el organismo usando ingeniería genómica de alta capacidad. Se determina que este grupo de genes tiene una alta probabilidad de tener impacto en el rasgo de interés basándose en uno cualquiera de una serie de métodos. Estos podrían incluir la selección basada en la función conocida o el impacto en el rasgo de interés, o la selección algorítmica basada en la diversidad genética beneficiosa previamente determinada.
40

45 Después se evalúa en la biblioteca microbiana de diseño genético de HTP resultante de organismos que contienen una secuencia de marcador de solubilidad de proteínas unido a un gen, el rendimiento en un modelo de cribado de alta capacidad, y se determinan las uniones de marcador de solubilidad de proteínas-gen que conducen a un mayor rendimiento y la información se almacena en una base de datos. La colección de perturbaciones genéticas (es decir, marcador de solubilidad de proteínas *x* dado unido a un gen *y* dado) forma una "biblioteca de intercambio de marcadores de solubilidad de proteínas", que se puede utilizar como una fuente de alteraciones genéticas potenciales que se utilizarán en el procesamiento de ingeniería microbiana. Con el tiempo, a medida que se implementa un mayor conjunto de perturbaciones genéticas contra una mayor diversidad de contextos genéticos microbianos, cada biblioteca se vuelve más poderosa como un corpus de datos confirmados experimentalmente que se pueden usar para diseñar de manera más precisa y predecible cambios específicos frente a cualquier contexto de interés. Es decir, en algunos ejemplos, las presentes descripciones enseñan la introducción de uno o más cambios genéticos en una célula hospedante basándose en resultados experimentales previos insertados dentro de los metadatos asociados con
50 cualquiera de las bibliotecas de diseño genético de la descripción.
55

Por lo tanto, en ejemplos particulares, el intercambio de marcadores de solubilidad de proteínas es un procedimiento de múltiples etapas que comprende:

1. Seleccionar un conjunto de marcadores de solubilidad de proteínas "x" para que actúen como una "escalera". "Idealmente, se ha mostrado que estos marcadores de solubilidad de proteínas conducen a una mayor solubilidad de proteínas a lo largo de múltiples locus genómicos, pero el único requisito es que perturben la solubilidad de alguna manera.

5 2. Seleccionar un conjunto de genes "n" para direccionamiento. Este conjunto puede ser todos los ORF de un genoma o un subconjunto de ORF. El subconjunto se puede elegir usando anotaciones en los ORF relacionados con la función, por relación con perturbaciones beneficiosas previamente demostradas (intercambios de promotores previos, intercambios de STOP, intercambios de MARCADORES DE DEGRADACIÓN o intercambios de SNP), por selección algorítmica basada en interacciones epistáticas entre perturbaciones generadas previamente, otros criterios de selección basados en hipótesis con respecto al ORF beneficioso para dirigir, o mediante selección aleatoria.

10 3. Ingeniería de cepas de alta capacidad para llevar a cabo rápidamente y en paralelo las siguientes modificaciones genéticas: Cuando existe un marcador de solubilidad de proteínas nativo dentro de un gen diana n y se conoce su secuencia, reemplazar el marcador de solubilidad de proteínas nativo con cada uno de los marcadores de solubilidad de proteínas x de la escalera. Cuando el marcador de solubilidad de proteínas nativo no existe, o su secuencia es desconocida, insertar cada uno de los marcadores de solubilidad de proteínas x en la escalera. De esta manera se construye una "biblioteca" (también denominada biblioteca de diseño genético HTP) de cepas, en donde cada miembro de la biblioteca es un caso de marcador de solubilidad de proteínas x unido a la diana n, en un contexto genético por lo demás idéntico. Como se ha descrito previamente, se pueden insertar combinaciones de marcadores de solubilidad de proteínas, ampliando la variedad de posibilidades combinatorias sobre las que se construye la biblioteca.

15 20 4. Cribado de alta capacidad de la biblioteca de cepas en un contexto donde su rendimiento frente a uno o más indicadores es indicativo del rendimiento que se está optimizando.

25 Este procedimiento fundamental se puede extender para proporcionar mejoras adicionales en el rendimiento de las cepas, entre otras cosas: (1) Consolidando múltiples perturbaciones beneficiosas en un solo contexto genético de cepa, ya sea una a la vez en un procedimiento interactivo, o como múltiples cambios en una sola etapa. Las múltiples perturbaciones pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatorizada de cambios. Por ejemplo, si el conjunto de dianas son todos los genes de una ruta, la regeneración secuencial de la biblioteca de perturbaciones en un miembro o miembros mejorados de la biblioteca de cepas previa puede optimizar el nivel de expresión de todos los genes en una ruta, independientemente de qué genes son limitantes de velocidad en cualquier iteración dada; (2) Introduciendo los datos de rendimiento resultantes de la generación individual y combinatoria de la biblioteca en un algoritmo que usa esos datos para predecir un conjunto óptimo de perturbaciones basado en la interacción de cada perturbación; y (3) Implementando una combinación de los dos enfoques anteriores.

30 35 El enfoque se ilustra en la presente descripción con microorganismos industriales, pero es aplicable a cualquier organismo donde se puedan identificar los rasgos deseados en una población de mutantes genéticos. Por ejemplo, esto podría usarse para mejorar el rendimiento de células CHO, levaduras, células de insectos, algas, así como organismos multicelulares, tales como plantas.

40 **7. Intercambio de MARCADORES DE DEGRADACIÓN: una herramienta molecular para la obtención de bibliotecas de cepas microbianas de intercambio de MARCADORES DE DEGRADACIÓN**

45 Además de los ejemplos anteriores con respecto a los métodos para mejorar la productividad de las células hospedantes a través de la optimización de los mecanismos postraduccionales, una estrategia de eliminación de genes también puede tener inconvenientes que pueden ser abordados por los marcadores de degradación de proteínas (así como terminadores y marcadores de solubilidad de proteínas) de la presente descripción. La eliminación en su totalidad de un gen y su correspondiente producto proteico puede, en algunos casos, imponer una modificación drástica a la célula. Se puede lograr una respuesta más precisa y ajustable a través de bibliotecas de marcadores de degradación de proteínas que se dirigen a una proteína para la degradación a velocidades variables. Este enfoque también puede tener el beneficio de permitir la modulación de productos proteicos que pueden ser esenciales para la supervivencia celular y no sería viable si se eliminaran por completo. Puesto que estos marcadores de degradación también funcionan a un nivel postraduccional, pueden ser capaces de abordar escenarios donde los niveles de transcripción de ARNm alterados no dan como resultado niveles de proteínas alterados como se ha descrito anteriormente.

50 **Intercambio de marcadores de degradación de proteínas (intercambio de MARCADORES DE DEGRADACIÓN)**

55 En algunos ejemplos, la presente descripción enseña métodos de selección de secuencias de marcadores de degradación de proteínas ("marcadores de degradación") con propiedades de degradación de proteínas óptimas o modulación del nivel de proteínas para producir efectos beneficiosos en la productividad general de las cepas hospedantes.

Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos de identificación de uno o más marcadores de degradación de proteínas y/o generación de variantes de uno o más marcadores de degradación de proteínas dentro de una célula hospedante, que presentan un intervalo de fuerzas de degradación o modulan los niveles de proteínas diana (p. ej., marcadores de degradación de proteínas descritos más adelante). Una combinación particular de estos marcadores de degradación de proteínas identificados y/o generados se pueden agrupar entre sí como una escalera de marcadores de degradación de proteínas, lo cual se explica con más detalle a continuación.

La escalera de marcadores de degradación de proteínas en cuestión después se asocia con un gen de interés dado. Por lo tanto, si se tienen los marcadores de degradación de proteínas PDT₁-PDT₈ (véase la Tabla 18) que representan un subconjunto de marcadores de degradación de proteínas que se han identificado a partir de diversas fuentes como se detalla en la Tabla 18) y se asocia la escalera de marcadores de degradación de proteínas con un solo gen de interés en una célula hospedante (es decir, transformar genéticamente una célula hospedante con un marcador de degradación de proteínas dado operativamente unido a un gen diana dado), entonces el efecto de cada combinación del marcador de degradación de proteínas se puede determinar caracterizando cada una de las cepas transformadas resultantes de cada esfuerzo combinatorio, dado que las células hospedantes transformadas tienen un contexto genético por lo demás idéntico, excepto el o los marcadores de degradación particulares asociadas con el gen diana. Las células hospedantes resultantes que se transforman mediante este procedimiento forman bibliotecas de diseño genético de HTP.

La biblioteca de diseño genético de HTP se puede referir a la colección de cepas microbianas físicas reales que se forma a través de este procedimiento, siendo cada cepa miembro representativa de una marcador de degradación de proteínas dado operativamente unido a una proteína diana particular, en un contexto genético por lo demás idéntico, denominándose dicha biblioteca una "biblioteca de cepas microbianas de intercambio de marcadores de degradación" o "biblioteca de cepas microbianas de intercambio de MARCADORES DE DEGRADACIÓN". En el contexto específico de *E. coli*, la biblioteca se puede denominar "biblioteca de cepas de *E. coli* de intercambio de MARCADORES DE DEGRADACIÓN" o "biblioteca de cepas de *E. coli* de intercambio de MARCADORES DE DEGRADACIÓN", pero los términos se pueden usar como sinónimos, puesto que *E. coli* es un ejemplo específico de un microbio.

Además, la biblioteca de diseño genético de HTP se puede referir a la colección de perturbaciones genéticas, en este caso un marcador de degradación de proteínas *x* dado operativamente unido a un gen *y* dado, denominándose dicha colección una "biblioteca de intercambio de marcadores de degradación de proteínas" o "biblioteca de intercambio de MARCADORES DE DEGRADACIÓN".

Además, se puede utilizar la misma escalera de marcadores de degradación de proteínas que comprende el marcador de degradación de proteínas PDT₁-PDT₈ para diseñar microbios, en donde cada uno de las ocho marcadores de degradación de proteínas está operativamente unido a 10 genes diana diferentes. El resultado de este procedimiento serían 80 cepas de células hospedantes que por lo demás se suponen genéticamente idénticas, excepto por los marcadores de degradación de proteínas particulares operativamente unidos a un gen diana de interés. Estas 80 cepas de células hospedantes se podrían cribar y caracterizar adecuadamente y dar lugar a otra biblioteca de diseño genético de HTP. La caracterización de las cepas microbianas en la biblioteca de diseño genético de HTP produce información y datos que se pueden almacenar en cualquier base de datos, incluyendo, sin limitación, una base de datos relacional, una base de datos orientada a objetos o una base de datos NoSQL altamente distribuida. Estos datos/información podrían incluir, por ejemplo, un efecto del marcador de degradación de proteínas dado (p. ej., PDT₁-PDT₈) cuando está operativamente unido a un gen diana dado. Estos datos/información también pueden ser el conjunto más amplio de efectos combinatorios que resultan de unir operativamente dos o más marcadores de degradación (p. ej., PDT₁-PDT₈) a un gen diana dado.

Los ejemplos antes mencionados de ocho marcadores de degradación de proteínas y 10 genes diana son meramente ilustrativos, ya que el concepto se puede aplicar con cualquier número dado de marcadores de degradación de proteínas que se han agrupado entre sí basándose en la presentación de un intervalo de fuerzas de degradación y cualquier número dado de genes diana.

En resumen, la utilización de diversos marcadores de degradación de proteínas para modular la degradación de diversas proteínas en un organismo es una herramienta poderosa para optimizar un rasgo de interés. La herramienta molecular de intercambio de marcadores de degradación de proteínas, desarrollada por los inventores, utiliza una escalera de secuencias de marcadores de degradación de proteínas que han demostrado que varían la degradación (p. ej., mejoran) de al menos una proteína en al menos una condición. Después, esta escalera se aplica sistemáticamente a un grupo de genes en el organismo usando la ingeniería genómica de alta capacidad. Se determina que este grupo de genes tiene una alta probabilidad de tener impacto en el rasgo de interés basándose en uno cualquiera de una serie de métodos. Estos podrían incluir la selección basada en la función conocida o el impacto en el rasgo de interés, o la selección algorítmica basada en la diversidad genética beneficiosa previamente determinada.

Después, se evalúa el rendimiento de la biblioteca microbiana de diseño genético HTP resultante de organismos que contienen una secuencia de marcador de degradación de proteínas unida a un gen en un modelo de cribado de alta capacidad, y se determinan las uniones de marcador de degradación de proteínas-gen que conducen a un mayor rendimiento y la información se almacena en una base de datos. La colección de perturbaciones genéticas (es decir, marcador de degradación de proteínas *x* dado unido a un gen *y* dado) forma una "biblioteca de intercambio de

marcadores de degradación de proteínas", que se puede utilizar como una fuente de potenciales alteraciones genéticas que se utilizarán en el procesamiento de transformación microbiana. Con el tiempo, a medida que se implementa un mayor conjunto de perturbaciones genéticas contra una mayor diversidad de contextos genéticos microbianos, cada biblioteca se vuelve más poderosa como un corpus de datos confirmados experimentalmente que se puede utilizar para diseñar de manera más precisa y predecible cambios concretos frente a cualquier contexto genético de interés. Es decir, en algunos ejemplos, las presentes descripciones enseñan la introducción de uno o más cambios genéticos en una célula hospedante basándose en resultados experimentales previos insertados dentro de los metadatos asociados con cualquiera de las bibliotecas de diseño genético de la descripción.

Por lo tanto, en ejemplos particulares, el intercambio de marcadores de degradación de proteínas es un procedimiento de múltiples etapas que comprende:

1. Seleccionar un conjunto de marcadores de degradación de proteínas "x" para que actúen como una "escalera". Idealmente, se ha mostrado que estos marcadores de degradación de proteínas conducen a una mayor degradación de proteínas en múltiples locus genómicos, pero el único requisito es que perturben la degradación de alguna manera.

2. Seleccionar un conjunto de genes "n" para direccionamiento. Este conjunto puede ser todos los ORF de un genoma o un subconjunto de ORF. El subconjunto se puede elegir utilizando anotaciones en ORF relacionados con la función, por relación con perturbaciones beneficiosas previamente demostradas (intercambios de promotores previos, intercambios de STOP, intercambios de MARCADORES DE SOLUBILIDAD o intercambios de SNP), por selección algorítmica basada en interacciones epistáticas entre perturbaciones generadas previamente, otros criterios de selección basados en hipótesis con respecto al ORF beneficioso como diana, o mediante selección aleatoria.

3. Ingeniería de cepas de alta capacidad para llevar a cabo rápidamente y en paralelo las siguientes modificaciones genéticas: Cuando existe un marcador de degradación de proteínas nativo dentro de un gen diana n y se conoce su secuencia, reemplazar el marcador de degradación de proteínas nativo con cada uno de los marcadores de degradación de proteínas x de la escalera. Cuando el marcador de degradación de proteínas nativo no existe, o se desconoce su secuencia, insertar cada uno de los marcadores de degradación de proteínas x de la escalera. De esta manera se construye una "biblioteca" (también denominada biblioteca de diseño genético de HTP) de cepas, en donde cada miembro de la biblioteca es un caso de marcador de degradación de proteínas x unido a la diana n, en un contexto genético por lo demás idéntico. Como se ha descrito previamente, se pueden insertar combinaciones de marcadores de degradación de proteínas, ampliando la variedad de posibilidades combinatorias sobre las que se construye la biblioteca.

4. Cribado de alta capacidad de la biblioteca de cepas en un contexto donde su rendimiento frente a uno o más indicadores es indicativo del rendimiento que se está optimizando.

Este procedimiento fundamental se puede extender para proporcionar mejoras adicionales en el rendimiento de las cepas, entre otras cosas: (1) Consolidando múltiples perturbaciones beneficiosas en un solo contexto genético de cepa, ya sea una a la vez en un procedimiento interactivo, o como múltiples cambios en una sola etapa. Las múltiples perturbaciones pueden ser un conjunto específico de cambios definidos o una biblioteca combinatoria parcialmente aleatorizada de cambios. Por ejemplo, si el conjunto de objetivos son todos los genes de una ruta, la regeneración secuencial de la biblioteca de perturbaciones en un miembro o miembros mejorados de la biblioteca de cepas previa puede optimizar el nivel de expresión de todos los genes en una ruta, independientemente de qué genes son limitantes de velocidad en cualquier iteración dada; (2) Introduciendo los datos de rendimiento resultantes de la generación individual y combinatoria de la biblioteca en un algoritmo que usa esos datos para predecir un conjunto óptimo de perturbaciones basado en la interacción de cada perturbación; y (3) Implementar una combinación de los dos enfoques anteriores.

El enfoque se ilustra en la presente descripción con microorganismos industriales, pero es aplicable a cualquier organismo donde se puedan identificar rasgos deseados en una población de mutantes genéticos. Por ejemplo, esto se podría usar para mejorar el rendimiento de células CHO, levaduras, células de insectos, algas, así como organismos multicelulares, tales como plantas.

8. Mapeo de epistasis - Una herramienta analítica predictiva que permite consolidaciones genéticas beneficiosas

En algunos ejemplos, la presente descripción enseña métodos de mapeo de epistasis para predecir y combinar alteraciones genéticas beneficiosas en una célula hospedante. Las alteraciones genéticas pueden ser creadas por cualquiera de los conjuntos de herramientas moleculares de HTP antes mencionados (p. ej., Intercambios de promotores, intercambios de SNP, intercambios de codones de inicio/parada, optimización de secuencia, intercambios de marcadores de solubilidad de proteínas, intercambios de marcadores de degradación de proteínas e intercambios de STOP) y el efecto de esas alteraciones genéticas se conocerían a partir de la caracterización de las bibliotecas de cepas microbianas de diseño genético de HTP obtenidas. Por lo tanto, como se usa en el presente documento, la expresión mapeo de epistasis incluye métodos para identificar combinaciones de alteraciones genéticas (p. ej., SNP

beneficiosos o asociaciones de promotores/genes diana beneficiosas) que es probable que produzcan aumentos en el rendimiento del hospedante.

5 En ejemplos, los métodos de mapeo de epistasis de la presente descripción se basan en la idea de que es más probable que la combinación de mutaciones beneficiosas de dos grupos funcionales diferentes mejore el rendimiento del hospedante, en comparación con una combinación de mutaciones del mismo grupo funcional. Véase, p. ej., Costanzo, The Genetic Landscape of a Cell, *Science*, vol. 327, Número 5964, 22 de enero de 2010, págs. 425-431.

10 Es más probable que las mutaciones del mismo grupo funcional operen por el mismo mecanismo y, por lo tanto, es más probable que presenten epistasis negativa o neutra en el rendimiento general del hospedante. Por el contrario, es más probable que las mutaciones de diferentes grupos funcionales operen por mecanismos independientes, lo que puede conducir a un mejor rendimiento del hospedante y, en algunos casos, a efectos sinérgicos. Por ejemplo, en referencia a la Figura 19, *lysA* y *zwf* son genes que operan en diferentes rutas para lograr la producción de lisina. Basado en la diferencia en el rendimiento individual de esos genes, los cambios genéticos que usan esos genes deberían dar como resultado efectos de consolidación aditivos. Esto se confirmó en la medición real de los efectos consolidados de la combinación de *lysA* y *zwf*, como se muestra en la Figura 16B y el Ejemplo 6.

15 Por lo tanto, en algunos ejemplos, la presente descripción enseña métodos de análisis de mutaciones de SNP para identificar SNP que se predice que pertenecen a diferentes grupos funcionales. En algunos ejemplos, la similitud de los grupos funcionales de SNP se determina calculando la similitud del coseno de los perfiles de interacción de mutación (similar a un coeficiente de correlación, véase la Figura 16A). La presente descripción también ilustra la comparación de SNP mediante una matriz de similitud de mutaciones (véase la Figura 15, para el análisis de ejemplo realizado en *Corynebacterium*) o dendrograma (véase la Figura 16A, para el análisis de ejemplo realizado en *Corynebacterium*).

20 Por lo tanto, el procedimiento de mapeo de epistasis proporciona un método para agrupar y/o clasificar una diversidad de mutaciones genéticas aplicadas en uno o más contextos genéticos con el propósito de consolidaciones eficientes y efectivas de dichas mutaciones en uno o más contextos genéticos.

25 En los ejemplos, la consolidación se realiza con el objetivo de crear cepas novedosas que están optimizadas para la producción de biomoléculas diana. Mediante el procedimiento de mapeo de epistasis enseñado, es posible identificar agrupaciones funcionales de mutaciones, y dichas agrupaciones funcionales permiten una estrategia de consolidación que minimiza los efectos epistáticos indeseables.

30 Como se ha explicado previamente, la optimización de microbios para usar en la fermentación industrial es un problema importante y difícil, con amplias implicaciones para la economía, la sociedad y el mundo natural. Tradicionalmente, la ingeniería microbiana se ha realizado mediante un procedimiento lento e incierto de mutagénesis aleatoria. Dichos enfoques aprovechan la capacidad evolutiva natural de las células para adaptarse a la presión de selección impuesta artificialmente. Dichos enfoques también están limitados por la rareza de las mutaciones beneficiosas, la solidez del entorno de conveniencia subyacente y, de manera más general, infrautilizan el estado de la técnica en biología celular y molecular.

35 Los enfoques modernos aprovechan la nueva comprensión de la función celular en el nivel mecanístico y nuevas herramientas de biología molecular para llevar a cabo manipulaciones genéticas dirigidas a fines fenotípicos específicos. En la práctica, dichos enfoques racionales se confunden por la complejidad subyacente de la biología. Los mecanismos causales se entienden mal, en particular cuando se intentan combinar dos o más cambios los cuales tienen cada uno un efecto beneficioso observado. A veces, dichas consolidaciones de cambios genéticos producen resultados positivos (medidos por aumentos en la actividad fenotípica deseada), aunque el resultado positivo neto puede ser menor de lo esperado y en algunos casos mayor de lo esperado. En otros casos, dichas combinaciones producen un efecto neto neutro o un efecto neto negativo. Este fenómeno se conoce como epistasis y es uno de los desafíos fundamentales para la ingeniería microbiana (y la ingeniería genética en general).

45 Como se ha mencionado antes, la plataforma de ingeniería genómica de HTP actual resuelve muchos de los problemas asociados con los enfoques tradicionales de ingeniería microbiana. La plataforma de HTP actual usa tecnologías de automatización para realizar cientos o miles de mutaciones genéticas de una vez. En ejemplos particulares, a diferencia de los enfoques racionales descritos antes, la plataforma de HTP descrita permite la construcción paralela de miles de mutantes para explorar de manera más efectiva grandes subconjuntos del espacio genómico relevante, como se describe en la solicitud de EE.UU. N° 15/140.296, titulada "Diseño de cepas microbianas Sistema y métodos para la producción mejorada a gran escala de secuencias de nucleótidos diseñadas". Al intentar "todo", la presente plataforma de HTP evita las dificultades inducidas por nuestra comprensión biológica limitada.

50 Sin embargo, al mismo tiempo, la presente plataforma de HTP se enfrenta al problema de estar fundamentalmente limitada por el tamaño explosivo combinatorio del espacio genómico y la efectividad de las técnicas computacionales para interpretar los conjuntos de datos generados dada la complejidad de las interacciones genéticas. Se necesitan técnicas para explorar subconjuntos de espacios combinatorios vastos de manera que maximicen la selección no aleatoria de combinaciones que produzcan los resultados deseados.

- Enfoques de HTP algo similares han demostrado ser efectivos en el caso de la optimización enzimática. En este nicho del problema, una secuencia genómica de interés (del orden de 1000 bases) codifica una cadena de proteínas con alguna configuración física complicada. La configuración precisa está determinada por las interacciones electromagnéticas colectivas entre sus componentes atómicos constituyentes. Esta combinación de secuencia genómica corta y problema de plegamiento físicamente restringido se presta específicamente a estrategias de optimización ambiciosas. Es decir, es posible mutar individualmente la secuencia en cada resto y barajar los mutantes resultantes para muestrear eficazmente el espacio de secuencia local con una resolución compatible con la modelización de respuesta de actividad de secuencia.
- Sin embargo, para optimizaciones genómicas completas para biomoléculas, estos enfoques centrados en restos son insuficientes por algunas razones importantes. Primero, debido al aumento exponencial en el espacio de secuencia relevante asociado con optimizaciones genómicas para biomoléculas. Segundo, debido a la complejidad adicional de regulación, expresión e interacciones metabólicas en la síntesis de biomoléculas. Los autores de la presente invención han resuelto estos problemas mediante el procedimiento de mapeo de epistasis enseñado.
- El método enseñado para modelar interacciones epistáticas, entre una colección de mutaciones con el propósito de una consolidación más eficiente y eficaz de dichas mutaciones en uno o más contextos genéticos, es innovador y muy necesario en la técnica.
- Cuando se describe el procedimiento de mapeo de la epistasis, las expresiones "más eficiente" y "más efectivo" se refieren a evitar interacciones epistáticas indeseables entre cepas de consolidación con respecto a objetivos fenotípicos particulares.
- Puesto que el procedimiento se ha desarrollado de forma general más arriba, ahora se describirá un ejemplo de flujo de trabajo más específico.
- Primero*, se comienza con una biblioteca de mutaciones M y uno o más contextos genéticos (p. ej., cepas bacterianas parentales). Ni la elección de la biblioteca ni la elección de los contextos genéticos son específicas del método descrito en el presente documento. Pero en una implementación particular, una biblioteca de mutaciones puede incluir exclusivamente, o en combinación: bibliotecas de intercambio de SNP, bibliotecas de intercambio de promotores o cualquier otra biblioteca de mutaciones descrita en el presente documento.
- En una implementación, solo se proporciona un único contexto genético. En este caso, primero se generará una colección de distintos contextos genéticos (mutantes microbianos) a partir de este único contexto. Esto se puede lograr aplicando la biblioteca primaria de mutaciones (o algún subconjunto de la misma) al contexto dado, por ejemplo, la aplicación de una biblioteca de diseño genético de HTP de SNP particulares o una biblioteca de diseño genético de HTP de promotores particulares para el contexto genético dado, para crear una población (quizás cientos o miles) de mutantes microbianos con un contexto genético idéntico excepto por la alteración genética particular de la biblioteca de diseño genético de HTP dada incorporada en el mismo. Como se detalla a continuación, este ejemplo puede conducir a una biblioteca combinatoria o una biblioteca por pares.
- En otra implementación, se puede dar simplemente una colección de distintos contextos genéticos conocidos. Como se detalla a continuación, este ejemplo puede conducir a un subconjunto de una biblioteca combinatoria.
- En una implementación particular, se determina el número de contextos genéticos y la diversidad genética entre estos contextos (medidos en número de mutaciones o distancia de edición de secuencia o similar) para maximizar la eficacia de este método.
- Un contexto genético puede ser una cepa natural, nativa o no manipulada genéticamente o una cepa transformada mutada. N cepas de distinto contexto genético se pueden representar por un vector \mathbf{b} . En un ejemplo, el contexto de \mathbf{b} puede representar contextos transformados, formados aplicando N mutaciones primarias $\mathbf{m}_0 = (m_1, m_2, \dots, m_N)$ a una cepa de contexto de tipo natural b_0 para formar las N cepas de contexto mutado $\mathbf{b} = \mathbf{m}_0 b_0 = (m_1 b_0, m_2 b_0, \dots, m_N b_0)$, donde $m_i b_0$ representa la aplicación de la mutación m_i a la cepa de contexto b_0 .
- En cualquier caso (es decir, un único contexto genético proporcionado o una colección de contextos genéticos), el resultado es una colección de N contextos genéticamente distintos. Se miden los fenotipos relevantes para cada contexto genético.
- Segundo*, cada mutación en una colección de M mutaciones \mathbf{m}_1 se aplica a cada contexto dentro de la colección de N cepas de contexto \mathbf{b} para formar una colección de mutantes $M \times N$. En la implementación donde los N contextos se obtenían ellos mismos aplicando el conjunto primario de mutaciones \mathbf{m}_0 (como se ha descrito antes), el conjunto resultante de mutantes se denominará a veces una biblioteca combinatoria o una biblioteca por pares. En otra implementación, en la que se ha proporcionado explícitamente una colección de contextos conocidos, el conjunto resultante de mutantes se puede denominar un subconjunto de una biblioteca combinatoria. Similar a la generación de vectores de contexto transformado, en los ejemplos, la interfaz de entrada 202 (véase la Figura 31) recibe el vector de mutación \mathbf{m}_1 y el vector de contexto \mathbf{b} , y una operación especificada tal como el producto cruzado.

Continuando con el ejemplo de contexto modificado anterior, la formación de la biblioteca combinatoria $M \times N$ se puede representar por la matriz formada por $m_1 \times m_0 b_0$, el producto cruzado de m_1 aplicado a los N contextos de $b = m_0 b_0$, donde cada mutación en m_1 se aplica a cada cepa de contexto dentro de b . Cada i -ésima fila de la matriz $M \times N$ resultante representa la aplicación de la i -ésima mutación dentro de m_1 a todas las cepas dentro de la colección de contexto b . En un ejemplo, $m_1 = m_0$ y la matriz representa la aplicación por pares de las mismas mutaciones a la cepa inicial b_0 . En ese caso, la matriz es simétrica con respecto a su diagonal ($M=N$), y la diagonal se puede ignorar en cualquier análisis ya que representa la aplicación de la misma mutación dos veces.

En los ejemplos, la formación de la matriz $M \times N$ se puede lograr introduciendo en la interfaz de entrada 202 (véase la Figura 31) la expresión compuesta $m_1 \times m_0 b_0$. Los vectores componentes de la expresión se pueden introducir directamente con sus elementos especificados explícitamente, mediante una o más especificaciones de ADN, o como llamadas a la biblioteca 206 para permitir la recuperación de los vectores durante la interpretación por el intérprete 204. Como se describe en la solicitud de patente de EE.UU., Número de serie 15/140.296, titulada "Sistema y métodos de diseño de cepas microbianas para la producción mejorada a gran escala de secuencias de nucleótidos transformadas", a través del intérprete 204, el motor de ejecución 207, el motor de colocación de pedidos 208 y la fábrica 210, el sistema LIMS 200 genera las cepas microbianas especificadas por la expresión de entrada.

Tercero, con referencia a la Figura 42, el equipo de análisis 214 (véase la Figura 31) mide las respuestas fenotípicas para cada mutante dentro de la matriz de biblioteca combinatoria $M \times N$ (4202). Así pues, la colección de respuestas se puede interpretar como una Matriz R de respuesta $M \times N$. Cada elemento de R se puede representar como $r_{ij} = y(m_i, m_j)$, donde y representa la respuesta (rendimiento) de la cepa de contexto b_j dentro de la colección transformada b como mutada por la mutación m_i . Por simplicidad y practicidad, se asumen mutaciones por pares donde $m_1 = m_0$. Cuando, como aquí, el conjunto de mutaciones representa una biblioteca de mutaciones por pares, la matriz resultante también puede denominarse matriz de interacción de genes o, más particularmente, matriz de interacción de mutaciones.

Los expertos en la técnica reconocerán que, en algunos ejemplos, las operaciones relacionadas con los efectos epistáticos y el diseño predictivo de cepas se pueden realizar enteramente a través de medios automatizados del sistema LIMS 200, p. ej., mediante el equipo de análisis 214 (véase la Figura 31), o por implementación humana, o mediante una combinación de medios automatizados y manuales. Cuando una operación no está totalmente automatizada, los elementos del sistema LIMS 200, p. ej., el equipo de análisis 214, pueden recibir, por ejemplo, los resultados de rendimiento humanos de las operaciones en lugar de generar resultados a través de sus propias capacidades operativas. Como se describe en otra parte del presente documento, los componentes del sistema LIMS 200, tales como el equipo de análisis 214, se pueden implementar total o parcialmente mediante uno o más sistemas informáticos. En algunos ejemplos, en particular cuando las operaciones relacionadas con el diseño predictivo de cepas se realizan por una combinación de medios automatizados y manuales, el equipo de análisis 214 puede incluir no solo hardware, software o firmware informático (o una combinación de los mismos), sino también equipos operados por un operador humano tal como el que se citan en la Tabla 5 más adelante, p. ej., el equipo citado bajo la categoría de "Evaluar el rendimiento".

Cuarto, el equipo de análisis 212 (véase la Figura 31) normaliza la matriz de respuesta. La normalización consiste en procesos manuales y/o, en este ejemplo, automatizados para ajustar los valores de respuesta medidos con el fin de eliminar el sesgo y/o aislar las partes relevantes del efecto específico de este método. Con respecto a la Figura 42, la primera etapa 4202 puede incluir la obtención de datos medidos normalizados. En general, en las reivindicaciones dirigidas al diseño predictivo de cepas y al mapeo de epistasis, las expresiones "medida de rendimiento" o "rendimiento medido" o similares se pueden usar para describir un indicador que refleja los datos medidos, ya sea en bruto o procesados de alguna manera, p. ej., datos normalizados. En una implementación particular, la normalización se puede realizar restando una respuesta de contexto previamente medida del valor de respuesta medido. En esa implementación, los elementos de respuesta resultantes se pueden formar como $r_{ij} = y(m_i, m_j) - y(m_j)$, donde $y(m_j)$ es la respuesta de la cepa de contexto transformado b_j dentro de la colección transformada b causada por aplicación de la mutación primaria m_j a la cepa parental b_0 . Obsérvese que cada fila de la matriz de respuestas normalizadas se trata como un perfil de respuesta para su correspondiente mutación. Es decir, la i -ésima fila describe el efecto relativo de la mutación correspondiente m_i aplicada a todas las cepas de contexto b_j para $j=1$ a N .

Con respecto al ejemplo de mutaciones por pares, el rendimiento/respuesta combinado de las cepas resultantes de dos mutaciones puede ser mayor, menor o igual que el rendimiento/respuesta de la cepa para cada una de las mutaciones individualmente. Este efecto se conoce como "epistasis" y, en algunos ejemplos, se puede representar como $e_{ij} = y(m_i, m_j) - (y(m_i) + y(m_j))$. Las variaciones de esta representación matemática son posibles y pueden depender, por ejemplo, de cómo interaccionan biológicamente los cambios individuales. Como se ha indicado antes, es más probable que las mutaciones del mismo grupo funcional operen por el mismo mecanismo y, por lo tanto, es más probable que presenten una epistasis negativa o neutra en el rendimiento general del hospedante. Por el contrario, es más probable que las mutaciones de diferentes grupos funcionales operen por mecanismos independientes, lo que puede conducir a un mejor rendimiento del hospedante al reducir los efectos mutantes redundantes, por ejemplo. Por lo tanto, es más probable que las mutaciones que producen respuestas diferentes se combinen de manera aditiva que las mutaciones que producen respuestas similares. Esto conduce al cálculo de la similitud en la siguiente etapa.

Quinto, el equipo de análisis 214 mide la similitud entre las respuestas, en el ejemplo de mutación por pares, la similitud entre los efectos de la i -ésima mutación y la j -ésima mutación (p. ej., primaria) dentro de la matriz de respuestas (4204). Recuerde que la i -ésima fila de \mathbf{R} representa los efectos en el rendimiento de la i -ésima mutación m_i en las N cepas de contexto, cada una de las cuales puede ser ella misma el resultado de mutaciones diseñadas como se ha descrito antes. Por lo tanto, la similitud entre los efectos de las mutaciones i -ésima y j -ésima se puede representar por la similitud S_{ij} entre las filas i -ésima y j -ésima, p_i y p_j , respectivamente, para formar una matriz de similitud \mathbf{S} , un ejemplo de la cual se ilustra en la Figura 15. La similitud se puede medir usando muchas técnicas conocidas, tales como la correlación cruzada o la similitud absoluta del coseno, p. ej., $S_{ij} = \text{abs}(\cos(p_i, p_j))$.

Como alternativa o complemento a un indicador como la similitud del coseno, se pueden agrupar los perfiles de respuesta para determinar el grado de similitud. El agrupamiento se puede realizar mediante el uso de algoritmos de agrupamiento basados en la distancia (p. ej., k -media, aglomeración jerárquica, etc.) junto con una medida de distancia adecuada (p. ej. Euclidiana, Hamming, etc.). Alternativamente, el agrupamiento se puede realizar usando algoritmos de agrupamiento basados en similitud (p. ej., espectral, corte mínimo, etc.) con una medida de similitud adecuada (p. ej., coseno, correlación, etc.). Por supuesto, las medidas de distancia se pueden equiparar a medidas de similitud y viceversa por una serie de operaciones funcionales estándar (p. ej., la función exponencial). En una implementación, el agrupamiento aglomerativo jerárquico se puede usar junto con una similitud de coseno absoluta. (Véase la Figura 16A para ver un ejemplo de análisis realizado en *Corynebacterium*).

Como ejemplo de agrupamiento, sea C un agrupamiento de mutaciones m_i en k grupos distintos. Sea \mathbf{C} la matriz de pertenencia al grupo, donde c_{ij} es el grado en que la mutación i pertenece al grupo j , un valor entre 0 y 1. La similitud basada en grupos entre las mutaciones i y j viene dada entonces por $C_i \times C_j$ (el producto escalar de las filas i -ésima y j -ésima de \mathbf{C}). En general, la matriz de similitud basada en grupos viene dada por $\mathbf{C}\mathbf{C}^T$ (es decir, \mathbf{C} multiplicado por \mathbf{C} -transposición). En el caso del agrupamiento duro (una mutación pertenece exactamente a un grupo), la similitud entre dos mutaciones es 1 si pertenecen al mismo grupo y 0 si no.

Como se describe en Costanzo, The Genetic Landscape of a Cell, *Science*, vol. 327, Número 5964, 22 de enero de 2010, págs. 425-431, dicho agrupamiento de perfiles de respuesta a mutaciones se relaciona con un mapeo aproximado de la organización funcional subyacente de una célula. Es decir, las mutaciones que se agrupan entre sí tienden a estar relacionadas por un proceso biológico subyacente o una ruta metabólica. Dichas mutaciones se denominan en el presente documento "grupo funcional". La observación clave de este método es que si dos mutaciones operan por el mismo proceso biológico o ruta, entonces los efectos observados (y en especial los beneficios observados) pueden ser redundantes. A la inversa, si dos mutaciones operan por un mecanismo distante, es menos probable que los efectos beneficiosos sean redundantes.

Sexto, basándose en el efecto epistático, el equipo de análisis 214 selecciona pares de mutaciones que conducen a respuestas diferentes, p. ej., su indicador de similitud de coseno cae por debajo de un umbral de similitud, o sus respuestas se encuentran dentro de grupos suficientemente separados (p. ej., en la Figura 15 y Figura 16A, por ejemplo, análisis realizados en *Corynebacterium*) como se muestra en la Figura 42 (4206). Basándose en su diferencia, los pares de mutaciones seleccionados deberían consolidarse en cepas de contexto mejor que pares similares.

Basado en los pares seleccionados de mutaciones que conducen a respuestas suficientemente diferentes, el sistema LIMS (p. ej., todas o alguna combinación del intérprete 204, motor de ejecución 207, colocador de pedidos 208 y fábrica 210) se puede usar para diseñar cepas microbianas que tengan esas mutaciones seleccionadas (4208). En los ejemplos, como se describe a continuación y en otra parte del presente documento, los efectos epistáticos se pueden incorporar o usar junto con el modelo predictivo para ponderar o filtrar la selección de la cepa.

Se supone que es posible estimar el rendimiento (también conocido como puntuación) de una cepa hipotética obtenida por la consolidación de una colección de mutaciones de la biblioteca en un contexto particular a través de algún modelo predictivo preferido. Un modelo predictivo representativo usado en los métodos enseñados se proporciona en la siguiente sección titulada "Diseño predictivo de cepas" que se encuentra en la sección más grande de: "Análisis computacional y predicción de efectos de los criterios de diseño genético de todo el genoma".

Cuando se emplea una técnica de diseño predictivo de cepas tal como la regresión lineal, el equipo de análisis 214 puede restringir el modelo a mutaciones que tienen medidas de similitud baja, p. ej., filtrando los resultados de la regresión para mantener solo mutaciones suficientemente diferentes. Alternativamente, el modelo predictivo se puede ponderar con la matriz de similitud. Por ejemplo, algunos ejemplos pueden emplear una regresión de mínimos cuadrados ponderados usando la matriz de similitud para caracterizar las interdependencias de las mutaciones propuestas. Como ejemplo, la ponderación se puede hacer aplicando el truco del "núcleo (*Kernel*)" al modelo de regresión. (En la medida en que el "truco del núcleo" es general para muchos enfoques de modelización de aprendizaje automático, esta estrategia de reponderación no se limita a la regresión lineal).

Los expertos en la técnica conocen dichos métodos. En los ejemplos, el núcleo es una matriz que tiene elementos $1 - w \cdot s_{ij}$ donde 1 es un elemento de la matriz identidad y w es un valor real entre 0 y 1. Cuando $w = 0$, esto se reduce a un modelo de regresión estándar. En la práctica, el valor de w estará ligado a la precisión (valor r^2 o error cuadrático medio (RMSE)) del modelo predictivo cuando se evalúe frente a las construcciones combinatorias por pares y sus

efectos asociados $y(m_i, m_j)$. En una implementación simple, w se define como $w = 1 - r^2$. En este caso, cuando el modelo es completamente predictivo, $w = 1 - r^2 = 0$ y la consolidación se basa únicamente en el modelo predictivo y el procedimiento de mapeo epistático no juega ningún papel. Por otro lado, cuando el modelo predictivo no es predictivo en absoluto, $w = 1 - r^2 = 1$ y la consolidación se basa únicamente en el procedimiento de mapeo epistático. Durante cada iteración, se puede evaluar la precisión para determinar si el rendimiento del modelo está mejorando.

Debería estar claro que el procedimiento de mapeo epistático descrito en el presente documento no depende de qué modelo usa el equipo de análisis 214. Dado dicho modelo predictivo, es posible puntuar y clasificar todas las cepas hipotéticas accesibles a la biblioteca de mutaciones a través de la consolidación combinatoria.

En algunos ejemplos, para tener en cuenta los efectos epistáticos, el equipo de análisis 214 puede usar los perfiles de respuesta de mutación diferentes para aumentar la puntuación y el rango asociados con cada cepa hipotética del modelo predictivo. Este procedimiento puede considerarse de manera general como una ponderación de nuevo de las puntuaciones, para así favorecer las cepas candidatas con perfiles de respuesta diferentes (p. ej., cepas extraídas de una diversidad de grupos). En una implementación simple, una cepa puede tener su puntuación reducida por el número de mutaciones constituyentes que no satisfacen el umbral de disimilitud o que se extraen del mismo grupo (con ponderación adecuada). En una implementación particular, la estimación del rendimiento de una cepa hipotética se puede reducir por la suma de términos en la matriz de similitud asociada con todos los pares de mutaciones constituyentes asociadas con la cepa hipotética (nuevamente con la ponderación adecuada). Las cepas hipotéticas pueden volver a clasificarse utilizando estas puntuaciones aumentadas. En la práctica, dichos cálculos de reponderación se pueden realizar junto con la estimación de puntuación inicial.

El resultado es una colección de cepas hipotéticas con puntuación y rango aumentados para evitar de manera más efectiva interacciones epistáticas de confusión. Las cepas hipotéticas se pueden construir en este momento, o se pueden pasar a otro método computacional para su posterior análisis o uso.

Los expertos en la técnica reconocerán que el mapeo de epistasis y el diseño predictivo iterativo de cepas como se describe en el presente documento no se limitan a emplear solamente mutaciones por pares, sino que pueden expandirse a la aplicación simultánea de muchas más mutaciones a una cepa de contexto. En otro ejemplo, se pueden aplicar mutaciones adicionales secuencialmente a cepas que ya han sido mutadas usando mutaciones seleccionadas de acuerdo con los métodos predictivos descritos en el presente documento. En otro ejemplo, los efectos epistáticos se imputan aplicando la misma mutación genética a varios contextos de cepas que difieren ligeramente entre sí, y observando cualquier diferencia significativa en los perfiles de respuesta positiva entre los contextos de cepas modificadas.

Organismos susceptibles de diseño genético

La plataforma de ingeniería genómica HTP descrita se ilustra con cultivos de células microbianas industriales (p. ej., *Corynebacterium*), pero es aplicable a cualquier organismo celular hospedante donde se pueden identificar los rasgos deseados en una población de mutantes genéticos.

Además, como se expone en la introducción, la descripción actual proporciona una plataforma de ingeniería genómica HTP para mejorar las características de la célula hospedante en sistemas de *E. coli* y resuelve muchos problemas que previamente han impedido el desarrollo de dicho sistema en *E. coli*.

Por lo tanto, como se usa en el presente documento, el término "microorganismo" se debe tomar de manera amplia. Incluye, pero no se limita a los dos dominios procariontes, bacterias y arqueas, así como ciertos protistas y hongos eucariotes. Sin embargo, en ciertos ejemplos, se pueden usar organismos eucariotes "superiores" tales como insectos, plantas y animales en los métodos enseñados en el presente documento.

Las células hospedantes adecuadas incluyen, pero no se limitan a: células bacterianas, células de algas, células vegetales, células fúngicas, células de insectos y células de mamíferos. En un ejemplo ilustrativo, las células hospedantes adecuadas incluyen *E. coli* (p. ej., *E. coli* competente SHuffle™ disponible de New England BioLabs en Ipswich, Mass.). El genoma de *E. coli* tiene un tamaño de 4.646.332 pb (véase la Figura 52).

Las cepas hospedantes adecuadas de la especie *E. coli* comprenden: *E. coli* enterotoxigénica (ETEC), *E. coli* enteropatógena (EPEC), *E. coli* enteroinvasiva (EIEC), *E. coli* enterohemorrágica (EHEC), *E. coli* uropatógena (UPEC), *E. coli* productora de verotoxina, *E. coli* O157:H7, *E. coli* O104:H4, Escherichia coli O121, Escherichia coli O104:H21, Escherichia coli K1 y Escherichia coli NC101. En algunos ejemplos, la presente descripción enseña la transformación genómica de *E. coli* K12, *E. coli* B y *E. coli* C.

En algunos ejemplos, la presente descripción enseña la transformación genómica de las cepas de *E. coli* NCTC 12757, NCTC 12779, NCTC 12790, NCTC 12796, NCTC 12811, ATCC 11229, ATCC 25922, ATCC 8739, DSM 30083, BC 5849, BC 8265, BC 8267, BC 8268, BC 8270, BC 8271, BC 8272, BC 8273, BC 8276, BC 8277, BC 8278, BC 8279, BC 8312, BC 8317, BC 8319, BC 8320, BC 8321, BC 8322, BC 8326, BC 8327, BC 8331, BC 8335, BC 8338, BC 8341, BC 8344, BC 8345, BC 8346, BC 8347, BC 8348, BC 8863, y BC 8864.

En algunos ejemplos, la presente descripción enseña *E. coli* verocitotoxigénicas (VTEC), tal como las cepas BC 4734 (O26:H11), BC 4735 (O157:H-), BC 4736 , BC 4737 (n.d.), BC 4738 (O157:H7), BC 4945 (O26:H-), BC 4946 (O157:H7), BC 4947 (O111:H-), BC 4948 (O157:H), BC 4949 (O5), BC 5579 (O157:H7), BC 5580 (O157:H7), BC 5582 (O3:H), BC 5643 (O2:H5), BC 5644 (O128), BC 5645 (O55:H-), BC 5646 (O69:H-), BC 5647 (O101:H9), BC 5648 (O103:H2), BC 5850 (O22:H8), BC 5851 (O55:H-), BC 5852 (O48:H21), BC 5853 (O26:H11), BC 5854 (O157:H7), BC 5855 (O157:H-), BC 5856 (O26:H-), BC 5857 (O103:H2), BC 5858 (O26:H11), BC 7832, BC 7833 (O forma cruda:H-), BC 7834 (ONT:H-), BC 7835 (O103:H2), BC 7836 (O57:H-), BC 7837 (ONT:H-), BC 7838, BC 7839 (O128:H2), BC 7840 (O157:H-), BC 7841 (O23:H-), BC 7842 (O157:H-), BC 7843, BC 7844 (O157:H-), BC 7845 (O103:H2), BC 7846 (O26:H11), BC 7847 (O145:H-), BC 7848 (O157:H-), BC 7849 (O156:H47), BC 7850, BC 7851 (O157:H-), BC 7852 (O157:H-), BC 7853 (O5:H-), BC 7854 (O157:H7), BC 7855 (O157:H7), BC 7856 (O26:H-), BC 7857, BC 7858, BC 7859 (ONT:H-), BC 7860 (O129:H-), BC 7861, BC 7862 (O103:H2), BC 7863, BC 7864 (O forma cruda:H-), BC 7865, BC 7866 (O26:H-), BC 7867 (O forma cruda:H-), BC 7868, BC 7869 (ONT:H-), BC 7870 (O113:H-), BC 7871 (ONT:H-), BC 7872 (ONT:H-), BC 7873, BC 7874 (O forma cruda:H-), BC 7875 (O157:H-), BC 7876 (O111:H-), BC 7877 (O146:H21), BC 7878 (O145:H-), BC 7879 (O22:H8), BC 7880 (O forma cruda:H-), BC 7881 (O145:H-), BC 8275 (O157:H7), BC 8318 (O55:K:H-), BC 8325 (O157:H7), y BC 8332 (ONT), BC 8333.

En algunos ejemplos, la presente descripción enseña *E. coli* enteroinvasivas (EIEC), tal como las cepas BC 8246 (O152:K:H-), BC 8247 (O124:K(72):H3), BC 8248 (O124), BC 8249 (O112), BC 8250 (O136:K(78):H-), BC 8251 (O124:H-), BC 8252 (O144:K:H-), BC 8253 (O143:K:H-), BC 8254 (O143), BC 8255 (O112), BC 8256 (O28a.e), BC 8257 (O124:H-), BC 8258 (O143), BC 8259 (O167:K:H5), BC 8260 (O128a.c.:H35), BC 8261 (O164), BC 8262 (O164:K:H-), BC 8263 (O164), y BC 8264 (O124).

En algunos ejemplos, la presente descripción enseña *E. coli* enterotoxigénicas (ETEC), tal como las cepas BC 5581 (O78:H11), BC 5583 (O2:K1), BC 8221 (O118), BC 8222 (O148:H-), BC 8223 (O111), BC 8224 (O110:H-), BC 8225 (O148), BC 8226 (O118), BC 8227 (O25:H42), BC 8229 (O6), BC 8231 (O153:H45), BC 8232 (O9), BC 8233 (O148), BC 8234 (O128), BC 8235 (O118), BC 8237 (O111), BC 8238 (O110:H17), BC 8240 (O148), BC 8241 (O6H16), BC 8243 (O153), BC 8244 (O15:H-), BC 8245 (O20), BC 8269 (O125a.c:H-), BC 8313 (O6:H6), BC 8315 (O153:H-), BC 8329, BC 8334 (O118:H12), y BC 8339.

En algunos ejemplos, la presente descripción enseña *E. coli* enteropatógenas (EPEC), tal como las cepas BC 7567 (O86), BC 7568 (O128), BC 7571 (O114), BC 7572 (O119), BC 7573 (O125), BC 7574 (O124), BC 7576 (O127a), BC 7577 (O126), BC 7578 (O142), BC 7579 (O26), BC 7580 (OK26), BC 7581 (O142), BC 7582 (O55), BC 7583 (O158), BC 7584 (O-), BC 7585 (O-), BC 7586 (O-), BC 8330, BC 8550 (O26), BC 8551 (O55), BC 8552 (O158), BC 8553 (O26), BC 8554 (O158), BC 8555 (O86), BC 8556 (O128), BC 8557 (OK26), BC 8558 (O55), BC 8560 (O158), BC 8561 (O158), BC 8562 (O114), BC 8563 (O86), BC 8564 (O128), BC 8565 (O158), BC 8566 (O158), BC 8567 (O158), BC 8568 (O111), BC 8569 (O128), BC 8570 (O114), BC 8571 (O128), BC 8572 (O128), BC 8573 (O158), BC 8574 (O158), BC 8575 (O158), BC 8576 (O158), BC 8577 (O158), BC 8578 (O158), BC 8581 (O158), BC 8583 (O128), BC 8584 (O158), BC 8585 (O128), BC 8586 (O158), BC 8588 (O26), BC 8589 (O86), BC 8590 (O127), BC 8591 (O128), BC 8592 (O114), BC 8593 (O114), BC 8594 (O114), BC 8595 (O125), BC 8596 (O158), BC 8597 (O26), BC 8598 (O26), BC 8599 (O158), BC 8605 (O158), BC 8606 (O158), BC 8607 (O158), BC 8608 (O128), BC 8609 (O55), BC 8610 (O114), BC 8615 (O158), BC 8616 (O128), BC 8617 (O26), BC 8618 (O86), BC 8619, BC 8620, BC 8621, BC 8622, BC 8623, BC 8624 (O158), y BC 8625 (O158).

En algunos ejemplos, la presente descripción también enseña métodos para a transformación de organismos *Shigella*, que incluyen *Shigella flexneri*, *Shigella dysenteriae*, *Shigella boydii* y *Shigella sonnei*.

Generación de grupos de diversidad genética para usar en el diseño genético y la plataforma de ingeniería microbiana de HTP

En algunos ejemplos, los métodos de la presente descripción se caracterizan como diseño genético. Como se usa en el presente documento, la expresión diseño genético se refiere a la reconstrucción o alteración del genoma de un organismo hospedante a través de la identificación y selección de las variantes más óptimas de un gen particular, parte de un gen, promotor, codón de parada, 5'UTR, 3'UTR u otra secuencia de ADN para diseñar y crear nuevas células hospedantes superiores.

En algunos ejemplos, una primera etapa en los métodos de diseño genético de la presente descripción es obtener una población de grupo de diversidad genética inicial con una pluralidad de variaciones de secuencia a partir de la cual se puede reconstruir un nuevo genoma de hospedante.

En algunos ejemplos, una etapa posterior en los métodos de diseño genético enseñados en el presente documento es usar uno o más de los conjuntos de herramientas moleculares de HTP antes mencionados (p. ej., intercambio de SNP, intercambio de promotor, intercambio de terminador, intercambio de marcadores de solubilidad de proteínas o intercambio de marcadores de degradación de proteínas) para construir bibliotecas de diseño genético de HTP, que luego funcionan como conductores del procedimiento de ingeniería genómica, proporcionando bibliotecas de alteraciones genómicas particulares para ensayar en una célula hospedante.

Aprovechamiento de grupos de diversidad de cepas de tipo natural existentes

5 En algunos ejemplos, la presente descripción enseña métodos para identificar la diversidad de secuencias presente entre los microbios de una población de tipo natural dada. Por lo tanto, un grupo de diversidad puede ser un número n dado de microbios de tipo natural usados para el análisis, representando dichos genomas de microbios el "grupo de diversidad".

10 En algunos ejemplos, los grupos de diversidad pueden ser el resultado de la diversidad existente presente en la variación genética natural entre dichos microbios de tipo natural. Esta variación puede resultar de variantes de cepas de una célula hospedante dada o puede ser el resultado de que los microbios sean especies completamente diferentes. Las variaciones genéticas pueden incluir cualquier diferencia en la secuencia genética de las cepas, se encuentren de forma natural o no. En algunos ejemplos, las variaciones genéticas pueden incluir intercambios de SNP, intercambios de PRO, intercambios de codón de inicio/detención, intercambios de MARCADORES DE SOLUBILIDAD, intercambios de MARCADORES DE DEGRADACIÓN o intercambios de STOP, entre otros.

Aprovechamiento de grupos de diversidad de variantes de cepas industriales existentes

15 En otros ejemplos de la presente descripción, los grupos de diversidad son variantes de cepas creadas durante los procedimientos tradicionales de mejora de cepas (p. ej., una o más cepas de organismos hospedantes generadas por mutación aleatoria y seleccionadas para obtener mejores rendimientos a lo largo de los años). Por tanto, en algunos ejemplos, el grupo de diversidad o los organismos hospedantes pueden comprender una colección de cepas de producción históricas.

20 En ejemplos particulares, un grupo de diversidad puede ser una cepa microbiana parental original (S_1) con una secuencia genética "base" en un punto de tiempo particular ($S_1\text{Gen}_1$) y después cualquier número de cepas descendientes posteriores (S_2, S_3, S_4, S_5 , etc., generalizable a S_{2-n}) que se derivaron/desarrollaron a partir de dicha cepa S_1 y que tienen un genoma diferente ($S_{2-n}\text{Gen}_{2-n}$), en relación con el genoma base de S_1 .

25 Por ejemplo, en algunos ejemplos, la presente descripción enseña la secuenciación de los genomas microbianos en un grupo de diversidad para identificar los SNP presentes en cada cepa. En un ejemplo, las cepas del grupo de diversidad son cepas de producción microbiana históricas. Por lo tanto, un grupo de diversidad de la presente descripción puede incluir, por ejemplo, una cepa base industrial y una o más cepas industriales mutadas producidas por programas tradicionales de mejora de cepas.

30 Una vez que se identifican todos los SNP en el grupo de diversidad, la presente descripción enseña métodos de intercambio de SNP y métodos de cribado para delinear (es decir, cuantificar y caracterizar) los efectos (p. ej., creación de un fenotipo de interés) de los SNP individualmente y en grupos. Por lo tanto, como se ha mencionado antes, una etapa inicial en la plataforma enseñada puede ser obtener una población de grupo de diversidad genética inicial con una pluralidad de variaciones de secuencia, p. ej. SNP. Después, una etapa posterior en la plataforma enseñada puede ser usar uno o más de los conjuntos de herramientas moleculares de HTP antes mencionados (p. ej. intercambio de SNP) para construir bibliotecas de diseño genético de HTP, que después funcionan como conductores del procedimiento de ingeniería genómica, proporcionando bibliotecas de alteraciones genómicas particulares para ensayar en un microbio.

35 En algunos ejemplos, los métodos de intercambio de SNP de la presente descripción comprenden la etapa de introducir uno o más SNP identificados en una cepa mutada (p. ej., una cepa de entre $S_{2-n}\text{Gen}_{2-n}$) en una cepa base ($S_1\text{Gen}_1$) o cepa de tipo natural.

40 En otros ejemplos, los métodos de intercambio de SNP de la presente descripción comprenden la etapa de eliminar uno o más SNP identificados en una cepa mutada (p. ej., una cepa de entre $S_{2-n}\text{Gen}_{2-n}$).

Creación de grupos de diversidad por mutagénesis

45 En algunos ejemplos, las mutaciones de interés en una población de células de un grupo de diversidad dado se pueden generar artificialmente por cualquier medio para las mutaciones de cepas, incluidas sustancias químicas mutagénicas o radiación. El término "mutagenizante" se usa en el presente documento para referirse a un método para inducir una o más modificaciones genéticas en material de ácido nucleico celular.

50 La expresión "modificación genética" se refiere a cualquier alteración del ADN. Las modificaciones de genes representativas incluyen inserciones, deleciones, sustituciones de nucleótidos y combinaciones de las mismas, y pueden ser tan pequeñas como una sola base o tan grandes como decenas de miles de bases. Por lo tanto, la expresión "modificación genética" abarca inversiones de una secuencia de nucleótidos y otros reordenamientos cromosómicos, de modo que se altera la posición u orientación del ADN que comprende una región de un cromosoma. Un reordenamiento cromosómico puede comprender un reordenamiento intracromosómico o un reordenamiento intercromosómico.

55 En un ejemplo, los métodos mutagenizantes empleados en la materia objeto actualmente reivindicada son sustancialmente aleatorios, de modo que se puede producir una modificación genética en cualquier posición de

nucleótido disponible dentro del material de ácido nucleico que se va a mutagenizar. Dicho de otra manera, en un ejemplo, la mutagenización no muestra una preferencia o mayor frecuencia de aparición en secuencias de nucleótidos particulares.

5 Los métodos de la descripción pueden emplear cualquier agente mutagénico que incluye, pero no se limita a: luz ultravioleta, radiación de rayos X, radiación gamma, N-etil-N-nitrosourea (ENU), metinitrosourea (MNU), procarbazina (PRC), trietilen-melamina (TEM), monómero de acrilamida (AA), clorambucilo (CHL), melfalán (MLP), ciclofosfamida (CPP), sulfato de dietilo (DES), metanosulfonato de etilo (EMS), metanosulfonato de metilo (MMS), 6-mercaptopurina (6-MP), mitomicina-C (MMC), N-metil-N'-nitro-N-nitrosoguanidina (MNNG), ³H₂O y uretano (UR) (véase, p. ej., Rinchik, 1991; Marker et al., 1997; y Russell, 1990). Los agentes mutagénicos adicionales son bien conocidos por los expertos en la técnica, incluidos los descritos en http://www.iephb.nw.ru/~spirov/hazard/mutagen_lst.html.

10 El término "mutagenizar" también abarca un método para alterar (p. ej., por mutación dirigida) o modular una función celular, para mejorar así la velocidad, la calidad o la extensión de la mutagénesis. Por ejemplo, una célula se puede alterar o modular para que de ese modo sea disfuncional o deficiente en la reparación del ADN, metabolismo mutágeno, sensibilidad a mutágenos, estabilidad genómica o combinaciones de los mismos. Por lo tanto, la alteración de las funciones génicas que normalmente mantienen la estabilidad genómica se puede usar para potenciar la mutagénesis. Los objetivos representativos de la alteración incluyen, pero no se limitan a ADN ligasa I (Bentley et al., 2002) y caseína quinasa I (patente de Estados Unidos N° 6.060.296).

15 En algunos ejemplos, la mutagénesis específica de sitio (p. ej., mutagénesis dirigida por cebador usando un kit disponible en el mercado tal como el kit Transformer Site Directed mutagenesis (Clontech)) se usa para realizar una pluralidad de cambios a lo largo de una secuencia de ácido nucleico con el fin de generar ácido nucleico que codifica una enzima de escisión de la presente descripción.

20 La frecuencia de modificación genética tras la exposición a uno o más agentes mutagénicos se puede modular variando la dosis y/o la repetición del tratamiento, y se puede adaptar para una aplicación particular.

25 Por lo tanto, en algunos ejemplos, la "mutagénesis", como se usa en el presente documento, comprende todas las técnicas conocidas en la técnica para inducir mutaciones, incluida la mutagénesis por PCR propensa a errores, mutagénesis dirigida por oligonucleótidos, mutagénesis dirigida al sitio y recombinación de secuencias iterativa por cualquiera de las técnicas descritas en el presente documento.

Mutaciones de un solo locus para generar diversidad

30 En algunos ejemplos, la presente descripción enseña la mutación de poblaciones de células introduciendo, eliminando o reemplazando porciones seleccionadas de ADN genómico. Por tanto, en algunos ejemplos, la presente descripción enseña métodos para dirigir mutaciones a un locus específico. En otros ejemplos, la presente descripción enseña el uso de tecnologías de edición de genes tales como ZFN, TALENS, Lambda Red o CRISPR, para editar selectivamente regiones de ADN diana.

35 En otros ejemplos, la presente descripción enseña la mutación de regiones de ADN seleccionadas fuera del organismo hospedante y después la inserción de la secuencia mutada de nuevo en el organismo hospedante. Por ejemplo, en algunos ejemplos, la presente descripción enseña la mutación de promotores nativos o sintéticos para producir una gama de variantes de promotores con diversas propiedades de expresión (véase la escalera de promotores más adelante). En otros ejemplos, la presente descripción es compatible con técnicas de optimización de un solo gen, tales como ProSAR (Fox et al. 2007. "Improving catalytic function by ProSAR-driven enzyme evolution". *Nature Biotechnology* Vol 25 (3) 338-343).

40 En algunos ejemplos, las regiones seleccionadas de ADN se producen in vitro por el barajado de genes de variantes naturales o barajado con oligos sintéticos, recombinación plásmido-plásmido, recombinación de plásmido virus, recombinación de virus-virus. En otros ejemplos, las regiones genómicas se producen por PCR propensa a errores.

45 En algunos ejemplos, la generación de mutaciones en regiones genéticas seleccionadas se logra por "PCR de reensamblaje". Brevemente, se sintetizan los cebadores oligonucleótidos (oligos) para la amplificación por PCR de segmentos de una secuencia de ácido nucleico de interés, de manera que las secuencias de los oligonucleótidos se solapan con las uniones de dos segmentos. La región de solapamiento tiene típicamente una longitud de aproximadamente 10 a 100 nucleótidos. Cada uno de los segmentos se amplifica con un conjunto de dichos cebadores. Después, los productos de PCR se "reensamblan" de acuerdo con los protocolos de ensamblaje. En resumen, en un protocolo de ensamblaje, los productos de la PCR primero se purifican de los cebadores, por ejemplo, mediante electroforesis en gel o cromatografía de exclusión por tamaños. Los productos purificados se mezclan entre sí y se someten a aproximadamente 1-10 ciclos de desnaturalización, reasociación y extensión en presencia de polimerasa y desoxinucleósido trifosfatos (dNTP) y sales tampón adecuadas en ausencia de cebadores adicionales ("autocebado"). La PCR posterior con cebadores que flanquean el gen se usa para amplificar el rendimiento de los genes completamente reensamblados y barajados.

55 En algunos ejemplos de la descripción, las regiones de ADN mutado, tal como las descritas antes, se enriquecen con secuencias mutantes de modo que el espectro de mutantes múltiples, es decir, las posibles combinaciones de

mutaciones, se muestrea de manera más eficiente. En algunos ejemplos, las secuencias mutadas se identifican mediante una matriz de afinidad de proteína mutS (Wagner et al., *Nucleic Acids Res.* 23(19): 3944-3948 (1995); Su et al., *Proc. Natl. Acad. Sci.* (U.S.A.), 83: 5057-5061 (1986)) con una etapa preferida de amplificación del material purificado por afinidad in vitro antes de una reacción de ensamblaje. Este material amplificado después se pone en una reacción de PCR de ensamblaje o reensamblaje como se describe en partes posteriores de esta solicitud.

Escalera de promotores

Los promotores regulan la velocidad a la que se transcriben los genes y pueden influir en la transcripción en una variedad de formas. Los promotores constitutivos, por ejemplo, dirigen la transcripción de sus genes asociados a una velocidad constante independientemente de las condiciones celulares internas o externas, mientras que los promotores regulables aumentan o disminuyen la velocidad a la que se transcribe un gen dependiendo de las condiciones celulares internas y/o externas, p. ej., velocidad de crecimiento, temperatura, respuestas a sustancias químicas ambientales específicas y similares. Los promotores se pueden aislar de sus entornos celulares normales y transformar para regular la expresión de prácticamente cualquier gen, lo que permite la modificación eficaz del crecimiento celular, el rendimiento del producto y/u otros fenotipos de interés.

En algunos ejemplos, la presente descripción enseña métodos para producir bibliotecas de escalera de promotores para usar en métodos de diseño genético posteriores. Por ejemplo, en algunos ejemplos, la presente descripción enseña métodos para identificar uno o más promotores y/o generar variantes de uno o más promotores dentro de una célula hospedante, que presentan un intervalo de fuerzas de expresión o propiedades reguladoras superiores. Una combinación particular de estos promotores identificados y/o generados se puede agrupar como una escalera de promotores, que se explica con más detalle a continuación.

En algunos ejemplos, la presente descripción enseña el uso de escaleras de promotores. En algunos ejemplos, las escaleras de promotores de la presente descripción comprenden promotores que presentan un intervalo continuo de perfiles de expresión. Por ejemplo, en algunos ejemplos, las escaleras de promotores se crean por: identificación de promotores naturales, nativos o de tipo natural que presentan un intervalo de fuerzas de expresión en respuesta a un estímulo, o por expresión constitutiva (véase, p. ej., la Figura 20 y Figuras 28- 30). Estos promotores identificados se pueden agrupar entre sí como una escalera de promotores.

En otros ejemplos, la presente descripción enseña la creación de escaleras de promotores que presentan un intervalo de perfiles de expresión en diferentes condiciones. Por ejemplo, en algunos ejemplos, la presente descripción enseña la creación de una escalera de promotores con máximos de expresión extendidos a lo largo de las diferentes etapas de una fermentación (véase, p. ej., la Figura 28). En otros ejemplos, la presente descripción enseña la creación de una escalera de promotores con diferentes dinámicas de máximos de expresión en respuesta a un estímulo específico (véase, p. ej., la Figura 29). Los expertos en la técnica reconocerán que las escaleras de promotores reguladores de la presente descripción pueden ser representativas de uno o más perfiles reguladores.

En algunos ejemplos, las escaleras de promotores de la presente descripción están diseñadas para perturbar la expresión génica de una manera predecible a lo largo de un intervalo continuo de respuestas. En algunos ejemplos, la naturaleza continua de una escalera de promotores confiere a los programas de mejora de cepas un poder predictivo adicional. Por ejemplo, en algunos ejemplos, el intercambio de promotores o secuencias de terminación de una ruta metabólica seleccionada puede producir una curva de rendimiento de la célula hospedante, que identifica la relación o perfil de expresión más óptimo; producir una cepa en la que el gen diana ya no es un factor limitante para una reacción particular o cascada genética, mientras que también evita la sobreexpresión innecesaria o la expresión errónea en circunstancias inadecuadas. En algunos ejemplos, las escaleras de promotores se crean por: identificación de promotores naturales, nativos o de tipo natural que presentan los perfiles deseados. En la Tabla 1.4 se pueden encontrar ejemplos de promotores nativos para usar en los métodos proporcionados en el presente documento. En otros ejemplos, las escaleras de promotores se crean mutando promotores que se encuentran de forma natural para derivar múltiples secuencias de promotores mutadas. En cada uno de estos promotores mutados se analiza su efecto en la expresión del gen diana. En algunos ejemplos, en los promotores editados se analiza la actividad de expresión a través de una variedad de condiciones, de modo que la actividad de cada variante del promotor se documenta/caracteriza/anota y se almacena en una base de datos. Las variantes de promotores editadas resultantes se organizan posteriormente en escaleras de promotores dispuestas en función de la fuerza de su expresión (p. ej., con variantes de expresión alta cerca de la parte superior y expresión atenuada cerca de la parte inferior, conduciendo así al término "escalera"). En la Tabla 1.4 se pueden encontrar ejemplos de promotores sintéticos para usar en los métodos proporcionados en el presente documento.

En algunos ejemplos, la presente descripción enseña escaleras de promotores que son una combinación de promotores que se encuentran de forma natural identificados y promotores variantes mutados.

En algunos ejemplos, la presente descripción enseña métodos para identificar promotores naturales, nativos o de tipo natural que satisfacen los dos criterios siguientes: 1) representan una escalera de promotores constitutivos; y 2) podrían ser codificados por secuencias de ADN cortas, idealmente menos de 100 pares de bases. En algunos ejemplos, los promotores constitutivos de la presente descripción presentan expresión génica constante en dos condiciones de crecimiento seleccionadas (típicamente comparadas entre las condiciones experimentadas durante el

cultivo industrial). En el Ejemplo 12 se puede encontrar un ejemplo de examen de la expresión génica usando diferentes promotores proporcionados en el presente documento. En algunos ejemplos, los promotores de la presente descripción consistirán en un promotor central de ~60 pares de bases y una 5' UTR de entre 26 y 40 pares de bases de longitud.

5 Los promotores nativos para incluir en las escaleras de promotores para usar en los métodos PROSWP proporcionados en el presente documento se pueden seleccionar basándose en dicho promotor nativo que muestra una variación mínima en la expresión de un gen asociado. Además, los promotores nativos pueden tener una longitud de 60-90 pb y pueden consistir en una secuencia que se encuentra 50 pb delante de un sitio de inicio de transcripción putativo y, opcionalmente, la secuencia hasta pero sin incluir un codón de inicio putativo. En la Tabla 1.4 se pueden encontrar ejemplos de promotores nativos para usar en los métodos proporcionados en el presente documento. En particular, los promotores nativos para usar en los métodos proporcionados en el presente documento se pueden seleccionar de los ácidos nucleicos de SEQ ID NO 71-131 de la Tabla 1.4.

15 En algunos ejemplos, una o más de las secuencias de promotores que se encuentran de forma natural identificadas mencionadas anteriormente se eligen para la edición de genes. En algunos ejemplos, los promotores naturales se editan por cualquiera de los métodos de mutación descritos más arriba. En otros ejemplos, los promotores de la presente descripción se editan sintetizando nuevas variantes de promotores con la secuencia deseada.

20 Los promotores sintéticos para incluir en escaleras de promotores para usar en los métodos PROSWP proporcionados en el presente documento pueden ser secuencias quiméricas de 60-90 pb de longitud. Las bibliotecas de promotores sintéticos para usar en el presente documento pueden comprender un conjunto o pluralidad de promotores sintéticos que se pueden diseñar y construir de manera que es probable que se expresen constitutivamente y/o representen un intervalo de fuerzas de expresión en comparación entre sí. Además, los promotores sintéticos se pueden diseñar y construir de manera que no sea probable que se unan a elementos reguladores presentes en *E. coli* y, por lo tanto, dirijan la expresión génica de forma constitutiva.

25 Para lograr estos objetivos de diseño, los promotores sintéticos quiméricos pueden comprender todos o una combinación de los elementos que se encuentran en la Tabla 1.5. En particular, en relación con un sitio de inicio de la transcripción, los promotores sintéticos pueden comprender o consistir en una región distal, una región -35, una región central, una región -10 y una región 5'UTR/sitio de unión del ribosoma (RBS), como se muestra en la **Figura 54**. La región distal puede estar situada justo secuencia arriba de la región -35, mientras que la región central puede estar situada entre las regiones -35 y -10, como se muestra en la **Figura 54**. Tanto la región distal como la central pueden ser importantes para la unión de elementos reguladores (véase Cox et al., *Mol Syst Biol.* 2007; 3: 145). Puesto que se espera que el promotor p_R del fago lambda dirija la expresión de manera constitutiva, las regiones distal y central de este promotor se pueden usar en la estrategia de diseño. La región central del promotor p_L del fago lambda también se incluirá por la misma razón, así como para añadir variedad adicional a la biblioteca.

35 Las regiones -35 y -10 se pueden incluir porque se sabe que son particularmente importantes en procariontes para la unión de la ARN polimerasa y, por lo tanto, pueden ser críticas para modular el grado de expresión. En un ejemplo, se usan las regiones -35 y -10 del promotor p_R y el promotor p_L del fago lambda. Las regiones -35 y -10 de p_R y p_L se pueden usar ya que se espera que dirijan una expresión fuerte. Además, se pueden usar las regiones -35 y -10 que se encuentran en muchos promotores de *E. coli* nativos, de modo que dichas regiones -35 y -10 representan variaciones pequeñas de p_R y p_L y se puede esperar que disminuyan la fuerza del promotor en comparación con p_R y p_L . La secuencia variable de 6 pb que constituye las regiones -35 y -10 se puede seleccionar de las secuencias -35 y -10 que se encuentran en la Tabla 1.5.

45 Además de los elementos anteriores, los promotores sintéticos quiméricos pueden comprender una región no traducida 5' (5'-UTR) que incluye un sitio de unión al ribosoma (RBS), que puede ser particularmente importante en procariontes para la unión al ribosoma y, por lo tanto, ser crítico para modular el grado de expresión de proteínas. En un ejemplo, el 5'-UTR/RBS del gen *acs* nativo de *E. coli* se puede usar para añadir variedad adicional a la biblioteca. En otro ejemplo, se puede usar el 5'UTR/RBS del promotor p_R del fago lambda.

En la Tabla 1.4 se pueden encontrar ejemplos de promotores sintéticos para usar en los métodos proporcionados en el presente documento. En particular, los promotores sintéticos para usar en los métodos proporcionados en el presente documento se pueden seleccionar de los ácidos nucleicos de SEQ ID NO 132-207 de la Tabla 1.4.

50 La descripción completa de la solicitud de patente de EE.UU. No. 62/264.232, presentada el 07 de diciembre de 2015.

En la Tabla 1 y/o Tabla 1.4 a continuación se proporciona una lista no exhaustiva de los promotores de la presente descripción. Cada una de las secuencias de promotores se puede denominar promotor heterólogo o polinucleótido promotor heterólogo.

Tabla 1. Secuencias de promotores seleccionadas de la presente descripción.

SEQ ID NO.	Nombre corto del promotor	Nombre del promotor
1	P1	Pcg0007_lib_39
2	P2	Pcg0007
3	P3	Pcg1860
4	P4	Pcg0755
5	P5	Pcg0007_265
6	P6	Pcg3381
7	P7	Pcg0007_119
8	P8	Pcg3121

Tabla 1.4 Secuencias de promotores adicionales de la presente descripción.

Nombre del promotor	SEQ ID NO.	Tipo*
b0904_promotor	71	Nativo
b2405_promotor	72	Nativo
b0096_promotor	73	Nativo
b0576_promotor	74	Nativo
b2017_promotor	75	Nativo
b1278_promotor	76	Nativo
b4255_promotor	77	Nativo
b0786_promotor	78	Nativo
b0605_promotor	79	Nativo
b1824_promotor	80	Nativo
b1061_promotor	81	Nativo
b0313_promotor	82	Nativo
b0814_promotor	83	Nativo
b4133_promotor	84	Nativo
b4268_promotor	85	Nativo
b0345_promotor	86	Nativo
b2096_promotor	87	Nativo
b1277_promotor	88	Nativo
b1646_promotor	89	Nativo
b4177_promotor	90	Nativo
b0369_promotor	91	Nativo
b1920_promotor	92	Nativo

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*
b3742_promotor	93	Nativo
b3929_promotor	94	Nativo
b3743_promotor	95	Nativo
b1613_promotor	96	Nativo
b1749_promotor	97	Nativo
b2478_promotor	98	Nativo
b0031_promotor	99	Nativo
b2414_promotor	100	Nativo
b1183_promotor	101	Nativo
b0159_promotor	102	Nativo
b2837_promotor	103	Nativo
b3237_promotor	104	Nativo
b3778_promotor	105	Nativo
b2349_promotor	106	Nativo
b1434_promotor	107	Nativo
b3617_promotor	108	Nativo
b0237_promotor	109	Nativo
b4063_promotor	110	Nativo
b0564_promotor	111	Nativo
b0019_promotor	112	Nativo
b2375_promotor	113	Nativo
b1187_promotor	114	Nativo
b2388_promotor	115	Nativo
b1051_promotor	116	Nativo
b4241_promotor	117	Nativo
b4054_promotor	118	Nativo
b2425_promotor	119	Nativo
b0995_promotor	120	Nativo
b1399_promotor	121	Nativo
b3298_promotor	122	Nativo
b2114_promotor	123	Nativo
b2779_promotor	124	Nativo
b1114_promotor	125	Nativo
b3730_promotor	126	Nativo

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*
b3025_promotor	127	Nativo
b0850_promotor	128	Nativo
b2365_promotor	129	Nativo
b4117_promotor	130	Nativo
b2213_promotor	131	Nativo
pMB029_promotor	132	Sintético
pMB023_promotor	133	Sintético
pMB025_promotor	134	Sintético
pMB019_promotor	135	Sintético
pMB008_promotor	136	Sintético
pMB020_promotor	137	Sintético
pMB022_promotor	138	Sintético
pMB089_promotor	139	Sintético
pMB001_promotor	140	Sintético
pMB051_promotor	141	Sintético
pMB070_promotor	142	Sintético
pMB074_promotor	143	Sintético
pMB046_promotor	144	Sintético
pMB071_promotor	145	Sintético
pMB013_promotor	146	Sintético
pMB080_promotor	147	Sintético
pMB03 8_promotor	148	Sintético
pMB060_promotor	149	Sintético
pMB064_promotor	150	Sintético
pMB058_promotor	151	Sintético
pMB085_promotor	152	Sintético
pMB081_promotor	153	Sintético
pMB091_promotor	154	Sintético
pMB027_promotor	155	Sintético
pMB048_promotor	156	Sintético
pMB055_promotor	157	Sintético
pMB006_promotor	158	Sintético
pMB012_promotor	159	Sintético
pMB014_promotor	160	Sintético

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*
pMB028_promotor	161	Sintético
pMB059_promotor	162	Sintético
pMB061_promotor	163	Sintético
pMB043_promotor	164	Sintético
pMB066_promotor	165	Sintético
pMB079_promotor	166	Sintético
pMB032_promotor	167	Sintético
pMB068_promotor	168	Sintético
pMB082_promotor	169	Sintético
pMB030_promotor	170	Sintético
pMB067_promotor	171	Sintético
pMB050_promotor	172	Sintético
pMB069_promotor	173	Sintético
pMB017_promotor	174	Sintético
pMB039_promotor	175	Sintético
pMB011_promotor	176	Sintético
pMB072_promotor	177	Sintético
pMB016_promotor	178	Sintético
pMB077_promotor	179	Sintético
pMB047_promotor	180	Sintético
pMB052_promotor	181	Sintético
pMB090_promotor	182	Sintético
pMB035_promotor	183	Sintético
pMB073_promotor	184	Sintético
pMB004_promotor	185	Sintético
pMB054_promotor	186	Sintético
pMB024_promotor	187	Sintético
pMB007_promotor	188	Sintético
pMB005_promotor	189	Sintético
pMB003_promotor	190	Sintético
pMB088_promotor	191	Sintético
pMB065_promotor	192	Sintético
pMB037_promotor	193	Sintético
pMB009_promotor	194	Sintético

Nombre del promotor	SEQ ID NO.	Tipo*
pMB041_promotor	195	Sintético
pMB036_promotor	196	Sintético
pMB049_promotor	197	Sintético
pMB044_promotor	198	Sintético
pMB042_promotor	199	Sintético
pMB086_promotor	200	Sintético
pMB053_promotor	201	Sintético
pMB057_promotor	202	Sintético
pMB018_promotor	203	Sintético
pMB002_promotor	204	Sintético
pMB015_promotor	205	Sintético
pMB087_promotor	206	Sintético
pMB063_promotor	207	Sintético
*Promotores nativos de Escherichia coli		

Tabla 1.5. Partes de secuencias usadas en la biblioteca de promotor sintético-5'UTR combinatoria

Nombre de la parte	Secuencias de la parte	Origen
distal	ACCGTGCGTG (SEQ ID NO. 208)	fago λ , promotor P_R
-35	TTTACA	variación
	TTGACT	fago λ , promotor P_R
	TTGACA	fago λ , promotor P_L
	TAGGCT	variación
	TAGACT	variación
Centro	ATTTTACCTCTGGCGGT (SEQ ID NO. 209)	fago λ , promotor P_R
	TAAATACCACTGGCGGT (SEQ ID NO. 210)	fago λ , promotor P_L
-10	GATAAT	fago λ , promotor P_R
	GATACT	fago λ , promotor P_L
	TAGAGT	variación
	TATAAT	variación
	TATTTT	variación

Nombre de la parte	Secuencias de la parte	Origen
5'-UTR/RBS	GGTTGCATGTACTAAGGAGGTTGT (SEQ ID NO. 211)	fago λ , promotor P_R
	TAACATCCTACAAGGAGAACAAAAGC (SEQ ID NO. 212)	<i>E. coli</i> , promotor del gen <i>acs</i>

En algunos ejemplos, los promotores de la presente descripción presentan una identidad de secuencia de al menos 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81 %, 80%, 79%, 78%, 77%, 76%, o 75% con un promotor de la tabla 1 y/o tabla 4.

5 **Diseño de elementos reguladores bicistrónicos**

Una de las barreras para un diseño genético de HTP eficiente y escalable es la falta de piezas estándar que se puedan reutilizar de forma fiable en combinaciones novedosas. Muchos ejemplos dentro de *E. coli* destacan cómo las funciones genéticas aparentemente simples se comportan de manera diferente en diferentes entornos. Por ejemplo, en algunos ejemplos, un elemento del sitio de unión al ribosoma (RBS) procariota que inicia la traducción de una secuencia codificante podría no funcionar en absoluto con otra secuencia codificante (véase Salis, H.M., et al. "Automated design of synthetic ribosome binding sites to control protein expression" *Nat. Biotechnol.* Vol. 27, 946-950 (2009)). Si los elementos genéticos que codifican el control de los procesos celulares centrales, tales como la transcripción y la traducción, no se pueden reutilizar de manera fiable, entonces hay pocas posibilidades de que objetos de orden superior codificados a partir de dichos elementos básicos sean fiables en sistemas a mayor escala. En algunos ejemplos, los métodos de la presente descripción superan estos desafíos mencionados antes mediante el uso de secuencias reguladoras de diseño bicistrónico.

Los diseños bicistrónicos de la presente descripción, en algunos ejemplos, pueden reducir en gran medida la variabilidad dependiente del marco en la fuerza de expresión de un promotor dado para una variedad de genes codificantes (Mutalik, et al. "Precise and reliable gene expression via standard transcription and translation initiation elements" *Nat. Biotechnol.* Vol 10 (4) págs. 354-368 (2013)). En algunos ejemplos, la presente descripción enseña que un diseño bicistrónico (BCD) es una secuencia de nucleótidos en la que un promotor dirige la expresión de dos secuencias codificantes, donde la primera secuencia codificante (Cistrón 1) termina y la segunda secuencia codificante empieza en la misma base nucleotídica (Cistrón 2/gen diana). Esta estrategia proporciona un medio para evitar la variabilidad en la fuerza de expresión de las segundas secuencias codificantes debido a interacciones impredecibles entre el promotor y la segunda secuencia codificante.

En algún ejemplo, los promotores de la presente descripción son secuencias reguladoras de compuestos que siguen el diseño bicistrónico. Es decir, en algunos ejemplos, los promotores en las escaleras de promotores de la presente descripción son secuencias reguladoras más grandes que comprenden **i)** un promotor operativamente unido a **ii)** un primer sitio de unión al ribosoma (SD1), que está operativamente unido a **iii)** una primera secuencia cistrónica (Cis1), en donde Cis1 se solapa con **iv)** un segundo sitio de unión al ribosoma (SD2), que después se une operativamente a **v)** una secuencia codificante del gen diana (Cis2) (véase la Figura 43). En algunos ejemplos, la presente descripción se refiere a la combinación de los elementos i)-iv) como un "diseño bicistrónico" o "secuencia reguladora de diseño bicistrónico" (BCD).

En algunos ejemplos, los BCD de la presente descripción se pueden unir operativamente a cualquier gen diana. Por lo tanto, en algunas realizaciones, los BCD de la presente descripción se pueden usar en lugar de los promotores tradicionales. En algunos ejemplos, la presente descripción enseña que el uso de BCD en la caja de herramientas de intercambio de PRO aumenta la coherencia con la que los transcritos expresados se traducen. Sin desear estar ligados a ninguna teoría, los autores de la presente invención creen que la presencia de secuencias líder SD1 y Cis1 unidas operativamente al gen diana recluta complejos ribosomales activos, que después son capaces de reiniciar regularmente la traducción del gen diana a través del sitio de unión al ribosoma SD2.

Se ha descrito una colección de promotores y elementos de diseño bicistrónico que se pueden usar para la ingeniería genómica de HTP (véase Mutalik, et al. "Precise and reliable gene expression via standard transcription and translation initiation elements" *Nat. Biotechnol.* Vol 10 (4) págs. 354-368 (2013)). Sin embargo, todas estas secuencias descritas contienen secuencias de ADN idénticas en los primeros 35 nt de los 48 nt de la secuencia de diseño bicistrónico reguladora (véase la secuencia del estado actual de Mutalik en la Figura 43).

En algunos ejemplos, la presente descripción enseña que el BCD de Mutalik et al., no se podría usar para transformar eficazmente múltiples genes diana en un solo organismo. Es decir, en algunos ejemplos, la presente descripción enseña contra la integración múltiple de Mutalik BCD en el genoma de una célula hospedante. Sin desear estar ligados por ninguna teoría, los autores de la presente invención creen que el uso repetido de del BCD de Mutalik et al. daría como resultado tasas crecientes de recombinación homóloga (HR) no deseada desencadenada por la presencia de secuencias altamente homólogas en todo el genoma.

En algunos ejemplos, la presente descripción resuelve este problema describiendo nuevos BCD con secuencias de nucleótidos no idénticas. Estos nuevos BCD se pueden usar para la ingeniería genómica de HTP en *E. coli* para

proporcionar cambios predecibles en la expresión de múltiples genes dentro de un solo genoma, independientemente de las secuencias codificantes de estos genes, sin inducir recombinación homóloga indeseable.

5 En algunos ejemplos, la presente descripción enseña métodos de expresión de dos proteínas génicas diana en un organismo hospedante en niveles relativamente similares. Por lo tanto, en algunos ejemplos, la presente descripción enseña la expresión de dos o más proteínas de genes diana dentro de 0,2, 0,4, 0,6, 0,8, 1, 1,2, 1,4, 1,6, 1,8, 2, 2,2, 2,4, 2,6, 2,8 o 3 veces el uno del otro.

10 En algunos ejemplos, la presente descripción enseña métodos de expresión dos proteínas génicas diana en un organismo hospedante en niveles similares, mientras que se reduce el riesgo de sucesos de recombinación homóloga (HR) no deseables desencadenados por el uso de secuencias reguladoras idénticas. Por lo tanto, en algunos ejemplos, la presente descripción enseña métodos para variar la secuencia de BCD de una manera que mantiene los niveles de expresión, a la vez que reduce el riesgo de HR. Es decir, en algunos ejemplos, la presente descripción enseña la expresión de dos o más genes que codifican proteínas a través de BCD que son idénticos salvo por sus secuencias Cis1.

Promotores de BCD

15 En algunos ejemplos, los BCD de la presente descripción comprenden una secuencia de promotor. En algunos ejemplos, los promotores comprendidos en los BCD pueden ser cualquier promotor capaz de expresarse en la célula hospedante. Por lo tanto, en algunos ejemplos, los promotores pueden ser cualquier promotor descrito en la memoria descriptiva. En algunos ejemplos, el promotor puede ser cualquier promotor que se sepa que funciona en *E. coli*. En otros ejemplos, los promotores pueden ser cualquier promotor descrito en la Tabla 1 y/o la Tabla 1.4.

20 **Primer y segundo sitio de unión al ribosoma (SD1 y SD2)**

En algunos ejemplos, los BCD de la presente descripción comprenden un primer y un segundo sitio de unión al ribosoma, denominados SD1 y SD2, respectivamente. En algunos ejemplos, las secuencias de SD1 y SD2 pueden ser las mismas. En otros ejemplos, las secuencias de SD1 y SD2 pueden ser diferentes.

25 En algunos ejemplos, las secuencias de los SD pueden ser cualquier sitio de unión al ribosoma conocido funcional en el hospedante sometido a la ingeniería genómica de HTP. En otros ejemplos, la presente descripción enseña una secuencia de SD de NNNGGANNN, en donde N se refiere a cualquier nucleótido. En otros ejemplos, la presente descripción enseña secuencias de SD seleccionadas de las secuencias descritas en la Tabla 1.1.

Tabla 1.1- Lista no limitante de sitios de unión al ribosoma, susceptibles para usar de SD1 y SD2.

SD nº	Secuencia	SD nº	Secuencia	SD nº	Secuencia
1	TGCGGAGGG	50	AGTGAACC	99	GGAGGAGAC
2	CACGGAGGC	51	GCCGGAGTG	100	TAGGGAACG
3	TGGGGAGGG	52	GCGGGACAG	101	TGAGGACGT
4	AGGGGAGGC	53	CCGGGATAA	102	ATCGGAGGT
5	GGGGGAGGG	54	CCAGGAACG	103	CCGGGAGGG
6	TATGGAGGT	55	GACGGAGCA	104	ACGGGACGG
7	GACGGAGCG	56	AGCGGATTG	105	GTGGGAGAG
8	AGGGGAGTC	57	ATGGGATAT	106	CGAGGATAG
9	ACAGGAGGC	58	CTGGGAAGT	107	TGGGGAGCC
10	GGGGGAGCG	59	TGGGGAGCC	108	CGTGGAGTG
11	ACAGGAGTC	60	GTGGGACAT	109	TGGGGACTG
12	ACAGGAGAC	61	GCCGGAGCG	110	GGTGAAGC
13	TGGGGAGGC	62	AATGGAGGC	111	GATGGACGG
14	TGGGGAGAT	63	GGAGGATTG	112	AATGGAGGT
15	GGGGGAGAA	64	GATGGACTC	113	GGAGGATCC

ES 2 875 579 T3

SD nº	Secuencia	SD nº	Secuencia	SD nº	Secuencia
16	AGGGGAGGC	65	GAGGGATGG	114	ACAGGATCT
17	AGAGGAGTC	66	GAAGGACAG	115	CCCGGACAG
18	AGGGGATAT	67	TAGGGAAGG	116	ACGGGAAAC
19	CGTGGAGTG	68	TGGGGAACC	117	AGTGGACCG
20	GGGGGAAGG	69	TTAGGAGTC	118	CATGGATCA
21	AGGGGAATC	70	GGAGGAGGA	119	CTGGGATGT
22	CCCGGAGGT	71	CCGGGATCT	120	TTAGGATGG
23	AGGGGAGGG	72	GGGGGATGA	121	GTAGGATTC
24	TTTGGAGTC	73	CTGGGAGTG	122	TTTGGAGTT
25	AGGGGACAC	74	CAAGGAACC	123	CGCGGACTC
26	CGGGGAGAC	75	GCTGGAGGC	124	CTCGGACAG
27	GGGGGAGGG	76	ATGGGACCT	125	CAAGGAGTC
28	AAGGAGATC	77	GGAGGAGGG	126	CTTGGACAG
29	TAAGGAGGT	78	CTGGGATGC	127	GCGGATCG
30	GGGGGAGTC	79	ACAGGATAC	128	GGGGGACAG
31	GGAGGATCG	80	GAGGGAAGG	129	ATTGGATGG
32	CTGGGATCG	81	AGTGGATCT	130	CCTGGATAT
33	ATCGGACCG	82	CCGGGAGTT	131	CTCGGATAC
34	GGGGGAGTG	83	GGTGGAGGC	132	GGGGGAGCC
35	TAGGGAGCA	84	CCAGGAAGA	133	TGGGGATTG
36	GGTGGAGGG	85	GGGGGATTT	134	CCAGGATCA
37	GCCGGAGGT	86	CGCGGAGTA	135	CGGGGACGG
38	TGAGGAAAG	87	CCAGGAATC	136	TCAGGACAA
39	CTTGGAGGG	88	ATTGGAGTG	137	GTAGGATGG
40	TGAGGAGAT	89	CTAGGAAGT	138	TCCGGACTA
41	CGTGGAGGG	90	GAAGGATAG	139	GTCGGATCA
42	GCGGGAGGG	91	AAAGGACAC	140	TGCGGAGTT
43	GGGGGATAG	92	CCGGGACGT	141	GGTGGACGG
44	GGGGGAGCG	93	ATGGGAGTG	142	CTTGGACGA
45	CCGGGAGCA	94	AGCGGACAG	143	CTGGGACTT
46	GCGGGAGTA	95	GGCGGATCT	144	TATGGAGTA
47	GTAGGACCG	96	ATAGGAGGG	145	ACAGGAGGC
48	CCGGGAAGG	97	TATGGAGGG		
49	GACGGAGTC	98	GGTGGACTC		

Debería estar claro que el procedimiento de mapeo epistático descrito en el presente documento no depende de qué modelo usa el equipo de análisis 214. Dado dicho modelo predictivo, se pueden puntuar y clasificar todas las cepas hipotéticas accesibles a la biblioteca de mutaciones a través de la consolidación combinatoria.

5 En algunos ejemplos, la presente descripción enseña que la variación de secuencias de SD individuales en el BCD afectará a la expresión general del BCD. Algunas secuencias de SD pueden servir para aumentar o disminuir el potencial de expresión general del BCD. Sin embargo, se espera que cada BCD presente resultados de expresión consistentes cuando se combinen con diferentes secuencias Cis2 de genes diana.

10 En algunos ejemplos, la secuencia de SD2 está completamente insertada dentro de la secuencia codificante de la primera secuencia cistrónica. Es decir, en algunos ejemplos, la secuencia de SD2 está integrada en las secuencias codificantes de Cis2. Sin desear estar ligados a ninguna teoría, los autores de la presente invención creen que las disposiciones del BCD en las que el sitio de unión al ribosoma del gen diana (SD2) está completamente insertado en la secuencia codificante del gen secuencia arriba (Cis1) da como resultado el acoplamiento de la traducciones de los péptidos Cis1 y Cis2. Más específicamente, los autores de la presente invención plantean la hipótesis de que la actividad de helicasa intrínseca de los ribosomas que llegan al codón de parada de una secuencia Cis1 secuencia arriba elimina las estructuras de ARN inhibitoras que de otro modo interrumpirían el inicio de la traducción del gen diana de Cis2 secuencia abajo.

Primera secuencia cistrónica (Cis1)

20 En algunos ejemplos, la primera secuencia cistrónica Cis1 de la presente descripción puede ser cualquier secuencia que codifique un péptido continuo. Por ejemplo, en algunos ejemplos, la secuencia de Cis1 codifica un péptido que tiene 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, o más aminoácidos de longitud, incluyendo cualquier intervalo y subintervalo de los mismos. En algunos ejemplos, no es necesario que Cis1 codifique un péptido funcional.

25 En algunos ejemplos, la secuencia de Cis1 codifica un péptido líder de 16 aminoácidos. En algunos ejemplos la secuencia de nucleótidos de Cis1 es:

5' -ATGAAAGCAATTTTCGTACTGAAACATCTTAATCATGCACAGGAGACTTTCTAA-3' (SEQ ID No. 17).

En otros ejemplos, la presente descripción enseña de que secuencia de Cis1 puede ser 5'-ATGNN - 3' , donde N puede ser cualquier ácido nucleico, con la condición de que Cis1 codifique un péptido.

30 En algunos ejemplos, la presente descripción enseña que el codón de parada de Cis1 y el codón de inicio de Cis2 deben estar próximos o solaparse. Por ejemplo, en algunos ejemplos, el codón de parada de Cis1 debe estar dentro de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, o 50 nucleótidos del codón de inicio de Cis2, incluyendo todos los intervalos y subintervalos en los mismos.

35 En algunos ejemplos, la secuencia de Cis1 solapa con la secuencia de Cis 2 en 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 100 o más nucleótidos, incluyendo todos los intervalos y subintervalos en los mismos. En otros ejemplos, el BCD de la presente descripción está diseñado de manera que la secuencia Cis1 solapa en 1 nucleótido con la secuencia Cis2 codificante del gen diana, de modo que los últimos nucleótidos codifican tanto un codón de parada como de inicio a través de un desplazamiento de marco de -1 (véase la Figura 43). En algunos ejemplos, las secuencias Cis1 y Cis 2 deben ser codificadas en diferentes marcos de lectura abiertos para evitar así la formación de una proteína quimérica que combine las secuencias de Cis1 y Cis2.

45 En algunos ejemplos, la presente descripción enseña que el codón de inicio de la secuencia de Cis1 puede ser cualquier codón de inicio funcional. En algunos ejemplos, la presente descripción enseña que los procarionotas usan ATG (AUG) en el codón de inicio más común, seguido de GTG (GUG) y TTG (UUG).

50 En algunos ejemplos, la presente descripción enseña que la secuencia de Cis1 no tiene ningún codón de parada prematuro. En otros ejemplos, la presente descripción enseña que codones raros en la secuencia de Cis1 pueden reducir la eficacia de traducción de Cis2. Por lo tanto, en algunos ejemplos, Cis1 codificará un péptido sin ningún codón raro para lograr la máxima expresión. En otros ejemplos, Cis1 codificará un péptido con uno o más codones raros con el fin de modular la expresión de Cis2.

En otros ejemplos, la presente descripción enseña que múltiples codones repetidos en la secuencia de Cis1 pueden reducir la eficacia de traducción de Cis2. Por lo tanto, en algunos ejemplos, Cis1 codificará un péptido sin ninguna repetición de codones para lograr la máxima expresión. En otros ejemplos, Cis1 codificará un péptido con una o más repeticiones de codones con el fin de modular la expresión de Cis2.

Segunda secuencia cistrónica (Cis2- Gen diana)

En algunos ejemplos, la presente descripción enseña que los BCD de la presente descripción están operativamente unidos a una secuencia del gen diana Cis2, de la misma manera en que los promotores de las bibliotecas de intercambio de PRO están unidos operativamente a secuencias diana. Es decir, en algunos ejemplos, los BCD de la presente descripción pueden tomar el lugar de los promotores tradicionales en las bibliotecas y métodos de intercambio de PRO de la presente descripción. Las secuencias de Cis2, en algunos ejemplos, pueden ser cualquier secuencia de interés.

La presente descripción enseña que, en algunos ejemplos, los genes diana que codifican un polipéptido serán regulados más eficazmente por BCD que por un promotor. Es decir, en algunos ejemplos, los BCD no modularán la expresión de ARN no codificante más de lo que sería posible mediante un promotor.

10 Escaleras de terminadores

En algunos ejemplos, la presente descripción enseña métodos para mejorar las cepas hospedantes transformadas genéticamente proporcionando una o más secuencias de terminación de la transcripción en una posición 3' hasta el final del elemento codificante de ARN. En algunos ejemplos, la presente descripción enseña que la adición de secuencias de terminación mejora la eficacia de la transcripción de ARN de un gen seleccionado en el hospedante genéticamente transformado. En otros ejemplos, la presente descripción enseña que la adición de secuencias de terminación reduce la eficacia de la transcripción de ARN de un gen seleccionado en el hospedante genéticamente transformado. Por lo tanto, en algunos ejemplos, las escaleras de terminadores de la presente descripción comprenden una serie de secuencias de terminadores que presentan un intervalo de eficiencias de transcripción (p. ej., un terminador débil, un terminador medio y un promotor fuerte).

Una secuencia de terminación de la transcripción puede ser cualquier secuencia de nucleótidos, que cuando se pone transcripcionalmente secuencia abajo de una secuencia de nucleótidos que codifica un marco de lectura abierto, produce el final de la transcripción del marco de lectura abierto. Dichas secuencias son conocidas en la técnica y pueden ser de origen procariota, eucariota o fago. Los ejemplos de secuencias de terminadores incluyen, pero no se limitan al terminador de PTH, terminador de pET-T7, terminador de T3-T Φ , terminador de pBR322-P4, terminador del virus de estomatitis vesicular, terminador de rrmB-T1, terminador de rrmC, terminador transcripcional TTadc y secuencias de terminación reconocidas por levaduras, tales como el terminador de la transcripción Mata (factor α), secuencia de terminación de la transcripción del factor α nativo, secuencia de terminación de la transcripción ADR1, secuencia de terminación de la transcripción ADH2 y secuencia de terminación de la transcripción GAPD. Se puede encontrar una lista no exhaustiva de secuencias terminadoras de la transcripción en el registro iGEM, que está disponible en: <http://partsregistry.org/Terminators/Catalog>.

En algunos ejemplos, las secuencias de terminación de la transcripción pueden ser específicas de la polimerasa o no específicas, sin embargo, los terminadores de la transcripción seleccionados para usar en los presentes ejemplos deben formar una "combinación funcional" con el promotor seleccionado, lo que significa que la secuencia del terminador debe ser capaz de terminar la transcripción por el tipo de ARN polimerasa que comienza en el promotor.

Por ejemplo, en algunos ejemplos, la presente descripción enseña que un promotor de ARN pol II eucariota y terminadores de ARN pol II eucariotas, un promotor T7 y terminadores T7, un promotor T3 y terminadores T3, un promotor reconocido por levadura y secuencias de terminación reconocidas por levadura, etc., generalmente formarían una combinación funcional. La identidad de las secuencias de terminación de la transcripción usadas también se puede seleccionar basándose en la eficacia con la que se termina la transcripción a partir de un promotor dado. Por ejemplo, una secuencia de terminador de la transcripción heteróloga se puede proporcionar transcripcionalmente secuencia abajo del elemento codificante del ARN para lograr una eficiencia de terminación de al menos 60%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% a partir de un promotor dado.

En algunos ejemplos, la eficacia de la transcripción de ARN a partir de la construcción de expresión transformada se puede mejorar proporcionando una secuencia de ácido nucleico que forme una estructura secundaria que comprenda dos o más horquillas en una posición 3' hasta el extremo del elemento codificante de ARN. Sin desear estar ligado a una teoría particular, la estructura secundaria desestabiliza el complejo de elongación de la transcripción y conduce a que la polimerasa se disocie del molde de ADN, minimizando así la transcripción improductiva de la secuencia no funcional y aumentando la transcripción del ARN deseado. Por consiguiente, se puede proporcionar una secuencia de terminación que forme una estructura secundaria que comprenda dos o más horquillas adyacentes. Generalmente, una horquilla puede estar formada por una secuencia de nucleótidos palindrómica que puede plegarse sobre sí misma para formar una región de tallo emparejada cuyos brazos están conectados por un bucle de una sola hebra. En algunos ejemplos, la secuencia de terminación comprende 2, 3, 4, 5, 6, 7, 8, 9, 10 o más horquillas adyacentes. En algunos ejemplos, las horquillas adyacentes están separadas por 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 nucleótidos no emparejados. En algunos ejemplos, un tallo de la horquilla comprende 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más pares de bases de longitud. En ciertos ejemplos, un tallo de horquilla tiene de 12 a 30 pares de bases de longitud. En ciertos ejemplos, la secuencia de terminación comprende dos o más horquillas de tamaño mediano que tienen una región de tallo que comprende aproximadamente de 9 a 25 pares de bases. En algunos ejemplos, la horquilla comprende una región que forma bucle de 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 nucleótidos. En algunos ejemplos, la región que forma bucle comprende 4-8 nucleótidos. Sin desear estar ligados

a una teoría particular, la estabilidad de la estructura secundaria se puede correlacionar con la eficiencia de la terminación. La estabilidad de la horquilla está determinada por su longitud, el número de emparejamientos erróneos o protuberancias que contiene y la composición de bases de la región emparejada. Los emparejamientos entre guanina y citosina tienen tres enlaces de hidrógeno y son más estables en comparación con los emparejamientos de adenina-timina, que solo tienen dos. El contenido de G/C de una secuencia de nucleótidos palindrómica formadora de horquilla puede ser al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90% o más. En algunos ejemplos, el contenido de G/C de una secuencia de nucleótidos palindrómica formadora de horquilla es al menos de 80%. En algunos ejemplos, la secuencia de terminación deriva de una o más secuencias terminadoras de la transcripción de origen procariota, eucariota o fago. En algunos ejemplos, se proporciona una secuencia de nucleótidos que codifica una serie de 4, 5, 6, 7, 8, 9, 10 o más adeninas (A) en 3' de la secuencia de terminación.

En algunos ejemplos, la presente descripción enseña el uso de una serie de secuencias de terminación en tándem. En algunos ejemplos, la primera secuencia de terminador de la transcripción de una serie de 2, 3, 4, 5, 6, 7 o más se puede poner directamente en 3' del nucleótido final del elemento que codifica el ARNbc o a una distancia de al menos 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-100, 100-150, 150-200, 200-300, 300-400, 400-500, 500-1,000 o más nucleótidos 3' hasta el nucleótido final del elemento codificante de ARNbc. El número de nucleótidos entre las secuencias de terminadores de la transcripción en tándem puede variar, por ejemplo, las secuencias de terminadores de la transcripción pueden estar separadas por 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50 o más nucleótidos. En algunos ejemplos, las secuencias de terminador de la transcripción se pueden seleccionar basándose en su estructura secundaria predicha según lo determinado por un algoritmo de predicción de estructura. Los programas de predicción de estructuras son bien conocidos en la técnica e incluyen, por ejemplo, CLC Main Workbench.

Los expertos en la técnica reconocerán que los métodos de la presente descripción son compatibles con cualquier secuencia de terminación. En algunos ejemplos, la presente descripción enseña el uso de terminadores de *Corynebacterium glutamicum* anotados como se describe en Pfeifer-Sancar et al. 2013. "Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique" Pfeifer-Sancar et al. *BMC Genomics* 2013, 14:888). En otros ejemplos, la presente descripción enseña el uso de secuencias de terminadores de la transcripción que se encuentran en el registro iGEM, que está disponible en: <http://partsregistry.org/Terminators/Catalog>. En la Tabla 1.2 a continuación se proporciona una lista no exhaustiva de secuencias de terminadores de la transcripción de la presente descripción.

Tabla 1.2. Lista no exhaustiva de secuencias de terminación de la presente descripción.

<i>E. coli</i>			
Nombre	Descripción	Dirección	Longitud
BBa_B0010	T1 de rrnB de <i>E. coli</i>	Directa	80
BBa_B0012	TE de colifago T7	Directa	41
BBa_B0013	TE de colifago T7 (+/-)	Directa	47
BBa_B0015	Terminador doble (B0010-B0012)	Directa	129
BBa_B0017	Terminador doble (B0010-B0010)	Directa	168
BBa_B0053	Terminador (His)	Directa	72
BBa_B0055	-- Sin descripción --		78
BBa_B1002	Terminador (artificial, pequeño, %T~85%)	Directa	34
BBa_B1003	Terminador (artificial, pequeño, %T~80)	Directa	34
BBa_B1004	Terminador (artificial, pequeño, %T~55)	Directa	34
BBa_B1005	Terminador (artificial, pequeño, %T~25%)	Directa	34
BBa_B1006	Terminador (artificial, grande, %T~>90)	Directa	39
BBa_B1010	Terminador (artificial, grande, %T~<10)	Directa	40
BBa_I11013	Modificación de partes biológicas BBa_B0015		129
BBa_I51003	-- Sin descripción --		110

ES 2 875 579 T3

BBa_J61048	Terminador [rnpB-T1]	Directa	113
BBa_K1392970	Terminador+ Promotor Tetr + Endolisina de T4		623
BBa_K1486001	promotor de arabinosa + CpxR	Directa	1924
BBa_K1486005	promotor de arabinosa + sfGFP-CpxR [Cterm]	Directa	2668
BBa_K1486009	CpxR y Split IFP1.4 [Nterm + Nterm]	Directa	3726
BBa_K780000	Terminador para Bacillus subtilis		54
BBa_K864501	Terminador tardío de T22, P22	Directa	42
BBa_K864600	Terminador transcripcional de T0 (21 imm)	Directa	52
BBa_K864601	Terminador transcripcional de Lambda t1	Directa	
BBa_B0011	LuxICDABEG (+/-)	Bidireccional	46
BBa_B0014	Terminador doble (B0012-B0011)	Bidireccional	95
BBa_B0021	LuxICDABEG (+/-), inverso	Bidireccional	46
BBa_B0024	Terminador doble (B0012-B0011), inverso	Bidireccional	95
BBa_B0050	Terminador (pBR322, +/-)	Bidireccional	33
BBa_B0051	Terminador (yciA/tonA, +/-)	Bidireccional	35
BBa_B1001	Terminador (artificial, pequeño, %T~90)	Bidireccional	34
BBa_B1007	Terminador (artificial, grande, %T~80)	Bidireccional	40
BBa_B1008	Terminador (artificial, grande, %T~70)	Bidireccional	40
BBa_B1009	Terminador (artificial, grande, %T~40%)	Bidireccional	40
BBa_K187025	Terminador en pAB, plásmido BioBytes		60
BBa_K259006	GFP- Terminador	Bidireccional	823
BBa_B0020	Terminador (inverso B0010)	Inversa	82
BBa_B0022	TE de colifago T7, Inverso	Inversa	41
BBa_B0023	TE de colifago T7, Inverso	Inversa	47
BBa_B0025	Terminador doble (B0015), Inverso	Inversa	129
BBa_B0052	Terminador (rrnC)	Directa	41
BBa_B0060	Terminador (Inverso B0050)	Bidireccional	33
BBa_B0061	Terminador (Inverso B0051)	Bidireccional	35
BBa_B0063	Terminador (Inverso B0053)	Inversa	72
Spy	Terminador (SEQ ID NO. 225)		90
pheA	Terminador (SEQ ID NO. 226)		51
osmE	Terminador (SEQ ID NO. 227)		42
rpoH	Terminador (SEQ ID NO. 228)		41
vibE	Terminador (SEQ ID NO. 229)		71
Thr1_ABC	Terminador (SEQ ID NO. 230)		57

<i>Corynebacterium</i>					
Terminador	Inicio de terminador	Final de terminador	Hebra	Final de transcrito	Secuencia de ADN
cg0001 T1	1628	1647	+	bucle	SEQ ID NO. 9
cg0007 T2	7504	7529	+	tallo 1	SEQ ID NO. 10
cg0371 T3	322229	322252	+	tallo 1	SEQ ID NO. 11
cg0480 T4	421697	421720	-	tallo 1	SEQ ID NO. 12
cg0494 T5	436587	436608	+	bucle	SEQ ID NO. 13
cg0564 T6	499895	499917	+	tallo 1	SEQ ID NO. 14
cg0610 T7	541016	541039	+	tallo 2	SEQ ID NO. 15
cg0695 T8	613847	613868	-	bucle	SEQ ID NO. 16

Escalera de marcadores de solubilidad de proteínas

En algunos ejemplos, la presente descripción enseña métodos para mejorar cepas hospedantes genéticamente transformadas proporcionando una o más secuencias de marcadores de solubilidad de proteínas operativamente unidas con una proteína diana derivada de un gen diana. Los marcadores de solubilidad pueden ser parejas de fusión operativamente unidas con la proteína diana en el extremo N o en el extremo C de la proteína diana. En algunos ejemplos, la presente descripción enseña que la adición de secuencias de marcadores de solubilidad mejora la solubilidad de una proteína traducida de un gen seleccionado en el hospedante genéticamente transformado. En otros ejemplos, los marcadores de solubilidad también se pueden usar para ayudar en la purificación de la proteína diana.

Los marcadores eficaces para usar en las escaleras de marcadores de solubilidad de proteínas de la presente descripción pueden ser cualquier marcador de solubilidad conocido en la técnica que forme dominios independientes, bien plegados, dominios altamente solubles. Estos dominios pueden contribuir a la solubilidad de su proteína diana a través de un efecto aditivo, o cuando se usan como un marcador N-terminal se pueden plegar rápidamente después de salir del ribosoma y bloquear estéricamente la cadena de aminoácidos emergente de la proteína diana para que no interaccione con otros componentes celulares que pueden causar un plegado erróneo. Además, los marcadores de solubilidad para incluir en las escalas de marcadores de solubilidad pueden tener propiedades en común tales como ser dominios pequeños, estrechamente plegados o ser secuencias líder de proteínas que se sabe que son altamente solubles. Las secuencias de marcadores de solubilidad de proteínas pueden ser cualquiera de los marcadores conocidos en la técnica tales como, por ejemplo, cualquiera de los marcadores encontrados en Costa et al., *Front Microbiol.* 2014; 5: 63. En un ejemplo, las secuencias de marcadores de solubilidad incluyen los marcadores que se encuentran en la Tabla 17.

En un ejemplo, el marcador de solubilidad de proteínas es una pareja de fusión. El gen que codifica la pareja de fusión puede estar presente en cualquiera de los vectores (p. ej., vectores lanzadera) proporcionados en el presente documento de manera que la integración del gen para una proteína diana en el vector une operativamente el gen que codifica la pareja de fusión con el gen diana. Los vectores de expresión de *E. coli* que comprenden marcadores de solubilidad para usar en el presente documento pueden comprender una secuencia de reconocimiento de proteasa entre el gen que codifica la pareja de fusión del marcador de solubilidad y el gen que codifica la proteína diana, que puede permitir la eliminación del marcador según sea necesario. La elección de una pareja de fusión para usar en los métodos de intercambio de solubilidad proporcionados en el presente documento puede depender de:

(i) El propósito de la fusión: ¿es para mejorar la solubilidad o para la purificación por afinidad? Se encuentran disponibles una variedad de marcadores de fusión que tienen diferentes propósitos, y los sistemas que contienen marcadores tanto de solubilidad como de afinidad como, por ejemplo, el marca doble de hexahistina (His6)-MBP, se pueden diseñar con el fin de obtener una producción de proteína rápida "en una sola etapa". Algunos marcadores de proteínas también pueden funcionar tanto en funciones de afinidad como de solubilidad, como por ejemplo, la MBP o la glutatión-S-transferasa (GST; Esposito y Chatterjee, *Curr Opin Biotechnol.* 2006 Agosto; 17(4): 353-8. Publicación electrónica 15 de junio de 2006).

(ii) La composición y tamaño de los aminoácidos: las proteínas diana pueden requerir marcadores más grandes o más pequeños dependiendo de su aplicación. Los marcadores más grandes pueden presentar una gran diversidad en el contenido de aminoácidos y pueden imponer una carga metabólica en la célula hospedante diferente de la impuesta por los marcadores pequeños.

(iii) Los niveles de producción requeridos: los estudios estructurales pueden requerir niveles de producción de proteínas más altos que se pueden lograr rápidamente con un marcador de fusión más grande, que tiene señales de inicio de la traducción fuertes, mientras que el estudio de interacciones fisiológicas puede exigir niveles de producción más bajos y marcadores pequeños.

5 (iv) Ubicación del marcador: las parejas de fusión pueden promover diferentes efectos cuando se ubican en el extremo N o el extremo C de la proteína diana. Los marcadores N-terminales a menudo pueden ser ventajosas frente a los marcadores C-terminales porque: (1) proporcionan un entorno fiable para el inicio de la traducción eficiente, en el que las proteínas de fusión aprovechan los sitios de inicio de la traducción eficientes en el marcador; (2) se pueden eliminar dejando pocos restos adicionales o ninguno en la secuencia N-terminal nativa de la proteína diana, puesto que la mayoría de las endoproteasas escinden en o cerca del extremo C de sus sitios de reconocimiento.

Tabla 17. Lista no exhaustiva de secuencias de marcadores de solubilidad de proteínas de la presente descripción.

Nombre del marcador de solubilidad (nº de marcador de solubilidad de proteínas)	Descripción	Organismo	SEQ ID NO. de ácido nucleico	SEQ ID NO. de aminoácidos	Tamaño (aa)
GB1 (PST1)	IgG dominio B1 de Proteína G	Streptococcus sp.	231	235	56
FH8 (PST2)	antígeno de 8 kDa de Fasciola hepatica	F. hepatica	232	236	69
Ubiquitina (PST3)			233	237	
SUMO (PST4)	Ubiquitina modificada pequeña	Homo sapiens	234	238	~100

Escaleras de marcadores de degradación de proteínas

15 En algunos ejemplos, la presente descripción enseña métodos para mejorar cepas hospedantes genéticamente transformadas proporcionando una o más secuencias de marcadores de degradación de proteínas operativamente unidas con una proteína diana derivada de un gen diana. La adición de una secuencia de marcador de degradación usando los métodos proporcionados en el presente documento puede marcar la proteína diana para la degradación. Marcar la proteína diana para la degradación puede reducir o modular la abundancia de proteína diana dentro de una célula. Al reducir o modular los niveles o la abundancia de proteína diana en la célula, la adición de secuencias de marcadores de degradación a una proteína diana puede afectar finalmente al fenotipo general de las cepas resultantes.

Los marcadores efectivos para usar en las escaleras de marcadores de degradación de proteínas de la presente descripción pueden ser cualquier marcador de degradación conocido en la técnica que sea parte de una ruta de degradación conocida en el organismo hospedante (p. ej., *E. coli*). Por ejemplo, las rutas de degradación conocidas en *E. coli* pueden incluir el sistema clpXP/clpAP, el sistema HflB, el sistema ftsH y el sistema lon. Por consiguiente, los marcadores de degradación para usar en los métodos de intercambio de marcadores de degradación proporcionados en el presente documento pueden incluir cualquier marcador que se sepa que funciona en cualquiera de estos sistemas de degradación de proteínas de *E. coli*. En algunos casos, los marcadores de degradación se pueden mutar de manera que se confiera la capacidad del marcador mutante resultante para tener su actividad sintonizada. Por ejemplo, la clase de marcadores *ssrA* se puede mutar de modo que los marcadores de degradación *ssrA* mutados marcan una proteína marcada para la degradación a través de la ruta de degradación de ClpXP con diferentes grados de eficacia. En un ejemplo, los marcadores *ssrA* pueden contener mutaciones de un solo aminoácido en los últimos tres restos de la secuencia consenso de AANDENYALAA, de modo que las proteínas diana que comprenden un marcador *ssrA* mutado C-terminal pueden ser degradadas en diferentes niveles de eficacia por determinadas proteasas intracelulares específicas de la cola. (p. ej., proteasa Tsp) dependiendo de qué aminoácido se ha mutado en los últimos tres restos de la secuencia consenso del marcador *ssrA* (véase Keiler K C, Sauer R T. "Sequence determinants of C-terminal substrate recognition by the Tsp protease". *J Biol Chem.* 1996;271:2589-2593. Por consiguiente, usando los métodos de intercambio de marcadores de degradación de la presente descripción, es posible obtener cepas de células hospedantes que tienen proteínas diana de estabilidad variable mediante la construcción de variantes que portan marcadores peptídicos C-terminales con alteraciones menores en la secuencia consenso de Tsp. Los ejemplos de marcadores *ssrA* mutantes para usar en los métodos del presente documento pueden comprender la SEQ ID NO: 248, 249 o 250 de aminoácidos.

Otro ejemplo de marcadores *ssrA* mutados para usar en los métodos proporcionados en el presente documento pueden ser los marcadores DAS encontrados en McGinness et al., "Engineering Controllable Protein Degradation" *Mol. Cell*, Vol 22 (5), junio de 2006. En los marcadores DAS, dos restos en el marcador *ssrA* se reemplazaban dando como resultado marcadores *ssrA* mutados que presentan una unión de ClpX debilitada sin disminuir el reconocimiento

de SspB. Así pues, las proteínas diana que llevan los marcadores DAS pueden ser degradadas de manera eficiente por ClpXP solo cuando está presente SspB, permitiendo regular la degradación intracelular controlando los niveles de SspB.

5 En la Tabla 18 se puede encontrar una lista no exhaustiva de secuencias de marcadores de degradación de proteínas de la presente descripción.

Tabla 18. Lista no exhaustiva de secuencias de marcadores de degradación de proteínas de la presente descripción.

Nombre de marcador de degradación (nº de marcador de degradación de proteínas)	Descripción	Organismo	SEQ ID NO. de ácido nucleico	SEQ ID NO. de aminoácidos	Fuente
ssrA_LAA (PDT1)	nativo	<i>E. coli</i>	239	247	Andersen et al., <i>Appl Environ Microbiol.</i> Junio 1998; 64(6): 2240-2246.
ssrA_LVA (PDT2)	mutante	<i>E. coli</i>	240	248	Andersen et al., <i>Appl Environ Microbiol.</i> Junio 1998; 64(6): 2240-2246.
ssrA_AAV (PDT3)	mutante	<i>E. coli</i>	241	249	Andersen et al., <i>Appl Environ Microbiol.</i> Junio 1998; 64(6): 2240-2246.
ssrA_ASV (PDT4)	mutante	<i>E. coli</i>	242	250	Andersen et al., <i>Appl Environ Microbiol.</i> Junio 1998; 64(6): 2240-2246.
ftsH-cll89-97 (PDT5)	nativo	<i>E. coli</i>	243	251	Kobiler O, Koby S, Teff D, Court D, Oppenheim AB. <i>PNAS.</i> 23 Oct 2002, 99(23): 14964-14969.
cl108 (PDT6)	nativo	<i>E. coli</i>	244	252	Herman et al., <i>Genes Dev.</i> 1 mayo 1998; 12(9): 1348-1355.
su120 (PDT7)	nativo	<i>E. coli</i>	245	253	Wohlever et al., <i>Protein Eng Des Sel.</i> Abril 2013; 26(4): 299-305.
β20 (PDT8)	nativo	<i>E. coli</i>	246	254	Wohlever et al., <i>Protein Eng Des Sel.</i> Abril 2013; 26(4): 299-305.

10 Los marcadores de degradación pueden ser parejas de fusión operativamente unidas con la proteína diana en el extremo N o en el extremo C de la proteína diana. Por consiguiente, el gen que codifica la pareja de fusión puede estar presente en cualquiera de los vectores (p. ej., vectores lanzadera) proporcionados en el presente documento de manera que la integración del gen para una proteína diana en el vector une operativamente el gen que codifica la

15 pareja de fusión con el gen diana de manera que la traducción de la construcción genera una proteína de fusión con el marcador de degradación presente en el extremo N o en el extremo C de la proteína diana, según se desee. En un ejemplo, la colocación de los marcadores de degradación (o sus mutantes) en el extremo N o el extremo C de una proteína diana puede depender del marcador usado. Por ejemplo, los marcadores de degradación (o sus mutantes) asociados con el sistema clpXP/clpAP, el sistema Hf1B, el sistema ftsH o el marcador sul20 del sistema lon se pueden unir operativamente a una proteína diana en el extremo C, mientras que los marcadores de degradación β20 (o sus mutantes) del sistema de lon se pueden unir operativamente a una proteína diana en el extremo N o internamente. En un ejemplo, el marcador de degradación es el marcador N-degron (Ntag) para *E. coli* como se encuentra en Sekar K, Gentile AM, Bostick JW, Tyo KEJ (2016) N-Terminal-Based Targeted, Inducible Protein Degradation in Escherichia coli. *PLoS ONE* 11 (2): e0149746. Ntag se puede colocar en el extremo N de una proteína diana de interés y puede servir para marcar la proteína diana de interés para la degradación en la célula hospedante de *E. coli* a través del sistema clpXP/clpAP. En otro ejemplo, el marcador de degradación es el marcador RepA que se puede situar en el extremo N de una proteína diana como se describe en Butz et al., *Biochemistry*, 2011, 50 (40), págs. 8594-8602. El marcador RepA N-terminal puede servir para marcar la proteína diana de interés para la degradación en la célula hospedante de *E. coli* a través del sistema clpXP/clpAP.

20

25

Grupos de diversidad basados en hipótesis y ascenso de colina

La presente descripción enseña que los métodos de ingeniería genómica de HTP de la presente descripción no requieren conocimiento genético previo para lograr ganancias significativas en el rendimiento de la célula hospedante. De hecho, la presente descripción enseña métodos para generar grupos de diversidad (p. ej., Figura 1) a través de varios enfoques funcionalmente agnósticos, incluyendo la mutagénesis aleatoria y la identificación de la diversidad genética entre las variantes de células hospedantes preexistentes (p. ej., tal como la comparación entre una célula hospedante de tipo natural y una variante industrial).

Sin embargo, en algunos ejemplos, la presente descripción también enseña métodos basados en hipótesis de diseño de mutaciones de diversidad genética que se usarán para la transformación de HTP secuencia abajo. Es decir, en algunos ejemplos, la presente descripción enseña el diseño dirigido de mutaciones seleccionadas. En algunos ejemplos, las mutaciones dirigidas se incorporan en las bibliotecas de ingeniería de la presente descripción (p. ej., intercambio de SNP, intercambio de PRO, intercambio de STOP, intercambio de MARCADORES DE SOLUBILIDAD o intercambio de MARCADORES DE DEGRADACIÓN).

En algunos ejemplos, la presente descripción enseña la creación de mutaciones dirigidas basadas en la anotación de genes, la función génica hipotética (o confirmada) o la ubicación dentro de un genoma. Los grupos de diversidad de la presente descripción pueden incluir mutaciones en genes que se plantea que están implicadas en una ruta metabólica o genética específica asociada en la bibliografía con un rendimiento mayor de una célula hospedante. En otros ejemplos, el grupo de diversidad de la presente descripción también puede incluir mutaciones de genes presentes en un operón asociado con un rendimiento mejorado del hospedante. En otros ejemplos más, el grupo de diversidad de la presente descripción también puede incluir mutaciones de genes basadas en la función predicha algorítmica u otra anotación genética.

En algunos ejemplos, la presente descripción enseña un enfoque basado en "cubiertas" para priorizar los objetivos de mutaciones basadas en hipótesis. La metáfora de las cubiertas para priorizar objetivos se basa en la hipótesis de que solo un puñado de genes primarios son responsables de la mayor parte de un ejemplo particular del rendimiento de una célula hospedante (p. ej., la producción de una sola biomolécula). Estos genes primarios se encuentran en el núcleo de la cubierta, seguidos por los genes de efectos secundarios en la segunda capa, los efectos terciarios en la tercera cubierta, y ... etc. Por ejemplo, en un ejemplo, el núcleo de la cubierta podría comprender genes que codifican enzimas biosintéticas críticas dentro de una ruta metabólica seleccionada (p. ej., producción de ácido cítrico). Los genes ubicados en la segunda cubierta podrían comprender genes que codifican otras enzimas dentro de la ruta biosintética responsable de la desviación del producto o la señalización por retroalimentación. Los genes de tercer nivel bajo esta metáfora ilustrativa probablemente comprenderían genes reguladores responsables de modular la expresión de la ruta biosintética, o de regular el flujo de carbono general dentro de la célula hospedante.

La presente descripción también enseña métodos de "ascenso de colina (*Hill climb*)" para optimizar las ganancias de rendimiento de cada mutación identificada. En algunos ejemplos, la presente descripción enseña que las mutaciones aleatorias, naturales o basadas en hipótesis en bibliotecas de diversidad de HTP pueden dar como resultado la identificación de genes asociados con el rendimiento de la célula hospedante. Por ejemplo, los presentes métodos pueden identificar uno o más SNP beneficiosos ubicados en, o cerca de, una secuencia codificante de gen. Este gen podría estar asociado con el rendimiento de la célula hospedante, y su identificación se puede comparar con el descubrimiento de una "colina" de rendimiento en el espacio de mutaciones genéticas combinatorias de un organismo.

En algunos ejemplos, la presente descripción enseña métodos para explorar el espacio combinatorio alrededor del ascenso identificado incorporado en la mutación SNP. Es decir, en algunos ejemplos, la presente descripción enseña la perturbación del gen identificado y las secuencias reguladoras asociadas con el fin de optimizar las ganancias de rendimiento obtenidas de ese nodo genético (es decir, ascenso de colina). Por lo tanto, de acuerdo con los métodos de la presente descripción, un gen podría identificarse primero en una biblioteca de diversidad procedente de mutagénesis aleatoria, pero posteriormente se podría mejorar para usar en el programa de mejora de cepas mediante la mutación dirigida de otra secuencia dentro del mismo gen.

El concepto de ascenso de colina también se puede ampliar más allá de la exploración del espacio combinatorio que rodea a una sola secuencia de gen. En algunos ejemplos, una mutación en un gen específico podría revelar la importancia de una ruta metabólica o genética particular para el rendimiento de la célula hospedante. Por ejemplo, en algunos ejemplos, el descubrimiento de que una mutación en un solo gen de degradación de ARN daba como resultado ganancias significativas en el rendimiento del hospedante se podría usar como base para la mutación de genes de degradación de ARN relacionados como un medio para extraer ganancias de rendimiento adicionales del organismo hospedante. Las personas con experiencia en la técnica reconocerán variantes de los enfoques de las cubiertas y el ascenso de colina descritos anteriormente para el diseño genético dirigido.

Armazón de la ruta biosintética

En algunos ejemplos, la presente descripción enseña que la productividad de algunos procedimientos bioindustriales está limitada por la difusión aleatoria de sustratos, compuestos intermedios y enzimas biosintéticas dentro de una célula hospedante. En algunos ejemplos, la presente descripción enseña que la productividad de los cultivos de células

hospedantes se puede aumentar mediante la co-localización de enzimas biosintéticas en una ruta. Por lo tanto, en algunos ejemplos, la presente descripción enseña la unión de enzimas biosintéticas a un armazón, tal como un armazón de ADN o de proteínas.

5 En algunos ejemplos, la colocalización se logra mediante fusiones recombinantes de dominios de unión de ADN a las enzimas biosintéticas en la ruta, que después se unen a una región de armazón de ADN, restringiendo así a las enzimas de la ruta cerca unas de otras en la célula. En otros ejemplos, la colocalización se logra mediante fusiones recombinantes de dominios de unión de proteínas a las enzimas biosintéticas en la ruta, que después se unen a una región de armazón de proteína, restringiendo así las enzimas de la ruta cerca unas de otras en la célula. En algunos ejemplos, la colocalización aumenta la tasa de producción y disminuye la concentración de los compuestos intermedios de la ruta en la célula (véase la Figura 44).

10 En algunos ejemplos, la presente descripción enseña un método de alta capacidad para transformar el genoma de *Escherichia coli*, en donde las secuencias de nucleótidos que codifican dominios de unión de ADN o de unión de proteínas se insertan en genes que codifican enzimas en una ruta biosintética, y se introduce un plásmido de armazón de ADN o una proteína armazón en la célula. De acuerdo con un ejemplo de la descripción, se cree que los dominios de unión de ADN o proteína ligados a los genes biosintéticos localizarán las enzimas recombinantes de la ruta juntas en el plásmido o péptido armazón, conduciendo así a una productividad mejorada del producto diana.

15 En algunos ejemplos, esta descripción resuelve el problema de la productividad limitada por difusión de moléculas pequeñas en células de *E. coli* con genomas transformados por métodos de alta capacidad. Actualmente, los únicos ejemplos descritos de armazón de ADN para localizar enzimas biosintéticas han sido procedimientos de baja capacidad en los que las enzimas recombinantes de la ruta son codificadas en plásmidos (Lee, et al., "Improved Production of L-Threonine in Escherichia coli by Use of a DNA Scaffold System" *App. And Environ. Microbiol.* Vol 79(3), pág. 774-782 (2013)). En algunos ejemplos, la descripción actual proporciona un medio para incorporar dominios de unión de ADN en enzimas de la ruta que están codificadas cromosómicamente, de una manera de alta capacidad.

20 En algunos ejemplos, la presente descripción enseña enzimas biosintéticas quiméricas y ADN y proteínas de armazón. Los diversos ejemplos de esta tecnología se discuten con más detalle a continuación.

Proteínas quiméricas de unión de ADN

25 En algunos ejemplos, la presente descripción enseña proteínas quiméricas que comprenden enzimas biosintéticas seleccionadas que están ligadas al dominio de unión de ADN. De acuerdo con estos ejemplos, se espera que las enzimas biosintéticas quiméricas sean reclutadas a un armazón de ADN por sus dominios de unión de ADN, concentrando así las diversas actividades biosintéticas en un área de la célula hospedante.

30 En algunos ejemplos, las enzimas biosintéticas y los dominios de unión de ADN se unen covalentemente. En algunos ejemplos, las enzimas biosintéticas se fusionan traduccionalmente a los dominios de unión de ADN. Por lo tanto, en algunos ejemplos, las enzimas biosintéticas quiméricas se forman acoplando el dominio de unión de ADN al extremo amino, al extremo carboxi o a un sitio interno dentro de la proteína de la ruta biosintética. Los expertos en la técnica reconocerán la necesidad de asegurarse de que la adición del dominio de unión de ADN no reduzca sustancialmente la actividad de la enzima biosintética.

35 En algunos ejemplos de la presente descripción, la enzima biosintética se acopla a su dominio de unión de ADN por una secuencia de conector polipeptídico corto. Los conectores adecuados incluyen péptidos de entre aproximadamente 6 y aproximadamente 40 aminoácidos de longitud. Las secuencias de conectores preferidas incluyen secuencias de conectores ricas en glicina (p. ej. G3-5), ricas en serina (p. ej. GSG, GSGS (SEQ ID NO. 18), GSGSG (SEQ ID NO. 19), GSNG (SEQ ID NO. 20), o ricas en alanina (p. ej., TSAAA (SEQ ID NO. 21)). Otras secuencias de conectores de ejemplo tienen una combinación de restos de glicina, alanina, prolina y metionina tales como AAAGGM (SEQ ID NO. 22); AAAGGMPPAAAGGM (SEQ ID NO. 23); AAAGGM (SEQ ID NO. 24); y PPAAAGGMM (SEQ ID NO. 25). Los conectores pueden tener prácticamente cualquier secuencia que dé como resultado una proteína de ruta biológica quimérica generalmente flexible.

40 En algunos ejemplos, los métodos de la presente descripción son compatibles con cualquier dominio de unión de ADN capaz de funcionar en cis con la enzima biosintética. En algunos ejemplos, los dominios de unión de ADN son preferiblemente exógenos al organismo hospedante. En otros ejemplos, la presente descripción enseña la selección de dominios de unión de ADN que son suficientemente selectivos para evitar una unión excesiva fuera del ADN armazón diseñado.

45 Se conocen en la técnica varios dominios de unión de ADN junto con sus correspondientes sitios de reconocimiento de nucleótidos en el ADN (es decir, sitios de unión de ADN) y son adecuados para su uso en el sistema y los métodos de la presente descripción. Por ejemplo, en un ejemplo de la presente descripción, la parte de unión de ADN de una proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de cremallera de leucina en donde el armazón comprende la secuencia de unión de ADN de cremallera de leucina correspondiente. En otro ejemplo de la presente descripción, la parte de unión de ADN de una proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de hélice-bucle-hélice en donde el armazón comprende la correspondiente secuencia de unión de ADN de hélice-bucle-hélice. En otro ejemplo, la parte de unión de ADN de una proteína de la ruta biológica quimérica

comprende un dominio de unión de ADN de hélice alada en donde el almacén comprende la secuencia de unión de ADN de hélice alada correspondiente. En otro ejemplo, la parte de unión de ADN de una proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de hélice-giro-hélice alada en donde el almacén comprende la correspondiente secuencia de unión de ADN de hélice-giro-hélice alada. En otro ejemplo, la parte de unión de ADN de una proteína de la ruta biológica quimérica comprende una unión de ADN de hélice-giro-hélice en donde el almacén comprende la correspondiente secuencia de unión de ADN de hélice-giro-hélice. En otro ejemplo, la parte de unión de ADN de la proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de HMG-caja en donde el almacén comprende la correspondiente secuencia de unión de ADN de HMG-caja. En otro ejemplo, la parte de unión de ADN de la proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de TALE diseñado a medida en donde el almacén comprende la correspondiente secuencia de unión de ADN de TALE diseñada. En otro ejemplo de la presente descripción, la parte de unión de ADN de una proteína de la ruta biológica quimérica comprende un dominio de unión de ADN de dedo de zinc en donde el almacén comprende la correspondiente secuencia de unión de ADN de dedo de zinc.

En la Tabla 1.3 se proporcionan ejemplos de secuencias de dominio de unión de ADN de dedos de zinc y los correspondientes sitios de unión de ADN. Otros dominios de unión de ADN de dedos de zinc y sus correspondientes secuencias de unión de ADN diana conocidas en la técnica también son adecuados para usar en la presente descripción (véase, por ejemplo, Greisman H A y Pabo C O, "A General Strategy for Selecting High-Affinity Zinc Finger Proteins for Diverse DNA Target Sites," *Science* 275:657-661 (1997), Rebar E J y Pabo C O, "Zinc Finger Phage: Affinity Selection of Fingers with New DNA-Binding Specificities," *Science* 263:671-673 (1994); Maeder et al., "Rapid "Open-Source" Engineering of Customized Zinc-Finger Nucleases for Highly Efficient Gene Modification," *Mol. Cell.* 31:294-301 (2008), Sander et al., "Selection-Free Zinc-Finger-Nuclease Engineering by Context-Dependent Assembly (CoDA)," *Nat. Methods* 8:67-69 (2011), patente de EE.UU. nº 5.5789.538 de Rebar, patente de EE.UU. nº 6.410.248 de Greisman, patente de EE.UU. nº 7.605.140 de Rebar, patente de EE.UU. nº 6.140.081 de Barbas, patente de EE.UU. nº 7.067.617 de Barbas, patente de EE.UU. nº 6.205.404 de Michaels, y publicación de solicitud de patente de EE.UU. nº 20070178454 de Joung.

También se conocen en la técnica métodos de optimización de las especificidades de unión de ADN de los dominios de dedos de zinc y métodos de transformación de sitios de unión de ADN sintéticos, y se pueden usar en la presente descripción para generar nuevas parejas de unión de dedos de zinc (véase p. ej., Bulyk et al., "Exploring the DNA-binding Specificities of Zinc Fingers with DNA Microarrays," *Proc. Nat'l Acad. Sci. U.S.A* 98(13): 7158-63 (2001) y "Hurt et al., "Highly Specific Zinc Finger Proteins Obtained by Directed Domain Shuffling and Cell-based Selection," *Proc. Nat'l Acad. Sci. U.S.A.* 100(21): 12271-6 (2003), patente de EE.UU. nº 5.5789.538 de Rebar, patente de EE.UU. nº 6.410.248 de Greisman, patente de EE.UU. nº 7.605.140 de Rebar, patente de EE.UU. nº 6.140.081 de Barbas, patente de EE.UU. nº 7.067.617 de Barbas, patente de EE.UU. nº 6.205.404 de Michaels, y publicación de solicitud de patente de EE.UU. nº 20070178454 de Joung.

Tabla 1.3 - Lista no limitante de dominios de unión de ADN

Dedo de Zinc	Secuencia del dominio de unión de ADN	Secuencia de unión de ADN (5'→3')
Zif268	PGEKPYACPVESCDRRFRSDELTRHIRIHTGQKPFQC RICMRNFSRSDHLTTHIRHTHTGKPFACDICGRKFARS DERKRHTKIHT (SEQ ID NO. 26)	GCGTGGGCG GCG GGG GCG
PBSII	PGEKPYACPECGKSFSQRANLRAHQRTHTGKPYKC PECGKSFSRSDHLTTHQRTHTGKPYKCPECGKSFSR SDVLRHQRTHT (SEQ ID NO. 27)	GTGTGGAAA
ZFa	PGERPFQCRICMRNFSDSPTLRRHTRTHTGKPFQCRI CMRNFSVRHNLTRHLRHTHTGKPFQCRICMRNFSVRT SLARHLKTH (SEQ ID NO. 28)	GTTCGATGCC

ES 2 875 579 T3

Dedo de Zinc	Secuencia del dominio de unión de ADN	Secuencia de unión de ADN (5'→3')
ZFb	<p>PGERPFQCRICMRNFSSKDDHLHRHTRTHTGEKPFQCR</p> <p>ICMRNFSLSQTLKRHLRTHHTGEKPFQCRICMRNFSRL</p> <p>DMLARHLKTH (SEQ ID NO. 29)</p>	GCGGCTGGG
ZFc	<p>PGERPFQCRICMRNFSSPSKLRHTRTHTGEKPFQCRIC</p> <p>MRNFSGDGSLARHLRTHHTGEKPFQCRICMRNFSRVD</p> <p>NLPRHLKTH (SEQ ID NO. 30)</p>	GAGGACGGC
Tyr123	<p>EKPYKCPECGKSFSDRSNLTRHQRTHHTGEKPYKCPEC</p> <p>GKSFSSTSNLARHQRTHHTGEKPFKCPECGKSFSRSDA</p> <p>LTRHQRTHHT (SEQ ID NO. 31)</p>	GTGGATGAC
Tyr456	<p>EKPYKCPECGKSFSQSSNLARHQRTHHTGEKPYKCPEC</p> <p>GKSFSRSDHLTKHQRTHHTGEKPFKCPECGKSFSQSSN</p> <p>LARHQRTHHT (SEQ ID NO. 32)</p>	GAAGGGGAA
Blues	<p>ASDDRPYACPVESCDRRFSRRDVLNMNHIRIHTGQKPF</p> <p>QCRICMRNFSRSDHLTTHIRTHHTGEKPFACDICGRKFA</p> <p>NRDTRLRHSKIHLRQNDLE (SEQ ID NO. 33)</p>	GTTTGGATG
Jazz	<p>ASDDRPYACPVESCDRRFSRSDVLRHRIHTGQKPFQ</p> <p>CRICMRNFSRSDVLRHNRTHHTGEKPFACDICGRKFA</p> <p>SRDVLRRHNRIHLRQNDLE (SEQ ID NO. 34)</p>	GCTGCTGCG
Bagly	<p>EFMTGDRPYACPVESCDRRFSRSDVLRHRIHTGQKP</p> <p>FQCRICMRNFSRSDVLRHNRTHHTGEKPFACDICGRK</p> <p>FASRDVLRHNRHRIHLRQGRSHVCAECGKAFVSSKLK</p> <p>RHQLVHTGEKPFQLE (SEQ ID NO. 35)</p>	CGGGCTGCTGC (SEQ ID NO. 36)

Dedo de Zinc	Secuencia del dominio de unión de ADN	Secuencia de unión de ADN (5'→3')
Gli1	KREPESVYE TDCRWDGCSQEFDSSQEQLVHHINSEHIIH GERKEFVCHWGGCSRELRFKAQYMLVVHMRRHTG EKPHKCTFEGCRKSYSRLENLKTHLRSHTGEKPYMCE HEGCSKAFSNASDRAKHQNRTHSNEKPYVCKLPGCT KRYTDPSSLRKHVKTVHGPDAAHVTKRHRGD (SEQ ID NO. 37)	GACCACCCAAG ACGA (SEQ ID NO. 38)
HIVC	PFQCRICMRNFSRLRTDLDRHTRTHTGEKPFQCRICMR NFSLSQTLRRHLRTHHTGEKPFQCRICMRNFSRLSNLG RHLKTHHTGEK (SEQ ID NO. 39)	GATGCTGCA
B3	AQAALPEKPYACPECGKSFSDPGNLVRHQRTHTG EKPYKCPECGKSFSRSDKLVHRHQRTHTGEKPYKCPEC GKSFSQSSHLVRHQRTHTGKKTSGQAG (SEQ ID NO. 40)	GACGGGGG
N1	AQAALPEKPYACPECGKSFSQSSSLVRHQRTHTGE KPYKCPECGKSFSQSSNLVRHQRTHTGEKPYKCPEC KSFSRSDKLVHRHQRTHTGKKTSGQAG (SEQ ID NO. 41)	GTAGAAGGG
Sp-1	PGKKKQHICHIQCGKVGKTSHLRAHLRWHTGERP FMCTWSYCGKRFRSDELQRHKRTHHTGEKKFACPEC PKRFMRSDHLSKHIKTHQNKKG (SEQ ID NO. 42) PGKKKQHACPECGKSFSKSSHLRAHQRTHTGERPYK CPECGKSFSRSDDELQRHQRTHTGEKPYKCPECGKSFS RSDHLSKHQRTHTQNKKG (SEQ ID NO. 43)	GGGGCGGGG

Secuencia de armazón de ácido nucleico

5 En algunos ejemplos, la presente descripción enseña un armazón de ADN que comprende una o más de las secuencias de unión de ADN correspondientes a los dominios de unión de ADN contenidos dentro de las enzimas biosintéticas quiméricas. En algunos ejemplos, el armazón de ADN es un plásmido extracromosómico u otro vector. En otros ejemplos, el armazón de ADN está codificado dentro del genoma de una célula hospedante.

Los vectores de ácido nucleico adecuados incluyen, sin limitación, plásmidos, vectores de baculovirus, vectores de bacteriófagos, fagémidos, cósmidos, fósidos, cromosomas artificiales bacterianos, vectores virales (por ejemplo, vectores virales basados en virus vaccinia, poliovirus, adenovirus, virus adenoasociado, SV40, virus del herpes simple

y similares), cromosomas artificiales, plásmidos de levadura, cromosomas artificiales de levadura y otros vectores. En algunos ejemplos de la presente descripción, se prefieren los vectores adecuados para usar en células hospedantes procariontas. Por consiguiente, los vectores de ejemplo para usar en procariontas tales como *Escherichia coli* incluyen, pero no se limitan a pACYC184, pBeloBacll, pBR332, pBAD33, pBBR1MCS y sus derivados, pSC101, SuperCos (cosmid), pWE15 (cosmid), pTrc99A, pBAD24, vectores que contiene un origen de replicación ColE1 y sus derivados, pUC, pBluescript, pGEM, Ori_Plsmid27 (SEQ ID NO. 213), cadena principal del vector 1 (SEQ ID NO. 214), cadena principal del vector 2 (SEQ ID NO. 215), cadena principal del vector 3 (SEQ ID NO. 216), cadena principal del vector 4 (SEQ ID NO. 217) y vectores pTZ.

En algunos ejemplos, la presente descripción enseña que una subunidad de armazón de ácido nucleico puede comprender 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o más sitios de unión de ADN diferentes acoplados entre sí para facilitar la unión e inmovilización de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 o más proteínas de la ruta biosintética diferentes. En algunos ejemplos, la presente descripción enseña que los armazones de ADN tienen un único sitio de unión de ADN para cada proteína biosintética quimérica correspondiente.

En otros ejemplos, el armazón de ácido nucleico puede comprender dos o más copias del mismo sitio de unión de ADN. Esta arquitectura permite optimizar la estequiometría de la proteína biológica que se va a conseguir. De acuerdo con este ejemplo de la presente descripción, los mismos sitios de unión de ADN se pueden acoplar entre sí para crear centros enzimáticos para una reacción química particular. Por lo tanto, en algunos ejemplos, el armazón de ADN comprende grupos de múltiples sitios de unión de ADN, correspondiendo cada grupo a un gen/enzima biosintética quimérica específica.

En algunos ejemplos de la presente descripción, el método de ensamblar una ruta biológica sintética implica inmovilizar al menos un primer gen biosintético quimérico (p. ej., una enzima) y una segunda enzima biosintética quimérica sobre el armazón de ácido nucleico. La primera enzima biosintética quimérica produce un primer producto que es un sustrato para la segunda proteína quimérica de la ruta biológica. La segunda enzima biosintética quimérica se inmoviliza sobre la construcción de armazón de manera que se coloque adyacente o muy cerca de la primera enzima biosintética quimérica. De esta forma, la concentración eficaz del primer producto es alta y la segunda enzima biosintética quimérica puede actuar eficazmente sobre el primer producto. Como un ejemplo, un armazón de ácido nucleico sintético ha inmovilizado sobre el mismo, en el orden de 3'→ 5' o 5'→ 3' de la construcción de armazón a) la primera enzima biosintética quimérica, y b) la segunda enzima biosintética quimérica para formar una "subunidad de armazón". La subunidad de armazón se puede repetir dos o más veces dentro del armazón de ácido nucleico sintético.

De acuerdo con este y todos los ejemplos de la presente descripción, se pueden inmovilizar dos o más copias (p. ej., dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más moléculas) de cada enzima biosintética quimérica sobre una subunidad de armazón. Por ejemplo, en algunos ejemplos, una subunidad de armazón ha inmovilizado sobre la misma, a) una molécula (copia) de la primera enzima biosintética quimérica y b) una molécula de la segunda enzima biosintética quimérica. En otros ejemplos, una subunidad de armazón ha inmovilizado sobre la misma, a) una molécula de la primera enzima biosintética quimérica y b) dos o más moléculas (p. ej., dos, tres, cuatro, cinco, seis o más moléculas) de la segunda enzima biosintética quimérica. Por consiguiente, se puede variar la relación de cualquier proteína dada en una ruta biológica con respecto a cualquier otra proteína en la ruta. Solo a modo de ejemplo, la relación de una primera proteína de la ruta biológica quimérica a una segunda proteína de la ruta biológica quimérica puede variar de aproximadamente 0,1:10 hasta aproximadamente 10:0,1, por ejemplo, de aproximadamente 0,1:10 a aproximadamente 0,5:10, de aproximadamente 0,5:10 a aproximadamente 1,0:10, de aproximadamente 1,0:10 a aproximadamente 2:10, de aproximadamente 2:10 a aproximadamente 5:10, de aproximadamente 5:10 a aproximadamente 7:10, de aproximadamente 7:10 a aproximadamente 10:10, de aproximadamente 10:7 a aproximadamente 10:5, de aproximadamente 10:5 a aproximadamente 10:2, de aproximadamente 10:2 a aproximadamente 10:1, de aproximadamente 10:1 a aproximadamente 10:0,5, o de aproximadamente 10:0,5 a aproximadamente 10:1.

En algunos ejemplos, se inmovilizan al menos tres enzimas biosintéticas quiméricas sobre el armazón de ácido nucleico sintético para comprender una subunidad de armazón. De acuerdo con este ejemplo de la presente descripción, la primera enzima biosintética quimérica produce un primer producto que es un sustrato para la segunda enzima biosintética quimérica, y la segunda proteína quimérica de la ruta biológica produce un segundo producto que es un sustrato para la tercera enzima biosintética quimérica. En estos ejemplos, una subunidad de armazón ha inmovilizado sobre la misma, en el orden de 3'→ 5' o 5'→ 3' del armazón a) la primera enzima biosintética quimérica, b) la segunda enzima biosintética quimérica, y c) la tercera enzima biosintética enzima. La unidad de armazón se puede repetir dos o más veces en la construcción de ácido nucleico como se ha descrito antes.

En otro ejemplo de la presente descripción, se inmovilizan al menos cuatro enzimas biosintéticas quiméricas sobre el armazón de ácido nucleico. En otro ejemplo de la presente descripción, se inmovilizan al menos cinco enzimas biosintéticas quiméricas sobre el armazón de ácido nucleico. Será evidente a partir de estos ejemplos que se puede inmovilizar una sexta, séptima, octava, novena, décima, etc., enzima biosintética quimérica sobre el armazón de ácido nucleico, que las proteínas quiméricas están inmovilizadas espacialmente en el orden en que funcionan en un ruta, y que cada proteína se puede inmovilizar sobre el armazón en una dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más copias (o moléculas).

En algunos ejemplos, la presente descripción enseña el espaciado de cada sitio de unión de ADN dentro del ácido nucleico armazón. De acuerdo con este ejemplo de la presente descripción, los dos o más sitios de unión de ADN están ubicados adyacentes entre sí dentro de una subunidad de armazón, acoplados entre sí en tándem o separados por al menos un nucleótido espaciador. Los dos o más sitios de unión al ADN pueden estar separados entre sí por 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, o más nucleótidos espaciadores. El espacio entre diferentes sitios de unión de ADN puede variar dentro de una unidad de armazón (es decir, el espacio entre un primer y un segundo sitio de unión de ADN puede diferir del espacio entre el segundo y tercer sitio de unión de ADN). El espaciado óptimo entre diferentes sitios de unión de ADN dentro de una subunidad de armazón variará dependiendo de los requisitos de la enzima biosintética y la ruta biológica que se está reconstruyendo, y debe optimizarse para lograr una productividad óptima de la ruta biológica.

Armazón peptídico

En algunos ejemplos, los métodos de formación de armazón de la presente descripción también se pueden aplicar a armazones proteicos/estructurales dentro de la célula. En algunos ejemplos, la presente descripción enseña la aplicación de métodos descritos en la solicitud de patente publicada de EE.UU. nº 20110008829.

Proteínas quiméricas que se unen a proteínas

En algunos ejemplos, la presente descripción enseña proteínas quiméricas que comprenden enzimas biosintéticas seleccionadas que están ligadas a uno o más dominios de unión de proteínas capaces de unirse a un péptido de reclutamiento. De acuerdo con estos ejemplos, se espera que las enzimas biosintéticas quiméricas se recluten en un péptido armazón interactuando con péptidos de reclutamiento contenidos dentro del péptido armazón.

En algunos ejemplos, las enzimas biosintéticas y los dominios de unión de proteínas están unidos covalentemente. En algunos ejemplos, las enzimas biosintéticas se fusionan traduccionalmente a los dominios de unión de proteínas. Por lo tanto, en algunos ejemplos, las enzimas biosintéticas quiméricas se forman por acoplamiento del dominio de unión de proteínas al extremo amino, al extremo carboxi o a un sitio interno dentro de la proteína de la ruta biosintética. Los expertos en la técnica reconocerán la necesidad de asegurar que la adición del dominio de unión de proteínas no reduce sustancialmente la actividad de la enzima biosintética.

En algunos ejemplos de la presente descripción, la enzima biosintética se acopla a su dominio de unión de proteínas por una secuencia de conector polipeptídico corto como se describe en partes anteriores de esta descripción.

Se conocen en la técnica varios dominios de unión de proteínas (PBD) junto con sus correspondientes secuencias de péptidos de reclutamiento y son adecuados para usar en el sistema y los métodos de la presente descripción. A continuación se presenta una discusión ilustrativa no limitante de los PBD adecuados.

SH3

Los PBD adecuados incluyen dominios SH3. Los dominios SH3 incluyen dominios SH3 de Clase I; dominios SH3 de clase II; y dominios SH3 no convencionales. Las secuencias de aminoácidos de los dominios SH3 se conocen en la técnica. Véanse, por ejemplo, los aminoácidos 136-189 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso NP.sub.--058431 (proteína Crk de Homo sapiens); aminoácidos 136-189 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso AAH31149 (proteína Crk de Mus musculus); y aminoácidos 4-77 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso P27986 (subunidad p85 de Homo sapiens de fosfatidilinositol 3-quinasa).

En algunos ejemplos, un dominio SH3 es un dominio SH3 de Clase I y comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: EGYQYRA LYDYKKEREE DIDLHLGDIL TVNKGSLVAL GFSDGQEARP EEIGWLNQYN ETTGERGDFFP GTYVEYI (SEQ ID NO. 44), incluyendo todos los intervalos y subintervalos intermedios.

En algunos ejemplos, un dominio SH3 es un dominio SH3 de Clase II y comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: YVRALDFDFNGNDEEDLPFKKGDILRIRDKPQQWNAEDSEGKRGKGMIPVPYVEK (SEQ ID NO. 45). Como un ejemplo no limitante, un dominio SH3 comprende la secuencia de aminoácidos: MAEYVRALDFDFNGNDEEDLPFKKGDILRIRDKPQQWNAEDSEGKRGKGMIPVPYVEKY (SEQ ID NO. 46), incluyendo todos los intervalos y subintervalos intermedios.

Un dominio SH3 se une a péptidos ricos en prolina que forman una hélice de tipo II de poliprolina levógira, donde dichos péptidos comprenden la secuencia consenso mínima Pro-X-X-Pro. En algunos ejemplos, cada Pro está precedida por un resto alifático. En algunos ejemplos, un péptido de reclutamiento es un ligando del dominio SH3. Un dominio SH3 une péptidos ricos en prolina que forman una hélice de tipo II de poliprolina levógira, donde dichos

péptidos comprenden la secuencia consenso mínima Pro-X-X-Pro. En algunos ejemplos, cada Pro está precedida por un resto alifático. Por ejemplo, los ejemplos no limitantes de secuencias de aminoácidos de péptidos que comprenden ligandos del dominio SH3 incluyen: RPLPVAP (SEQ ID NO. 47; unido por un dominio SH3 de Clase I); PPPALPPKRRRPG (SEQ ID NO. 48); y PPPALPPKKR (SEQ ID NO. 49; unido por un dominio SH3 de Clase II).

5 PDZ

Los PBD adecuados incluyen dominios PDZ. Las secuencias de aminoácidos de dominios PDZ son conocidas en la técnica. Véase, por ejemplo, los aminoácidos 108-191, aminoácidos 201-287, y aminoácidos 354-434 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso AAC52113 (proteína de densidad post-sináptica 95 de *Homo sapiens*); y los aminoácidos 80-161 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso NP_033254 (sintrofina de *Mus musculus*).

10 En algunos ejemplos, un dominio PDZ adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: EITLERGNSGLGFSIAGGTDNPHIGDDPSIFIT
15 KIIPGGAAAQDGRRLRVNDSILFVNEVDVREVTTHSAAVEALKEAGSIVRLYV (SEQ ID NO. 50), incluyendo todos los intervalos y subintervalos intermedios.

20 En algunos ejemplos, un dominio PDZ adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: VMEIKLIKGPKGLGFSIAGGVGNQHIPGDN SIYVTKIIEGGAHKDGRQLQ
IGDKILAVNSVGLDVMHEDAVAALKNTYDVVYLKVA (SEQ ID NO.51), incluyendo todos los intervalos y subintervalos intermedios.

25 En algunos ejemplos, un dominio PDZ adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: RVIHRGSTGLGFNIVGGEDGEGIFISFILAGGPA
DLSGELRKGDQILSVNGVDLRNASHEQAAIALKNAGQVTIIAQ (SEQ ID NO. 52), incluyendo todos los intervalos y subintervalos intermedios.

30 En algunos ejemplos, un dominio PDZ adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: RRVTVRKADAGGLGISIKGGRENKMPILISK
IFKGLAADQTEALFVGDAILSVNGEDLSSATHDEAVQALKKTKGEWLEVK (SEQ ID NO. 53), incluyendo todos los intervalos y subintervalos intermedios. Por ejemplo, un dominio PDZ puede comprender la secuencia de aminoácidos
35 MLQRRRVTVRKADAGGLGISIKGGRENKMPILISKIFKGLAADQTEALFVGDAILSVNGE DLSS
ATHDEAVQALKKTKGEVLEVVKYMKEVSPYFKGS (SEQ ID NO. 54).

40 En algunos ejemplos, un péptido de reclutamiento es un ligando del dominio PDZ. Un dominio PDZ se une a los restos 4-5 C-terminales de las proteínas diana. En algunos ejemplos, un ligando del dominio PDZ consenso comprende un resto hidrófobo, p. ej., Val o Ile, en el extremo carboxi. Por ejemplo, ejemplos no limitantes de secuencias de aminoácidos de péptidos que comprenden ligandos del dominio PDZ incluyen: IESDV (SEQ ID NO. 55); VKESLV (SEQ ID NO. 56); GVKESLV (SEQ ID NO. 57); GVKQSL (SEQ ID NO. 58); GVKESGA (SEQ ID NO. 59); YVKESLV (SEQ ID NO. 60); y VETDV (SEQ ID NO. 61).

GBD

45 Los PBD adecuados incluyen dominios de unión de GTPasa (GBD), también denominados en la técnica motivos CRIB (Cdc42/Rac-unión interactiva) motifs. En algunos ejemplos, un GBD se une a una GTPasa pequeña tipo Cdc42p y/o tipo Rho. Las secuencias de aminoácidos of GBD son conocidas en la técnica. Véase, p. ej., los aminoácidos 198-240 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso NP.sub.--001103835 (proteína similar al síndrome de Wiskott-Aldrich de *Rattus norvegicus* (WASP)); los aminoácidos 69-112 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso Q13177 (PAK-2 de *Homo sapiens*); y los aminoácidos
50 70-105 de la secuencia de aminoácidos proporcionada en GenBank con el nº de acceso P35465 (PAK-1 de *Rattus norvegicus*). Véanse también las secuencias de aminoácidos PAK (75-111), ACK (504-549) y WASP (232-274), presentadas en la FIG. 3A de Garrard et al. (2003) *EMBO J.* 22:1125. Véanse también las secuencias de aminoácidos ACK (505-531), WASP (236-258), PAK1 (70-94), PAK2 (71-91), PAK-4 (6-30), presentadas en la FIG. 1A de Bishop y Hall (2000) *Biochem. J.* 348:241.

55 En algunos ejemplos, un GBD adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la

secuencia de aminoácidos: ADI GTPSNFQHIG HVGWDPNTGF DLNNDPELK NLFDMCGISE (SEQ ID NO. 62), y todos los intervalos y subintervalos intermedios.

5 En algunos ejemplos, un GBD adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: KERPEISLPDFEHTIHVGFDAVTGEFTGMPEQWAR (SEQ ID NO. 63), y todos los intervalos y subintervalos intermedios.

10 En algunos ejemplos, un GBD adecuado comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: MTKADIGTPSNFQHIGHVWDPNTGFDLNNLDPELKNLFDLDMCGISEAQLKDRETSKVIY DFIEK TGGVEAVKNELRRQAP (SEQ ID NO. 64), y todos los intervalos y subintervalos intermedios.

En algunos ejemplos, un péptido de reclutamiento es un ligando de GBD. Un ligando de ejemplo no limitante de GBD comprende la secuencia de aminoácidos LVGALMHVMQKRSRAIHSSDEGEDQAGDEDED (SEQ ID NO. 65).

15 Péptidos de la cremallera de leucina

Los PBD adecuados incluyen péptidos de la cremallera de leucina. En algunos ejemplos, los péptidos de la cremallera de leucina son péptidos que interactúan por un dominio superenrollado. Las secuencias de aminoácidos de los dominios de la cremallera de leucina son conocidos en la técnica. Los péptidos de la cremallera de leucina incluyen un péptido de cremallera de leucina EE12RR345L; un péptido de cremallera de leucina RR12EE354L; y similares.

20 Un ejemplo de una secuencia de aminoácidos de un péptido de cremallera de leucina es una cremallera de leucina EE12RR345L de la secuencia de aminoácidos: LEIEAAFLERENTALETRVAELRQRVQRLR NRVSQYRTRYGPLGGGK (SEQ ID NO. 66).

25 En algunos ejemplos, un péptido de cremallera de leucina comprende una secuencia de aminoácidos que tiene al menos aproximadamente 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o 100% de similitud de secuencia de aminoácidos respecto a la secuencia de aminoácidos: LEIEAA FLERENTALETRVAELRQRVQRLRNRVSQYRTRYGPLGGGK (SEQ ID NO. 67), y todos los intervalos y subintervalos intermedios. Dicho péptido de cremallera de leucina puede servir como un PBD o como un péptido de reclutamiento.

30 Otro ejemplo no limitante de una secuencia de aminoácidos de un péptido de cremallera de leucina es un péptido de cremallera de leucina RR12EE345L de la secuencia de aminoácidos: LEIRAFLRQRNTALRT EVAELEQEVQRLNEVVSQYETRYGPLGGGK (SEQ ID NO. 68).

35 Las descripciones anteriores han descrito la producción de proteínas biosintéticas quiméricas que comprenden un dominio de unión de proteínas diseñado para dirigirse (unirse a) uno o más péptidos de reclutamiento ubicados en un polipéptido almacén. Los expertos en la técnica reconocerán otras variaciones compatibles de esta disposición. Por ejemplo, en algunos ejemplos, la presente descripción enseña la producción de proteínas biosintéticas quiméricas que comprenden péptidos de reclutamiento dirigidos por dominios de unión de proteínas ubicados en un polipéptido almacén. En otros ejemplos, el dominio de unión de proteínas y los péptidos de reclutamiento se incorporan cada uno en dos o más proteínas biosintéticas quiméricas, de manera que las proteínas quiméricas forman complejos (p. ej., dímeros o heterodímeros). Un ejemplo ilustrativo de formación de complejos es el uso de dominios de cremallera de leucina compatibles colocados en proteínas biosintéticas quiméricas, de modo que dos o más proteínas biosintéticas quiméricas pueden formar un complejo a través de los dominios de cremallera de leucina.

Polipéptido almacén

45 En algunos ejemplos, la presente descripción enseña un polipéptido almacén que organiza las enzimas de la ruta biosintética en un complejo funcional. En algunos ejemplos, los polipéptidos almacén de la presente descripción comprenden dos o más péptidos de reclutamiento. Es decir, en algunos ejemplos, los polipéptidos almacén de la presente descripción son capaces de reclutar dos proteínas biosintéticas quiméricas más.

50 En algunos ejemplos, el polipéptido almacén de la presente descripción es un péptido exógeno introducido en una célula hospedante (p. ej., por la transformación de una secuencia de ADN que codifica el polipéptido almacén o la introducción directa del péptido). En otros ejemplos, el polipéptido almacén es una estructura natural dentro de la célula hospedante. Es decir, en algunos ejemplos, el polipéptido almacén es un orgánulo o membrana (p. ej., el retículo endoplásmico o el aparato de Golgi). Por lo tanto, en algunos ejemplos, el polipéptido almacén de la presente descripción incluye estructuras de células hospedantes compuestas por más de una secuencia de péptidos.

55 En algunos ejemplos, las secuencias de péptidos de reclutamiento dentro del polipéptido almacén se organizan para así optimizar las rutas biosintéticas diana cuyas enzimas se están reclutando. En algunos ejemplos, la organización del polipéptido almacén es similar a la descrita antes para almacenos de ADN. Por lo tanto, en algunos ejemplos, los

polipéptidos armazón contienen agrupaciones de péptidos de reclutamiento con el fin de regular el orden y las proporciones de diversas proteínas biosintéticas químéricas.

Cultivo celular y fermentación

5 Las células de la presente descripción se pueden cultivar en medios nutritivos convencionales modificados según sea adecuado para cualquier reacción o selección biosintética deseada. En algunos ejemplos, la presente descripción enseña el cultivo en medios inductores para activar promotores. En algunos ejemplos, la presente descripción enseña medios con agentes de selección, incluyendo agentes de selección de transformantes (p. ej., antibióticos), o selección de organismos adecuados para crecer en condiciones de inhibición (p. ej., condiciones de alto contenido de etanol).
10 En algunos ejemplos, la presente descripción enseña el cultivo de cultivos celulares en medios optimizados para el crecimiento celular. En otros ejemplos, la presente descripción enseña el cultivo de cultivos celulares en medios optimizados para el rendimiento del producto. En algunos ejemplos, la presente descripción enseña cultivos en crecimiento en medios capaces de inducir el crecimiento celular y también contiene los precursores necesarios para la producción del producto final (p. ej., altos niveles de azúcares para la producción de etanol).

15 Las condiciones de cultivo, tales como temperatura, pH y similares, son las adecuadas para usar con la célula hospedante seleccionada para la expresión, y resultarán evidentes para los expertos en la técnica. Como se ha indicado, se encuentran disponibles muchas referencias para el cultivo y producción de muchas células, incluyendo células de origen bacteriano, vegetal, animal (incluyendo mamífero) y arqueobacteriano. Véase, p. ej., Sambrook, Ausubel (véase todo antes), así como Berger, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Inc., San Diego, CA; y Freshney (1994) Culture of Animal Cells, a Manual of Basic
20 Technique, tercera edición, Wiley-Liss, New York y las referencias citadas en el mismo; Doyle y Griffiths (1997) Mammalian Cell Culture: Essential Techniques John Wiley and Sons, NY; Humason (1979) Animal Tissue Techniques, cuarta edición W.H. Freeman y Company; y Ricciardelle et al., (1989) In Vitro Cell Dev. Biol. 25:1016-1024. Para el cultivo y regeneración de células vegetales, Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, N.Y.; Gamborg y Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental
25 Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg N.Y.); Jones, ed. (1984) Plant Gene Transfer and Expression Protocols, Humana Press, Totowa, N.J. y Plant Molecular Biology (1993) R. R. D. Croy, Ed. Bios Scientific Publishers, Oxford, U.K. ISBN 0 12 198370 6. Los medios de cultivo celular se exponen en general en Atlas and Parks (eds.) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, Fla.. Se encuentra información adicional para el cultivo celular en la bibliografía disponible en el mercado tal como el Life Science Research Cell
30 Culture Catalogue de Sigma-Aldrich, Inc (St Louis, Mo.) ("Sigma-LSRCCC") y, por ejemplo, The Plant Culture Catalogue and supplement también de Sigma-Aldrich, Inc (St Louis, Mo.) ("Sigma-PCCS").

El medio de cultivo para usar debe usarse de una forma adecuada que cumpla los requisitos de las respectivas cepas. Se presentan descripciones de medios de cultivo de diversos microorganismos en el "Manual of Methods for General Bacteriology" de la American Society for Bacteriology (Washington D.C., EE.UU., 1981).

35 La presente descripción además proporciona un procedimiento para la preparación fermentativa de un producto de interés, que comprende las etapas de: a) cultivar un microorganismo según la presente descripción en un medio adecuado, que da como resultado un caldo de fermentación; y b) concentrar el producto de interés en el caldo de fermentación de a) y/o en las células del microorganismo.

40 En algunos ejemplos, la presente descripción enseña que los microorganismos producidos se pueden cultivar de forma continua, como se describe, por ejemplo, en el documento WO 05/021772, o de forma discontinua en un procedimiento discontinuo (cultivo discontinuo) o en un procedimiento discontinuo alimentado o discontinuo alimentado repetido con el fin de producir el compuesto químico orgánico deseado. Un resumen de carácter general sobre los métodos de cultivo conocidos está disponible en el libro de texto de Chmiel (Bioprozeßtechnik. 1: Einführung in die Bioverfahrenstechnik (Gustav Fischer Verlag, Stuttgart, 1991)) o en el libro de texto de Storhas (Bioreaktoren and
45 periphäre Einrichtungen (Vieweg Verlag, Braunschweig/Wiesbaden, 1994)).

En algunos ejemplos, las células de la presente descripción se cultivan en condiciones de fermentación continua o discontinua.

50 La fermentación discontinua clásica es un sistema cerrado, en donde las composiciones del medio se establecen al comienzo de la fermentación y no están sujetas a alternancias artificiales durante la fermentación. Una variación del sistema discontinuo es una fermentación discontinua alimentada que también tiene uso en la presente descripción. En esta variación, el sustrato se añade en incrementos a medida que avanza la fermentación. Los sistemas discontinuos alimentados son útiles cuando es probable que la represión de catabolitos inhiba el metabolismo de las células y cuando es deseable tener cantidades limitadas de sustrato en el medio. Las fermentaciones discontinuas y discontinuas alimentadas son comunes y bien conocidas en la técnica.

55 La fermentación continua es un sistema donde se añade continuamente un medio de fermentación definido a un biorreactor y se retira una cantidad igual de medio acondicionado simultáneamente para procesar y recolectar los productos biomoléculas de interés deseados. En algunos ejemplos, la fermentación continua generalmente mantiene los cultivos a una densidad alta constante donde las células están principalmente en fase de crecimiento logarítmico.

En algunos ejemplos, la fermentación continua generalmente mantiene los cultivos en una fase de crecimiento estacionaria o logarítmica/estacionaria tardía. Los sistemas de fermentación continua se esfuerzan por mantener condiciones de crecimiento en estado estacionario.

5 Los métodos para modular los nutrientes y los factores de crecimiento para los procedimientos de fermentación continua, así como las técnicas para maximizar la velocidad de formación del producto, son bien conocidos en la técnica de la microbiología industrial.

10 Por ejemplo, una lista no limitante de fuentes de carbono para los cultivos de la presente descripción incluye azúcares y carbohidratos tales como, por ejemplo, glucosa, sacarosa, lactosa, fructosa, maltosa, melazas, soluciones que contienen sacarosa del procesamiento de la remolacha azucarera o caña de azúcar, almidón, hidrolizado de almidón y celulosa; aceites y grasas tales como, por ejemplo, aceite de soja, aceite de girasol, aceite de cacahuete y grasa de coco; ácidos grasos tales como, por ejemplo, ácido palmítico, ácido esteárico y ácido linoleico; alcoholes tales como, por ejemplo, glicerol, metanol y etanol; y ácidos orgánicos tales como, por ejemplo, ácido acético o ácido láctico.

15 Una lista no limitante de las fuentes de nitrógeno para los cultivos de la presente descripción incluye compuestos orgánicos que contienen nitrógeno tales como peptonas, extracto de levadura, extracto de carne, extracto de malta, licor de maceración de maíz, harina de soja y urea; o compuestos inorgánicos tales como sulfato de amonio, cloruro de amonio, fosfato de amonio, carbonato de amonio y nitrato de amonio. Las fuentes de nitrógeno se pueden usar individualmente o como una mezcla.

20 Una lista no limitante de las posibles fuentes de fósforo para los cultivos de la presente descripción incluye ácido fosfórico, dihidrogenofosfato de potasio o hidrogenofosfato de dipotasio o las correspondientes sales que contienen sodio.

El medio de cultivo puede comprender adicionalmente sales, por ejemplo en forma de cloruros o sulfatos de metales tales como, por ejemplo, sodio, potasio, magnesio, calcio y hierro, tales como por ejemplo sulfato de magnesio o sulfato de hierro, que son necesarios para el crecimiento.

25 Finalmente, se pueden usar factores de crecimiento esenciales tales como aminoácidos, por ejemplo homoserina y vitaminas, por ejemplo tiamina, biotina o ácido pantoténico, además de las sustancias antes mencionadas.

En algunos ejemplos, el pH del cultivo se puede controlar con cualquier ácido o base, o sal tampón, incluyendo, pero no limitado a hidróxido de sodio, hidróxido de potasio, amoniaco o amoniaco acuoso; o compuestos ácidos tales como ácido fosfórico o ácido sulfúrico de una manera adecuada. En algunos ejemplos, el pH se ajusta generalmente a un valor de 6,0 a 8,5, preferiblemente de 6,5 a 8.

30 En algunos ejemplos, los cultivos de la presente descripción pueden incluir un agente antiespumante tal como, por ejemplo, ésteres poliglicólicos de ácidos grasos. En algunos ejemplos, los cultivos de la presente descripción se modifican para estabilizar los plásmidos de los cultivos añadiendo sustancias selectivas adecuadas tales como, por ejemplo, antibióticos.

35 En algunos ejemplos, el cultivo se realiza en condiciones aeróbicas. Con el fin de mantener estas condiciones, se introducen en el cultivo oxígeno o mezclas de gases que contienen oxígeno tales como, por ejemplo, aire. Asimismo, se pueden usar líquidos enriquecidos con peróxido de hidrógeno. La fermentación se lleva a cabo, cuando sea adecuado, a presión elevada, por ejemplo a una presión elevada de 0,03 a 0,2 MPa. La temperatura del cultivo es normalmente de 20°C a 45°C y preferiblemente de 25°C a 40°C, en particular preferiblemente de 30°C a 37°C. En los procedimientos discontinuos o discontinuos alimentados, el cultivo preferiblemente se continúa hasta que se haya formado una cantidad del producto deseado de interés (p. ej., un compuesto químico orgánico) suficiente para ser recuperado. Este objetivo normalmente se puede lograr en un plazo de 10 horas a 160 horas. En procedimientos continuos, son posibles tiempos de cultivo más largos. La actividad de los microorganismos da como resultado una concentración (acumulación) del producto de interés en el medio de fermentación y/o en las células de dichos microorganismos.

45 En algunos ejemplos, el cultivo se lleva a cabo en condiciones anaeróbicas.

Cribado

En algunos ejemplos, la presente descripción enseña cribados iniciales de alta capacidad. En otros ejemplos, la presente descripción también enseña validaciones robustas basadas en tanques de datos de rendimiento (véase la Figura 6B).

50 En algunos ejemplos, el procedimiento de cribado de alta capacidad está diseñado para predecir el rendimiento de las cepas en biorreactores. Como se ha descrito previamente, las condiciones de cultivo se seleccionan para que sean adecuadas para el organismo y reflejen las condiciones del biorreactor. Las colonias individuales se recogen y se transfieren a placas de 96 pocillos y se incuban durante un periodo de tiempo adecuado. Posteriormente, las células se transfieren a nuevas placas de 96 pocillos para cultivos de siembra adicionales o para cultivos de producción. Los cultivos se incuban durante periodos de tiempo variables, donde que se pueden realizar múltiples mediciones. Estas

55

pueden incluir mediciones de producto, biomasa u otras características que predicen el rendimiento de las cepas en biorreactores. Los resultados de los cultivos de alta capacidad se usan para predecir el rendimiento del biorreactor.

5 En algunos ejemplos, la validación del rendimiento basada en tanques se usa para confirmar el rendimiento de las cepas aisladas por cribado de alta capacidad. Las cepas candidatas se criban usando reactores de fermentación a escala de laboratorio (p. ej., los reactores descritos en la Tabla 5 de la presente descripción) según características relevantes de rendimiento de la cepa tales como productividad o rendimiento.

Recuperación y cuantificación de productos

10 Los expertos en la técnica conocen métodos de cribado para la producción de productos de interés y se describen a lo largo de la presente memoria descriptiva. Dichos métodos se pueden emplear cuando se criban las cepas de la descripción.

15 En algunos ejemplos, la presente descripción enseña métodos para mejorar cepas diseñadas para producir productos intracelulares no secretados. Por ejemplo, la presente descripción enseña métodos para mejorar la solidez, el rendimiento, la eficiencia o la conveniencia general de los cultivos celulares que producen enzimas intracelulares, aceites, productos farmacéuticos u otras moléculas pequeñas o péptidos valiosos. La recuperación o aislamiento de productos intracelulares no secretados se puede lograr mediante técnicas de lisis y recuperación que son bien conocidas en la técnica, incluidas las descritas en el presente documento.

20 Por ejemplo, en algunos ejemplos, las células de la presente descripción se pueden recolectar por centrifugación, filtración, sedimentación u otro método. A continuación, las células recolectadas se rompen mediante cualquier método conveniente, que incluyen ciclos de congelación-descongelación, sonicación, rotura mecánica o el uso de agentes de lisis celular u otros métodos, que son bien conocidos por los expertos en la técnica.

25 El producto de interés resultante, p. ej. un polipéptido, se puede recuperar/aislar y opcionalmente purificar por cualquiera de una serie de métodos conocidos en la técnica. Por ejemplo, un producto polipéptido se puede aislar del medio nutriente por procedimientos convencionales que incluyen, pero no se limitan a: centrifugación, filtración, extracción, secado por pulverización, evaporación, cromatografía (p. ej., intercambio iónico, afinidad, interacción hidrofóbica, cromatoenfoco y exclusión por tamaños) o precipitación. Finalmente, se puede usar cromatografía líquida de alto rendimiento (HPLC) en las etapas finales de purificación. (Véase, por ejemplo, Purification of intracellular protein as described in Parry et al., 2001, *Biochem. J.* 353:117, y Hong et al., 2007, *Appl. Microbiol. Biotechnol.* 73:1331).

30 Además de las referencias indicadas antes, se conocen bien en la técnica una variedad de métodos de purificación, que incluyen, por ejemplo, los expuestos en: Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; Bollag et al. (1996) *Protein Methods*, 2ª Edición, Wiley-Liss, NY; Walker (1996) *The Protein Protocols Handbook Humana Press*, NJ; Harris y Angal (1990) *Protein Purification Applications: A Practical Approach*, IRL Press at Oxford, Oxford, Inglaterra; Harris y Angal *Protein Purification Methods: A Practical Approach*, IRL Press at Oxford, Oxford, Inglaterra; Scopes (1993) *Protein Purification: Principles and Practice* 3ª Edición, Springer Verlag, NY; Janson y Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications*, Segunda Edición, Wiley-VCH, NY; y Walker (1998) *Protein Protocols on CD-ROM*, Humana Press, NJ.

35 En algunos ejemplos, la presente descripción enseña los métodos para mejorar las cepas diseñadas para producir productos secretados. Por ejemplo, la presente descripción enseña métodos para mejorar la solidez, rendimiento, eficiencia o la conveniencia general de los cultivos celulares que producen moléculas pequeñas o péptidos valiosos.

40 En algunos ejemplos, se pueden usar métodos inmunológicos para detectar y/o purificar productos secretados o no secretados producidos por las células de la presente descripción. En un enfoque de ejemplo, el anticuerpo generado contra una molécula de producto (p. ej., contra un polipéptido de insulina o un fragmento inmunogénico del mismo) usando métodos convencionales se inmoviliza sobre perlas, se mezcla con medio de cultivo celular en condiciones en las que se une la endoglucanasa y se hace precipitar. En algunos ejemplos, la presente descripción enseña el uso de ensayos de inmunoabsorción ligados a enzimas (ELISA).

45 En otros ejemplos relacionados, se usa inmunocromatografía, como se describe en la patente de EE.UU. N° 5.591.645, patente de EE.UU. N° 4.855.240, patente de EE.UU. N° 4.435.504, patente de EE.UU. N° 4.980.298, y Se-Hwan Paek, et al., "Development of rapid One-Step Immunochromatographic assay, Methods", 22, 53-60, 2000). Una inmunocromatografía general detecta una muestra mediante el uso de dos anticuerpos. Existe un primer anticuerpo en una solución de ensayo o en una parte en un extremo de una pieza de ensayo en una forma aproximadamente rectangular hecha de una membrana porosa, donde se deja caer la solución de ensayo. Este anticuerpo está marcado con partículas de látex o partículas coloidales de oro (en lo sucesivo este anticuerpo se denominará como anticuerpo marcado). Cuando la solución de ensayo vertida incluye una muestra que se va a detectar, el anticuerpo marcado reconoce la muestra para unirse con la muestra. Un complejo de la muestra y el anticuerpo marcado fluye por capilaridad hacia un absorbente, que está hecho de un papel de filtro y unido a un extremo opuesto al extremo que tiene incluido el anticuerpo marcado. Durante el flujo, el complejo de la muestra y el anticuerpo marcado es reconocido y capturado por un segundo anticuerpo (en lo sucesivo, se denominará anticuerpo de golpeteo) que existe en el medio

de la membrana porosa y, como resultado de esto, aparece el complejo en una parte de detección de la membrana porosa como señal visible y se detecta.

5 En algunos ejemplos, los métodos de cribado de la presente descripción se basan en técnicas de detección fotométrica (absorción, fluorescencia). Por ejemplo, en algunos ejemplos, la detección se puede basar en la presencia de un detector de fluoróforo como GFP unida a un anticuerpo. En otros ejemplos, la detección fotométrica se puede basar en la acumulación del producto deseado del cultivo celular. En algunos ejemplos, el producto puede ser detectable por UV del cultivo o extractos de dicho cultivo.

10 Los expertos en la técnica reconocerán que los métodos de la presente descripción son compatibles con las células hospedantes que producen cualquier producto biomolécula deseable de interés. La Tabla 2 a continuación presenta una lista no limitante de las categorías de productos, biomoléculas y células hospedante, incluidas dentro del alcance de la presente descripción. Estos ejemplos se proporcionan con fines ilustrativos y no pretenden limitar la aplicabilidad de la tecnología descrita actualmente de ninguna manera.

Tabla 2. - Una lista no limitante de células hospedantes y productos de interés de la presente descripción.

Categoría de producto	Productos	Categoría de hospedante	Hospedantes
Aminoácidos	Lisina	Bacteria	<i>Escherichia coli</i>
Aminoácidos	Metionina	Bacteria	<i>Escherichia coli</i>
Aminoácidos	MSG	Bacteria	<i>Escherichia coli</i>
Aminoácidos	Treonina	Bacteria	<i>Escherichia coli</i>
Aminoácidos	Treonina	Bacteria	<i>Escherichia coli</i>
Aminoácidos	Triptófano	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Madera de agar	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Ambrox	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Nootkatona	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Aceite de pachuli	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Azafrán	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Aceite de sándalo	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Valenceno	Bacteria	<i>Escherichia coli</i>
Sabor y Fragancia	Vanillina	Bacteria	<i>Escherichia coli</i>
Alimento	CoQ10/Ubiquinol	Bacteria	<i>Escherichia coli</i>
Alimento	Ácidos grasos Omega 3	Bacteria	<i>Escherichia coli</i>
Alimento	Ácidos grasos Omega 6	Bacteria	<i>Escherichia coli</i>
Alimento	Vitamina B12	Bacteria	<i>Escherichia coli</i>
Alimento	Vitamina B2	Bacteria	<i>Escherichia coli</i>
Alimento	Vitamina B2	Bacteria	<i>Escherichia coli</i>
Alimento	Eritritol	Bacteria	<i>Escherichia coli</i>
Alimento	Eritritol	Bacteria	<i>Escherichia coli</i>
Alimento	Eritritol	Bacteria	<i>Escherichia coli</i>
Alimento	Glicósidos de esteviol	Bacteria	<i>Escherichia coli</i>

Categoría de producto	Productos	Categoría de hospedante	Hospedantes
Hidrocoloides	Goma diutan	Bacteria	<i>Escherichia coli</i>
Hidrocoloides	Goma gellan	Bacteria	<i>Escherichia coli</i>
Hidrocoloides	Goma xantana	Bacteria	<i>Escherichia coli</i>
Compuestos intermedios	1,3-PDO	Bacteria	<i>Escherichia coli</i>
Compuestos intermedios	1,4-BDO	Bacteria	<i>Escherichia coli</i>
Compuestos intermedios	Butadieno	Bacteria	<i>Escherichia coli</i>
Compuestos intermedios	n-butanol	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido cítrico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido cítrico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido glucónico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido itacónico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido láctico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	Ácido láctico	Bacteria	<i>Escherichia coli</i>
Ácidos orgánicos	LCDA - DDDA	Bacteria	<i>Escherichia coli</i>
Policétidos/Ag	Spinosad	Bacteria	<i>Escherichia coli</i>
Policétidos/Ag	Spinetoram	Bacteria	<i>Escherichia coli</i>

Genes heterólogos de interés

En un ejemplo, se proporcionan en el presente documento métodos para expresar genes heterólogos en una célula hospedante microbiana. El gen heterogéneo se puede introducir en la célula hospedante microbiana usando los métodos proporcionados en el presente documento y/o conocidos en la técnica, de manera que la célula hospedante microbiana use el gen heterólogo para producir un producto de interés. En un ejemplo, la célula hospedante microbiana es una cepa de *E. coli*. La cepa de *E. coli* puede ser cualquier cepa de *E. coli* conocida en la técnica y/o proporcionada en el presente documento. El gen heterólogo puede ser una versión de tipo natural de dicho gen o un mutante del mismo. El gen heterólogo se puede unir operativamente a un promotor, terminador, marcador de solubilidad de proteínas, marcador de degradación de proteínas o cualquier combinación de los mismos. La unión operativa del gen heterólogo al promotor, terminador, marcador de solubilidad de proteínas o marcador de degradación de proteínas se puede lograr usando los métodos de intercambio de promotores, intercambio de terminadores, intercambio de marcadores de solubilidad y/o intercambio de degradación proporcionados a lo largo de esta descripción.

En un ejemplo, el gen heterólogo se une operativamente a un promotor seleccionado de la Tabla 1. En un ejemplo, el gen heterólogo se une operativamente a una secuencia de promotor sintético quimérico de 60-90 pb, en donde el promotor quimérico sintético consiste en una parte distal del promotor p_R del fago lambda, regiones -35 y -10 variables de los promotores p_L y p_R del fago lambda, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda o una parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli*. Las secuencias de ácido nucleico de la parte distal del promotor p_R del fago lambda, regiones -35 y -10 variables de los promotores p_L y p_R del fago lambda, porciones centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda o una parte de 5' UTR/sitio de unión ribosómico (RBS) del promotor del gen *acs* de *E. coli* para usar en el promotor quimérico sintético se puede seleccionar de las secuencias de ácido nucleico que se encuentran en la Tabla 1.5. En un ejemplo, el gen heterólogo se puede unir operativamente a un promotor quimérico sintético que tiene una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 132-207 que se encuentran en la Tabla 1.4.

En un ejemplo, el gen heterólogo está operativamente unido a un terminador seleccionado de la tabla 1.2. En otro ejemplo, el gen heterólogo está operativamente unido a una secuencia de terminador seleccionada de la tabla 19.

En un ejemplo, el gen heterólogo está operativamente unido a un marcador de solubilidad seleccionado de la tabla 17.

En un ejemplo, el gen heterólogo está operativamente unido a una secuencia de marcador de degradación seleccionada de la tabla 18.

Además de los ejemplos anteriores, el gen heterólogo puede ser cualquiera de los genes requeridos para generar los productos de interés encontrados en la tabla 2 o cualquier gen conocido en la técnica que pueda ser expresado como un gen heterólogo en la célula hospedante microbiana (p. ej., *E. coli*) para producir un producto de interés.

En un ejemplo, el gen heterólogo es un gen que es parte de la ruta biosintética de la lisina como se ilustra en la figura 19. Además de este ejemplo, el gen heterólogo se puede seleccionar del gen *asd*, el gen *ask*, el gen *hom*, el gen *dapA*, el gen *dapB*, el gen *dapD*, el gen *ddh*, el gen *argD*, el gen *dapE*, el gen *dapF*, el gen *lysA*, el gen *lysE*, el gen *zwf*, el gen *pgi*, el gen *ktk*, el gen *fbp*, el gen *ppc*, el gen *pck*, el gen *ddx*, el gen *pyc* o el gen *icd*. En un ejemplo, el gen heterólogo que es parte de la ruta de la lisina como se proporciona en el presente documento está operativamente unido a un promotor sintético quimérico con una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 132-207.

En un ejemplo, el gen heterólogo es un gen que es parte de la ruta biosintética del licopeno como se ilustra, por ejemplo, en la figura 59. Además de este ejemplo, el gen heterólogo se puede seleccionar del gen *dxs*, el gen *ispC*, el gen *ispE*, el gen *ispD*, el gen *ispF*, el gen *ispG*, el gen *ispH*, el gen *idi*, el gen *ispA*, el gen *ispB*, el gen *crtE*, el gen *crtB*, el gen *crtI*, el gen *crtY*, el gen *ymgA*, el gen *dxr*, el gen *elbA*, el gen *gdhA*, el gen *appY*, el gen *elbB* gene, o el gen *ymgB*. En un ejemplo, un gen heterólogo que es parte de la ruta del licopeno como se proporciona en el presente documento está operativamente unido a un promotor sintético quimérico con una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 132-207.

En un ejemplo, el gen heterólogo es un gen que codifica un producto biofarmacéutico o un gen en la ruta para generar un producto biofarmacéutico. En un ejemplo, la célula hospedante microbiana es *E. coli* y el producto biofarmacéutico es cualquier producto biofarmacéutico que se haya mostrado que es producido en *E. coli*. El producto biofarmacéutico se puede seleccionar entre humulina (insulina rh), intrónA (interferón alfa2b), roferon (interferón alfa2a), humatrope (hormona de crecimiento somatropina rh), neupogen (filgrastim), detaferon (interferón beta-lb), lispro (insulina de acción rápida), Rapilysin (reteplasa), infergen (interferón alfacon-1), glucagón, beromun (tasonermina), ontak (denileucina diftitox), lantus (insulina glargina de acción prolongada), kineret (anakinra), natreacor (nesiritida), somavert (pegvisomant), calcitonina (calcitonina recombinante de salmón), lucentis (ranibizumab), preotact (hormona paratiroidea humana), kyrstexal (urato oxidasa rh, PEGilada), nivestim (filgrastim, rhGCSF), voraxaze (glucarpidasa) u preos (hormona paratiroidea). En un ejemplo, un gen heterólogo que codifica un producto biofarmacéutico o un gen en una ruta que genera un producto biofarmacéutico como se proporciona en el presente documento está operativamente unido a un promotor quimérico sintético con una secuencia de ácido nucleico seleccionada de las SEQ ID NO. 132-207.

Criterios de selección y objetivos

Los criterios de selección aplicados a los métodos de la presente descripción variarán con los objetivos específicos del programa de mejora de cepas. La presente descripción puede adaptarse para cumplir cualquier objetivo del programa. Por ejemplo, en algunos ejemplos, el objetivo del programa puede ser maximizar los rendimientos de lotes individuales de reacciones sin límites de tiempo inmediatos. En otros ejemplos, el objetivo del programa puede ser reequilibrar los rendimientos biosintéticos para producir un producto específico o para producir una relación particular de productos. En otros ejemplos, el objetivo del programa puede ser modificar la estructura química de un producto, tal como alargar la cadena de carbonos de un polímero. En algunos ejemplos, el objetivo del programa puede ser mejorar las características de rendimiento tales como el rendimiento, título, productividad, eliminación de subproductos, tolerancia a las variaciones del proceso, temperatura de crecimiento óptima y tasa de crecimiento. En algunos ejemplos, el objetivo del programa es mejorar el rendimiento del hospedante medido por la productividad volumétrica, productividad específica, rendimiento o título de un producto de interés producido por un microbio.

En otros ejemplos, el objetivo del programa puede ser optimizar la eficiencia de síntesis de una cepa comercial en términos de rendimiento del producto final por cantidad de entrada (p. ej., cantidad total de etanol producido por libra de sacarosa). En otros ejemplos, el objetivo del programa puede ser optimizar la velocidad de síntesis, medida, por ejemplo, en términos de tasas de finalización de lotes o tasas de rendimiento en sistemas de cultivo continuo. En otros ejemplos, el objetivo del programa puede ser aumentar la resistencia de la cepa a un fago particular, o aumentar de otro modo el vigor/robustez de la cepa en condiciones de cultivo.

En algunos ejemplos, los proyectos de mejora de cepas pueden estar sujetos a más de un objetivo. En algunos ejemplos, el objetivo del proyecto de cepas puede girar en torno a la calidad, fiabilidad o la rentabilidad general. En algunos ejemplos, la presente descripción enseña métodos de mutaciones o grupos de mutaciones seleccionadas asociadas con una o más de las propiedades de cepas descritas anteriormente.

Las personas que tengan conocimientos ordinarios en la técnica reconocerán cómo adaptar los criterios de selección de cepas para cumplir el objetivo particular del proyecto. Por ejemplo, las selecciones del rendimiento máximo de un lote individual de una cepa en la saturación de la reacción pueden ser adecuadas para identificar cepas con altos

rendimientos de lotes individuales. La selección basada en la consistencia del rendimiento en un intervalo de temperaturas y condiciones puede ser adecuada para identificar cepas con mayor robustez y fiabilidad.

5 En algunos ejemplos, los criterios de selección para la fase inicial de alta capacidad y la validación basada en tanque serán idénticos. En otros ejemplos, la selección basada en tanques puede operar con criterios de selección adicionales y/o diferentes. Por ejemplo, en algunos ejemplos, la selección de cepas de alta capacidad se podría basar en los rendimientos de finalización de la reacción de lotes individuales, mientras que la selección basada en el tanque puede ampliarse para incluir selecciones basadas en los rendimientos de la velocidad de reacción.

Secuenciación

10 En algunos ejemplos, la presente descripción enseña la secuenciación del genoma completo de los organismos descritos en el presente documento. En otros ejemplos, la presente descripción también enseña la secuenciación de plásmidos, productos de PCR y otros oligonucleótidos como controles de calidad para los métodos de la presente descripción. Los expertos en la técnica conocen bien los métodos de secuenciación para proyectos grandes y pequeños.

15 En algunos ejemplos, se puede usar cualquier técnica de alta capacidad para secuenciar ácidos nucleicos en los métodos de la descripción. En algunos ejemplos, la presente descripción enseña la secuenciación del genoma completo. En otros ejemplos, la presente descripción enseña secuenciación de amplicones secuenciación ultra profunda para identificar variaciones genéticas. En algunos ejemplos, la presente descripción también enseña métodos novedosos para la preparación de bibliotecas, que incluyen el marcaje (véase el documento WO/2016/073690). Las técnicas de secuenciación de ADN incluyen reacciones clásicas de secuenciación de dideoxi (método de Sanger) utilizando terminadores o cebadores marcados y separación en gel en soporte plano o capilar; secuenciación por síntesis usando nucleótidos marcados terminados de forma reversible, pirosecuenciación; secuenciación 454; hibridación específica de alelos con una biblioteca de sondas de oligonucleótidos marcadas; secuenciación por síntesis usando hibridación específica de alelos con una biblioteca de clones marcados que va seguida de ligación; monitorización en tiempo real de la incorporación de nucleótidos marcados durante una etapa de polimerización; 25 secuenciación de polony; y secuenciación SOLiD.

En un ejemplo de la descripción, se emplean métodos de secuenciación de alta capacidad que comprenden una etapa de aislamiento espacial de moléculas individuales en una superficie sólida donde se secuencian en paralelo. Dichas superficies sólidas pueden incluir superficies no porosas (tales como en la secuenciación de Solexa, p. ej. Bentley et al, *Nature*, 456: 53-59 (2008) o secuenciación de Complete Genomics, p. ej. Drmanac et al, *Science*, 327: 78-81 (2010)), matrices de pocillos, que pueden incluir moldes unidos a perlas o partículas (tal como con 454, p. ej. Margulies et al, *Nature*, 437: 376-380 (2005) o secuenciación de Ion Torrent, publicación de patente de EE.UU. 2010/0137143 o 2010/0304982), membranas micromecanizadas (tales como con secuenciación SMRT, p. ej. Eid et al, *Science*, 323: 133-138 (2009)), o matrices de perlas (tal como con secuenciación SOLiD o secuenciación polony, p. ej. Kim et al., *Science*, 316: 1481-1414 (2007)).

35 En otro ejemplo, los métodos de la presente descripción comprenden amplificar las moléculas aisladas antes o después de que se aislen espacialmente en una superficie sólida. La amplificación previa puede comprender amplificación basada en emulsión, tal como PCR en emulsión o amplificación por círculo rodante. También se enseña la secuenciación basada en Solexa donde las moléculas molde individuales se aíslan espacialmente sobre una superficie sólida, después de lo cual se amplifican en paralelo por PCR puente para formar poblaciones clonales separadas, o agrupamientos, y luego se secuencian, como se describe en Bentley et al. (citado antes) y en las instrucciones del fabricante (p. ej. Kit de preparación de muestras TruSeq™ y hoja de datos, Illumina, Inc., San Diego, California, 2010); y además en las siguientes referencias: patentes de EE.UU. N° 6.090.592; 6.300.070; 7.115.400; y EP0972081B1.

45 En un ejemplo, las moléculas individuales dispuestas y amplificadas sobre una superficie sólida forman agrupamientos en una densidad de al menos 10^5 agrupamientos por cm^2 ; o en una densidad de al menos 5×10^5 por cm^2 ; o en una densidad de al menos 10^6 agrupamientos por cm^2 . En un ejemplo, se emplean químicas de secuenciación que tienen tasas de error relativamente altas. En dichos ejemplos, las puntuaciones de calidad promedio producidas por dichas químicas son funciones decrecientes monótonas de longitudes de lecturas de secuencias. En un ejemplo, dicha disminución corresponde a que el 0,5 por ciento de las lecturas de secuencias tienen al menos un error en las posiciones 1-75; 1 por ciento de las lecturas de secuencias tiene al menos un error en las posiciones 76-100; y 2 por ciento de las lecturas de secuencias tienen al menos un error en las posiciones 101-125.

Análisis computacional y predicción de los efectos de los criterios de diseño genético en todo el genoma

55 En algunos ejemplos, la presente descripción enseña métodos para predecir los efectos de alteraciones genéticas particulares que se han incorporado en una cepa hospedante determinada. En otros ejemplos, la descripción proporciona métodos para generar alteraciones genéticas propuestas que deberían incorporarse en una cepa hospedante determinada, con el fin de que dicho hospedante posea un rasgo fenotípico particular o parámetro de cepa. En ejemplos dados, la descripción proporciona modelos predictivos que se pueden usar para diseñar nuevas cepas hospedantes.

En algunos ejemplos, la presente descripción enseña métodos para analizar los resultados de rendimiento de cada ronda de cribado y métodos para generar nuevas modificaciones de secuencia de todo el genoma propuestas que se prevé que mejoren el rendimiento de la cepa en la siguiente ronda de cribado.

5 En algunos ejemplos, la presente descripción enseña que el sistema genera modificaciones de secuencia propuestas para las cepas hospedantes basándose en resultados de cribado previos. En algunos ejemplos, las recomendaciones del presente sistema se basan en los resultados del cribado inmediatamente anterior. En otros ejemplos, las recomendaciones del presente sistema se basan en los resultados acumulados de uno o más de los cribados precedentes.

10 En algunos ejemplos, las recomendaciones del presente sistema se basan en bibliotecas de diseño genético de HTP desarrolladas previamente. Por ejemplo, en algunos ejemplos, el presente sistema está diseñado para guardar resultados de cribados previos y aplicar esos resultados a un proyecto diferente, en el mismo organismo hospedante o en uno diferente.

15 En otros ejemplos, las recomendaciones del presente sistema se basan en conocimientos científicos. Por ejemplo, en algunos ejemplos, las recomendaciones se basan en propiedades conocidas de genes (de fuentes tales como bases de datos de genes anotados y la bibliografía relevante), optimización de codones, deslizamiento transcripcional, uORF u otras optimizaciones de secuencias y hospedantes basadas en hipótesis.

20 En algunos ejemplos, las modificaciones de secuencia propuestas para una cepa hospedante recomendadas por el sistema, o el modelo predictivo, se llevan a cabo mediante el uso de uno o más de los conjuntos de herramientas moleculares descritos que comprenden: (1) intercambios de promotores, (2) intercambios de SNP, (3) Intercambios de codones de inicio/detención, (4) optimización de secuencia, (5) intercambios de detención, (6) intercambios de marcadores de solubilidad, (7) intercambios de marcadores de degradación y (8) mapeo de epistasias.

25 La plataforma de ingeniería genética de HTP descrita en el presente documento es agnóstica con respecto a cualquier rasgo fenotípico o microbio particular (p. ej., producción de un compuesto particular). Es decir, la plataforma y las lecciones enseñadas en el presente documento se pueden usar con cualquier célula hospedante para modificar dicha célula hospedante para que tenga cualquier rasgo fenotípico deseado. Además, las lecciones aprendidas de un procedimiento de ingeniería genética de HTP dado usado para crear una nueva célula hospedante, se pueden aplicar a cualquier número de otras células hospedantes, como resultado del almacenamiento, caracterización y análisis de una gran cantidad de parámetros de procedimiento que ocurren durante los métodos enseñados.

30 Como se mencionó en la sección de mapeo epistático, es posible estimar el rendimiento (también conocido como puntuación) de una cepa hipotética obtenida consolidando una colección de mutaciones de una biblioteca de diseño genético de HTP en un contexto particular a través de algún modelo predictivo preferido. Dado dicho modelo predictivo, es posible puntuar y clasificar todas las cepas hipotéticas accesibles a la biblioteca de mutaciones a través de la consolidación combinatoria.

La siguiente sección describe los modelos particulares usados en la presente plataforma de HTP.

35 **Diseño predictivo de cepas**

En el presente documento se describe un enfoque para el diseño predictivo de cepas, que incluye: métodos para describir los cambios genéticos y el rendimiento de la cepa, predecir el rendimiento de la cepa basado en la composición de los cambios en la cepa, recomendar diseños candidatos con un rendimiento previsto alto y filtrar las predicciones para optimizar por consideraciones de segundo orden, p. ej. similitud con cepas existentes, epistasias o confianza en las predicciones.

Entradas al modelo de diseño de cepas

45 En un ejemplo, para facilitar la ilustración, los datos de entrada pueden comprender dos componentes: (1) conjuntos de cambios genéticos y (2) rendimiento relativo de la cepa. Los expertos en la técnica reconocerán que este modelo se puede extender fácilmente para considerar una amplia variedad de entradas, mientras se tiene presente la consideración compensatoria del sobreajuste. Además de los cambios genéticos, algunos de los parámetros de entrada (variables independientes) que se pueden ajustar son los tipos de células (género, especie, cepa, caracterización filogenética, etc.) y los parámetros del procedimiento (p. ej., condiciones ambientales, equipo de manipulación, técnicas de modificación, etc.) con los cuales se lleva a cabo la fermentación con las células.

50 Los conjuntos de cambios genéticos pueden provenir de las colecciones discutidas previamente de perturbaciones genéticas denominadas bibliotecas de diseño genético HTP. El rendimiento relativo de la cepa se puede evaluar basándose en cualquier parámetro o rasgo fenotípico de interés dado (p. ej., producción de un compuesto, molécula pequeña o producto de interés).

55 Los tipos de células se pueden especificar en categorías generales como sistemas procariontes y eucariotes, género, especie, cepa, cultivos de tejidos (frente a células dispersas), etc. Los parámetros del procedimiento que se pueden ajustar incluyen temperatura, presión, configuración del reactor y composición del medio. Los ejemplos de

configuración del reactor incluyen el volumen del reactor, si el procedimiento es discontinuo o continuo y, si es continuo, el caudal volumétrico, etc. También se puede especificar la estructura de soporte, si hay alguna, en la que residen las células. Los ejemplos de composición del medio incluyen las concentraciones de electrolitos, nutrientes, productos de desecho, ácidos, pH y similares.

5 **Conjuntos de cambios genéticos de bibliotecas de diseño genético de HTP seleccionados para usar en el modelo de regresión lineal inicial que posteriormente se usa para crear el modelo de diseño de cepas predictivo**

10 Se muestra a continuación un ejemplo, un conjunto de entradas de una tabla de cambios genéticos en *Corynebacterium*, en la Tabla 3. Cada fila indica un cambio genético en la cepa 7000051473, así como metadatos sobre el mecanismo de cambio, p. ej. intercambio de promotores o intercambio de SNP. aceE, zwf y pyc están todos relacionados con el ciclo del ácido cítrico.

15 En este caso, la cepa 7000051473 tiene un total de 7 cambios. "Último cambio" significa que el cambio en esta cepa representa la modificación más reciente en este linaje de cepas. Por lo tanto, la comparación del rendimiento de esta cepa con el rendimiento de la parental representa un punto de datos relativo al rendimiento de la mutación del "último cambio".

Tabla 3 - Tabla de entrada de diseño de cepas para la cepa 7000051473

cepa	nombre	biblioteca	cambio	de	a	último cambio
7000051473	dlc19_42	proswp	pcg3121	cg1144	pcg3121_cg1144	1
7000051473	dlc19_42	scswp	acee atg>ttg	ttg	acee_atg	0
7000051473	dlc19_42	snpswp	dss_033	NA	na	0
7000051473	dlc19_42	snpswp	dss_084	NA	t	0
7000051473	dlc19_42	snpswp	dss_316	NA	na	0
7000051473	dlc19_42	proswp	pcg0007_39	zwf	pcg0007_39_zwf	0
7000051473	dlc19_42	proswp	pcg1860	pyc	pcg1860_pyc	0

Evaluación del rendimiento de la cepa construida

20 El objetivo del modelo enseñado es predecir el rendimiento de la cepa en función de la composición de los cambios genéticos introducidos en la cepa. Para construir una referencia para la comparación, el rendimiento de la cepa se calcula con respecto a una cepa de referencia común, calculando primero el rendimiento mediano por cepa, por placa de ensayo. Después se calcula el rendimiento relativo como la diferencia en el rendimiento medio entre una cepa diseñada y la cepa de referencia común dentro de la misma placa. La restricción de los cálculos a comparaciones dentro de la placa asegura que todas las muestras consideradas recibían las mismas condiciones experimentales.

25 La Figura 23 muestra la distribución de los rendimientos relativos de la cepa de *Corynebacterium* para los datos de entrada en consideración. Un rendimiento relativo de cero indica que la cepa diseñada funcionaba igual de bien que la cepa de "referencia" o base en placa. Es de interés la capacidad del modelo predictivo para identificar las cepas que es probable que funcionen significativamente por encima de cero. Además, y de manera más general, es de interés si alguna cepa determinada tiene mayor rendimiento que la progenitora por algún criterio. En la práctica, el criterio puede ser un título de producto que cumpla o supere algún umbral por encima del nivel parental, aunque también se podría usar en su lugar o además el tener una diferencia estadísticamente significativa con la progenitora en la dirección deseada. El papel de la cepa base o de "referencia" es simplemente servir como un factor de normalización añadido para hacer comparaciones dentro o entre placas.

35 Un concepto a tener en cuenta es el de las diferencias entre: cepa parental y cepa de referencia. La cepa parental es el contexto que se usó para una ronda actual de mutagénesis. La cepa de referencia es una cepa de control llevada en cada placa para facilitar las comparaciones, especialmente entre placas, y es típicamente la "cepa base" como se ha mencionado antes. Pero puesto que la cepa base (p. ej., la cepa de tipo natural o industrial que se usa para medir el rendimiento general) no es necesariamente una "base" en el sentido de ser un objetivo de mutagénesis en una ronda determinada de mejora de la cepa, un término más descriptivo es "cepa de referencia".

40 En resumen, una cepa base/de referencia se usa para medir el rendimiento de las cepas construidas, en general, mientras que la cepa parental se usa para medir el rendimiento de un cambio genético específico en el contexto genético relevante.

Clasificación del rendimiento de las cepas construidas con regresión lineal

El objetivo del modelo descrito es clasificar el rendimiento de las cepas construidas, describiendo el rendimiento relativo de la cepa, en función de la composición de los cambios genéticos introducidos en las cepas construidas. Como se ha discutido a lo largo de la descripción, las diversas bibliotecas de diseño genético de HTP proporcionan el repertorio de posibles cambios genéticos (p. ej., perturbaciones/alteraciones genéticas) que se introducen en las cepas transformadas. La regresión lineal es la base del modelo predictivo de ejemplo descrito actualmente.

La siguiente tabla (es decir, la Tabla 4) contiene entrada de ejemplo para la modelización basada en regresión. Los rendimientos de la cepa se clasifican en relación con una cepa de base común, en función de la composición de los cambios genéticos contenidos en la cepa.

Cada encabezado de columna representa un cambio genético, un "1" representa la presencia del cambio, mientras que un "0" representa la ausencia de un cambio. "DSS" se refiere a intercambios de SNP de una biblioteca particular (las primeras 3 columnas después de *relativ_perf*). Las 3 últimas columnas son intercambios de promotores, donde *pcgXXXX* indica el promotor particular y las 3 últimas letras representan el gen al que se está aplicando el promotor. Los genes están relacionados con el metabolismo central. Los promotores son de *Corynebacterium glutamicum* (de ahí la notación "cg"). Se puede encontrar más información sobre los promotores usados en la Tabla 1, que enumera los promotores P1-P8 y la lista de secuencias de la presente solicitud. Además, se puede encontrar información detallada sobre cada promotor P1-P8 en la solicitud provisional de EE.UU. N° 62/264.232, presentada el 7 de diciembre de 2015, y titulada "Promotores de *Corynebacterium glutamicum*". Para facilitar la referencia, en la siguiente tabla, *pcg3121* = P8; *pcg0755* = P4; y *pcg1860* = P3.

Tabla 4 - Resumen de cambios genéticos y su efecto sobre el rendimiento relativo.

<i>relativo_perf</i>	<i>dss_033</i>	<i>dss_034</i>	<i>dss_056</i>	<i>pcg3121_pgi</i>	<i>pcg0755_zwf</i>	<i>pcg1860_pyc</i>
0,1358908	0	0	0	0	0	1
-1,8946985	1	0	0	1	0	1
-0,0222045	0	0	0	1	0	0
0,6342183	1	0	1	0	0	0
-0,0803285	1	1	0	0	0	0
2,6468117	0	0	0	1	0	0

Regresión lineal para caracterizar cepas construidas

La regresión lineal es un método atractivo para la plataforma de ingeniería genómica HTP descrita, debido a la facilidad de implementación e interpretación. Los coeficientes de regresión resultantes se pueden interpretar como el aumento o disminución promedio en el rendimiento relativo de la cepa atribuible a la presencia de cada cambio genético.

Por ejemplo, como se ve en la Figura 24, esta técnica permite concluir que cambiar el promotor *pgi* a *pcg3121* mejora el rendimiento relativo de la cepa en aproximadamente 5 unidades en promedio y, por lo tanto, es un cambio potencialmente muy deseable, en ausencia de cualquier interacción epistática negativa (nota: la entrada es un valor normalizado sin unidades).

Por lo tanto, el método enseñado usa modelos de regresión lineal para describir/caracterizar y clasificar cepas construidas, que tienen diversas perturbaciones genéticas introducidas en sus genomas a partir de las diversas bibliotecas enseñadas.

Modelización de diseño predictivo

El modelo de regresión lineal descrito antes, que usaba datos de cepas construidas, se puede usar para hacer predicciones de rendimiento para cepas que todavía no se han construido.

El procedimiento se puede resumir como sigue: se generan *in silico* todas las configuraciones posibles de cambios genéticos → se usa el modelo de regresión para predecir el rendimiento relativo de la cepa → se ordenan los diseños de cepas candidatas por rendimiento. Por lo tanto, usando el modelo de regresión para predecir el rendimiento de las cepas todavía no construidas, el método permite la producción de cepas de mayor rendimiento y, al mismo tiempo, llevar a cabo menos experimentos.

Generación de configuraciones

5 Cuando se construye un modelo para predecir el rendimiento de cepas todavía no construidas, la primera etapa es producir una secuencia de candidatos de diseño. Esto se hace fijando el número total de cambios genéticos en la cepa y después definiendo todas las posibles combinaciones de cambios genéticos. Por ejemplo, se puede establecer el número total de potenciales cambios/perturbaciones genéticas en 29 (p. ej., 29 posibles SNP, o 29 promotores diferentes, o cualquier combinación de los mismos siempre que el universo de perturbaciones genéticas sea 29) y luego decidir diseñar todos los posibles combinaciones de 3 miembros de los 29 cambios genéticos potenciales, que darán como resultado 3.654 diseños de cepas candidatas.

10 Para proporcionar el entorno a las 3.654 cepas candidatas mencionadas anteriormente, considere que se puede calcular el número de agrupaciones no redundantes de tamaño r a partir de n miembros posibles usando $n! / ((n - r)! * r!)$. Si $r = 3$, $n = 29$ da 3654. Por lo tanto, si se diseñan todas las combinaciones posibles de 3 miembros de 29 cambios potenciales, el resultado es 3.654 cepas candidatas. Los 29 cambios genéticos potenciales están presentes en el eje x de la Figura 25.

Predicción del rendimiento de los nuevos diseños de cepas

15 Usando la regresión lineal construida antes con las configuraciones combinatorias como entrada, se puede entonces predecir el rendimiento relativo esperado de cada diseño candidato. La Figura 25 resume la composición de los cambios para los 100 diseños superiores de cepas previstos para *Corynebacterium*. El eje x da el conjunto de potenciales cambios genéticos (29 cambios genéticos posibles) y el eje y muestra el orden jerárquico. Las celdas negras indican la presencia de un cambio particular en el diseño candidato, mientras que las celdas blancas indican la ausencia de ese cambio. En este ejemplo en particular, todos los 100 diseños superiores contienen los cambios `pcg3121_pgi`, `pcg1860_pyc`, `dss_339` y `pcg0007_39_lysa`. Además, el diseño del candidato superior contiene los cambios `dss_034`, `dss_009`.

20 La precisión predictiva debería aumentar con el tiempo a medida que se usan nuevas observaciones para reentrenar y reajustar iterativamente el modelo. Los resultados de un estudio de los autores de la invención ilustran los métodos mediante los cuales el modelo predictivo se puede volver a entrenar y mejorar de forma iterativa. La Figura 46 compara las predicciones del modelo con los valores de medición observados. La calidad de las predicciones del modelo se puede evaluar a través de varios métodos, incluyendo un coeficiente de correlación que indica la fuerza de la asociación entre los valores predichos y observados, o el error cuadrático medio, que es una medición del error promedio del modelo. Usando un indicador elegido para la evaluación del modelo, el sistema puede definir reglas sobre cuándo debe ser reentrenado el modelo.

25 Un par de suposiciones no expuestas del modelo anterior incluyen: (1) no hay interacciones epistáticas; y (2) los cambios/perturbaciones genéticas utilizadas para construir el modelo predictivo (p. ej., a partir de datos de cepas construidas como se ilustra en la Figura 24, o cualquier conjunto de datos que se use como referencia para construir el modelo) se hicieron todos en el mismo contexto de *Corynebacterium*, como las combinaciones propuestas de cambios genéticos (p. ej., como se ilustra en la Figura 25).

Filtrado por características de segundo orden

30 El ejemplo ilustrativo anterior se centró en predicciones de regresión lineal basadas en el rendimiento previsto de la célula hospedante. En algunos ejemplos, los presentes métodos de regresión lineal también se pueden aplicar a factores que no son biomoléculas, tales como biomasa de saturación, resistencia u otras características medibles de la célula hospedante. Por lo tanto, los métodos de la presente descripción también enseñan a considerar otras características fuera del rendimiento previsto cuando se priorizan los candidatos a construir. Suponiendo que hay datos relevantes adicionales, los términos no lineales también se incluyen en el modelo de regresión.

Cercanía con las cepas existentes

Las cepas previstas que son similares a las que ya se han construido podrían dar como resultado ahorro de tiempo y coste a pesar de no ser un candidato previsto superior

Diversidad de cambios

45 Cuando se construyen los modelos mencionados antes, no se puede estar seguro de que los cambios genéticos sean realmente aditivos (como se asume por regresión lineal y se menciona como una suposición más arriba) debido a la presencia de interacciones epistáticas. Por lo tanto, el conocimiento de la disimilitud del cambio genético se puede usar para aumentar la probabilidad de aditividad positiva. Si se sabe, por ejemplo, que los cambios `dss_034` y `dss_009` (que son intercambios de SNP) de la cepa clasificada superior antes están en la misma ruta metabólica y tienen características de rendimiento similares, entonces esa información se podría usar para seleccionar otra cepa de clasificación superior con una composición diferente de cambios. Como se describe en la sección anterior en relación con el mapeo de la epistasis, los mejores cambios genéticos previstos se pueden filtrar para restringir la selección a mutaciones con perfiles de respuesta suficientemente diferentes. Alternativamente, la regresión lineal puede ser una regresión de mínimos cuadrados ponderados usando la matriz de similitud para ponderar las predicciones.

Diversidad del rendimiento previsto

Finalmente, se puede optar por diseñar cepas con un rendimiento previsto medio o pobre, con el fin de validar y posteriormente mejorar los modelos predictivos.

Optimización del diseño de cepas iterativo

5 Como se describe en el ejemplo anterior, todos los 100 diseños superiores de cepas contienen los cambios pcg3121_pgi, pcg1860_pyc, dss_339 y pcg0007_39_lysa. Además, el diseño de la cepa candidata superior contiene los cambios dss_034, dss_009.

10 En ejemplos, el motor de colocación de pedidos 208 pone un pedido de fábrica a la fábrica 210 para fabricar cepas microbianas que incorporan las mutaciones candidatas superiores. En forma de bucle de retroalimentación, los resultados pueden ser analizados por el equipo de análisis 214 para determinar qué microbios presentan propiedades fenotípicas deseadas (314). Durante la fase de análisis, se evalúan los cultivos de cepas modificadas para determinar su rendimiento, es decir, su expresión de las propiedades fenotípicas deseadas, que incluyen la capacidad para producir a escala industrial. Por ejemplo, la fase de análisis usa, entre otras cosas, datos de imágenes de placas para medir el crecimiento de colonias microbianas como indicador de la salud de las colonias. El equipo de análisis 214 se usa para correlacionar los cambios genéticos con el rendimiento fenotípico y guardar los datos de correlación genotipo-fenotipo resultantes en bibliotecas, que se pueden almacenar en la biblioteca 206, para comunicar la producción microbiana futura.

20 En particular, los cambios candidatos que realmente dan como resultado un rendimiento medido suficientemente alto, se pueden añadir como filas en la base de datos a tablas como la Tabla 4 anterior. De esta manera, las mutaciones de mejor rendimiento se añaden al modelo de diseño de cepas predictivo en una forma de aprendizaje automático supervisado.

25 LIMS itera el ciclo de diseño/construcción/ensayo/análisis basado en las correlaciones desarrolladas a partir de los experimentos de fábrica anteriores. Durante un ciclo subsiguiente, el equipo de análisis 214 solo, o en conjunto con operadores humanos, puede seleccionar los mejores candidatos como cepas base para la entrada de nuevo en la interfaz de entrada 202, usando los datos de correlación para ajustar las modificaciones genéticas para lograr un mejor rendimiento fenotípico con granularidad más fina. De esta manera, el sistema de gestión de la información del laboratorio de ejemplos de la descripción implementa un bucle de retroalimentación de mejora de la calidad.

En resumen, con referencia al diagrama de flujo de la Figura 33, el flujo de trabajo de diseño de cepas predictivo iterativo se puede describir de la siguiente manera:

- 30 - Se genera un conjunto de entrenamiento de variables de entrada y salida, p. ej., cambios genéticos como entradas y características de rendimiento como salidas (3302). La generación se puede realizar mediante el equipo de análisis 214 basándose en cambios genéticos previos y el correspondiente rendimiento medido de las cepas microbianas que incorporan esos cambios genéticos.
- Se desarrolla un modelo inicial (p. ej., modelo de regresión lineal) basado en el conjunto de entrenamiento (3304). Esto puede ser realizado por el equipo de análisis 214.
- 35 - Se generan cepas candidatas del diseño (3306)
 - 40 o En un ejemplo, el equipo de análisis 214 puede fijar el número de cambios genéticos que se van a hacer en el contexto de una cepa, en forma de combinaciones de cambios. Para representar estos cambios, el equipo de análisis 214 puede proporcionar al intérprete 204 una o más expresiones de especificación de ADN que representan esas combinaciones de cambios. (Estos cambios genéticos o las cepas microbianas que incorporan esos cambios se pueden denominar "entradas de ensayo"). El intérprete 204 interpreta la una o más especificaciones de ADN, y el motor de ejecución 207 ejecuta las especificaciones de ADN para completar la especificación de ADN con salidas resueltas que representan las cepas de diseño candidatas individuales para esos cambios.
- 45 - Basándose en el modelo, el equipo de análisis 214 predice el rendimiento esperado de cada cepa de diseño candidata (3308).
- El equipo de análisis 214 selecciona un número limitado de diseños candidatos, p. ej., 100, con el rendimiento previsto más alto (3310).
 - 50 o Como se describe en otra parte del presente documento con respecto al mapeo de epistasis, el equipo de análisis 214 puede tener en cuenta efectos de segundo orden tales como epistasis, p. ej., filtrando diseños superiores para efectos epistáticos o factorizando epistasis en el modelo predictivo.
- Se construyen las cepas candidatas filtradas (en la fábrica 210) según el pedido de fábrica generado por el motor de colocación de pedidos 208 (3312).
- El equipo de análisis 214 mide el rendimiento real de las cepas seleccionadas, selecciona un número limitado de esas cepas seleccionadas basándose en su rendimiento real superior (3314) y añade los cambios de diseño

y su rendimiento resultante para modelo predictivo (3316). En el ejemplo de regresión lineal, se añaden los conjuntos de cambios de diseño y su rendimiento asociado como nuevas filas en la Tabla 4.

- 5 - El equipo de análisis 214 después itera de nuevo a la generación de nuevas cepas candidatas de diseño (3306) y continúa iterando hasta que se satisface una condición de parada. La condición de parada puede comprender, por ejemplo, el rendimiento medido de al menos una cepa microbiana que satisfaga un indicador de rendimiento, tal como rendimiento, velocidad de crecimiento o título.

En el ejemplo anterior, la optimización iterativa del diseño de cepas usa la retroalimentación y regresión lineal para implementar el aprendizaje automático. En general, el aprendizaje automático se puede describir como la optimización de los criterios de rendimiento, p. ej., parámetros, técnicas u otras características, en el rendimiento de una tarea informativa (tal como clasificación o regresión) usando un número limitado de ejemplos de datos marcados, y luego realizando la misma tarea en datos desconocidos. En el aprendizaje automático supervisado, tal como el del ejemplo de regresión lineal anterior, la máquina (p. ej., un dispositivo informático) aprende, por ejemplo, identificando patrones, categorías, relaciones estadísticas u otros atributos, presentados por los datos de entrenamiento. El resultado del aprendizaje después se usa para predecir si los nuevos datos presentarán los mismos patrones, categorías, relaciones estadísticas u otros atributos.

Los ejemplos de la descripción pueden emplear otras técnicas de aprendizaje automático supervisado cuando los datos de entrenamiento están disponibles. En ausencia de datos de entrenamiento, los ejemplos pueden emplear aprendizaje automático sin supervisión. Alternativamente, los ejemplos pueden emplear aprendizaje automático semi-supervisado, usando una pequeña cantidad de datos marcados y una gran cantidad de datos no marcados. Los ejemplos también pueden emplear una selección de características para seleccionar el subconjunto de las características más relevantes para optimizar el rendimiento del modelo de aprendizaje automático. Dependiendo del tipo de enfoque de aprendizaje automático seleccionado, como alternativas o además de la regresión lineal, los ejemplos pueden emplear, por ejemplo, regresión logística, redes neuronales, máquinas de vectores de soporte (SVM), árboles de decisión, modelos de Markov ocultos, redes bayesianas, Gram Schmidt, aprendizaje basado en refuerzo, aprendizaje basado en agrupamientos que incluye agrupamiento jerárquico, algoritmos genéticos y cualquier otra máquina de aprendizaje adecuada conocida en la técnica. En particular, los ejemplos pueden emplear regresión logística para proporcionar probabilidades de clasificación (p. ej., clasificación de genes en diferentes grupos funcionales) junto con las propias clasificaciones. Véase, p. ej., "A simple and efficient algorithm for gene selection using sparse logistic regression", *Bioinformatics*, Vol. 19, No. 17 2003, pp. 2246-2253, Leng, et al., "Classification using functional data analysis for temporal gene expression data", *Bioinformatics*, Vol. 22, No. 1, Oxford University Press (2006), pág. 68-76.

Los ejemplos pueden usar arquitecturas aceleradas de unidad de procesamiento de gráficos (GPU) que han encontrado una popularidad creciente en la realización de tareas de aprendizaje automático, en particular en la forma conocida como redes neuronales profundas (DNN). Los ejemplos de la descripción pueden emplear el aprendizaje automático basado en GPU, tal como el que se describe en "GPU-Based Deep Learning Inference: A Performance and Power Analysis", NVidia Whitepaper, Noviembre de 2015, Dahl, et al., "Multi-task Neural Networks for QSAR Predictions", Dept. of Computer Science, Univ. de Toronto, Junio de 2014 (arXiv:1406.1231 [stat.ML]). También se pueden encontrar técnicas de aprendizaje automático aplicables a ejemplos de la descripción en, entre otras referencias, Libbrecht, et al., "Machine learning applications in genetics and genomics", *Nature Reviews: Genetics*, Vol. 16, Junio de 2015, Kashyap, et al., "Big Data Analytics in Bioinformatics: A Machine Learning Perspective", *Journal of Latex Class Files*, Vol. 13, No. 9, Sept. 2014, Prompramote, et al., "Machine Learning in Bioinformatics", Capítulo 5 de *Bioinformatics Technologies*, pág. 117-153, Springer Berlin Heidelberg 2005.

Diseño de cepas iterativo predictivo: Ejemplo

5 A continuación se proporciona una aplicación de ejemplo del flujo de trabajo del diseño de cepas predictivo iterativo descrito antes.

Se preparó un conjunto inicial de variables de entrada y salida de entrenamiento. Este conjunto comprendía 1864 cepas transformadas únicas con una composición genética definida. Cada cepa contenía entre 5 y 15 cambios de ingeniería. Estaban presentes un total de 336 cambios genéticos únicos en el entrenamiento.

50 Se desarrolló un modelo computacional predictivo inicial. La implementación usaba un modelo lineal generalizado (Kernel Ridge Regression con kernel polinomial de orden 4). La implementación modeliza dos fenotipos distintos (*rendimiento* y *productividad*). Estos fenotipos se combinaron como suma ponderada para obtener una única puntuación para la clasificación, como se muestra a continuación. Varios parámetros del modelo, p. ej. factor de regularización, se ajustaron mediante una validación cruzada de k iteraciones de los datos de entrenamiento designados.

55 La implementación no incorpora ningún análisis explícito de los efectos de interacción como se describe en la sección de Mapeo de epistasia anterior. Sin embargo, como comprenderán los expertos en la técnica, el modelo lineal generalizado implementado puede capturar efectos de interacción implícitamente a través de los términos de segundo, tercer y cuarto orden del núcleo.

El modelo se entrenó contra el conjunto de entrenamiento. El modelo ajustado tiene un valor de R^2 (*coeficiente de determinación*) de 0,52 con respecto al *rendimiento* y un valor de R^2 de 0,67 con respecto a la *productividad*. La Figura 46 demuestra un ajuste de calidad significativo del modelo de *rendimiento* a los datos de entrenamiento.

5 Se generaron cepas candidatas. Este ejemplo incluye una restricción de construcción en serie asociada con la introducción de nuevos cambios genéticos en una cepa parental (en este ejemplo, solo se diseñó una nueva mutación en una cepa a la vez). Aquí, los candidatos no se consideran simplemente como una función del número deseado de cambios. En cambio, el equipo de análisis 214 seleccionó, como punto de partida, una colección de cepas previamente diseñadas que se sabe que tienen indicadores de alto rendimiento ("cepas de semillas"). El equipo de análisis 214 aplicó cambios genéticos individualmente a cada una de las cepas de semillas. Los cambios genéticos introducidos
10 no incluían los que ya estaban presentes en la cepa de semillas. Por diversas razones técnicas, biológicas o de otro tipo, se requerían explícitamente ciertas mutaciones, p. ej., *opca_4*, o se excluían explícitamente, p. ej., *dss_422*. Usando 166 cepas de semillas disponibles y los 336 cambios caracterizados por el modelo, se diseñaron 6239 nuevas cepas candidatas.

15 Basándose en el modelo, el equipo de análisis 214 predecía el rendimiento de los diseños de cepas candidatas. El equipo de análisis 214 clasificaba los candidatos de "mejor" a "peor" basándose en el rendimiento previsto con respecto a dos fenotipos de interés (*rendimiento* y *productividad*). Específicamente, el equipo de análisis 214 usó una suma ponderada para puntuar una cepa candidata:

$$\text{Puntuación} = 0,8 * \text{rendimiento}/\text{máx}(\text{rendimientos}) + 0,2 * \text{prod}/\text{máx}(\text{prods}),$$

donde *rendimiento* representa el rendimiento previsto para la cepa candidata,

20 *máx*(*rendimientos*) representa el rendimiento máximo de todas las cepas candidatas,

prod representa la productividad de la cepa candidata, y

máx.(*prods*) representa el rendimiento máximo de todas las cepas candidatas.

25 El equipo de análisis 214 generó un conjunto final de recomendaciones a partir de la lista clasificada de candidatos imponiendo tanto restricciones de capacidad como restricciones operativas. En este ejemplo, el límite de capacidad se estableció en 48 cepas de diseño candidatas generadas por ordenador. Debido a restricciones operativas, en este ejemplo solo se usó una cepa de semillas por columna de una placa de 96 pocillos. Esto significa que después de que se eligiera una variedad de semillas, se podían generar hasta 8 cambios en esa cepa, pero solo se podían elegir 6 cepas de semillas en una semana dada.

30 El modelo entrenado (descrito antes) se usó para predecir el rendimiento esperado (para *rendimiento* y *productividad*) de cada cepa candidata. El equipo de análisis 214 clasificaba las cepas candidatas usando la función de puntuación dada antes. Se aplicaron restricciones de capacidad y operativas para proporcionar un conjunto filtrado de 48 cepas candidatas. Este conjunto de cepas candidatas filtradas se representa en la Figura 47.

35 Las cepas candidatas filtradas se construyeron (en la fábrica 210) basándose en un pedido de fábrica generado por el motor de colocación de pedidos 208 (3312). La orden se basaba en las especificaciones de ADN correspondientes a las cepas candidatas. En la práctica, el procedimiento de construcción tiene una tasa de fallos esperada por lo que no se construye un conjunto aleatorio de cepas. Para este ciclo de construcción, aproximadamente 20% de las cepas candidatas fallaron en la compilación, lo que resultó en 37 cepas construidas.

40 El equipo de análisis 214 se usó para medir el *rendimiento* real y el *rendimiento* de la *productividad* de las cepas seleccionadas. El equipo de análisis 214 evaluaba el modelo y recomendaba las cepas basándose en tres criterios: precisión del modelo; mejora en el *rendimiento* de la cepa; y equivalencia (o mejora) con diseños generados por expertos humanos.

45 Se midieron los fenotipos de *rendimiento* y *productividad* para las cepas recomendadas y se compararon con los valores previstos por el modelo. Como se muestra en la Figura 48, el modelo demuestra una utilidad predictiva útil. En particular, los valores de *rendimiento* previstos para las cepas recomendadas tienen un coeficiente de correlación de Pearson-r de 0,59 con las observaciones correspondientes.

50 A continuación, el equipo de análisis 214 calculaba el cambio de *rendimiento* en porcentaje de la cepa parental para cada una de las cepas recomendadas. Estos datos se muestran en la Figura 49 (en gris claro). Los autores de la invención encontraron que muchas de las cepas previstas de hecho presentaban las ganancias de *rendimiento* esperadas con respecto a sus progenitoras inmediatas. En particular, la cepa mejor prevista mostró una mejora de 6% en el *rendimiento* con respecto a su progenitora inmediata.

En paralelo al procedimiento de diseño de cepas basado en modelos descrito anteriormente, un experto humano diseñó de forma independiente una colección de 48 cepas. De estas cepas, 37 se construyeron y ensayaron con éxito. Estos datos demostraban que los diseños de cepas basados en modelos se comportaban de manera comparable a las cepas diseñadas por expertos humanos. Estos expertos son científicos altamente capacitados (p. ej., nivel de

doctorado) empleados o contratados por el cesionario de la presente descripción, y están familiarizados con los ejemplos de esta descripción. Para comparar los dos métodos, los autores de la invención primero inspeccionaron las distribuciones de rendimiento de cada grupo (Figura 50). En este experimento, el rendimiento medio de las cepas basadas en modelos mostró un aumento de 1% con respecto a los diseños generados por expertos humanos.

5 Después los autores de la invención compararon cepas diseñadas por expertos humanos y diseñadas por modelos computacionales agrupadas por contextos, es decir, cepas nuevas con el mismo progenitor (Figura 51). De nuevo, los autores de la invención encontraron que los diseños generados por ordenador funcionan de manera comparable, y en algunos casos mejor que los diseños generados por expertos humanos, y además tienden a producir menos variabilidad. Finalmente, los autores de la invención compararon el cambio en porcentaje con respecto a las cepas parentales del experto humano y las cepas diseñadas por modelos (Figura 49). Una vez más, estas poblaciones mostraron ganancias comparables.

Véase la Tabla 4.1 para las estadísticas resumidas tabuladas.

Tabla 4.1. Estadísticas de rendimiento medidas para cepas diseñadas por el modelo predictivo y por una referencia humana experta.

		Rendimiento [UA]	Cambio de rendimiento del progenitor [%]	Productividad [UA]	Cambio de productividad del progenitor [%]
método de diseño					
modelo computacional	recuento	37	37	37	37
	media	1,058068108	0,3578340	0,737928919	-2,5428848
	est.	0,017811031	1,8293665	0,083619804	9,6743873
	mín.	1,015310000	-4,5346677	0,572780000	-23,3626353
	mediana	1,058710000	0,005007939	0,766870000	-1,1824159
	máx.	1,093510000	6,0097309	0,872790000	26,6124119
Experto humano	recuento	37	37	37	37
	media	1,038804595	-0,0005237	0,748320811	-1,6126436
	est.	0,032053625	1,9227716	0,120527468	9,8530758
	mín.	0,964910000	-3,1043233	0,535980000	-21,4589256
	mediana	1,045530000	0,0449168	0,760300000	-1,9241048
	máx.	1,094790000	7,8487174	0,984110000	21,7335193

Al final de cada ronda del ciclo de predicción → construcción → ensayo, los autores de la invención estaban interesados en evaluar la calidad de las predicciones del modelo e incorporar de forma iterativa nuevos datos en el modelo previo. Para la evaluación del modelo anterior, los autores de la invención se centraron en medir la precisión predictiva comparando las predicciones del modelo con las mediciones experimentales. La precisión predictiva se puede evaluar mediante varios métodos, que incluyen un coeficiente de correlación que indica la fuerza de la asociación entre los valores previstos y observados, o el error cuadrático medio, que es una medida del error promedio del modelo.

A lo largo de muchas rondas de experimentación, las predicciones del modelo pueden derivar y se pueden añadir nuevos cambios genéticos a las entradas de entrenamiento para mejorar la precisión predictiva. Para este ejemplo, se añadieron cambios de diseño y su rendimiento resultante al modelo predictivo (3316).

Diseño e ingeniería genómica como servicio

En ejemplos de la descripción, el software 3210 del sistema LIMS de la Figura 32 se puede implementar en un sistema informático en la nube 3202 de la Figura 32, para permitir que múltiples usuarios diseñen y construyan cepas microbianas de acuerdo con los ejemplos de la presente descripción. La figura 32 ilustra un entorno de computación en la nube 3204 de acuerdo con ejemplos de la presente descripción. Los ordenadores clientes 3206, tales como los

ilustrados en la Figura 32, acceden al sistema LIMS a través de una red 3208, tal como Internet. En ejemplos, el software de aplicación del sistema LIMS 3210 reside en el sistema informático en la nube 3202. El sistema LIMS puede usar uno o más sistemas informáticos que usan uno o más procesadores, del tipo ilustrado en la Figura 32. El propio sistema informático en la nube incluye una interfaz de red 3212 para interconectar las aplicaciones del sistema LIMS 3210 con los ordenadores clientes 3206 a través de la red 3208. La interfaz de red 3212 puede incluir una interfaz de programación de aplicaciones (API) para permitir que las aplicaciones cliente en los ordenadores cliente 3206 accedan al software del sistema LIMS 3210. En particular, a través de la API, los ordenadores clientes 3206 pueden acceder a componentes del sistema LIMS 200, incluyendo, sin limitación, el software que ejecuta la interfaz de entrada 202, el intérprete 204, el motor de ejecución 207, el motor de colocación de pedidos 208, la fábrica 210, así como el equipo de ensayo 212 y equipo de análisis 214. Un software como módulo de software de servicio (SaaS) 3214 ofrece el software del sistema LIMS 3210 como un servicio a los ordenadores clientes 3206. Un módulo de gestión de la nube 3216 gestiona el acceso al sistema LIMS 3210 por los ordenadores clientes 3206. El módulo de gestión de la nube 3216 puede permitir una arquitectura de nube que emplea aplicaciones de múltiples usuarios, virtualización u otras arquitecturas conocidas en la técnica para servir a múltiples usuarios.

15 **Automatización genómica**

La automatización de los métodos de la presente descripción permite el cribado fenotípico de alta capacidad y la identificación de productos diana de múltiples variantes de cepas de ensayo simultáneamente.

La plataforma de modelización predictiva de ingeniería genómica antes mencionada se basa en el hecho de que se construyen cientos y miles de cepas mutantes de una manera de alta capacidad. Los sistemas robóticos e informáticos que se describen a continuación son los mecanismos estructurales mediante los cuales se puede llevar a cabo dicho procedimiento de alta capacidad.

En algunos ejemplos, la presente descripción enseña métodos para mejorar las productividades de la célula hospedante o rehabilitar cepas industriales. Como parte de este procedimiento, la presente descripción enseña métodos para ensamblar ADN, construir nuevas cepas, cribar cultivos en placas y cribar cultivos en modelos para fermentación en tanque. En algunos ejemplos, la presente descripción enseña que uno o más de los métodos antes mencionados para crear y ensayar nuevas cepas hospedantes es asistido por robótica automatizada.

En algunos ejemplos, la presente descripción enseña una plataforma de transformación de cepas de alta capacidad como se muestra en la Figura 6A-B o Figura 26.

Sistemas robóticos de HTP

30 En algunos ejemplos, los métodos automatizados de la descripción comprenden un sistema robótico. Los sistemas descritos en el presente documento se dirigen en general al uso de placas de microtitulación de 96 o 384 pocillos, pero como apreciarán los expertos en la técnica, se puede usar cualquier número de placas o configuraciones diferentes. Además, cualquiera o todas las etapas descritas en el presente documento se pueden automatizar; por lo tanto, por ejemplo, los sistemas pueden estar total o parcialmente automatizados.

35 En algunos ejemplos, los sistemas automatizados de la presente descripción comprenden uno o más módulos de trabajo. Por ejemplo, en algunos ejemplos, el sistema automatizado de la presente descripción comprende un módulo de síntesis de ADN, un módulo de clonación de vectores, un módulo de transformación de cepas, un módulo de cribado y un módulo de secuenciación (véase la Figura 7).

40 Como apreciarán los expertos en la técnica, un sistema automatizado puede incluir una amplia variedad de componentes, que incluyen, pero no se limitan a: manipuladores de líquidos; uno o más brazos robóticos; manipuladores de placas para el posicionamiento de las microplacas; selladores de placas, perforadores de placas, manipuladores automatizados de tapas para quitar y reemplazar las tapas para los pocillos en placas sin contaminación cruzada; conjuntos de puntas desechables para distribución de muestras con puntas desechables; conjuntos de puntas lavables para distribución de muestras; bloques de carga de 96 pocillos; termocicladores integrados; gradillas de reactivos refrigerados; posiciones de las pipetas de la placa de microtitulación (opcionalmente refrigeradas); torres de apilamiento de placas y puntas; estaciones de procesamiento de perlas magnéticas; sistemas de filtración; agitadores de placas; lectores y aplicadores de códigos de barras; y sistemas informáticos.

45 En algunos ejemplos, los sistemas robóticos de la presente descripción incluyen manipuladores automatizados de líquidos y partículas que permite el pipeteo de alta capacidad para realizar todas las etapas en el procedimiento de aplicaciones de recombinación y direccionamiento de genes. Esto incluye manipulaciones de líquidos y partículas tales como aspiración, dispensación, mezcla, dilución, lavado, transferencias volumétricas precisas; recuperar y descartar puntas de pipeta; y pipeteo repetitivo de volúmenes idénticos para múltiples suministros a partir de una única aspiración de muestra. Estas manipulaciones son transferencias de líquidos, partículas, células y organismos exentas de contaminación cruzada. Los instrumentos realizan la replicación automatizada de muestras de microplacas en filtros, membranas y/o placas hija, transferencias de alta densidad, diluciones en serie de placa completa y operación de alta capacidad.

En algunos ejemplos, el sistema de manipulación de líquidos automatizado personalizado de la descripción es una máquina TECAN (p. ej., una TECAN Freedom Evo personalizada).

5 En algunos ejemplos, los sistemas automatizados de la presente descripción son compatibles con plataformas para placas de múltiples pocillos, placas de pocillos profundos, placas de pocillos cuadrados, depósitos de reactivos, tubos de ensayo, minitubos, tubos de microcentrífuga, crioviales, filtros, chips de micromatrices, fibras ópticas, perlas, geles de agarosa y acrilamida y otras matrices o plataformas en fase sólida se alojan en una plataforma modular actualizable. En algunos ejemplos, los sistemas automatizados de la presente descripción contienen al menos una plataforma modular para superficies de trabajo de múltiples posiciones para poner muestras fuente y de salida, reactivos, dilución de muestras y reactivos, placas de ensayo, depósitos de muestras y reactivos, puntas de pipeta y un estación de lavado de puntas activa.

10 En algunos ejemplos, los sistemas automatizados de la presente descripción incluyen sistemas de electroporación de alta capacidad. En algunos ejemplos, los sistemas de electroporación de alta capacidad son capaces de transformar células en placas de 96 o 384 pocillos. En algunos ejemplos, los sistemas de electroporación de alta capacidad incluyen los sistemas de electroporación de alta capacidad VWR®, BTX™, Bio-Rad® Gene Pulser MXcell™ u otro sistema de electroporación de múltiples pocillos.

15 En algunos ejemplos, el termociclador integrado y/o los reguladores térmicos se usan para estabilizar la temperatura de intercambiadores de calor, tales como bloques o plataformas controlados, para proporcionar control de temperatura preciso de las muestras en incubación de 0°C a 100°C.

20 En algunos ejemplos, los sistemas automatizados de la presente descripción son compatibles con cabezales de máquina intercambiables (de uno o múltiples canales) con una o múltiples sondas magnéticas, sondas de afinidad, replicadores o pipeteadores, capaces de manipular de manera robótica líquidos, partículas, células y organismos multicelulares. Las separadores magnéticos de múltiples pocillos o múltiples tubos y las estaciones de filtración manipulan líquidos, partículas, células y organismos en formatos de muestra individuales o múltiples.

25 En algunos ejemplos, los sistemas automatizados de la presente descripción son compatibles con sistemas de visión por cámara y/o espectrómetro. Por lo tanto, en algunos ejemplos, los sistemas automatizados de la presente descripción son capaces de detectar y registrar cambios de color y absorción en cultivos celulares en curso.

30 En algunos ejemplos, el sistema automatizado de la presente descripción está diseñado para ser flexible y adaptable con múltiples complementos de hardware para permitir que el sistema lleve a cabo múltiples aplicaciones. Los módulos del programa de software permiten la creación, modificación y ejecución de métodos. Los módulos de diagnóstico del sistema permiten la configuración, alineamiento de instrumentos y operaciones del motor. Las herramientas personalizadas, el material de laboratorio y los patrones de transferencia de líquidos y partículas permiten programar y realizar diferentes aplicaciones. La base de datos permite el almacenamiento de métodos y parámetros. Las interfaces robóticas y de ordenador permiten la comunicación entre instrumentos.

35 Por lo tanto, en algunos ejemplos, la presente descripción enseña una plataforma de transformación de cepas de alta capacidad, como se muestra en la Figura 26.

Las personas con experiencia en la técnica reconocerán las diversas plataformas robóticas capaces de llevar a cabo los métodos de ingeniería de HTP de la presente descripción. La Tabla 5 a continuación proporciona una lista no exclusiva de equipo científico capaz de llevar a cabo cada etapa de las etapas de ingeniería de HTP de la presente descripción como se describe en la Figura 26.

40 Tabla 5 - Lista no exclusiva de equipo científico compatible con los métodos de ingeniería de HTP de la presente descripción

	Tipo de equipo	Operación(es) realizadas	Equipo compatible Hecho/Modelo/Configuración
Adquisición y construcción de trozos de ADN	manipuladores de líquidos	Selección de aciertos (combinando por transferencia) cebadores/moles para amplificación por PCR de partes de ADN	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Termocicladores	Amplificación por PCR de partes de ADN	Inheco Cycler, ABI 2720, ABI Proflex 384, ABI Veriti, o equivalentes

	Tipo de equipo	Operación(es) realizadas	Equipo compatible Hecho/Modelo/Configuración
CC partes de ADN	Analizadores de fragmentos (electroforesis capilar)	electroforesis en gel para confirmar los productos de PCR de tamaño adecuado	Agilent Bioanalyzer, AATI Fragment Analyzer, o equivalentes
	Secuenciador (Sanger: Beckman)	Verificación de secuencia de partes/moldes	Beckman Ceq-8000, Beckman GenomeLab™, o equivalentes
	instrumento de NGS (secuenciación de próxima generación)	Verificación de secuencia de partes/moldes	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio u otros equivalentes
	nanogota/lector de placa	evaluación de la concentración de las muestras de ADN	Molecular Devices SpectraMax MS, Tecan M1000, o equivalentes
Generación de ensamblaje de ADN	manipuladores de líquidos	Selección de aciertos (combinando por transferencia) de partes de ADN para ensamblaje junto con el vector de clonación, adición de reactivos para la reacción/proceso de ensamblaje	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
CC ensamblaje de ADN	Selector de colonias	para inocular colonias en el medio líquido	Scirobotics Pickolo, Molecular Devices QPix 420
	manipuladores de líquidos	Selección de aciertos cebadores/moldes, dilución de muestras	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Analizadores de fragmentos (electroforesis capilar)	electroforesis en gel para confirmar los productos ensamblados de tamaño adecuado	Agilent Bioanalyzer, AATI Fragment Analyzer
	Secuenciador (Sanger: Beckman)	Verificación de secuencia de plásmidos ensamblados	ABI3730 Thermo Fisher, Beckman Ceq-8000, Beckman GenomeLab™, o equivalentes
	instrumento de NGS (secuenciación de próxima generación)	Verificación de secuencia de plásmidos ensamblados	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio u otros equivalentes
Preparación de cepa base y ensamblaje de ADN	centrífuga	centrifugación/sedimentación de células	Centrífuga de suelo Beckman Avanti, centrífuga Hettich
Transformación del ADN en cepa base	Electroporadores	transformación electroporativa de células	Electroporador BTX Gemini X2, BIO-RAD MicroPulser
	Transformación balística	transformación balística de células	BIO-RAD PDS1000
	Incubadores, termocicladores	para transformación química/choque térmico	Inheco Cyclor, ABI 2720, ABI Proflex 384, ABI Veriti, o equivalentes

	Tipo de equipo	Operación(es) realizadas	Equipo compatible Hecho/Modelo/Configuración
	manipuladores de líquidos	para combinación de ADN, células, tampón	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
Integración de ADN en el genoma de la cepa base	Selector de colonias	para inocular colonias en medio líquido	Scirobotics Pickolo, Molecular Devices QPix 420
	manipuladores de líquidos	Para transferir células a agar, transferir desde placas de cultivo a placas de cultivo diferentes (inoculación en otros medios selectivos)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Agitadores-incubadores de plataforma	incubación con agitación de cultivos en placas de microtitulación	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro
CC de cepa transformada	Selector de colonias	para inocular colonias en el medio líquido	Scirobotics Pickolo, Molecular Devices QPix 420
	manipuladores de líquidos	Selección de aciertos cebadores/moldes, dilución de muestras	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	termocicladores	verificación por cPCR de cepas	Inheco Cycler, ABI 2720, ABI Proflex 384, ABI Veriti, o equivalentes
	Analizadores de fragmentos (electroforesis capilar)	electroforesis en gel para confirmar los productos de cPCR de tamaño adecuado	Infors-ht Multitron Pro, Kuhner Shaker ISF4-X
	Secuenciador (Sanger: Beckman)	Verificación de secuencia de modificación introducida	Beckman Ceq-8000, Beckman GenomeLab™, o equivalentes
	instrumento de NGS (secuenciación de próxima generación)	Verificación de secuencia de modificación introducida	Illumina MiSeq series sequences, illumina Hi-Seq, Ion torrent, pac bio u otros equivalentes
Selección y consolidación de cepas CC en la placa de ensayo	manipuladores de líquidos	Para transferir desde placas de cultivo a placas de cultivo diferentes (inoculación en medios de producción)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Selector de colonias	para inocular colonias en medio líquido	Scirobotics Pickolo, Molecular Devices QPix 420
	Agitadores-incubadores de plataforma	incubación con agitación de cultivos en placas de microtitulación	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro
Cultivar cepas en placas de semillas	manipuladores de líquidos	Para transferir desde placas de cultivo a placas de cultivo diferentes (inoculación en medios de producción)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Agitadores-incubadores de plataforma	incubación con agitación de cultivos en placas de microtitulación	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro

	Tipo de equipo	Operación(es) realizadas	Equipo compatible Hecho/Modelo/Configuración
	dispensadores de líquidos	Dispensan medio de cultivo líquido en placas de microtitulación	Well mate (Thermo), Benchcel2R (velocidad 11), Plateloc (velocidad 11)
	marcadores de microplacas	aplican códigos de barras a placas	Marcador de microplacas (a2+ cabagilent), Benchcell 6R (velocidad 11)
Generación de producto de las cepas	manipuladores de líquidos	Para transferir desde placas de cultivo a placas de cultivo diferentes (inoculación en medios de producción)	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	Agitadores-incubadores de plataforma	incubación con agitación de cultivos en placas de microtitulación	Kuhner Shaker ISF4-X, Infors-ht Multitron Pro
	dispensadores de líquidos	Dispensan medio de cultivo líquido en placas de microtitulación múltiples y sellan las placas	Well mate (Thermo), Benchcel2R (velocidad 11), Plateloc (velocidad 11)
	marcadores de microplacas	aplican códigos de barras a placas	Marcador de microplacas (a2+ cabagilent), Benchcell 6R (velocidad 11)
Evaluar rendimiento	manipuladores de líquidos	Para procesar el caldo de cultivo para analítica posterior	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200, Beckman Coulter Biomek FX, o equivalentes
	UHPLC, HPLC	análisis cuantitativo de compuestos precursores y diana	Agilent 1290 Series UHPLC y 1200 Series HPLC con detectores de UV y RI o equivalente; también cualquier LC/MS
	LC/MS	análisis altamente específico de compuestos precursores y diana así como productos secundarios y de degradación	Agilent 6490 QQQ y 6550 QTOF acoplado con 1290 Series UHPLC
	Espectrofotómetro	Cuantificación de diferentes compuestos usando ensayos basados en espectrofotómetro	Tecan M1000, spectramax M5, Genesys 10S
Cultivar cepas en matraces	Fermentadores;	incubación con agitación	Sartorius, DASGIPs (Eppendorf), BIO-FLO (Sartotius-stedim). Applikon
	Agitadores de plataforma		Innova 4900, o cualquier equivalente
Generar producto a partir de la cepa	Fermentadores: DASGIPs (Eppendorf), BIO-FLO (Sartotius-stedim)		
Eva luar ren dim ient	manipuladores de líquidos	Para transferir desde placas de cultivo a placas de cultivo	Hamilton Microlab STAR, Labcyte Echo 550, Tecan EVO 200,

	Tipo de equipo	Operación(es) realizadas	Equipo compatible Hecho/Modelo/Configuración
		diferentes (inoculación en medios de producción)	Beckman Coulter Biomek FX, o equivalentes
	UHPLC, HPLC	análisis cuantitativo de compuestos precursores y diana	Agilent 1290 Series UHPLC y 1200 Series HPLC con detectores de UV y RI o equivalente; también cualquier LC/MS
	LC/MS	análisis altamente específico de compuestos precursores y diana así como productos secundarios y de degradación	Agilent 6490 QQQ y 6550 QTOF acoplado con 1290 Series UHPLC
	Citómetro de flujo	Caracteriza el rendimiento de la cepa (mide la viabilidad)	BD Accuri, Millipore Guava
	Espectrofotómetro	Caracteriza el rendimiento de la cepa (mide la biomasa)	Tecan M1000, spectramax M5, u otros equivalentes

Hardware del sistema informático

La Figura 34 ilustra un ejemplo de un sistema informático 800 que se puede usar para ejecutar código de programa almacenado en un medio legible por ordenador no transitorio (p. ej., memoria) de acuerdo con ejemplos de la descripción. El sistema informático incluye un subsistema de entrada/salida 802, que se puede usar para interactuar con usuarios humanos y/u otros sistemas informáticos dependiendo de la aplicación. El subsistema de E/S 802 puede incluir, p. ej., un teclado, ratón, interfaz gráfica de usuario, pantalla táctil u otras interfaces para la entrada y, p. ej., un LED u otra pantalla plana, u otras interfaces para la salida, incluyendo interfaces de programas de aplicación (API). Otros elementos de los ejemplos de la descripción, tales como los componentes del sistema LIMS, se pueden implementar con un sistema informático tal como el del sistema informático 800.

El código de programa se puede almacenar en medios no transitorios tales como el almacenamiento persistente en la memoria secundaria 810 o en la memoria principal 808 o en ambas. La memoria principal 808 puede incluir memoria volátil tal como memoria de acceso aleatorio (RAM) o memoria no volátil tal como memoria de solo lectura (ROM), así como diferentes niveles de memoria caché para acceso más rápido a instrucciones y datos. La memoria secundaria puede incluir almacenamiento persistente tales como unidades de estado sólido, unidades de disco duro o discos ópticos. Uno o más procesadores 804 leen el código de programa de uno o más medios no transitorios y ejecutan el código para permitir que el sistema informático lleve a cabo los métodos realizados por los ejemplos del presente documento. Los expertos en la técnica comprenderán que el o los procesadores pueden asimilar el código fuente e interpretar o compilar el código fuente en código de máquina que sea comprensible en el nivel de puerta de hardware del o de los procesadores 804. El o los procesadores 804 pueden incluir unidades de procesamiento de gráficos (GPU) para manejar tareas computacionalmente intensas. En particular en el aprendizaje automático, una o más CPU 804 pueden descargar el procesamiento de grandes cantidades de datos a una o más GPU 804.

El o los procesadores 804 se pueden comunicar con redes externas a través de una o más interfaces de comunicación 807, tal como una tarjeta de interfaz de red, un transceptor WiFi, etc. Un bus 805 acopla comunicativamente el subsistema de E/S 802, el o los procesadores 804, los dispositivos periféricos 806, las interfaces de comunicaciones 807, la memoria 808 y el almacenamiento persistente 810. Los ejemplos de la descripción no se limitan a esta arquitectura representativa. Ejemplos alternativos pueden emplear diferentes disposiciones y tipos de componentes, p. ej., buses separados para componentes de entrada-salida y subsistemas de memoria.

Los expertos en la técnica comprenderán que algunos o todos los elementos de los ejemplos de la descripción, y las operaciones que los acompañan, pueden ser implementados total o parcialmente por uno o más sistemas informáticos que incluyen uno o más procesadores y uno o más sistemas de memoria como los del sistema informático 800. En particular, los elementos del sistema LIMS 200 y cualquier robótica y otros sistemas o dispositivos automatizados descritos en el presente documento se pueden implementar por ordenador. Algunos elementos y funcionalidades se pueden implementar localmente y otros se pueden implementar de una forma distribuida a lo largo de una red a través de diferentes servidores, p. ej., en forma cliente-servidor, por ejemplo. En particular, se puede hacer que las operaciones del lado del servidor estén disponibles para múltiples clientes en un modo de software como servicio (SaaS), como se muestra en la Figura 32.

El término componente en este contexto se refiere ampliamente a un componente de software, hardware o firmware (o cualquier combinación de los mismos). Los componentes típicamente son componentes funcionales que pueden

generar datos útiles u otros resultados usando la o las entradas especificadas. Un componente puede ser autónomo o no. Un programa de aplicación (también llamado una "aplicación") puede incluir uno o más componentes, o un componente puede incluir uno o más programas de aplicación.

5 Algunos ejemplos incluyen algunos, todos o ninguno de los componentes junto con otros módulos o componentes de la aplicación. Sin embargo, varios ejemplos pueden incorporar dos o más de estos componentes en un solo módulo y/o asociar una parte de la funcionalidad de uno o más de estos componentes con un componente diferente.

10 El término "memoria" puede ser cualquier dispositivo o mecanismo usado para almacenar información. De acuerdo con algunos ejemplos de la presente descripción, la memoria pretende abarcar cualquier tipo de, pero no se limita a: memoria volátil, memoria no volátil y memoria dinámica. Por ejemplo, la memoria puede ser memoria de acceso aleatorio, dispositivos de almacenamiento de memoria, dispositivos de memoria óptica, medios magnéticos, disquetes, cintas magnéticas, discos duros, SIMM, SDRAM, DIMM, RDRAM, DDR RAM, SODIMMS, memorias de solo lectura programables y borrables (EPROM), memorias de solo lectura programables y borrables eléctricamente (EEPROM), discos compactos, DVD y/o similares. De acuerdo con algunos ejemplos, la memoria puede incluir una o más unidades de disco, unidades flash, bases de datos, memorias caché local, memorias caché de procesador, bases de datos relacionales, bases de datos planas, servidores, plataformas basadas en la nube y/o similares. Además, los expertos en la técnica apreciarán que se pueden usar como memoria muchos dispositivos y técnicas adicionales para almacenar información.

15 La memoria se puede usar para almacenar instrucciones para ejecutar una o más aplicaciones o módulos en un procesador. Por ejemplo, la memoria se podría usar en algunos ejemplos para albergar todas o algunas de las instrucciones necesarias para ejecutar la funcionalidad de uno o más de los módulos y/o aplicaciones que se describen en esta solicitud.

Ingeniería de cepas microbianas de HTP basada en predicciones de diseño genético: un flujo de trabajo de ejemplo

20 En algunos ejemplos, la presente descripción enseña la ingeniería dirigida de nuevos organismos hospedantes basada en las recomendaciones de los sistemas de análisis computacional de la presente descripción.

25 En algunos ejemplos, la presente descripción es compatible con todos los métodos de clonación y de diseño genético. Es decir, en algunos ejemplos, la presente descripción enseña el uso de técnicas de clonación tradicionales tales como reacción en cadena de la polimerasa, digestiones con enzimas de restricción, ligación, recombinación homóloga, RT PCR y otros generalmente conocidos en la técnica y se describen, por ejemplo, en: Sambrook et al. (2001) Molecular Cloning: A Laboratory Manual (3ª ed., Cold Spring Harbor Laboratory Press, Plainview, Nueva York).

30 En algunos ejemplos, las secuencias clonadas pueden incluir posibilidades de cualquiera de las bibliotecas de diseño genético de HTP que se enseñan en el presente documento, por ejemplo: promotores de una biblioteca de intercambio de promotores, SNP de una biblioteca de intercambio de SNP, codones de inicio o parada de una biblioteca de intercambio de codones de inicio/parada, terminadores de una biblioteca de intercambio de STOP, marcadores de solubilidad de proteínas de una biblioteca de intercambio MARCADORES DE SOLUBILIDAD, marcadores de degradación de proteínas de una biblioteca de intercambio de MARCADORES DE DEGRADACIÓN u optimizaciones de secuencia de una biblioteca de optimización de secuencias.

Además, las combinaciones de secuencias exactas que deberían incluirse en una construcción particular pueden ser instruidas por la función de mapeo epistático.

35 En otros ejemplos, las secuencias clonadas también pueden incluir secuencias basadas en un diseño racional (basado en hipótesis) y/o secuencias basadas en otras fuentes, tales como publicaciones científicas.

40 En algunos ejemplos, la presente descripción enseña métodos de ingeniería dirigida, que incluyen las etapas de i) generar ADN específico de SNP hecho a medida, ii) ensamblar plásmidos específicos de SNP, iii) transformar células hospedantes diana con ADN específico de SNP, y iv) hacer un bucle de salida con los marcadores de selección (véase la Figura 2).

45 La Figura 6A representa el flujo de trabajo general de los métodos de ingeniería de cepas de la presente descripción, que incluyen adquirir y ensamblar ADN, ensamblar vectores, transformar células hospedantes y eliminar marcadores de selección.

Construcción de oligonucleótidos de ADN específicos

50 En algunos ejemplos, la presente descripción enseña a insertar y/o reemplazar y/o alterar y/o eliminar un segmento de ADN del organismo célula hospedante. En algunos ejemplos, los métodos enseñados en el presente documento implican la construcción de un oligonucleótido de interés (es decir, un segmento de ADN diana), que se incorporará al genoma de un organismo hospedante. En algunos ejemplos, los segmentos de ADN diana de la presente descripción se pueden obtener por cualquier método conocido en la técnica, que incluye: copiar o cortar de un molde conocido, mutación o síntesis de ADN. En algunos ejemplos, la presente descripción es compatible con productos de síntesis

de genes disponibles en el mercado para producir secuencias de ADN diana (p. ej., GeneArt™, GeneMaker™, GenScript™, Anagen™, Blue Heron™, Entelechon™, GeNOsys, Inc. o Qiagen™).

5 En algunos ejemplos, el segmento de ADN diana se diseña para incorporar un SNP en una región de ADN seleccionada del organismo hospedante (p. ej., añadiendo un SNP beneficioso). En otros ejemplos, el segmento de ADN se diseña para eliminar un SNP del ADN de los organismos hospedantes (p. ej., eliminar un SNP perjudicial o neutro).

10 En algunos ejemplos, los oligonucleótidos usados en los métodos de la invención se pueden sintetizar usando cualquiera de los métodos de síntesis enzimática o química conocidos en la técnica. Los oligonucleótidos se pueden sintetizar sobre soportes sólidos tales como vidrio de poro controlado (CPG), perlas de poliestireno o membranas compuestas de polímeros termoplásticos que pueden contener CPG. Los oligonucleótidos también se pueden sintetizar en matrices, en una microescala paralela usando microfluidos (Tian et al., *Mol. BioSyst.*, 5, 714-722 (2009)), o tecnologías conocidas que ofrecen combinaciones de ambos (véase Jacobsen et al., solicitud de patente de EE.UU. nº 2011/0172127).

15 La síntesis en matrices o mediante microfluidos ofrece una ventaja sobre la síntesis de soporte sólido convencional al reducir costes por el menor uso de reactivos. La escala requerida para la síntesis de genes es baja, por lo que la escala del producto oligonucleótido sintetizado a partir de matrices o mediante microfluidos es aceptable. Sin embargo, los oligonucleótidos sintetizados son de menor calidad que cuando se usa la síntesis de soporte sólido (véase Tian, más adelante; véase también Staehler et al., solicitud de patente de EE.UU. nº 2010/0216648).

20 Se ha logrado un gran número de avances en la química tradicional de la fosforamidita de cuatro etapas desde que se describió por primera vez en la década de 1980 (véase, por ejemplo, Sierzchala, et al. *J. Am. Chem. Soc.*, 125, 13427-13441 (2003) usando desprotección de aniones peroxi; Hayakawa et al., patente de EE.UU. nº 6.040.439 para grupos protectores alternativos; Azhayev et al., *Tetrahedron* 57, 4977-4986 (2001) para soportes universales; Kozlov et al., *Nucleosides, Nucleotides, and Nucleic Acids*, 24 (5-7), 1037-1041 (2005) para una síntesis mejorada de oligonucleótidos más largos mediante el uso de CPG de poros grandes; y Damha et al., *NAR*, 18, 3813-3821 (1990) para una derivatización mejorada).

25 Independientemente del tipo de síntesis, los oligonucleótidos resultantes pueden formar entonces bloques de construcción más pequeños para oligonucleótidos más largos. En algunos ejemplos, se pueden unir entre sí oligonucleótidos más pequeños usando protocolos conocidos en la técnica, tales como el ensamblaje en cadena de la polimerasa (PCA), reacción en cadena de la ligasa (LCR) y la síntesis de dentro hacia fuera equilibrada termodinámicamente (TBIO) (véase Czar et al. *Trends in Biotechnology*, 27, 63-71 (2009)). En el PCA, los oligonucleótidos que abarcan toda la longitud del producto más largo deseado se reasocian y se extienden en múltiples ciclos (típicamente aproximadamente 55 ciclos) para lograr finalmente el producto de longitud completa. La LCR usa la enzima ligasa para unir dos oligonucleótidos que se reasocian con un tercer oligonucleótido. La síntesis TBIO empieza en el centro del producto deseado y se extiende progresivamente en ambas direcciones usando oligonucleótidos que se solapan que son homólogos de la cadena directa en el extremo 5' del gen y contra la cadena inversa en el extremo 3' del gen.

30 Otro método para sintetizar un fragmento de ADN bicatenario más grande es combinar oligonucleótidos más pequeños por PCR de cadena superior (TSP). En este método, una pluralidad de oligonucleótidos se extiende en toda la longitud de un producto deseado y contiene regiones que se solapan con el o los oligonucleótidos adyacentes. La amplificación se puede realizar con cebadores directos e inversos universales y, mediante múltiples ciclos de amplificación se forma un producto de ADN bicatenario de longitud completa. Este producto después se puede someter a corrección de errores opcional y amplificación adicional que da como resultado el fragmento de ADN bicatenario deseado producto final.

35 En un método de TSP, el conjunto de oligonucleótidos más pequeños que se combinarán para formar el producto deseado de longitud completa tiene entre 40-200 bases de largo y se solapan entre sí en al menos aproximadamente 15-20 bases. Para fines prácticos, la región de solapamiento debe ser como mínimo lo suficientemente larga para asegurar la reasociación específica de los oligonucleótidos y tener una temperatura de fusión (T_m) suficientemente alta para la reasociación a la temperatura de reacción empleada. El solapamiento se puede extender hasta el punto donde un oligonucleótido dado se solapa completamente con oligonucleótidos adyacentes. La cantidad de solapamiento no parece tener ningún efecto sobre la calidad del producto final. El primer y último bloque de construcción de oligonucleótidos en el ensamblaje deben contener sitios de unión para cebadores de amplificación directos e inversos. En un ejemplo, la secuencia del extremo terminal del primer y último oligonucleótido contiene la misma secuencia de complementariedad para permitir el uso de cebadores universales.

Ensamblaje/clonación de plásmidos a medida

55 En algunos ejemplos, la presente descripción enseña métodos para construir vectores capaces de insertar secciones de ADN diana deseadas (p. ej., que contienen un SNP particular) en el genoma de los organismos hospedantes. En algunos ejemplos, la presente descripción enseña métodos de clonación de vectores que comprenden el ADN diana, brazos de homología y al menos un marcador de selección (véase la Figura 3).

En algunos ejemplos, la presente descripción es compatible con cualquier vector adecuado para la transformación en el organismo hospedante. En algunos ejemplos, la presente descripción enseña el uso de vectores lanzadera compatibles con una célula hospedante. En un ejemplo, un vector lanzadera para usar en los métodos proporcionados en el presente documento es un vector lanzadera compatible con una célula hospedante de *E. coli* y/o *Corynebacterium*. Los vectores lanzadera para uso en los métodos proporcionados en el presente documento pueden comprender marcadores para selección y/o contraselección como se describe en el presente documento. Los marcadores pueden ser cualquier marcador conocido en la técnica y/o proporcionado en el presente documento. Los vectores lanzadera pueden comprender además cualquier secuencia o secuencias reguladoras y/o secuencias útiles en el ensamblaje de dichos vectores lanzadera como se conoce en la técnica. Los vectores lanzadera pueden comprender además cualquier origen de replicación que pueda ser necesario para la propagación en una célula hospedante como se proporciona en el presente documento, tal como, por ejemplo, *E. coli* o *C. glutamicum*. La secuencia reguladora puede ser cualquier secuencia reguladora conocida en la técnica o proporcionada en el presente documento tal como, por ejemplo, un promotor, inicio, parada, aislante, señal, secuencia de secreción y/o terminación usada por la maquinaria genética de la célula hospedante. En ciertos casos, el ADN diana se puede insertar en vectores, construcciones o plásmidos que se pueden obtener de cualquier repositorio o producto de catálogo, como un vector comercial (véase, p. ej., los vectores ADN2.0 a medida o GATEWAY®). En ciertos casos, el ADN diana se puede insertar en vectores, construcciones o plásmidos que se pueden obtener de cualquier repositorio o producto de catálogo, tal como un vector comercial (véase, por ejemplo, los vectores ADN2.0 a medida o GATEWAY®).

En algunos ejemplos, los métodos de ensamblaje/clonación de la presente descripción pueden emplear al menos una de las siguientes estrategias de ensamblaje: **i)** clonación convencional de tipo II, **ii)** clonación mediada por tipo IIS o "Golden Gate" (véase, p. ej., Engler, C., R. Kandzia y S. Marillonnet. "A one pot, one step, precision cloning method with high-throughput capability". PLoS One 3:e3647; Kotera, I., y T. Nagai. 2008 "A high-throughput and single-tube recombination of crude PCR products using a DNA polymerase inhibitor and type IIS restriction enzyme." *J Biotechnol* 137:1-7.; Weber, E., R. Gruetzner, S. Werner, C. Engler, y S. Marillonnet. 2011 "Assembly of Designer TAL Effectors by Golden Gate Cloning". *PLoS One* 6:e1722), **iii)** recombinación GATEWAY®, **iv)** clonación TOPO®, ensamblaje mediado por exonucleasa (Aslanidis y de Jong 1990. "Ligation-independent cloning of PCR products (LIC-PCR)." *Nucleic Acids Research*, Vol. 18, No. 20 6069), **v)** recombinación homóloga, **vi)** unión de extremos no homólogos, **vii)** ensamblaje de Gibson (Gibson et al., 2009 "Enzymatic assembly of DNA molecules up to several hundred kilobases" *Nature Methods* 6, 343-345) o una combinación de los mismos. Las estrategias de ensamblaje basadas en IIS tipo modular se describen en la publicación PCT WO 2011/154147.

En algunos ejemplos, la presente descripción enseña vectores de clonación con al menos un marcador de selección. Se conocen en la técnica varios genes marcadores de selección que a menudo codifican la función de resistencia a antibióticos para la selección en células procariontas (p. ej., contra ampicilina, kanamicina, tetraciclina, cloranfenicol, zeocina, espectinomicina/estreptomycin) o células eucariotas (p. ej., geneticina, neomicina, higromicina, puromicina, blasticidina, zeocina) bajo presión selectiva. Otros sistemas de marcadores permiten el cribado e identificación de células deseadas o no deseadas, tales como el conocido sistema de cribado azul/blanco usado en bacterias para seleccionar clones positivos en presencia de X-gal o indicadores fluorescentes tales como proteínas fluorescentes verdes o rojas expresadas en células hospedantes transducidas con éxito. Otra clase de marcadores de selección, la mayoría de los cuales solo son funcionales en sistemas procariontas, se refiere a genes de marcadores contraseleccionables a menudo denominados también "genes de muerte" que expresan productos génicos tóxicos que matan las células productoras. Los ejemplos de dichos genes incluyen *sacB*, *rpsL* (*strA*), *tetAR*, *pheS*, *thyA*, *gata-1* o *ccdB*, cuya función se describe en (Reyrat et al. 1998 "Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis". *Infect Immun.* 66(9): 4011-4017).

Ensamblaje de vectores de ADN, amplificación y edición del genoma

En algunos ejemplos, la presente descripción enseña etapas de ingeniería genómica de HTP específicas para *E. coli*. Por lo tanto, en algunos ejemplos, la presente descripción enseña métodos de construcción y amplificación de construcciones en *E. coli*, así como métodos de transformación de *E. coli*.

En algunos ejemplos, los vectores de ADN de la presente descripción comprenden **i)** un ori de replicación condicional (R6K), **ii)** un gen de resistencia a antibióticos, **iii)** uno o más genes de contraselección (p. ej., *sacB* y/o *PheS*) y **iv)** un ori de replicación para *S. cerevisiae*.

En algunos ejemplos, la presente descripción enseña métodos para ensamblar construcciones de ADN en *S. cerevisiae* auxotrófica. Por lo tanto, en algunos ejemplos, los vectores de la presente descripción comprenden un origen de replicación para *S. cerevisiae*. Esto permite que el vector se replique en *S. cerevisiae* durante el ensamblaje.

En algunos ejemplos, la presente descripción enseña que los métodos de propagación de ADN ensamblado se producen en *E. coli* que contiene proteína *pir*. Por lo tanto, en algunos ejemplos, los vectores de la presente descripción comprenden un origen de replicación de R6K. En algunos ejemplos, el origen de replicación de R6K está condicionado a la presencia de la proteína *pir*. Es decir, en algunos ejemplos, los vectores descritos actualmente que comprenden el origen de replicación de R6K solo se amplificarán en células hospedantes que comprenden el gen de *pir*. Esto permite a los investigadores amplificar los vectores de la presente descripción durante las etapas de construcción y

amplificación del vector, a la vez que se evita también la expresión cromosómica adicional de los vectores durante las etapas de transformación en la célula hospedante.

5 En algunos ejemplos, los vectores de la presente descripción comprenden un gen PheS. La fenilalanil-tRNA sintetasa (PheS) de *Escherichia coli* puede ser útil como un marcador de contraselección, ya que su variante A294G incorpora erróneamente 4-cloro-fenilalanina (4CP) en proteínas celulares durante la traducción, causando así la muerte celular. En algunos ejemplos, el gen PheS está diseñado para incorporarse temporalmente al genoma de la célula hospedante. En algunos ejemplos, la presente descripción enseña métodos de contraselección que comprenden cultivar las células hospedantes en medio mínimo con 4CP. Las células que todavía comprenden el vector incorporarán 4CP en proteínas y morirán. Las células que han sacado en un bucle la secuencia PheS sobreviven.

10 En algunos ejemplos, la presente descripción enseña métodos para construir, ensamblar e integrar vectores en células hospedantes. En algunos ejemplos, la presente descripción enseña que los brazos de homología (homL y homR) se amplifican por PCR. En algunos ejemplos, el cambio genético deseado (barra negra entre homL y homR) está presente en el cebador inverso de homL y el cebador directo de homR (véase la Figura 3). En algunos ejemplos, el cebador directo de homL y el cebador inverso de homR tienen homología de secuencia con el plásmido de la cadena principal. 15 En la figura 45 se muestran más ilustraciones del procedimiento de entrada de bucle y salida de bucle.

En algunos ejemplos, los vectores de la presente descripción comprenden una o más secuencias de aislantes. La secuencia de aislante puede ser cualquier secuencia de aislante conocida en la técnica. En un ejemplo, la secuencia de ácido nucleico del aislante es la secuencia del aislante 1 (SEQ ID NO. 218), secuencia del aislante 2 (SEQ ID NO. 219) o las secuencias tanto del aislante 1 como 2 proporcionadas en el presente documento. En un ejemplo, los 20 vectores de la presente descripción comprenden secuencias de aislantes que flanquean los brazos de homología (homL y homR). En un ejemplo, los vectores de la presente descripción comprenden secuencias de aislantes que flanquean los brazos de homología (homL y homR) y secuencia o secuencias de terminación. Las secuencias de aislantes se pueden generar para que estén exentas de secuencias de endonucleasas de restricción.

25 En algunos ejemplos, los vectores de la presente descripción comprenden una combinación de elementos proporcionados en el presente documento. En algunos casos, un vector para usar en los métodos proporcionados en el presente documento puede comprender un origen de replicación de R6K, un gen SacB, un gen PheS como marcador de contraselección y un marcador auxotrófico de levadura URA3 tal como, por ejemplo, el vector 1 (véase la Figura 55), que tiene la secuencia de ácido nucleico de SEQ ID NO. 214. En algunos casos, un vector para usar en los métodos proporcionados en el presente documento puede ser una versión alterada del vector 1, tal como, por ejemplo, 30 el vector 2 (véase la Figura 56), que tiene la secuencia de ácido nucleico de SEQ ID NO. 215. Además de los elementos previamente citados para el vector 1, el vector 2 puede comprender además los elementos que se encuentran en la Tabla 15. Los vectores adicionales para usar en los métodos proporcionados en el presente documento incluyen el vector 3 (ácido nucleico de SEQ ID NO. 216; Figura 57) y el vector 4 (ácido nucleico de SEQ ID NO. 217; Figura 58). Los vectores tanto 3 como 4 se construyeron a partir del contexto del vector 2. Sin embargo, en el vector 3, la secuencia de promotor de sacB se reemplazó con un promotor que contenía el promotor P2-MCD2 (véase Mutalik et al, *Nat Methods*. abril 2013; 10 (4): 354-60) y una versión optimizada de codones del gen pheS nativo que contiene las mutaciones T251A/A294G (véase Miyazaki, K. *Biotechniques*. 1 de febrero de 2015; 58 (2): 86-8), mientras que el 35 vector 4 comprende el contexto del vector 3 con el marcador de selección URA3 reemplazado por el marcador TRP1.

Tabla 15. Elementos de secuencias de selección del vector 2

Nombre del elemento	Secuencia de la parte
Aislante1	TATTCACACGCAATCAACAGGCAGGATAATCGCTGGTAAGGTCAGTGC TTTCTTCAGGTAGTAGAGATAACAATAGTTCCCAACGATAGGTGGCAGA TTTCACTTTACAGACCGACTGGTTCAGAAGCGTAGATAATAGCCCCTGT TTTCCAATAAGGGATAGTGTAGGTAAGTCAACTCCTCCGTCAGAGCCA ACCGTTT (SEQ ID NO. 218)
Aislante2	GCACTAGGACTTGCCGCGGATACTGCCCATACATGAATTGCAGCCT CAGGGACGTCAGTAGATCATGGAGGTAGGGCATAATGTCCTCTGTTGTT AAAATGTGAGTTCTCAACGAAGCACGAATCGGTCAGAACCTACTACTAA GGAGATTTGGTAGGTGCACGGTTTCTGTTCGCATAGACCAGTTCATTTC GATGTCT (SEQ ID NO. 219)
T1	GGCATCAAATAAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTT TTATCTGTTGTTTGTCTGGTGAACGCTCTCCTGAGTAGGACAAATCCGCC GCCCTAGA (SEQ ID NO. 220)

Nombre del elemento	Secuencia de la parte
B0015	CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAAGACTGGGCCTTTC GTTTTATCTGTTGTTTGTCCGGTGAACGCTCTCTACTAGAGTCACACTGG CTCACCTTCGGGTGGGCCTTCTGCGTTTATA (SEQ ID NO. 221)
promotor de SacB	TTGACAATTAATCATCCGGCTCGTAATTTATGTGGATCTTAATCATGCT AAGGAGGTTTTCTAATG (SEQ ID NO. 222)
promotor y secuencia de PheS	GGCGGTGTTGACATAAATACCACTGGCGGTGATACTGAGCACATCAGC AGGTCACACAGGAAAGTACTAGATGTCGCATCTTGCAGAATTAGTAGC TTCAGCGAAGGCCGCGATTTCTCAGGCGAGTGACGTCGCAGCACTGGA TAATGTACGTGTTGAGTACCTGGGAAAGAAGGGACACCTTACTCTTCA AATGACAACCCTGCGCGAACTGCCGCCGGAGGAACGCCCCGCAGCAG GAGCGGTAATCAATGAGGCAAAGGAGCAAGTACAACAGGCACTGAAC GCCCCGTAAGGCTGAGTTGGAATCCGCCGATTAACGCGCGCCTTGCT GCGGAAACCATTGATGTCTCGCTGCCCGGGCGCCGATTGAGAATGGA GGCTTACACCCAGTGACTCGTACCATCGACCGTATCGAATCTTCTTTG GCGAACTTGCTTCACTGTGGCAACTGGACCGGAGATTGAGGACGACT ACCACAATTTTCGATGCCTTGAACATTCCCAGTATCATCCTGCACGCGC CGATCATGATACATTTCTGGTTTGATACCACCCGTTTGTCTTCGTACCCAG ACAAGCGGTGTCCAAATCCGTACGATGAAGGCTCAGCAACCACCGATC CGTATCATTGCTCCAGGGCGCGTGTACCGTAACGATTATGACCAGACA CATAACCCGATGTTTCACCAAATGGAAGGGTTGATTGTGGATACGAAT ATCTCTTTACGAATCTGAAGGGCACCTTACATGATTTCTTACGCAACT TTTTCGAGGAGGACCTTCAAATTCGTTTTCGTCCATCGTACTTCCCTTTT GCAGAACCTTCGGCTGAAGTGGATGTAATGGGGAAAAACGGTAAGTG GCTGGAGGTTTTAGGTTGCGGGATGGTTTCATCCAAATGTGCTTCGCAAC GTCGGCATCGACCCCGAAGTCTACAGTGGATTGCGATTGCGGATGGGA ATGGAACGTCTGACTATGCTTCGTTACGGCGTAACGATTTGCGCTCCT TTTTGAGAACGATCTTCGTTTTCTGAAGCAATTCAAATAA (SEQ ID NO. 223)

En la figura 35 se muestra un flujo de trabajo general asociado con el ensamblaje de ADN según un ejemplo de la presente descripción.

5 En algunos ejemplos, el ensamblaje de ADN entre la cadena principal y los insertos (homL y homR) se realiza en levadura a través de recombinación de reparación de huecos de levaduras. En algunos ejemplos, el marcador auxotrófico (TRP o URA) está presente en el plásmido de cadena principal para la selección de ADN ensamblado en medio mínimo. Los plásmidos ensamblados después se extraen del cultivo de levadura.

10 En algunos ejemplos, los plásmidos extraídos se transforman a continuación en una cepa de *E. coli* que contiene el gen pir para propagar los plásmidos deseados para la siguiente transformación de la cepa de interés. Los transformantes se seleccionan por su resistencia a los antibióticos dados. Los transformantes se seleccionarán para secuenciar para seleccionar el plásmido de interés ensamblado correctamente.

15 En algunas realizaciones, los plásmidos ensamblados correctamente se transforman en células hospedantes para la transformación del genoma por electroporación. Dado que las células hospedantes diana no contienen la proteína pir, se espera que las colonias formadas en el medio selectivo de antibióticos tengan integración del plásmido en el locus deseado en el genoma. La correcta integración del plásmido se puede verificar por PCR con el cebador fuera de homL y homR, respectivamente, y la unión del cebador dentro del plásmido.

20 En algunos ejemplos, la presente descripción enseña métodos de salida de bucles para eliminar la cadena principal del plásmido del genoma del hospedante. Por lo tanto, en algunos ejemplos, la presente descripción enseña que la selección posterior en medios contraselectivos (sacarosa) y/o 4CP se puede usar para aislar los clones que no contienen la parte de la cadena principal del plásmido. Por lo tanto, en algunos ejemplos, las células hospedantes aisladas que comprenden el integrante correcto se inoculan en medio LB y el cultivo se siembra en placa de agar LB que contiene sacarosa y 4-p-cloro-fenilalanina (LB+suc+4CP). Debido a la sensibilidad de las células que expresan el gen sacB a la sacarosa, y la sensibilidad de las células que expresan PheS a 4CP, se espera que las colonias formadas en una placa de agar LB+suc+4CP tengan el gen mutante o de tipo natural de interés. La amplificación por PCR del o

de los nucleótidos diana y la secuenciación del producto de PCR permite aislar los nuevos clones con la modificación del genoma deseada.

En algunos ejemplos, los clones resultantes se secuencian para encontrar los clones con el o los cambios de nucleótidos deseados. En algunos ejemplos, todo el procedimiento anterior se puede realizar con el manipulador de líquidos.

Transformación de células hospedantes

En algunos ejemplos, los vectores de la presente descripción se pueden introducir en las células hospedantes usando cualquiera de una variedad de técnicas, que incluyen transformación, transfección, transducción, infección viral, pistolas de genes o transferencia de genes mediada por Ti (véase Christie, P.J., y Gordon, J.E., 2014 "The Agrobacterium Ti Plasmids" *Microbiol Spectr.* 2014; 2(6); 10.1128). Los métodos particulares incluyen transfección con fosfato cálcico, transfección mediada por DEAE-Dextrano, lipofección o electroporación (Davis, L., Dibner, M., Battey, I., 1986 "Basic Methods in Molecular Biology"). Otros métodos de transformación incluyen, por ejemplo, transformación de acetato de litio y electroporación. Véase, por ejemplo, Gietz et al., *Nucleic Acids Res.* 27:69-74 (1992); Ito et al., *J. Bacterol.* 153: 163-168 (1983); y Becker y Guarente, *Methods in Enzymology* 194: 182-187 (1991). En algunos ejemplos, las células hospedantes transformadas se denominan cepas hospedantes recombinantes.

En algunos ejemplos, la presente descripción enseña la transformación de alta capacidad de células usando la plataforma robótica de placas de 96 pocillos y las máquinas de manipulación de líquidos de la presente descripción.

En algunos ejemplos, la presente descripción enseña el cribado de células transformadas con uno o más marcadores de selección como se ha descrito antes. En uno de dichos ejemplos, las células transformadas con un vector que comprende un marcador de resistencia a la kanamicina (KanR) se siembran en placas en medios que contienen cantidades eficaces del antibiótico kanamicina. Se supone que las unidades formadoras de colonias visibles en medios con adición de kanamicina han incorporado el casete del vector en su genoma. La inserción de las secuencias deseadas se puede confirmar por PCR, análisis de enzimas de restricción y/o secuenciación del sitio de inserción relevante.

Salida de bucle de secuencias seleccionadas

En algunos ejemplos, la presente descripción enseña métodos para la salida de bucles de regiones seleccionadas de ADN de los organismos hospedantes. El método de salida de bucle puede ser como se describe en Nakashima et al. 2014 "Bacterial Cellular Engineering by Genome Editing and Gene Silencing." *Int. J. Mol. Sci.* 15(2), 2773-2793. En algunos ejemplos, la presente descripción enseña la salida de bucles de marcadores de selección de transformantes positivos. Las técnicas de eliminación por salida de bucle son conocidas en la técnica y se describen en (Tear et al. 2014 "Excision of Unstable Artificial Gene-Specific inverted Repeats Mediates Scar-Free Gene Deletions in Escherichia coli." *Appl. Biochem. Biotech* 175:1858-1867). Los métodos de salida de bucles usados en los métodos proporcionados en el presente documento se pueden llevar a cabo usando recombinación homóloga de cruzamiento simple o recombinación homóloga de cruzamiento doble. En un ejemplo, la salida de bucles de regiones seleccionadas como se describe en el presente documento puede implicar el uso de recombinación homóloga de cruzamiento simple como se describe en el presente documento.

Primero, los vectores de salida de bucle se insertan en regiones diana seleccionadas dentro del genoma del organismo hospedante (p. ej., mediante recombinación homóloga, CRISPR, recombinación mediada por lambda Red u otra técnica de edición de genes). En un ejemplo, la recombinación homóloga de un solo cruzamiento se usa entre un plásmido o vector circular y el genoma de la célula hospedante con el fin de introducir un bucle en el plásmido o vector circular tal como se muestra en la Figura 3. El vector insertado se puede diseñar con una secuencia que sea una repetición directa de una secuencia de hospedante cercana existente o introducida, de modo que las repeticiones directas flanqueen la región de ADN programada para bucle y eliminación. Una vez insertadas, las células que contienen el plásmido o vector del bucle de salida se pueden contraseleccionar para la eliminación de la región de selección (p. ej., véase la Figura 4; falta de resistencia al gen de selección). En la Figura 45 se muestran más ilustraciones del procedimiento de entrada y salida de bucle.

Los expertos en la técnica reconocerán que la descripción del procedimiento de salida de bucle representa solo un método ilustrativo para eliminar regiones no deseadas de un genoma. De hecho, los métodos de la presente descripción son compatibles con cualquier método para delecciones del genoma, que incluyen, pero no se limitan a edición de genes mediante lambda Red, CRISPR, TALENS, FOK u otras endonucleasas. Los expertos en la técnica también reconocerán la capacidad de reemplazar regiones no deseadas del genoma mediante técnicas de recombinación homóloga.

Edición de genes mediada por Lambda RED

Como se proporciona en el presente documento, la edición de genes como se describe en el presente documento se puede realizar usando recombinación homóloga mediada por Lambda Red como se describe por Datsenko y Wanner, *PNAS USA* 97: 6640-6645 (2000).

El sistema lambda Red deriva del bacteriófago lambda Red y su uso como herramienta de ingeniería genética se puede denominar ingeniería recombinogénica, abreviatura de ingeniería genética mediada por recombinación homóloga. Se puede usar para realizar una variedad de modificaciones: inserción y eliminación de secuencias seleccionables y no seleccionables, mutaciones puntuales u otros cambios pequeños de pares de bases, y la adición de marcadores de proteínas. También tiene la flexibilidad de modificar el cromosoma de *E. coli*, el ADN plasmídico o el ADN BAC. Para usar el sistema recombinogénico de lambda red para modificar el ADN diana, se puede electroporar un sustrato de ADN donante lineal (ya sea ADNbc o ADNmc - véase más abajo) en *E. coli* que expresa las enzimas de lambda red. Después estas enzimas catalizan la recombinación homóloga del sustrato con la secuencia de ADN diana. Esto significa que la clonación ocurre in vivo, en comparación con la clonación con enzimas de restricción donde los cambios genéticos ocurren en un tubo de ensayo. El sustrato de ADN donante solo requiere ~50 nucleótidos de homología con el sitio diana para la recombinación.

El sistema de recombinación lambda Red tiene tres componentes: 1) Exo, 2) Beta y 3) Gam. Los tres son necesarios para la recombinación con un sustrato de ADNbc; sin embargo, solo se requiere Beta cuando se genera una modificación con un sustrato de ADNmc.

Gam: Gam evita que las nucleasas endógenas tanto RecBCD como SbcCD digieran el ADN lineal introducido en una célula hospedante de *E. coli*.

Exo: Exo es una exonucleasa dependiente de ADNbc 5'→ 3'. Exo puede degradar el ADNbc lineal empezando en el extremo 5' y generar 2 productos posibles: 1) un dúplex de ADNbc parcialmente con extremos protuberantes 3' monocatenarios o 2) si el ADNbc era lo suficientemente corto, un ADNmc cuya cadena complementaria entera estaba degradada.

Beta: Beta puede proteger el ADNmc creado por Exo y promover su reasociación con un ADNmc complementario diana en la célula. Sólo se requiere la expresión de Beta para la recombinogénica con un sustrato oligo de ADNmc.

Para usar en el presente documento, un método de recombinogénica con lambda Red puede implicar diseñar y generar ADN sustrato; expresar los genes de recombinación de lambda Red; transformar (p. ej., electroporar) el ADN sustrato; desarrollar los transformantes; y seleccionar y confirmar clones recombinantes.

Diseño y generación de ADN sustrato

El que se use un sustrato de ADNbc o ADNmc lineal puede depender del objetivo del experimento. El sustrato de ADNbc puede ser mejor para inserciones o deleciones mayores de aproximadamente 20 nucleótidos, mientras que el sustrato de ADNmc puede ser mejor para mutaciones puntuales o cambios de solo unos pocos pares de bases.

Sustrato de ADNbc

Las inserciones de ADNbc se pueden hacer por PCR usando cebadores que amplifican la secuencia de ADN de interés y la flanquean con 50 pares de bases de homología con el sitio de inserción diana. Los cebadores pueden ser de ~70 nucleótidos de longitud (20 nucleótidos que se reasocian con la secuencia de ADN de interés y 50 nucleótidos de homología con las regiones que flanquean el sitio diana). Los insertos de ADNbc pueden incluir: inserciones o deleciones grandes, que incluyen fragmentos de ADN seleccionables, tales como genes de resistencia a antibióticos, así como fragmentos de ADN no seleccionables, tales como reemplazos de genes y marcadores.

Sustrato de ADNmc

Los sustratos de ADNmc pueden ser oligonucleótidos sintéticos o productos de PCR cortos. De cualquier manera, el sustrato sería de ~70-100 nucleótidos de longitud con la o las alteraciones deseadas ubicadas en el centro de la secuencia. Dado que el lambda Red tiene una frecuencia de recombinación más alta cuando se dirige a la hebra retrasada de ADN, es mejor determinar la dirección de replicación a través de una región diana de interés y diseñar un oligo que sea complementario a la hebra retrasada. En algunos casos, se diseñan oligos que se dirigen a ambas cadenas. Uno de los dos oligos se recombinará con mayor eficacia que el otro, lo que puede ayudar a identificar la hebra retrasada.

El sustrato de ADNmc puede ser más eficiente que el ADNbc con una frecuencia de recombinación entre 0,1% y 1%, y se puede aumentar tanto como al 25-50% diseñando oligos que eviten activar el sistema de reparación de emparejamientos erróneos dirigido por metilo (MMR). El trabajo de la MMR es corregir los emparejamientos erróneos de ADN que ocurren durante la replicación del ADN. La activación de la MMR se puede evitar: 1) usando una cepa de bacterias que tiene proteínas de MMR claves inactivadas, o 2) oligos de ADNmc especialmente diseñados para evitar la MMR: 1) *E. coli* con MMR inactivada: el uso de *E. coli* con MMR inactiva es definitivamente la más fácil de las dos opciones, pero estas células son propensas a mutaciones y pueden tener más cambios accidentales en sus genomas. 2) Diseñando oligos de ADNmc que evitan la activación de MMR: en un ejemplo, se introduce un emparejamiento erróneo de C/C en o dentro de 6 pares de bases del sitio de edición. En otro ejemplo, el cambio deseado está flanqueado por 4-5 cambios silenciosos en los codones de balanceo, es decir, hace cambios en el tercer par de bases de los 4-5 codones adyacentes que alteran la secuencia de nucleótidos pero no la secuencia de aminoácidos de la proteína traducida. Estos cambios pueden ser 5' o 3' del cambio deseado.

Expresión de genes de recombinación con lambda Red

El sistema de recombinogénica del lambda Red se puede expresar en una célula hospedante por: 1) A partir de una bacteria con profago defectuoso integrado 2) de un plásmido, 3) de mini-λ o 4) del propio fago lambda Red. El control de la expresión de las proteínas de Red es fundamental para minimizar los efectos tóxicos de la expresión de Gam y limitar las mutaciones espontáneas que se producen cuando Red se expresa de forma constitutiva. El sistema de recombinación que se use dependerá de qué tipo de ADN se quiere editar; sin embargo, el ADN de BAC se puede modificar con cualquiera de los enfoques descritos a continuación.

Cepa bacteriana con profago defectuoso integrado:

Existen una serie de cepas de *E. coli* que expresan de manera estable genes de ingeniería recombinogénica de lambda Red debido a la integración de un fago lambda Red defectuoso. Una de esas cepas es DY380, que deriva de la cepa DH10B de *E. coli*. Varias otras cepas bacterianas usadas habitualmente para la ingeniería recombinogénica se pueden encontrar en Thomason et al (Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination. Current Protocols in Molecular Biology. 106:V:1.16:1.16.1-1.16.39) y Sharan et al (Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. Nature protocols. 2009;4(2):206-223).

En algunas de estas cepas, la expresión de *exo*, *beta* y *gam* está estrechamente regulada por el promotor del fago endógeno pL y el represor CI. Para fines de ingeniería recombinogénica, se usa una versión sensible a la temperatura del gen represor, CI857. Este represor mutante evita la expresión de los genes de recombinación a bajas temperaturas (30-34°C). Cambiar las bacterias a 42°C durante 15 minutos inactiva rápidamente el represor y permite la expresión de los genes de recombinación. Después de esto, la disminución de la temperatura permite que el represor vuelva a la forma natural y reprima nuevamente la expresión de *exo*, *beta* y *gam*. Una ventaja principal de usar este método para la expresión del lambda Red es que no requiere selección de antibióticos para mantener la expresión del sistema recombinogénico. Esta configuración también se puede usar para modificar genes cromosómicos. Después del suceso de edición inicial, el profago defectuoso se puede eliminar del cromosoma de *E. coli* hospedante mediante un segundo suceso de recombinación con lambda Red. Alternativamente, si el alelo modificado es seleccionable, se puede transferir a un contexto genético diferente mediante la transducción P1.

Plásmido:

La expresión de genes de lambda Red a partir de un plásmido permite un sistema recombinogénico móvil, pero se requiere una regulación estricta de la expresión para un experimento con éxito. Los promotores habitualmente usados para controlar la expresión de Red incluyen el promotor lac inducible por IPTG, el promotor pBAD inducible por arabinosa y el promotor pL del fago endógeno. Los plásmidos que también expresan los represores asociados con estos promotores (*lacI*, *araC*, *ci857*) se pueden usar en algunos casos con el fin de limitar la expresión con fugas del sistema Red. El uso de un plásmido para expresar el sistema recombinogénico lambda se puede usar para editar el ADN cromosómico bacteriano porque es fácil eliminar el sistema recombinogénico una vez que se han generado los clones recombinantes. Una forma sencilla de hacer esto es expresar los genes de lambda Red a partir de un plásmido con un origen de replicación sensible al calor. Una vez que el sistema recombinogénico ya no es necesario, las bacterias se pueden "curar" cultivándolas a 42°C.

Mini-λ:

Un híbrido entre el uso de un plásmido y la integración estable de un profago defectuoso es usar mini-λ, un trozo circular de ADN de fago defectuoso que no se replica, que cuando se introduce en las bacterias, se integra en el genoma. Mini-λ usa el promotor de Red endógeno pL y el represor *ci857* para regular la expresión. Se puede usar un antibiótico para seleccionar clones positivos pero, debido a que mini-λ se integra de manera estable, no se requiere la selección con fármacos para el mantenimiento. Un cambio de temperatura a 42°C no solo permite la activación de los genes de Red necesarios para la recombinación, sino que también conduce a la expresión de los genes *int* y *xis* que son responsables de la escisión de mini-λ del cromosoma del hospedante. Después de esto, mini-λ se puede purificar fácilmente de las bacterias igual que un plásmido.

Fago:

Otra opción para expresar el sistema Red puede ser usar un fago lambda Red, λTetR, que lleva el gen de resistencia a la tetraciclina y el represor de lambda Red *ci857*. Una vez introducido, el profago es estable y ya no requiere la selección con fármacos. Un inconveniente de este enfoque es que requiere la generación de bacteriófagos, que no es una técnica común de biología molecular. Sin embargo, una ventaja de este método es que se puede integrar de manera estable el sistema Red en una cepa de interés y ya y podría usarse la transducción P1 para mover la modificación a un contexto diferente, si es necesario. Este enfoque es el más adecuado para modificar plásmidos o BAC porque da como resultado una integración estable del fago en el genoma de *E. coli*.

Selección y confirmación de clones recombinantes

Si se ha insertado un gen de resistencia a antibióticos, los recombinantes se pueden seleccionar primero mediante la resistencia a los antibióticos, pero todos los clones deben someterse a ensayos adicionales para confirmar la presencia

de la modificación deseada. La PCR de colonias se puede usar para cribar clones positivos en la mayoría de los casos, y la digestión con enzimas de restricción se puede usar para cribar los plásmidos con mutaciones adecuadas. Las mutaciones puntuales y otros cambios sutiles se pueden confirmar por secuenciación, que también se puede usar para la confirmación de todos los clones, independientemente del tipo de ADN que sea el objetivo de modificación: el cromosoma de *E. coli*, un plásmido o un BAC.

Edición de genes mediada por CRISPR

En un ejemplo proporcionado en el presente documento, el genoma de una célula hospedante se puede modificar por CRISPR. En los Ejemplos 18 y 19 se puede encontrar un ejemplo ilustrativo para usar el sistema CRISPR/Cas9 para la edición de genes en *E. coli*.

El sistema CRISPR/Cas es un sistema inmunitario procarionota que confiere resistencia a elementos genéticos extraños, tales como los presentes en plásmidos y fagos, y que proporciona una forma de inmunidad adquirida. CRISPR significa repeticiones palindrómicas cortas agrupadas y regularmente interespaciadas (por sus siglas en inglés: *Clustered Regularly Interspaced Short Palindromic Repeat*), y cas significa sistema asociado a CRISPR, y se refiere a los genes cas pequeños asociados con el complejo de CRISPR.

Los sistemas CRISPR-Cas se caracterizan más ampliamente como sistemas de clase 1 o clase 2. La principal característica distintiva entre estos dos sistemas es la naturaleza del módulo efector de Cas. Los sistemas de clase 1 requieren el ensamblaje de múltiples proteínas Cas en un complejo (denominado "complejo en cascada") para mediar la interferencia, mientras que los sistemas de clase 2 usan una sola enzima Cas grande para mediar la interferencia. Cada uno de los sistemas de clase 1 y clase 2 se divide además en múltiples tipos de CRISPR-Cas basados en la presencia de una proteína Cas específica. Por ejemplo, el sistema de clase 1 se divide en los siguientes tres tipos: sistemas de tipo I, que contienen la proteína Cas3; sistemas de tipo III, que contienen la proteína Cas10; y los sistemas putativos de tipo IV, que contienen la proteína Csf1, una proteína similar a Cas8. Los sistemas de clase 2 en general son menos comunes que los sistemas de clase 1 y se dividen además en los tres tipos siguientes: sistemas de tipo II, que contienen la proteína Cas9; sistemas de tipo V, que contienen proteína Cas12a (anteriormente conocida como Cpf1, y denominada Cpf1 en el presente documento), Cas12b (anteriormente conocida como C2c1), Cas12c (anteriormente conocida como C2c3), Cas12d (anteriormente conocida como CasY) y Cas12e (anteriormente conocido como CasX); y sistemas tipo VI, que contienen Cas13a (anteriormente conocida como C2c2), Cas13b y Cas13c. Pyzocha et al., *ACS Chemical Biology*, vol. 13 (2), págs. 347-356. En un ejemplo, el sistema CRISPR-Cas para usar en los métodos proporcionados en el presente documento es un sistema de clase 2. En un ejemplo, el sistema CRISPR-Cas para usar en los métodos proporcionados en el presente documento es un sistema de clase 2 de tipo II, tipo V o tipo VI. En un ejemplo, el sistema CRISPR-Cas para usar en los métodos proporcionados en el presente documento se selecciona de Cas9, Cas12a, Cas12b, Cas12c, Cas12d, Cas12e, Cas13a, Cas13b, Cas13c u homólogos, ortólogos o parálogos de los mismos.

Los sistemas CRISPR usados en los métodos descritos en el presente documento comprenden un módulo efector de Cas que comprende una o más nucleasas asociadas a CRISPR (Cas) guiadas por ácidos nucleicos, denominadas en el presente documento proteínas efectoras Cas. En algunos ejemplos, las proteínas Cas pueden comprender uno o varios dominios de nucleasa. Una proteína efectora Cas se puede dirigir a moléculas de ácido nucleico monocatenarias o bicatenarias (p. ej. ácidos nucleicos de ADN o ARN) y puede generar roturas de cadena doble o simple. En algunos ejemplos, las proteínas efectoras Cas son proteínas Cas de tipo natural o que se encuentran de forma natural. En algunos ejemplos, las proteínas efectoras Cas son proteínas Cas mutantes, en donde se hacen una o más mutaciones, inserciones o deleciones en una proteína Cas de origen que se encuentra de forma natural o WT (p. ej., una proteína Cas parental) para producir una proteína Cas con una o más características alteradas en comparación con la proteína Cas parental.

En algunos casos, la proteína Cas es una nucleasa de tipo natural (WT). Los ejemplos no limitantes de proteínas Cas adecuadas para usar en la presente descripción incluyen C2c1, C2c2, C2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 y Csx12), Cas10, Cpf1, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx100, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, MAD1-20, SmCsm1, homólogos de las mismas, ortólogos de las mismas, variantes de las mismas, mutantes de las mismas o versiones modificadas de las mismas. Las nucleasas guiadas por ácido nucleico adecuadas (p. ej., Cas 9) pueden ser de un organismo de un género, que incluye, pero no se limita a: *Thiomicrospira*, *Succinivibrio*, *Candidatus*, *Porphyromonas*, *Acidomonococcus*, *Prevotella*, *Smithella*, *Moraxella*, *Synergistes*, *Francisella*, *Leptospira*, *Catenibacterium*, *Kandleria*, *Clostridium*, *Dorea*, *Coprococcus*, *Enterococcus*, *Fructobacillus*, *Weissella*, *Pediococcus*, *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Roseburia*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma*, *Alicyclobacillus*, *Brevibacillus*, *Bacillus*, *Bacteroidetes*, *Brevibacillus*, *Carnobacterium*, *Clostridiaridium*, *Clostridium*, *Desulfonatronum*, *Desulfovibrio*, *Helcococcus*, *Leptotrichia*, *Listeria*, *Methanomethyophilus*, *Methylobacterium*, *Opatutaceae*, *Paludibacter*, *Rhodobacter*, *Sphaerochaeta*, *Tuberibacillus*, y *Campylobacter*. Las especies de organismos de dicho género pueden ser por lo demás como se describe en el presente documento.

- Las nucleasas guiadas por ácido nucleico adecuadas (p. ej., Cas9) pueden ser de un organismo de un filo, que incluye, pero no se limita a: Firmicutes, Actinobacterias, Bacteroidetes, Proteobacterias, Espiroquetas y Tenericutes. Las nucleasas guiadas por ácido nucleico adecuadas pueden ser de un organismo de una clase, que incluye, pero no se limita a: Erysipelotrichia, Clostridia, Bacilli, Actinobacteria, Bacteroidetes, Flavobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Epsilonproteobacteria, Spirochaetes y Mollicutes.
- Las nucleasas guiadas por ácido nucleico adecuadas pueden ser de un organismo de un orden, que incluye, pero no se limita a: Clostridiales, Lactobacillales, Actinomycetales, Bacteroidales, Flavobacteriales, Rhizobiales, Rhodospirillales, Burkholderiales, Neisseriales, Legionellales, Nautiliales, Campylobacterales, Mycoplasmachaeetales, y Thiotrichales. Las nucleasas guiadas por ácido nucleico adecuadas pueden ser de un organismo de dentro de una familia, que incluye, pero no se limita a: Lachnospiraceae, Enterococcaceae, Leuconostocaceae, Lactobacillaceae, Streptococcaceae, Peptostreptococcaceae, Staphylococcaceae, Eubacteriaceae, Corynebacterineae, Bacteroidaceae, Flavobacterium, Cryomorphaceae, Rhodobiaceae, Rhodospirillaceae, Acetobacteraceae, Sutterellaceae, Neisseriaceae, Legionellaceae, Nautiliaceae, Campylobacteraceae, Spirochaetaceae, Mycoplasmataceae, y Francisellaceae.
- Otras nucleasas guiadas por ácido nucleico (p. ej., Cas9) adecuadas para usar en los métodos, sistemas y composiciones de la presente descripción incluyen las derivadas de un organismo tal como, pero no limitado a: *Thiomicrospira* sp. XS5, *Eubacterium rectale*, *Succinivibrio dextrinosolvens*, *Candidatus Methanoplasma termitum*, *Candidatus Methanomethylophilus alvus*, *Porphyromonas crevioricanis*, *Flavobacterium branchiophilum*, *Acidomonococcus* sp., *Lachnospiraceae bacterium* COE1, *Prevotella brevis* ATCC 19188, *Smithella* sp. SCADC, *Moraxella bovoculi*, *Synergistes jonesii*, *Bacteroidetes* oral taxón 274, *Francisella tularensis*, *Leptospira inadai* serovar Lyme str. 10, *Acidomonococcus* sp. estructura cristalina (5B43) *S. mutans*, *S. agalactiae*, *S. equisimilis*, *S. sanguinis*, *S. pneumoniae*; *C. jejuni*, *C. coli*; *N. salsuginis*, *N. tergaricus*; *S. auricularis*, *S. carnosus*; *N. meningitidis*, *N. gonorrhoeae*; *L. monocytogenes*, *L. ivanovii*; *C. botulinum*, *C. difficile*, *C. tetani*, *C. sordellii*; *Francisella tularensis* I, *Prevotella albensis*, *Lachnospiraceae bacterium* MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria bacterium* GW2011_GWA2_33_10, *Parcubacteria bacterium* GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Microgenomates*, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae bacterium* MA2020, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae bacterium* ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, *Porphyromonas macacae*, *Catenibacterium* sp. CAG:290, *Kandleria vitulina*, Clostridiales bacterium KA00274, *Lachnospiraceae bacterium* 3-2, *Dorea longicatena*, *Coprococcus catus* GD/7, *Enterococcus columbae* DSM 7374, *Fructobacillus* sp. EFB-N1, *Weissella halotolerans*, *Pediococcus acidilactici*, *Lactobacillus curvatus*, *Streptococcus pyogenes*, *Lactobacillus versmoldensis*, y *Filifactor alocis* ATCC 35896. Véanse, las patentes de EE.UU. nº 8.697.359; 8.771.945; 8.795.965; 8.865.406; 8.871.445; 8.889.356; 8.895.308; 8.906.616; 8.932.814; 8.945.839; 8.993.233; 8.999.641; 9.822.372; 9.840.713; solicitud de patente de EE.UU. nº 13/842.859 (US 2014/0068797 A1); 9.260.723; 9.023.649; 9.834.791; 9.637.739; solicitud de patente de EE.UU. nº 14/683.443 (US 2015/0240261 A1); solicitud de patente de EE.UU. nº 14/743.764 (US 2015/0291961 A1); 9.790.490; 9.688.972; 9.580.701; 9.745.562; 9.816.081; 9.677.090; 9.738.687; solicitud de EE.UU. nº 15/632.222 (US 2017/0369879 A1); solicitud de EE.UU. nº 15/631.989; solicitud de EE.UU. nº 15/632.001; y patente de EE.UU. nº 9.896.696.

En algunos ejemplos, una proteína efectora Cas comprende una o más de las siguientes actividades:

- una actividad de nickasa, es decir, la capacidad de escindir una sola cadena de una molécula de ácido nucleico;
- una actividad nucleasa bicatenaria, es decir, la capacidad de escindir ambas cadenas de un ácido nucleico bicatenario y crear una rotura bicatenaria;
- una actividad de endonucleasa;
- una actividad de exonucleasa; y/o
- una actividad helicasa, es decir, la capacidad de desenrollar la estructura helicoidal de un ácido nucleico bicatenario.

En los ejemplos de la descripción, la expresión "ácido nucleico guía" se refiere a un polinucleótido que comprende 1) una secuencia guía capaz de hibridar con una secuencia diana (denominada en el presente documento como "segmento de direccionamiento") y 2) una secuencia de armazón capaz de interaccionar con (ya sea sola o en combinación con una molécula de ARNtracr) una nucleasa guiada por ácido nucleico como se describe en el presente documento (denominada en el presente documento "segmento de armazón"). Un ácido nucleico guía puede ser ADN. Un ácido nucleico guía puede ser ARN. Un ácido nucleico guía puede comprender tanto ADN como ARN. Un ácido nucleico guía puede comprender nucleótidos modificados que no se encuentran de forma natural. En los casos en donde el ácido nucleico guía comprende ARN, el ácido nucleico guía de ARN puede estar codificado por una secuencia de ADN en una molécula de polinucleótido tal como un plásmido, construcción lineal o casete de edición como se describe en el presente documento.

En algunos ejemplos, los ácidos nucleicos guía descritos en el presente documento son ácidos nucleicos guía de ARN ("ARN guías" o "ARNg") y comprenden un segmento de direccionamiento y un segmento de armazón. En algunos

ejemplos, el segmento de armazón de un ARNg está comprendido en una molécula de ARN y el segmento de direccionamiento está comprendido en otra molécula de ARN separada. Dichos ejemplos se denominan en el presente documento "ARNg de doble molécula" o "ARNg de dos moléculas" o "ARNg duales". En algunos ejemplos, el ARNg es una sola molécula de ARN y se denomina en el presente documento "ARN de guía única" o "ARNgu". El término "ARN guía" o "ARNg" es inclusivo y se refiere tanto a ARN guía de dos moléculas como a ARNgu.

El segmento de direccionamiento al ADN de un ARNg comprende una secuencia de nucleótidos que es complementaria a una secuencia en una secuencia de ácido nucleico diana. Así pues, el segmento de direccionamiento de un ARNg interacciona con un ácido nucleico diana de una manera específica de secuencia a través de hibridación (es decir, emparejamiento de bases), y la secuencia de nucleótidos del segmento de direccionamiento determina la ubicación dentro del ADN diana a la que se unirá el ARNg. El grado de complementariedad entre una secuencia guía y su secuencia diana correspondiente, cuando se alinean de manera óptima usando un algoritmo de alineamiento adecuado, es aproximadamente de o más de aproximadamente 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97,5%, 99% o más. El alineamiento óptimo se puede determinar con el uso de cualquier algoritmo adecuado para alinear secuencias. En algunos ejemplos, una secuencia de guía es de aproximadamente o más de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, 75, o más nucleótidos de longitud. En algunos ejemplos, una secuencia guía es de menos de aproximadamente 75, 50, 45, 40, 35, 30, 25, 20 nucleótidos de longitud. En ejemplos, la secuencia guía es de 10-30 nucleótidos de longitud. La secuencia guía puede ser de 15-20 nucleótidos de longitud. La secuencia guía puede ser de 15 nucleótidos de longitud. La secuencia guía puede ser de 16 nucleótidos de longitud. La secuencia guía puede ser de 17 nucleótidos de longitud. La secuencia guía puede ser de 18 nucleótidos de longitud. La secuencia guía puede ser de 19 nucleótidos de longitud. La secuencia guía puede ser de 20 nucleótidos de longitud.

El segmento de armazón de un ARN guía interacciona con una o más proteínas efectoras Cas para formar un complejo de ribonucleoproteína (denominado en el presente documento CRISPR-RNP o complejo RNP). El ARN guía dirige el polipéptido unido a una secuencia de nucleótidos específica dentro de una secuencia de ácido nucleico diana a través del segmento de direccionamiento descrito antes. El segmento de armazón de un ARN guía comprende dos tramos de nucleótidos que son complementarios entre sí y que forman un dúplex de ARN bicatenario. Una secuencia suficiente dentro de la secuencia de armazón para promover la formación de un complejo de nucleasa direccionable puede incluir un grado de complementariedad a lo largo de la longitud de dos regiones de secuencia dentro de la secuencia de armazón, tal como una o dos regiones de secuencia implicadas en la formación de una estructura secundaria. En algunos casos, la una o dos regiones de secuencia están comprendidas o codificadas en el mismo polinucleótido. En algunos casos, la una o dos regiones de secuencia están comprendidas o codificadas en polinucleótidos separados. El alineamiento óptimo se puede determinar por cualquier algoritmo de alineamiento adecuado y puede tener en cuenta además las estructuras secundarias, tales como la autocomplementariedad dentro de una o dos regiones de secuencia. En algunos ejemplos, el grado de complementariedad entre la una o dos regiones de secuencia a lo largo de la longitud de la más corta de las dos cuando están alineadas de manera óptima es de aproximadamente o más de aproximadamente 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97,5%, 99% o mayor. En algunos ejemplos, al menos una de las dos regiones de secuencia es de aproximadamente o más de aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50 o más nucleótidos de longitud.

Una secuencia de armazón de un ARNg objeto puede comprender una estructura secundaria. Una estructura secundaria puede comprender una región de pseudonudo o una estructura de tallo-bucle. En algunos ejemplos, la compatibilidad de un ácido nucleico guía y una nucleasa guiada por ácido nucleico está determinada al menos parcialmente por la secuencia dentro o adyacente a la región de estructura secundaria del ARN guía. En algunos casos, la cinética de unión de un ácido nucleico guía a una nucleasa guiada por ácido nucleico está determinada en parte por estructuras secundarias dentro de la secuencia de armazón. En algunos casos, la cinética de unión de un ácido nucleico guía a una nucleasa guiada por ácido nucleico se determina en parte por la secuencia de ácido nucleico con la secuencia de armazón.

Se puede encontrar una secuencia de armazón compatible para una combinación de ARNg-proteína efectora Cas escaneando secuencias adyacentes a locus de una nucleasa Cas nativa. En otras palabras, las nucleasas Cas nativas pueden ser codificadas en un genoma en las proximidades de un ácido nucleico guía compatible correspondiente o una secuencia de armazón.

Las nucleasas guiadas por ácido nucleico pueden ser compatibles con ácidos nucleicos guía que no se encuentran dentro de las nucleasas endógenas del hospedante. Dichos ácidos nucleicos guía ortogonales se pueden determinar por ensayos empíricos. Los ácidos nucleicos guía ortogonales pueden provenir de diferentes especies bacterianas o ser sintéticos o transformados de otro modo para que no sean de origen natural. Los ácidos nucleicos guía ortogonales que son compatibles con una nucleasa guiada por ácido nucleico común pueden comprender una o más características comunes. Las características comunes pueden incluir una secuencia fuera de una región de pseudonudo. Las características comunes pueden incluir una región de pseudonudo. Las características comunes pueden incluir una secuencia primaria o estructura secundaria

Se puede transformar un ácido nucleico guía para que dirija a una secuencia diana deseada alterando la secuencia guía de modo que la secuencia guía sea complementaria a la secuencia diana, permitiendo así la hibridación entre la

secuencia guía y la secuencia diana. Un ácido nucleico guía con una secuencia guía transformada se puede denominar ácido nucleico guía transformado. Los ácidos nucleicos guía transformados a menudo no son naturales y no se encuentran en la naturaleza.

- 5 En algunos ejemplos, la presente descripción proporciona un polinucleótido que codifica un ARNg. En algunos ejemplos, un ácido nucleico que codifica el ARNg está comprendido en un vector de expresión, p. ej., un vector de expresión recombinante. En algunos ejemplos, la presente descripción proporciona un polinucleótido que codifica un polipéptido modificador dirigido al sitio. En algunos ejemplos, el polinucleótido que codifica un polipéptido modificador dirigido al sitio está comprendido en un vector de expresión, p. ej., un vector de expresión recombinante.

EJEMPLOS

- 10 Los ejemplos 1-9 son demostraciones de la plataforma de ingeniería genómica de HTP en *Corynebacterium*. Sin embargo, se han adaptado procedimientos similares para *E. coli* y los autores de la invención los están llevando a cabo con éxito.

A continuación se proporciona un breve índice con el único fin de ayudar al lector. Nada en este índice pretende limitar el alcance de los ejemplos o la descripción de la solicitud.

15 **Tabla 5.1 - Índice para la sección de ejemplos**

Ejemplo nº	Título	Breve descripción
1	Transformación de HTP de <i>Corynebacterium</i> y demostración de la creación de bibliotecas de SNP	Describe ejemplos de los métodos de ingeniería genética de alta capacidad de la presente descripción
2	Ingeniería genómica de HTP - implementación de una biblioteca de SNP para rehabilitar/mejorar una cepa microbiana industrial	Describe enfoques para rehabilitar organismos industriales a través de métodos de intercambio de SNP de la presente descripción.
3	Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de SNP para mejorar el rendimiento de cepas en la producción de lisina en <i>Corynebacterium</i> .	Describe una implementación de técnicas de intercambio de SNP para mejorar el rendimiento de cepas de <i>Corynebacterium</i> que producen lisina. También describe consolidaciones de mutaciones seleccionadas de segundo y tercer orden.
4	Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de promotores para mejorar una cepa microbiana industrial	Describe métodos para mejorar el rendimiento de cepas de organismos hospedantes a través de bibliotecas de diseño genético de intercambio de PRO de la presente descripción.
5	Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de PRO para mejorar el rendimiento de la cepa para la producción de lisina	Describe una implementación de técnicas de intercambio de PRO para mejorar el rendimiento de cepas de <i>Corynebacterium</i> que producen lisina.
6	Mapeo de epistasias - una herramienta algorítmica para predecir consolidaciones de mutaciones beneficiosas	Describe un ejemplo de las herramientas/algoritmos automatizados de la presente descripción para predecir consolidaciones de mutaciones genéticas beneficiosas.
7	Ingeniería genómica de HTP - consolidación de mutaciones de intercambio de PRO y ensayos combinatorios multifactoriales	Describe e ilustra la capacidad de los métodos de HTP de la presente descripción para explorar de forma eficaz el gran espacio de soluciones creado por la consolidación combinatoria de múltiples combinaciones de genes/biblioteca de diseño genético.
8	Ingeniería genómica de HTP - implementación de una biblioteca de terminadores para mejorar una cepa hospedante industrial	Describe e ilustra una aplicación de las bibliotecas de diseño genético de intercambio de STOP de la presente descripción.

Ejemplo nº	Título	Breve descripción
9	Comparación de conjuntos de herramientas de HTP frente a mutaciones por UV tradicionales	Proporciona resultados experimentales que comparan los métodos de diseño genético de HTP de la presente descripción frente a los programas tradicionales de mejora de cepas mutacionales.
10	Ingeniería genómica de HTP - implementación de una biblioteca de mutagénesis de transposones para mejorar el rendimiento de la cepa de <i>Escherichia coli</i>	Describe ejemplos del uso de una biblioteca de mutagénesis de transposones con los métodos de ingeniería genética de alta capacidad de la presente descripción, como se aplica a células de <i>E. coli</i> .
11	Ingeniería genómica de HTP - generación de cadenas principales de vectores para usar en ingeniería genómica de HTP en <i>Escherichia coli</i>	Describe ejemplos de generación de cadenas principales de vectores para usar en los métodos de ingeniería genética de alta capacidad de la presente descripción, como se aplica a células de <i>E. coli</i> .
12	Ingeniería genómica de HTP - generación y ensayo de una biblioteca de intercambio de promotores adicional para usar en la mejora de una cepa microbiana industrial	Describe métodos para diseñar, generar y ensayar bibliotecas de diseño genético de intercambio de PRO de la presente descripción.
13	Ingeniería genómica de HTP - ensayo de integración de la biblioteca de intercambio de promotores de la tabla 1.4 en el genoma de <i>E. coli</i> usando la cadena principal del vector 2	Describe la prueba del concepto del uso de la cadena principal del vector 2 del Ejemplo 11 en combinación con promotores de forma de la biblioteca de promotores de la Tabla 1.4 para conducir la integración de una única copia de una construcción de promotor-gen heteróloga en el genoma de <i>E. coli</i>
14	Ingeniería genómica de HTP - implementación de métodos PROSWP que usan una biblioteca de promotores derivada de la Tabla 1.4.	Describe métodos para mejorar el rendimiento de la cepa de los organismos hospedantes mediante bibliotecas de diseño genético de intercambio de PRO derivadas de la Tabla 1.4 de la presente descripción.
15	Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de TERMINADORES para mejorar el rendimiento de la cepa para la producción de licopeno	Describe métodos para mejorar el rendimiento de cepas de organismos hospedantes a través de bibliotecas de diseño genético de intercambio de terminadores de la presente descripción.
16	Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de TERMINADORES o una biblioteca de intercambio de PRO en combinación con una biblioteca de intercambio de MARCADORES DE SOLUBILIDAD o una biblioteca de intercambio de marcadores de DEGRADACIÓN para mejorar el rendimiento de la cepa para la producción de licopeno	Describe métodos para mejorar el rendimiento de cepas de organismos hospedantes a través de bibliotecas de diseño genético de intercambio de terminadores, intercambio de marcadores de solubilidad e intercambio de marcadores de degradación de la presente descripción.

Ejemplo 1: Transformación de HTP de *Corynebacterium* y demostración de la creación de bibliotecas de SNP

Este ejemplo ilustra ejemplos de los métodos de ingeniería genética de HTP de la presente descripción. Las células hospedantes se transforman con una variedad de secuencias de SNP de diferentes tamaños, todas dirigidas a diferentes zonas del genoma. Los resultados demuestran que los métodos de la presente descripción pueden generar cambios genéticos rápidos de cualquier tipo, en todo el genoma de una célula hospedante.

5

A. Clonación de vectores de transformación

Se eligieron al azar una variedad de SNP de *Corynebacterium glutamicum* (ATCC21300) y se clonaron en vectores de clonación de *Corynebacterium* usando técnicas de clonación de recombinación homóloga de levaduras para

ensamblar un vector en el que cada SNP estaba flanqueado por regiones de repetición directa, como se describe más arriba en la sección "Ensamblaje/Clonación de plásmidos a medida", y como se ilustra en la Figura 3.

Los casetes de SNP para este ejemplo se diseñaron para incluir un intervalo de longitudes de brazo de repetición directa de homología que varían de 0,5 Kb, 1 Kb, 2 Kb y 5 Kb. Además, los casetes de SNP se diseñaron para la recombinación homóloga dirigida a varias regiones distintas del genoma, como se describe con más detalle a continuación.

El genoma de *C. glutamicum* tiene un tamaño de 3.282.708 pb (véase la Figura 9). El genoma se dividió arbitrariamente en 24 regiones genéticas de igual tamaño y se diseñaron casetes de SNP para dirigirse a cada una de las 24 regiones. Por lo tanto, se clonó un total de 96 plásmidos distintos para este Ejemplo (4 tamaños de inserto diferentes x 24 regiones genómicas distintas).

Cada inserto de ADN se produjo por amplificación por PCR de regiones homólogas usando oligos de origen comercial y el ADN genómico de la cepa hospedante descrito antes como molde. El SNP a introducir en el genoma se codificó en las colas de oligos. Los fragmentos de PCR se ensamblaron en la cadena principal del vector usando recombinación homóloga en levaduras.

La clonación de cada SNP y brazo de homología en el vector se llevó a cabo de acuerdo con el flujo de trabajo de ingeniería de HTP descrito en la Figura 6A-B, Figura 3 y Tabla 5.

B. Transformación de clones ensamblados en *E. coli*

Los vectores se transformaron inicialmente en *E. coli* usando técnicas estándar de transformación por choque térmico con el fin de identificar clones ensamblados correctamente y para amplificar el ADN del vector para la transformación de *Corynebacterium*.

Por ejemplo, se analizó en bacterias de *E. coli* transformadas el éxito del ensamblaje. Se cultivaron cuatro colonias de cada placa de transformación de *E. coli* y se analizó el ensamblaje correcto por PCR. Este procedimiento se repitió para cada una de las 24 ubicaciones de transformación y para cada uno de los 4 tamaños de inserto diferentes (es decir, para los 96 transformantes de este ejemplo). Los resultados de este experimento se representaron como el número de colonias correctas identificadas de las cuatro colonias que se analizaron para cada tratamiento (tamaño del inserto y ubicación genómica) (véase la Figura 12). Los insertos más largos de 5 kb mostraban una disminución en la eficiencia del ensamblaje en comparación con los homólogos más cortos (n = 96).

C. Transformación de clones ensamblados en *Corynebacterium*

Los clones validados se transformaron en células hospedantes de *Corynebacterium glutamicum* mediante electroporación. Para cada transformación, se determinó el número de Unidades Formadoras de Colonias (UFC) por µg de ADN en función del tamaño del inserto (véase la Figura 13). La integración del genoma de *Coryne* también se analizó en función de la longitud de los brazos de homología, y los resultados mostraron que los brazos más cortos tenían una menor eficiencia (véase la Figura 13).

La eficiencia de la integración genómica también se analizó con respecto a la ubicación del genoma diana en transformantes de *C. glutamicum*. Las posiciones genómicas 1 y 2 presentaban una eficiencia de integración ligeramente menor en comparación con el resto del genoma (véase la Figura 10).

D. Salida de bucles de marcadores de selección de

Los cultivos de *Corynebacterium* identificados por tener integraciones satisfactorias del casete de inserto se cultivaron en medio que contenía sacarosa al 5% para la contraselección para bucles de salida del gen de selección *sacB*. La frecuencia de resistencia a la sacarosa para varios brazos de repetición directa de homología no variaba significativamente con la longitud de los brazos (véase la Figura 14). Estos resultados sugerían que las eficiencias de salida de bucle permanecían estables en las longitudes de los brazos de homología de 0,5 kb a 5 kb.

Con el fin de validar adicionalmente los sucesos de salida de bucle, se cultivaron y analizaron por secuenciación las colonias que presentaban resistencia a la sacarosa.

Los resultados de la secuenciación de las regiones genómicas del inserto se resumen en la Tabla 6 a continuación.

Tabla 6 - Frecuencia de validación de salida de bucle

Resultado	Frecuencia (error de toma de muestra al 95% de confianza)
Salida de bucle con éxito	13% (9%/20%)

Bucle todavía presente	42% (34%/50%)
Lectura mixta	44% (36%/52%)

Los resultados de la secuenciación mostraban una eficiencia de 10- 20% en las salidas de bucles. La salida de bucle real probablemente depende en cierta medida de la secuencia del inserto. Sin embargo, la selección de 10-20 colonias resistentes a la sacarosa conduce a altas tasas de éxito.

5 E. Resumen

La Tabla 7 a continuación proporciona una evaluación cuantitativa de las eficiencias de los métodos de ingeniería genómica de HTP de la presente descripción. Las tasas de ensamblaje de construcciones para las metodologías de homología de levaduras proporcionaban construcciones de ADN esperadas en casi 9 de cada 10 colonias analizadas. Las transformaciones de *Coryne* de construcciones de SNP con brazos de homología de 2 kb proporcionaban un promedio de 51 unidades formadoras de colonias por microgramo de ADN (UFC/ μ g), presentando 98% de dichas colonias inserciones de SNP correctamente integradas (eficacia de direccionamiento). Las eficiencias de salida de bucles se mantuvieron en el 0,2% de las células que se volvieron resistentes cuando se exponían a la sacarosa, y el 13% de estas presentaban secuencias de salida de bucle correctas.

Tabla 7- Resumen de resultados para la ingeniería de cepas de *Corynebacterium glutamicum*

Etapas de CC	Resultados para brazos de homología de 2 kb
Éxito del ensamblaje de las construcciones	87%
Eficiencia de transformación de <i>Coryne</i>	51 CFU/ μ g DNA (+/- 15)
Eficiencia del direccionamiento	98%
Eficiencia de la salida de bucle	0,2% (+/- 0,03%)

15 Ejemplo 2: Ingeniería genómica de HTP - implementación de una biblioteca de SNP para rehabilitar/mejorar una cepa microbiana industrial

Este ejemplo ilustra varios ejemplos de las bibliotecas de intercambio de SNP de los programas de mejora de cepas de HTP de la presente descripción. Específicamente, el ejemplo ilustra varios enfoques previstos para rehabilitar cepas industriales actualmente existentes. Este ejemplo describe los enfoques de onda ascendente y descendente para explorar el espacio de soluciones fenotípicas creado por las múltiples diferencias genéticas que pueden estar presentes entre cepas "base", "intermedias" e industriales.

A. Identificación de SNP en el grupo de diversidad

Se llevó a cabo un programa de mejora de cepas de ejemplo usando los métodos de la presente descripción en una cepa microbiana de producción industrial, denominada en el presente documento "C". Las cepas del grupo de diversidad para este programa están representadas por A, B y C. La cepa A representaba la cepa hospedante de producción original, antes de cualquier mutagénesis. La cepa C representaba la cepa industrial actual, que ha experimentado muchos años de mutagénesis y selección a través de programas tradicionales de mejora de cepas. La cepa B representaba una cepa de "término medio", que había sufrido algunas mutagénesis y había sido la predecesora de la cepa C. (véase la Figura 17A).

Las cepas A, B y C se secuenciaron y se analizaron sus genomas para determinar las diferencias genéticas entre las cepas. Se identificaron un total de 332 SNP no sinónimos. De estos, 133 SNP eran exclusivos de C, 153 eran compartidos adicionalmente por B y C, y 46 eran exclusivos de la cepa B (véase la Figura 17B). Estos SNP se usarán como grupo de diversidad para los ciclos de mejora de cepas aguas abajo.

35 B. Análisis de intercambio de SNP

Los SNP identificados del grupo de diversidad en la Parte A del Ejemplo 2 se analizarán para determinar su efecto en el rendimiento de la células hospedantes. La ronda de "aprendizaje" inicial del rendimiento de las cepas se dividirá en seis etapas, como se describe a continuación, y se muestra en el diagrama en la Figura 18.

40 *Primero*, todos los SNP de C se clonarán de forma individual y/o combinatoria en la cepa base A. Esto representará un mínimo de 286 transformantes individuales. El propósito de estos transformantes será identificar los SNP beneficiosos.

5 *Segundo*, todos los SNP de C se eliminarán de forma individual y/o combinatoria de la cepa comercial C. Esto representará un mínimo de 286 transformantes individuales. El propósito de estos transformantes será identificar los SNP neutros y perjudiciales. Las etapas opcionales adicionales 3-6 también se describen a continuación. La primera y segunda etapa de sumar y restar los SNP de dos puntos de tiempo genéticos (cepa base A y cepa industrial C) se denominan en el presente documento "onda", que comprende una "onda ascendente" (adición de SNP a una cepa base, primera etapa) y una "onda descendente" (eliminación de SNP de la cepa industrial, segunda etapa). El concepto de onda se extiende a adiciones/sustracciones adicionales de SNP.

10 *Tercero*, todos los SNP de B se clonarán de forma individual y/o combinatoria en la cepa base A. Esto representará un mínimo de 199 transformantes individuales. El propósito de estos transformantes será identificar SNP beneficiosos. Varios de los transformantes también servirán como datos de validación para los transformantes producidos en la primera etapa.

15 *Cuarto*, todos los SNP de B se eliminarán de forma individual y/o combinatoria de la cepa comercial B. Esto representará un mínimo de 199 transformantes individuales. El propósito de estos transformantes será identificar los SNP neutros y perjudiciales. Varios de los transformantes también servirán como datos de validación para los transformantes producidos en la segunda etapa.

20 *Quinto*, todos los SNP *exclusivos* de C (es decir, no están también presentes en B) se clonarán de forma individual y/o combinatoria en la cepa B comercial. Esto representará un mínimo de 46 transformantes individuales. El propósito de estos transformantes será identificar SNP beneficiosos. Varios de los transformantes también servirán como datos de validación para los transformantes producidos en la primera y tercera etapas.

25 *Sexto*, todos los SNP *exclusivos* de C se eliminarán de forma individual y/o combinatoria de la cepa comercial C. Esto representará un mínimo de 46 transformantes individuales. El propósito de estos transformantes será identificar los SNP neutros y perjudiciales. Varios de los transformantes también servirán como datos de validación para los transformantes producidos en las etapas segunda y cuarta.

Los datos recogidos de cada una de estas etapas se usan para clasificar cada SNP como beneficioso, neutro o perjudicial a primera vista.

C. Uso de mapeo epistático para determinar combinaciones beneficiosas de SNP

Los SNP beneficiosos identificados en la Parte B del Ejemplo 2 se analizarán por métodos de mapeo de epistasis de la presente descripción, con el fin de identificar los SNP que es probable que mejoren el rendimiento del hospedante cuando se combinan.

30 Se crearán nuevas variantes de cepas transformadas usando los métodos de ingeniería del Ejemplo 1 para ensayar las combinaciones de SNP de acuerdo con las predicciones del mapeo de epistasis. La consolidación de SNP puede tener lugar secuencialmente, o puede tener lugar alternativamente a través de múltiples ramas, de modo que puede existir más de una cepa mejorada con un subconjunto de SNP beneficiosos. La consolidación de SNP continuará durante múltiples rondas de mejora de las cepas, hasta que se produzca una cepa final que contenga la combinación óptima de SNP beneficiosos, sin ninguno de los bagajes de SNP neutros o perjudiciales.

Ejemplo 3: Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de SNP para mejorar el rendimiento de la cepa en la producción de lisina en *Corynebacterium*

40 Este ejemplo proporciona una implementación ilustrativa de una parte del programa de mejora de cepas de diseño de HTP de intercambio de SNP del Ejemplo 2 con el objetivo de producir mejoras en el rendimiento y la productividad de la producción de lisina en *Corynebacterium*.

La sección B de este ejemplo ilustra además las etapas de consolidación de mutaciones del programa de mejora de cepas de HTP de la presente descripción. Por lo tanto, el ejemplo proporciona resultados experimentales para una primera, segunda y tercera ronda de consolidación de los métodos de mejora de cepas de HTP de la presente descripción.

45 Las mutaciones para las consolidaciones de la segunda y tercera ronda derivan de intercambios de bibliotecas genéticas independientes. Por lo tanto, estos resultados también ilustran la capacidad de los programas de cepas de HTP para llevarlos a cabo en rutas paralelas de múltiples ramas, y la "memoria" de las mutaciones beneficiosas que se pueden insertar en metadatos asociados con las diversas formas de bibliotecas de diseño genético de la presente descripción.

50 Como se ha descrito antes, se secuenciaron los genomas de una cepa de referencia base proporcionada (cepa A) y una segunda cepa "transformada" (cepa C), y se identificaron todas las diferencias genéticas. La cepa base era una variante de *Corynebacterium glutamicum* que no había sufrido mutagénesis por UV. La cepa transformada también era una cepa de *C. glutamicum* que se había producido a partir de la cepa base después de varias rondas de programas tradicionales de mejora por mutaciones. Este ejemplo proporciona los resultados de intercambio de SNP para 186 diferencias de SNP no sinónimos distintos identificadas entre las cepas A y C.

A. Ingeniería de HTP y cribado de alta capacidad

5 Cada uno de los 186 SNP identificados se volvió a añadir individualmente a la cepa base, de acuerdo con los métodos de clonación y transformación de la presente descripción. Se analizó en cada cepa recién creada que comprende un solo SNP el rendimiento de lisina en cultivos a pequeña escala diseñados para evaluar el rendimiento del título del producto. Se llevaron a cabo cultivos a pequeña escala usando medios de cultivos a escala industrial. El título del producto se midió ópticamente tras agotamiento del carbono (es decir, representativo del rendimiento de un solo lote) con un ensayo colorimétrico estándar. Brevemente, se preparó una mezcla de ensayo concentrada y se añadió a las muestras de fermentación de modo que las concentraciones finales de los reactivos eran tampón de fosfato de sodio 160 mM, Amplex Red 0,2 mM, peroxidasa de rábano picante 0,2 U/ml y 0,005 U/ml de lisina oxidasa. Se dejó que las reacciones avanzaran hasta un punto final y se midió la densidad óptica usando un espectrofotómetro de placa Tecan M1000 a una longitud de onda de 560 nm. Los resultados del experimento se resumen en la Tabla 8 a continuación y se representan en la Figura 38.

Tabla 8 - Resultados resumidos de la transformación de cepas de intercambio de SNP para la producción de lisina

SNP	N	Rendimiento medio de lisina (cambio en A ₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_033	4	0,1062	0,00888	11,54348	2,895652
DSS_311	2	0,03603	0,01256	3,916304	4,095652
DSS_350	1	0,03178	0,01777	3,454348	5,794565
DSS_056	3	0,02684	0,01026	2,917391	3,345652
DSS_014	4	0,02666	0,00888	2,897826	2,895652
DSS_338	3	0,02631	0,01026	2,859783	3,345652
DSS_128	1	0,02584	0,01777	2,808696	5,794565
DSS_038	4	0,02467	0,00888	2,681522	2,895652
DSS_066	4	0,02276	0,00888	2,473913	2,895652
DSS_108	2	0,02216	0,01256	2,408696	4,095652
DSS_078	4	0,02169	0,00888	2,357609	2,895652
DSS_017	3	0,02102	0,01026	2,284783	3,345652
DSS_120	3	0,01996	0,01026	2,169565	3,345652
DSS_064	4	0,01889	0,00888	2,053261	2,895652
DSS_380	4	0,01888	0,00888	2,052174	2,895652
DSS_105	3	0,0184	0,01026	2	3,345652
DSS_407	1	0,01831	0,01777	1,990217	5,794565
DSS_018	2	0,01825	0,01256	1,983696	4,095652
DSS_408	3	0,01792	0,01026	1,947826	3,345652
DSS_417	3	0,01725	0,01026	1,875	3,345652
DSS_130	3	0,01724	0,01026	1,873913	3,345652
DSS_113	4	0,0172	0,00888	1,869565	2,895652
DSS_355	3	0,01713	0,01026	1,861957	3,345652
DSS_121	3	0,01635	0,01026	1,777174	3,345652
DSS_097	2	0,0162	0,01256	1,76087	4,095652

ES 2 875 579 T3

SNP	N	Rendimiento medio de lisina (cambio en A₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_107	3	0,01604	0,01026	1,743478	3,345652
DSS_110	2	0,01524	0,01256	1,656522	4,095652
DSS_306	4	0,01501	0,00888	1,631522	2,895652
DSS_316	1	0,01469	0,01777	1,596739	5,794565
DSS_325	4	0,01436	0,00888	1,56087	2,895652
DSS_016	4	0,01416	0,00888	1,53913	2,895652
DSS_324	4	0,01402	0,00888	1,523913	2,895652
DSS_297	4	0,01391	0,00888	1,511957	2,895652
DSS_118	2	0,01371	0,01256	1,490217	4,095652
DSS_100	2	0,01326	0,01256	1,441304	4,095652
DSS_019	1	0,01277	0,01777	1,388043	5,794565
DSS_131	3	0,01269	0,01026	1,379348	3,345652
DSS_394	4	0,01219	0,00888	1,325	2,895652
DSS_385	3	0,01192	0,01026	1,295652	3,345652
DSS_395	1	0,01162	0,01777	1,263043	5,794565
DSS_287	4	0,01117	0,00888	1,21413	2,895652
DSS_418	2	0,01087	0,01256	1,181522	4,095652
DSS_290	3	0,01059	0,01026	1,151087	3,345652
DSS_314	2	0,01036	0,01256	1,126087	4,095652
DSS_073	4	0,00986	0,00888	1,071739	2,895652
DSS_040	4	0,00979	0,00888	1,06413	2,895652
DSS_037	4	0,00977	0,00888	1,061957	2,895652
DSS_341	1	0,00977	0,01777	1,061957	5,794565
DSS_302	4	0,00939	0,00888	1,020652	2,895652
DSS_104	4	0,00937	0,00888	1,018478	2,895652
DSS_273	2	0,00915	0,01256	0,994565	4,095652
DSS_322	4	0,00906	0,00888	0,984783	2,895652
DSS_271	3	0,00901	0,01026	0,979348	3,345652
DSS_334	2	0,00898	0,01256	0,976087	4,095652
DSS_353	4	0,00864	0,00888	0,93913	2,895652
DSS_391	4	0,00764	0,00888	0,830435	2,895652
DSS_372	1	0,00737	0,01777	0,801087	5,794565
DSS_007	1	0,00729	0,01777	0,792391	5,794565

ES 2 875 579 T3

SNP	N	Rendimiento medio de lisina (cambio en A₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_333	2	0,0072	0,01256	0,782609	4,095652
DSS_402	4	0,00718	0,00888	0,780435	2,895652
DSS_084	1	0,0069	0,01777	0,75	5,794565
DSS_103	3	0,00676	0,01026	0,734783	3,345652
DSS_362	1	0,00635	0,01777	0,690217	5,794565
DSS_012	2	0,00595	0,01256	0,646739	4,095652
DSS_396	2	0,00574	0,01256	0,623913	4,095652
DSS_133	3	0,00534	0,01026	0,580435	3,345652
DSS_065	3	0,00485	0,01026	0,527174	3,345652
DSS_284	2	0,00478	0,01256	0,519565	4,095652
DSS_301	3	0,00465	0,01026	0,505435	3,345652
DSS_281	4	0,00461	0,00888	0,501087	2,895652
DSS_405	2	0,00449	0,01256	0,488043	4,095652
DSS_361	3	0,00438	0,01026	0,476087	3,345652
DSS_342	4	0,00434	0,00888	0,471739	2,895652
DSS_053	3	0,00422	0,01026	0,458696	3,345652
DSS_074	4	0,00422	0,00888	0,458696	2,895652
DSS_079	4	0,00375	0,00888	0,407609	2,895652
DSS_381	3	0,0036	0,01026	0,391304	3,345652
DSS_294	1	0,00336	0,01777	0,365217	5,794565
DSS_313	2	0,00332	0,01256	0,36087	4,095652
DSS_388	2	0,00305	0,01256	0,331522	4,095652
DSS_392	4	0,00287	0,00888	0,311957	2,895652
DSS_319	4	0,00282	0,00888	0,306522	2,895652
DSS_310	4	0,00263	0,00888	0,28587	2,895652
DSS_344	3	0,00259	0,01026	0,281522	3,345652
DSS_025	4	0,00219	0,00888	0,238043	2,895652
DSS_412	1	0,00204	0,01777	0,221739	5,794565
DSS_300	3	0,00188	0,01026	0,204348	3,345652
DSS_299	2	0,00185	0,01256	0,201087	4,095652
DSS_343	4	0,00184	0,00888	0,2	2,895652
DSS_330	3	0,00153	0,01026	0,166304	3,345652
DSS_416	4	0,00128	0,00888	0,13913	2,895652

ES 2 875 579 T3

SNP	N	Rendimiento medio de lisina (cambio en A₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_034	3	0,00128	0,01026	0,13913	3,345652
DSS_291	2	0,00102	0,01256	0,11087	4,095652
DSS_115	4	0,00063	0,00888	0,068478	2,895652
DSS_288	4	0,00044	0,00888	0,047826	2,895652
DSS_309	4	0,00008	0,00888	0,008696	2,895652
DSS_125	3	0	0,01026	0	3,345652
DSS_358	3	-0,00015	0,01026	-0,0163	3,345652
DSS_099	2	-0,00015	0,01256	-0,0163	4,095652
DSS_111	4	-0,00017	0,00888	-0,01848	2,895652
DSS_359	3	-0,00022	0,01026	-0,02391	3,345652
DSS_015	4	-0,00043	0,00888	-0,04674	2,895652
DSS_060	3	-0,0007	0,01026	-0,07609	3,345652
DSS_098	2	-0,00088	0,01256	-0,09565	4,095652
DSS_379	4	-0,00089	0,00888	-0,09674	2,895652
DSS_356	4	-0,0009	0,00888	-0,09783	2,895652
DSS_278	4	-0,00095	0,00888	-0,10326	2,895652
DSS_368	4	-0,001	0,00888	-0,1087	2,895652
DSS_351	1	-0,0015	0,01777	-0,16304	5,794565
DSS_296	1	-0,0015	0,01777	-0,16304	5,794565
DSS_119	3	-0,00156	0,01026	-0,16957	3,345652
DSS_307	3	-0,00163	0,01026	-0,17717	3,345652
DSS_077	4	-0,00167	0,00888	-0,18152	2,895652
DSS_030	3	-0,00188	0,01026	-0,20435	3,345652
DSS_370	2	-0,00189	0,01256	-0,20543	4,095652
DSS_375	2	-0,00212	0,01256	-0,23043	4,095652
DSS_280	3	-0,00215	0,01026	-0,2337	3,345652
DSS_345	4	-0,00225	0,00888	-0,24457	2,895652
DSS_419	1	-0,00234	0,01777	-0,25435	5,794565
DSS_298	2	-0,00249	0,01256	-0,27065	4,095652
DSS_367	3	-0,0026	0,01026	-0,28261	3,345652
DSS_072	3	-0,00268	0,01026	-0,2913	3,345652
DSS_366	4	-0,00272	0,00888	-0,29565	2,895652
DSS_063	4	-0,00283	0,00888	-0,30761	2,895652

ES 2 875 579 T3

SNP	N	Rendimiento medio de lisina (cambio en A₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_092	3	-0,00292	0,01026	-0,31739	3,345652
DSS_347	4	-0,0033	0,00888	-0,3587	2,895652
DSS_114	4	-0,0034	0,00888	-0,36957	2,895652
DSS_303	3	-0,00396	0,01026	-0,43043	3,345652
DSS_276	4	-0,00418	0,00888	-0,45435	2,895652
DSS_083	1	-0,00446	0,01777	-0,48478	5,794565
DSS_031	2	-0,00456	0,01256	-0,49565	4,095652
DSS_328	3	-0,00463	0,01026	-0,50326	3,345652
DSS_039	4	-0,00475	0,00888	-0,5163	2,895652
DSS_331	4	-0,00475	0,00888	-0,5163	2,895652
DSS_117	4	-0,00485	0,00888	-0,52717	2,895652
DSS_382	4	-0,00506	0,00888	-0,55	2,895652
DSS_323	4	-0,00507	0,00888	-0,55109	2,895652
DSS_041	2	-0,00527	0,01256	-0,57283	4,095652
DSS_069	4	-0,00534	0,00888	-0,58043	2,895652
DSS_308	3	-0,00534	0,01026	-0,58043	3,345652
DSS_365	3	-0,00536	0,01026	-0,58261	3,345652
DSS_403	3	-0,00594	0,01026	-0,64565	3,345652
DSS_376	1	-0,00648	0,01777	-0,70435	5,794565
DSS_293	3	-0,00652	0,01026	-0,7087	3,345652
DSS_286	1	-0,00672	0,01777	-0,73043	5,794565
S A.2C	139	-0,00694	0,00151	-0,75435	0,492391
DSS_410	1	-0,00724	0,01777	-0,78696	5,794565
DSS_312	2	-0,00725	0,01256	-0,78804	4,095652
DSS_336	1	-0,00747	0,01777	-0,81196	5,794565
DSS_327	2	-0,00748	0,01256	-0,81304	4,095652
DSS_127	4	-0,00801	0,00888	-0,87065	2,895652
DSS_332	3	-0,0085	0,01026	-0,92391	3,345652
DSS_054	2	-0,00887	0,01256	-0,96413	4,095652
DSS_024	2	-0,00902	0,01256	-0,98043	4,095652
DSS_106	3	-0,0096	0,01026	-1,04348	3,345652
DSS_400	4	-0,00964	0,00888	-1,04783	2,895652
DSS_346	3	-0,00976	0,01026	-1,06087	3,345652

SNP	N	Rendimiento medio de lisina (cambio en A ₅₆₀ comparado con la cepa de referencia)	Error est.	% de cambio frente a la referencia	% Error de cambio
DSS_320	1	-0,01063	0,01777	-1,15543	5,794565
DSS_275	4	-0,01066	0,00888	-1,1587	2,895652
DSS_371	3	-0,01111	0,01026	-1,20761	3,345652
DSS_277	1	-0,01315	0,01777	-1,42935	5,794565
DSS_282	3	-0,01326	0,01026	-1,4413	3,345652
DSS_393	3	-0,01379	0,01026	-1,49891	3,345652
DSS_378	3	-0,01461	0,01026	-1,58804	3,345652
DSS_289	3	-0,01563	0,01026	-1,69891	3,345652
DSS_317	1	-0,01565	0,01777	-1,70109	5,794565
DSS_062	4	-0,01626	0,00888	-1,76739	2,895652
DSS_340	1	-0,01657	0,01777	-1,80109	5,794565
DSS_109	2	-0,01706	0,01256	-1,85435	4,095652
DSS_011	2	-0,0178	0,01256	-1,93478	4,095652
DSS_089	4	-0,01844	0,00888	-2,00435	2,895652
DSS_059	1	-0,01848	0,01777	-2,0087	5,794565
DSS_112	2	-0,01959	0,01256	-2,12935	4,095652
DSS_043	2	-0,0213	0,01256	-2,31522	4,095652
DSS_413	1	-0,02217	0,01777	-2,40978	5,794565
DSS_305	4	-0,0227	0,00888	-2,46739	2,895652
DSS_045	4	-0,02289	0,00888	-2,48804	2,895652
DSS_082	2	-0,0231	0,01256	-2,51087	4,095652
DSS_272	1	-0,02311	0,01777	-2,51196	5,794565
DSS_390	4	-0,02319	0,00888	-2,52065	2,895652
DSS_010	3	-0,02424	0,01026	-2,63478	3,345652
DSS_357	2	-0,02525	0,01256	-2,74457	4,095652
DSS_085	4	-0,03062	0,00888	-3,32826	2,895652
DSS_044	3	-0,04088	0,01026	-4,44348	3,345652
DSS_315	2	-0,0501	0,01256	-5,44565	4,095652
DSS_080	2	-0,13519	0,01256	-14,6946	4,095652

B. Ingeniería de HTP de segunda ronda y cribado de alta capacidad - Consolidación de la biblioteca de intercambio de SNP con aciertos de intercambios de PRO seleccionados

5 Una de las cualidades de los métodos de HTP de la presente descripción es su capacidad para almacenar bibliotecas de diseño genético de HTP junto con información asociada con los efectos de cada SNP/Promotor/Terminador/Codón de inicio en los fenotipos de la célula hospedante. Los autores de la presente invención habían llevado a cabo

previamente un experimento de intercambio de promotores que había identificado varios intercambios de promotor *zwf* en *C. glutamicum* con efectos positivos en los rendimientos biosintéticos (véanse, p. ej., los resultados para la diana "N" en la Figura 22).

5 Los autores de la presente invención modificaron la cepa base A de este ejemplo para incluir también uno de los intercambios de promotor *zwf* previamente identificados del ejemplo 5. Los 176 SNP superiores identificados en el cribado inicial descrito antes en la Tabla 8 se reintrodujeron en esta nueva cepa base para crear una nueva biblioteca microbiana de diseño genético de intercambio de SNP. Al igual que en la etapa anterior, se analizó el rendimiento de lisina de cada cepa recién creada que comprende un único SNP. También se analizó en cepas mutantes de SNP
 10 seleccionadas una señal de productividad, midiendo la producción de lisina a las 24 horas usando el método colorimétrico descrito más arriba. Los resultados de esta etapa se resumen en la Tabla 9 a continuación y se muestran en la Figura 39.

Tabla 9 - Segunda ronda de cribado para ingeniería de cepas de intercambio de SNP para la producción de lisina

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000006318	BS2C_P00 07_39zwf	20	2	0,49	0,82	0,00	0,02
7000008538	DSS_002	4	2	0,53	0,78	0,01	0,02
7000008539	DSS_003	4		0,56		0,01	
7000008541	DSS_005	4		0,27		0,01	
7000008542	DSS_006	4		0,49		0,01	
7000008547	DSS_011	4		0,55		0,01	
7000008548	DSS_012	4		0,58		0,01	
7000008549	DSS_013	4		0,56		0,01	
7000008550	DSS_014	4		0,52		0,01	
7000008551	DSS_015	4		0,54		0,01	
7000008552	DSS_016	4	2	0,50	0,84	0,01	0,02
7000008553	DSS_017	4		0,44		0,01	
7000008555	DSS_019	4	4	0,46	0,84	0,01	0,01
7000008557	DSS_021	4	4	0,46	0,86	0,01	0,01
7000008559	DSS_023	4	2	0,55	0,86	0,01	0,02
7000008561	DSS_025	4		0,54		0,01	
7000008562	DSS_026	2		0,46		0,01	
7000008564	DSS_028	4		0,51		0,01	
7000008565	DSS_029	4	4	0,48	0,87	0,01	0,01
7000008566	DSS_030	4	4	0,47	0,85	0,01	0,01
7000008567	DSS_031	4		0,56		0,01	
7000008569	DSS_033	4	4	0,46	0,86	0,01	0,01
7000008570	DSS_034	2	2	0,53	0,85	0,01	0,02
7000008573	DSS_037	4		0,54		0,01	
7000008574	DSS_038	4		0,53		0,01	

ES 2 875 579 T3

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000008575	DSS_039	4		0,55		0,01	
7000008576	DSS_040	4		0,57		0,01	
7000008577	DSS_041	4		0,45		0,01	
7000008578	DSS_042	4	4	0,52	0,87	0,01	0,01
7000008579	DSS_043	4	4	0,45	0,87	0,01	0,01
7000008580	DSS_044	4	2	0,50	0,85	0,01	0,02
7000008581	DSS_045	4		0,47		0,01	
7000008582	DSS_046	4	2	0,61	0,85	0,01	0,02
7000008583	DSS_047	4	2	0,61	0,82	0,01	0,02
7000008586	DSS_050	4		0,57		0,01	
7000008587	DSS_051	4		0,56		0,01	
7000008588	DSS_052	4	2	0,49	0,85	0,01	0,02
7000008589	DSS_053	4	4	0,45	0,85	0,01	0,01
7000008590	DSS_054	4	4	0,45	0,88	0,01	0,01
7000008592	DSS_056	4		0,42		0,01	
7000008596	DSS_060	4	2	0,55	0,87	0,01	0,02
7000008597	DSS_061	4	2	0,37	0,86	0,01	0,02
7000008598	DSS_062	4	4	0,45	0,87	0,01	0,01
7000008601	DSS_065	4	4	0,47	0,88	0,01	0,01
7000008602	DSS_066	4		0,47		0,01	
7000008604	DSS_068		2		0,51		0,02
7000008605	DSS_069	4	4	0,47	0,88	0,01	0,01
7000008606	DSS_070	4		0,55		0,01	
7000008607	DSS_071	4	2	0,56	0,84	0,01	0,02
7000008608	DSS_072	4	2	0,54	0,83	0,01	0,02
7000008609	DSS_073	4	2	0,47	0,84	0,01	0,02
7000008610	DSS_074	4	2	0,51	0,83	0,01	0,02
7000008612	DSS_076	4	4	0,48	0,76	0,01	0,01
7000008613	DSS_077	4	4	0,46	0,87	0,01	0,01
7000008614	DSS_078	4	2	0,44	0,87	0,01	0,02
7000008615	DSS_079	4	2	0,47	0,90	0,01	0,02
7000008616	DSS_080	4	2	0,48	0,81	0,01	0,02
7000008619	DSS_083	4	2	0,59	0,86	0,01	0,02
7000008620	DSS_084	4	2	0,70	0,89	0,01	0,02

ES 2 875 579 T3

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000008621	DSS_085	4	4	0,49	0,89	0,01	0,01
7000008622	DSS_086	4	2	0,48	0,82	0,01	0,02
7000008624	DSS_088	4	2	0,47	0,88	0,01	0,02
7000008625	DSS_089	4	4	0,45	0,89	0,01	0,01
7000008626	DSS_090	4	4	0,47	0,87	0,01	0,01
7000008627	DSS_091	4		0,46		0,01	
7000008629	DSS_093	4	4	0,50	0,87	0,01	0,01
7000008630	DSS_094	4	2	0,57	0,86	0,01	0,02
7000008634	DSS_098	4	2	0,53	0,85	0,01	0,02
7000008636	DSS_100	4		0,52		0,01	
7000008637	DSS_101	4	2	0,49	0,85	0,01	0,02
7000008640	DSS_104	4	2	0,51	0,84	0,01	0,02
7000008645	DSS_109	4		0,51		0,01	
7000008646	DSS_110	4	2	0,57	0,86	0,01	0,02
7000008648	DSS_112	4	2	0,54	0,86	0,01	0,02
7000008651	DSS_115	4		0,49		0,01	
7000008652	DSS_116	4	2	0,52	0,82	0,01	0,02
7000008653	DSS_117	4	2	0,50	0,84	0,01	0,02
7000008657	DSS_121	4	2	0,78	0,88	0,01	0,02
7000008659	DSS_123	4		0,54		0,01	
7000008663	DSS_127	4		0,58		0,01	
7000008665	DSS_129	4		0,48		0,01	
7000008666	DSS_130	4		0,56		0,01	
7000008669	DSS_133	4		0,50		0,01	
7000008670	DSS_271	4	2	0,52	0,86	0,01	0,02
7000008672	DSS_273	4		0,56		0,01	
7000008677	DSS_278	2		0,46		0,01	
7000008678	DSS_279	4		0,55		0,01	
7000008681	DSS_282	4		0,51		0,01	
7000008683	DSS_284	4		0,59		0,01	
7000008684	DSS_285	4		0,51		0,01	
7000008685	DSS_286	4		0,56		0,01	
7000008687	DSS_288	4		0,46		0,01	
7000008688	DSS_289	4		0,57		0,01	

ES 2 875 579 T3

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000008689	DSS_290	4		0,47		0,01	
7000008693	DSS_294	4	2	0,52	0,63	0,01	0,02
7000008696	DSS_297	4	2	0,52	0,86	0,01	0,02
7000008697	DSS_298	4		0,58		0,01	
7000008699	DSS_300	4		0,48		0,01	
7000008700	DSS_301	4		0,58		0,01	
7000008701	DSS_302	4		0,47		0,01	
7000008702	DSS_303	3		0,46		0,01	
7000008703	DSS_304	3		0,48		0,01	
7000008705	DSS_306	4	2	0,53	0,80	0,01	0,02
7000008708	DSS_309	4		0,56		0,01	
7000008709	DSS_310	4		0,56		0,01	
7000008711	DSS_312	4		0,55		0,01	
7000008712	DSS_313	4		0,51		0,01	
7000008718	DSS_319	4	2	0,50	0,82	0,01	0,02
7000008720	DSS_321	4		0,56		0,01	
7000008722	DSS_323	2	2	0,48	0,85	0,01	0,02
7000008723	DSS_324	4		0,55		0,01	
7000008724	DSS_325	4		0,50		0,01	
7000008725	DSS_326	3		0,46		0,01	
7000008726	DSS_327	3		0,47		0,01	
7000008730	DSS_331	4		0,56		0,01	
7000008731	DSS_332	4	4	0,47	0,89	0,01	0,01
7000008732	DSS_333	4	4	0,47	0,87	0,01	0,01
7000008733	DSS_334	4		0,45		0,01	
7000008734	DSS_335	2		0,47		0,01	
7000008735	DSS_336	4		0,47		0,01	
7000008739	DSS_340	4		0,46		0,01	
7000008740	DSS_341	4	2	0,46	0,89	0,01	0,02
7000008741	DSS_342	4		0,56		0,01	
7000008742	DSS_343	4		0,55		0,01	
7000008743	DSS_344	4	4	0,48	0,87	0,01	0,01
7000008746	DSS_347	4	4	0,48	0,85	0,01	0,01
7000008747	DSS_348	4	4	0,46	0,86	0,01	0,01

ES 2 875 579 T3

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000008749	DSS_350	4	2	0,29	0,74	0,01	0,02
7000008752	DSS_353	4	2	0,46	0,85	0,01	0,02
7000008753	DSS_354	4	4	0,45	0,87	0,01	0,01
7000008755	DSS_356	4	4	0,46	0,86	0,01	0,01
7000008756	DSS_357	4	4	0,46	0,86	0,01	0,01
7000008758	DSS_359	2	2	0,45	0,85	0,01	0,02
7000008760	DSS_361	4	2	0,46	0,84	0,01	0,02
7000008761	DSS_362	4		0,44		0,01	
7000008763	DSS_364	4		0,44		0,01	
7000008764	DSS_365	4		0,46		0,01	
7000008765	DSS_366	4		0,55		0,01	
7000008766	DSS_367	4		0,55		0,01	
7000008767	DSS_368	4	2	0,44	0,86	0,01	0,02
7000008770	DSS_371	4	2	0,47	0,88	0,01	0,02
7000008771	DSS_372	4	2	0,46	0,83	0,01	0,02
7000008772	DSS_373	4	2	0,46	0,88	0,01	0,02
7000008774	DSS_375	4		0,45		0,01	
7000008776	DSS_377	4		0,45		0,01	
7000008777	DSS_378	4		0,57		0,01	
7000008778	DSS_379	4		0,54		0,01	
7000008779	DSS_380	4	2	0,46	0,87	0,01	0,02
7000008781	DSS_382	4	2	0,46	0,84	0,01	0,02
7000008782	DSS_383	4		0,48		0,01	
7000008783	DSS_384	4	2	0,47	0,82	0,01	0,02
7000008784	DSS_385	4	2	0,46	0,83	0,01	0,02
7000008786	DSS_387	3		0,43		0,01	
7000008787	DSS_388	3		0,47		0,01	
7000008788	DSS_389	4	2	0,46	0,89	0,01	0,02
7000008790	DSS_391	4		0,57		0,01	
7000008791	DSS_392	4		0,44		0,01	
7000008795	DSS_396	4	2	0,46	0,82	0,01	0,02
7000008799	DSS_400	4		0,47		0,01	
7000008800	DSS_401	4	2	0,46	0,86	0,01	0,02
7000008801	DSS_402	4		0,54		0,01	

ID de cepa	SNP	N para 24 h	N para 96h	Media 24 h (A ₅₆₀)	Media 96 h (A ₅₆₀)	Error est. 24 h	Error est. 96 h
7000008805	DSS_406	4	2	0,47	0,85	0,01	0,02
7000008807	DSS_408	4		0,45		0,01	
7000008810	DSS_411	4	2	0,46	0,87	0,01	0,02
7000008812	DSS_413	3		0,47		0,01	
7000008813	DSS_414	4	2	0,45	0,84	0,01	0,02
7000008815	DSS_416	4	2	0,45	0,87	0,01	0,02
7000008816	DSS_417	4		0,46		0,01	
7000008818	DSS_419	4	2	0,47	0,84	0,01	0,02
7000008820	DSS_421	4	2	0,45	0,79	0,01	0,02
7000008821	DSS_422	4		0,44		0,01	

Los resultados de esta segunda ronda de intercambio de SNP identificaron varios SNP capaces de aumentar el rendimiento de la cepa base y la productividad de lisina en una cepa base que comprende la mutación de intercambio del promotor *zwf* (véase, p. ej., SNP 084 y SNP 121 en la esquina superior derecha de la Figura 39).

5 C. Validación del cultivo en tanque

Las cepas que contienen los SNP superiores identificados durante las etapas de HTP anteriores se cultivaron en tanques de fermentación de ensayo de tamaño mediano. Brevemente, se cultivaron pequeños cultivos de 100 ml de cada cepa durante la noche, y después se usaron para inocular cultivos de 5 litros en los tanques de fermentación de ensayo con cantidades iguales de inoculado. El inoculado se normalizó para contener la misma densidad celular después de una medición de DO600.

Se dejó que los cultivos en tanque resultantes procedieran durante 3 días antes de la recolección. Las mediciones de rendimiento y productividad se calcularon a partir de los títulos de sustrato y producto en muestras tomadas del tanque en varios puntos durante la fermentación. Las muestras se analizaron para concentraciones particulares de moléculas pequeñas por cromatografía líquida de alta presión usando las referencias adecuadas. Los resultados de este experimento se resumen en la Tabla 10 a continuación y se muestran en la Figura 40.

Tabla 10 - Validación en tanque de microbios de intercambio de SNP

Cepa	N	Rendimiento medio (%)(g de lisina producidos/g de glucosa consumidos)	Error est.	Productividad media (g/l/h)	Error est.
cepa base	1	41,1502	0,5940 1	3,29377	0,2450 8
cepa base + <i>zwf</i>	7	48,2952	0,2245 1	2,73474	0,1000 5
cepa base + <i>zwf</i> + SNP121	2	50,325	0,4200 3	4,51397	0,1733
cepa base + <i>zwf</i> + <i>pyc</i> + <i>lysA</i>	5	52,191	0,2656 5	4,15269	0,1225 4

Como se predijo por los cultivos de alta capacidad a pequeña escala, los cultivos en tanque más grandes para cepas que comprenden el intercambio de promotor *zwf* combinado y SNP 121, presentaban aumentos significativos en el rendimiento y la productividad frente a la cepa de referencia base. La productividad de esta cepa, por ejemplo, saltó a 4,5 g/l/h en comparación con la productividad de 3,29 g/l/h de la cepa base (un aumento de 37,0% en la productividad en solo 2 rondas de intercambio de SNP).

Ejemplo 4: Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de promotores para mejorar una cepa microbiana industrial

Los ejemplos previos han demostrado el poder de los programas de mejora de cepas de HTP de la presente descripción para rehabilitar cepas industriales. Los ejemplos 2 y 3 describían la implementación de técnicas y bibliotecas de intercambio de SNP que exploran la diversidad genética existente dentro de varias cepas base, intermedias e industriales.

- 5 Este ejemplo ilustra ejemplos de los programas de mejora de cepas de HTP que usan las técnicas de intercambio de PRO de la presente descripción. A diferencia del ejemplo 3, este ejemplo enseña métodos para la generación de novo de mutaciones mediante la generación de bibliotecas de intercambio de PRO.

A. Identificación de una diana para el intercambio de promotores

- 10 Como se ha mencionado antes, el intercambio de promotores es un procedimiento de múltiples etapas que comprende una etapa de: Seleccionar un conjunto de "n" genes a los que dirigirse.

En este ejemplo, los autores de la invención han identificado un grupo de 23 genes de rutas potenciales para modular mediante los métodos de escalera de promotores de la presente descripción (19 genes para sobreexpresar y 4+ genes de desvío para regular por disminución, en una ruta metabólica de ejemplo que produce la molécula lisina). (Véase la Figura 19).

15 **B. Creación de la escalera de promotores**

Otra etapa en la implementación de un procedimiento de intercambio de promotores es la selección de un conjunto de "x" promotores para que actúen como una "escalera". Idealmente, se ha demostrado que estos promotores conducen a una expresión muy variable en múltiples locus genómicos, pero el único requisito es que perturben la expresión génica de alguna manera.

- 20 Estas escaleras de promotores, en ejemplos particulares, se crean: identificando promotores naturales, nativos o de tipo natural asociados con el gen diana de interés y luego mutando dicho promotor para obtener múltiples secuencias de promotor mutadas. En cada uno de estos promotores mutados se analiza el efecto en la expresión del gen diana. En algunos ejemplos, se analiza en los promotores editados la actividad de expresión a través de una variedad de condiciones, de modo que la actividad de cada variante del promotor se documente/caracterice/anote y se almacene en una base de datos. Las variantes de promotor editadas resultantes se organizan posteriormente en "escaleras" dispuestas en función de la fuerza de su expresión (p. ej., con variantes de expresión alta cerca de la parte superior y expresión atenuada cerca de la parte inferior, conduciendo así al término "escalera").

En el presente ejemplo ilustrativo, los autores de la invención han creado combinaciones de escalera de promotores:ORF para cada uno de los genes diana identificados en la Figura 19.

30 **C. Asociación de promotores de la escalera con genes diana**

Otra etapa en la implementación de un procedimiento de intercambio de promotores es la ingeniería de HTP de varias cepas que comprenden un promotor dado de la escalera de promotores asociado con un gen diana particular.

- 35 Si existe un promotor nativo delante del gen diana n y se conoce su secuencia, entonces se puede llevar a cabo el reemplazo del promotor nativo con cada uno de los promotores x en la escalera. Cuando el promotor nativo no existe o se desconoce su secuencia, entonces se puede llevar a cabo la inserción de cada uno de los promotores x en la escalera delante del gen n. De esta manera, se construye una biblioteca de cepas, en donde cada miembro de la biblioteca es un caso de un promotor x operativamente unido a la diana n, en un contexto genético por lo demás idéntico (véase, por ejemplo, la Figura 20).

D. Cribado de HTP de las cepas

- 40 Una etapa final en el procedimiento de intercambio de promotores es el cribado de HTP de las cepas en la biblioteca antes mencionada. Cada una de las cepas derivadas representa un caso de promotor x unido a diana n, en un contexto genético por lo demás idéntico.

- 45 Al implementar un cribado de HTP de cada cepa, en un escenario en el que se caracteriza su rendimiento frente a uno o más indicadores, los autores de la invención pueden determinar qué asociación de promotor/gen diana es más beneficiosa para un indicador determinado (p. ej., optimización de la producción de una molécula de interés). Véase la Figura 20 (efecto de los promotores P1-P8 en el gen de interés). En el ejemplo ilustrativo ilustrado en las Figuras 19-22, los autores de la invención han usado el procedimiento de intercambio de promotores para optimizar la producción de lisina. Una aplicación de los métodos de intercambio de Pro descritos antes se describe en el ejemplo 5, a continuación.

50 **Ejemplo 5: Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de PRO para mejorar el rendimiento de cepas para la producción de lisina.**

La siguiente sección proporciona una implementación ilustrativa de las herramientas del programa de mejora de cepas de diseño de HTP de intercambio de PRO de la presente descripción, como se describe en el Ejemplo 4. En este

ejemplo, una cepa de *Corynebacterium* se sometió a los métodos de intercambio de PRO de la presente descripción con el fin de aumentar el rendimiento de lisina de la célula hospedante.

A. Intercambio de promotores

5 Los intercambios de promotores se realizaron como se describe en el Ejemplo 4. Los genes seleccionados de la ruta biosintética de la lisina en la Figura 19 fueron la diana para intercambios de promotores usando los promotores P1-P8.

B. Ingeniería de HTP y cribado de alta capacidad

10 La ingeniería de HTP de los intercambios de promotores se realizó como se describe en los Ejemplos 1 y 3. El cribado de HTP de las cepas de intercambio de promotores resultantes se llevó a cabo como se describe en el Ejemplo 3. En total se realizaron 145 intercambios de PRO. Los resultados del experimento se resumen en la Tabla 11 a continuación y se muestran en la Figura 41.

Tabla 11 - Cribado de HTP de bibliotecas de intercambio de PRO de lisina

Cepa	promotor-diana	N	Media (A ₅₆₀)	Error est.	% cambio de rendimiento de la base
7000007713	Pcg 1860-asd	8	0,84595	0,00689	3,927615
7000007736	Pcg0755-asd	4	0,84036	0,00974	3,240866
7000007805	Pcg0007_119-asd	8	0,82493	0,00689	1,345242
7000007828	Pcg3121-asd	8	0,8246	0,00689	1,3047
7000007759	Pcg0007_265-asd	8	0,81155	0,00689	-0,29853
7000007782	Pcg3381-asd	8	0,8102	0,00689	-0,46438
7000007712	Pcg 1860-ask	8	0,83958	0,00689	3,14504
7000007735	Pcg0755-ask	8	0,81673	0,00689	0,337846
7000007827	Pcg3121-ask	8	0,81498	0,00689	0,122853
7000007804	Pcg0007_119-ask	8	0,81492	0,00689	0,115482
7000007758	Pcg0007_265-ask	8	0,80381	0,00689	-1,24942
7000007781	Pcg3381-ask	8	0,80343	0,00689	-1,2961
7000007780	Pcg3381-aspB	8	0,84072	0,00689	3,285093
7000007803	Pcg0007_119-aspB	8	0,82106	0,00689	0,8698
7000007809	Pcg0007_119-cg0931	8	0,83446	0,00689	2,516032
7000007717	Pcg 1860-cg0931	4	0,83129	0,00974	2,126588
7000007763	Pcg0007265-cg0931	4	0,82628	0,00974	1,511094
7000007671	Pcg0007_39-cg0931	8	0,82554	0,00689	1,420182
7000007740	Pcg0755-cg0931	8	0,81921	0,00689	0,642522
7000007694	Pcg0007-cg0931	8	0,80444	0,00689	-1,17202
7000007691	Pcg0007-dapA	8	0,8299	0,00689	1,955822
7000007783	Pcg3381-dapA	8	0,80951	0,00689	-0,54915
7000007760	Pcg0007_265-dapA	8	0,76147	0,00689	-6,45102
7000007806	Pcg0007_119-dapA	8	0,35394	0,00689	-56,5174
7000007761	Pcg0007_265-dapB	8	0,84157	0,00689	3,389518

ES 2 875 579 T3

Cepa	promotor-diana	N	Media (A₅₆₀)	Error est.	% cambio de rendimiento de la base
7000007738	Pcg0755-dapB	4	0,84082	0,00974	3,297378
7000007692	Pcg0007-dapB	8	0,83088	0,00689	2,076218
7000007784	Pcg3381-dapB	8	0,82474	0,00689	1,3219
7000007715	Pcg 1860-dapB	8	0,82232	0,00689	1,024595
7000007830	Pcg3121-dapB	8	0,81236	0,00689	-0,19902
7000007807	Pcg0007_119-dapB	4	0,69622	0,00974	-14,4672
7000007762	Pcg0007_265-dapD	8	0,84468	0,00689	3,771591
7000007808	Pcg0007_119-dapD	8	0,83869	0,00689	3,035701
7000007785	Pcg3381-dapD	8	0,83397	0,00689	2,455834
7000007670	Pcg0007_39-dapD	8	0,81698	0,00689	0,368559
7000007831	Pcg3121-dapD	4	0,8155	0,00974	0,186737
7000007693	Pcg0007-dapD	8	0,8117	0,00689	-0,28011
7000007716	Pcg 1860-dapD	8	0,79044	0,00689	-2,89196
7000007739	Pcg0755-dapD	8	0,78694	0,00689	-3,32195
7000007787	Pcg3381-dapE	8	0,83814	0,00689	2,968132
7000007833	Pcg3121-dapE	8	0,83721	0,00689	2,853878
7000007741	Pcg0755-dapE	8	0,83263	0,00689	2,291211
7000007810	Pcg0007_119-dapE	8	0,83169	0,00689	2,175729
7000007718	Pcg1860-dapE	8	0,81855	0,00689	0,561439
7000007672	Pcg0007_39-dapE	8	0,80932	0,00689	-0,5725
7000007765	Pcg0007265-dapF	8	0,8327	0,00689	2,299811
7000007788	Pcg3381-dapF	8	0,82942	0,00689	1,896853
7000007811	Pcg0007_119-dapF	8	0,82926	0,00689	1,877196
7000007696	Pcg0007-dapF	8	0,82099	0,00689	0,861201
7000007719	Pcg 1860-dapF	8	0,82067	0,00689	0,821888
7000007673	Pcg0007_39-dapF	8	0,82062	0,00689	0,815745
7000007789	Pcg3381-ddh	8	0,84817	0,00689	4,200349
7000007835	Pcg3121-ddh	8	0,82141	0,00689	0,912799
7000007812	Pcg0007_119-ddh	8	0,82093	0,00689	0,853829
7000007674	Pcg0007_39-ddh	8	0,81494	0,00689	0,117939
7000007720	Pcg 1860-ddh	8	0,81473	0,00689	0,09214
7000007766	Pcg0007265-ddh	8	0,81427	0,00689	0,035627
7000007743	Pcg0755-ddh	8	0,80655	0,00689	-0,9128
7000007697	Pcg0007-ddh	8	0,80621	0,00689	-0,95457

ES 2 875 579 T3

Cepa	promotor-diana	N	Media (A ₅₆₀)	Error est.	% cambio de rendimiento de la base
7000007779	Pcg3381-fbp	8	0,85321	0,00689	4,819529
7000007802	Pcg0007_119-fbp	4	0,81425	0,00974	0,03317
7000007710	Pcg1860-fbp	4	0,40253	0,00974	-50,5479
7000007687	Pcg0007-fbp	8	0,14881	0,00689	-81,7182
7000007825	Pcg3121-fbp	4	0,12471	0,00974	-84,679
7000007733	Pcg0755-fbp	4	0,08217	0,00974	-89,9052
7000007746	Pcg0755-hom	8	0,81925	0,00689	0,647436
7000007792	Pcg3381-hom	4	0,77674	0,00974	-4,57505
7000007723	Pcg1860-hom	8	0,71034	0,00689	-12,7325
7000007838	Pcg3121-hom	8	0,559	0,00689	-31,3251
7000007800	Pcg0007_119-icd	8	0,83236	0,00689	2,258041
7000007823	Pcg3121-icd	8	0,83155	0,00689	2,15853
7000007777	Pcg3381-icd	8	0,82844	0,00689	1,776456
7000007708	Pcg1860-icd	8	0,82384	0,00689	1,211332
7000007662	Pcg0007_39-icd	12	0,82008	0,00562	0,749404
7000007685	Pcg0007-icd	8	0,81257	0,00689	-0,17322
7000007754	Pcg0007_265-icd	4	0,81172	0,00974	-0,27765
7000007698	Pcg0007-lysA	4	0,8504	0,00974	4,474311
7000007675	Pcg0007_39-lysA	8	0,84414	0,00689	3,705251
7000007836	Pcg3121-lysA	4	0,83545	0,00974	2,637657
7000007767	Pcg0007_265-lysA	8	0,83249	0,00689	2,274012
7000007813	Pcg0007_119-lysA	8	0,83096	0,00689	2,086046
7000007790	Pcg3381-lysA	8	0,8118	0,00689	-0,26782
7000007676	Pcg0007_39-lysE	8	0,84394	0,00689	3,68068
7000007699	Pcg0007-lysE	4	0,83393	0,00974	2,45092
7000007768	Pcg0007_265-lysE	8	0,83338	0,00689	2,383351
7000007837	Pcg3121-lysE	4	0,83199	0,00974	2,212585
7000007791	Pcg3381-lysE	8	0,81476	0,00689	0,095825
7000007814	Pcg0007_119-lysE	8	0,81315	0,00689	-0,10197
7000007775	Pcg3381-odx	8	0,82237	0,00689	1,030738
7000007752	Pcg0007265-odx	8	0,81118	0,00689	-0,34399
7000007729	Pcg0755-odx	8	0,81103	0,00689	-0,36242
7000007683	Pcg0007-odx	8	0,80507	0,00689	-1,09462
7000007706	Pcg 1860-odx	4	0,79332	0,00974	-2,53815

ES 2 875 579 T3

Cepa	promotor-diana	N	Media (A₅₆₀)	Error est.	% cambio de rendimiento de la base
7000007660	Pcg0007_39-odx	8	0,79149	0,00689	-2,76297
7000007798	Pcg0007_119-odx	8	0,77075	0,00689	-5,31094
7000007821	Pcg3121-odx	4	0,74788	0,00974	-8,12059
7000007822	Pcg3121-pck	8	0,85544	0,00689	5,093491
7000007776	Pcg3381-pck	8	0,8419	0,00689	3,43006
7000007799	Pcg0007_119-pck	8	0,83851	0,00689	3,013588
7000007753	Pcg0007_265-pck	8	0,82738	0,00689	1,646232
7000007730	Pcg0755-pck	4	0,81785	0,00974	0,475442
7000007661	Pcg0007_39-pck	8	0,80976	0,00689	-0,51844
7000007684	Pcg0007-pck	8	0,79007	0,00689	-2,93742
7000007707	Pcg1860-pck	8	0,71566	0,00689	-12,0789
7000007840	Pcg3121-pgi	4	1,01046	0,00974	24,13819
7000007817	Pcg0007_119-pgi	7	0,99238	0,00736	21,917
7000007794	Pcg3381-pgi	7	0,99008	0,00736	21,63444
7000007771	Pcg0007_265-pgi	8	0,94665	0,00689	16,29893
7000007725	Pcg1860-pgi	8	0,85515	0,00689	5,057864
7000007702	Pcg0007-pgi	4	0,8056	0,00974	-1,02951
7000007658	Pcg0007_39-ppc	4	0,85221	0,00974	4,696676
7000007750	Pcg0007265-ppc	8	0,84486	0,00689	3,793705
7000007727	Pcg0755-ppc	8	0,84166	0,00689	3,400575
7000007773	Pcg3381-ppc	4	0,82883	0,00974	1,824369
7000007796	Pcg0007_119-ppc	8	0,82433	0,00689	1,27153
7000007704	Pcg1860-ppc	8	0,81736	0,00689	0,415244
7000007819	Pcg3121-ppc	8	0,79898	0,00689	-1,8428
7000007732	Pcg0755-ptsG	8	0,84055	0,00689	3,264208
7000007709	Pcg1860-ptsG	8	0,81075	0,00689	-0,39682
7000007663	Pcg0007_39-ptsG	8	0,80065	0,00689	-1,63763
7000007778	Pcg3381-ptsG	8	0,23419	0,00689	-71,229
7000007801	Pcg0007_119-ptsG	8	0,17295	0,00689	-78,7525
7000007824	Pcg3121-ptsG	8	0,16035	0,00689	-80,3005
7000007705	Pcg1860-pyc	8	0,85143	0,00689	4,60085
7000007728	Pcg0755-pyc	8	0,79803	0,00689	-1,95951
7000007659	Pcg0007_39-pyc	8	0,75539	0,00689	-7,19797
7000007751	Pcg0007_265-pyc	8	0,73664	0,00689	-9,50146

Cepa	promotor-diana	N	Media (A ₅₆₀)	Error est.	% cambio de rendimiento de la base
7000007682	Pcg0007-pyc	4	0,73142	0,00974	-10,1428
7000007774	Pcg3381-pyc	4	0,66667	0,00974	-18,0975
7000007797	Pcg0007_119-pyc	4	0,52498	0,00974	-35,5046
7000007820	Pcg3121-pyc	8	0,52235	0,00689	-35,8277
7000007841	Pcg3121-ktk	8	0,82565	0,00689	1,433696
7000007818	Pcg0007_1 19-ktk	8	0,81674	0,00689	0,339075
7000007749	Pcg0755-ktk	8	0,81496	0,00689	0,120396
7000007703	Pcg0007-ktk	4	0,76763	0,00974	-5,69424
7000007795	Pcg3381-ktk	8	0,72213	0,00689	-11,2841
7000007772	Pcg0007_265-ktk	8	0,68884	0,00689	-15,3738
7000007701	Pcg0007-zwf	4	0,95061	0,00974	16,78542
7000007747	Pcg0755-zwf	8	0,92595	0,00689	13,75587
7000007770	Pcg0007265-zwf	8	0,9029	0,00689	10,9241
7000007724	Pcg1860-zwf	8	0,79309	0,00689	-2,5664
7000007839	Pcg3121-zwf	4	0,13379	0,00974	-83,5635
7000000017	-	11 6	0,92115	0,00181	13,16617
7000006284	-	12 8	0,81398	0,00172	0
7000005754	-	64	0,79489	0,00243	-2,34527

5 Cuando se visualizan, los resultados del cribado de la biblioteca de intercambio de promotores sirven para identificar genes dianas que están más estrechamente correlacionados con el indicador de rendimiento que se está midiendo. En este caso, los genes diana *pgi*, *zwf*, *ppc*, *pck*, *fbp* y *ddh* se identificaron como genes para los cuales los intercambios de promotores producen grandes ganancias en rendimiento frente a las cepas base.

Las cepas seleccionadas de la Tabla 11 se volvieron a cultivar en placas pequeñas y se analizó el rendimiento de lisina como se ha descrito antes. Los resultados de este cribado secundario se muestran en la Figura 22.

Ejemplo 6: Mapeo de epistasis - una herramienta algorítmica para predecir consolidaciones de mutaciones beneficiosas

10 Este ejemplo describe un ejemplo de las técnicas de modelización predictiva usadas como parte del programa de mejora de cepas de HTP de la presente descripción. Después de una identificación inicial de mutaciones potencialmente beneficiosas (mediante el uso de bibliotecas de diseño genético como se ha descrito antes), la presente descripción enseña métodos para consolidar mutaciones beneficiosas en la segunda, tercera, cuarta rondas y posteriores adicionales de mejora de cepas de HTP. En algunos ejemplos, la presente descripción enseña que las consolidaciones de mutaciones se pueden basar en el rendimiento individual de cada una de dichas mutaciones. En otros ejemplos, la presente descripción enseña métodos para predecir la probabilidad de que dos o más mutaciones presenten efectos aditivos o sinérgicos si se consolidan en una única célula hospedante. El siguiente ejemplo ilustra un ejemplo de las herramientas de predicción de la presente descripción.

20 Se analizaron mutaciones seleccionadas de las bibliotecas de intercambio de SNP y de intercambio de promotores (intercambio de PRO) de los Ejemplos 3 y 5 para identificar las combinaciones de intercambios de SNP/PRO que conducirían más probablemente a mejoras en el rendimiento de la cepa del hospedante.

25 Las secuencias de la biblioteca de intercambio de SNP se compararon entre sí usando una matriz de similitud de coseno, como se describe en la sección "Mapeo de epistasis" de la presente descripción. Los resultados del análisis proporcionaron puntuaciones de similitud funcional para cada combinación de intercambio de SNP/PRO. Una representación visual de las similitudes funcionales entre todos los intercambios de SNP/PRO se muestra en un mapa

de calor en la Figura 15. Las puntuaciones de similitud funcional resultantes también se usaron para desarrollar un dendrograma que representa la distancia de similitud entre cada uno de los intercambios de SNP/PRO (Figura 16A).

Las mutaciones del mismo grupo funcional o similar (es decir, intercambios de SNP/PRO con alta similitud funcional) tienen más probabilidades de operar por el mismo mecanismo y, por lo tanto, es más probable que presenten epistasis negativa o neutra en el rendimiento general del hospedante cuando se combinan. Por el contrario, es más probable que las mutaciones de diferentes grupos funcionales operen por mecanismos independientes y, por lo tanto, es más probable que produzcan efectos aditivos o combinatorios beneficiosos en el rendimiento del hospedante.

Con el fin de ilustrar los efectos de las rutas biológicas en la epistasis, se combinaron y ensayaron intercambios de SNP y de PRO que presentaban diversas similitudes funcionales en cepas hospedantes. Se diseñaron tres combinaciones de intercambio de SNP/PRO en el genoma de *Corynebacterium glutamicum* como se describe en el Ejemplo 1: **i)** Pcg0007::zwf intercambio de PRO + Pcg1860::pyc intercambio de PRO, **ii)** Pcg0007::zwf intercambio de PRO + SNP 309, y **iv)** Pcg0007::zwf intercambio de PRO + Pcg0007::lysA intercambio de PRO (véase la Figura 15 y 16A para las relaciones de similitud funcional).

El rendimiento de cada una de las células hospedantes que contienen las combinaciones de intercambios de SNP/PRO se ensayó como se describe en el Ejemplo 3, y se comparó con el de una célula hospedante de control que contenía solo intercambio de PRO zwf. Las Tablas 12 y 13 a continuación resumen los resultados del rendimiento de células hospedantes (mediciones de 96 h) y la productividad (mediciones de 24 h) de cada una de las cepas.

Tabla 12- Acumulación de lisina para el experimento de mapeo de epistasis a las 24 horas.

Intercambio de SNP/ PRO	Media de lisina (A ₅₆₀)	Des. est.
6318 (zwf)	0,51	0,03
8126 (zwf+ lysA)	0,88	0,06
8156 (zwf+pyc)	0,53	0,01
8708 (zwf+ SNP 309)	0,56	0,00

Tabla 13- Acumulación de lisina para el experimento de mapeo de epistasis a las 96 horas.

Intercambio de SNP/ PRO	Media de lisina (A ₅₆₀)	Des. est.
6318 (zwf)	0,83	0,01
8126 (zwf + lysA)	0,94	0,02
8156 (zwf+ pyc)	0,83	0,06

Los resultados de rendimientos para el rendimiento del hospedante para cada combinación de intercambios SNP/PRO también se muestran en la Figura 16B. Las cepas hospedantes que combinaban intercambios de SNP/PRO que presentaban menor similitud funcional superaron a las cepas en las que los SNP combinados habían presentado una mayor similitud funcional en las mediciones tanto a las 24 como a las 96 horas.

Por lo tanto, el procedimiento de mapeo epistático es útil para predecir/programar/informar consolidaciones efectivas y/o positivas de cambios genéticos diseñados. El conocimiento analítico del procedimiento de mapeo epistático permite la creación de conjuntos de reglas predictivas que pueden guiar las rondas posteriores de desarrollo de cepas microbianas. El conocimiento predictivo obtenido de la biblioteca epistática se puede usar en tipos microbianos y tipos de moléculas diana.

Ejemplo 7: Ingeniería genómica de HTP - consolidación de mutaciones de intercambio de Pro y ensayos combinatorios multifactoriales

Los ejemplos anteriores han ilustrado métodos para consolidar un pequeño número de mutaciones de intercambio de PRO preseleccionadas con bibliotecas de intercambio de SNP (Ejemplo 3). Otros ejemplos han ilustrado los métodos epistáticos para seleccionar consolidaciones de mutaciones que es más probable que proporcionen propiedades beneficiosas aditivas o sinérgicas de la célula hospedante (Ejemplo 6). Este ejemplo ilustra la capacidad de los métodos de HTP de la presente descripción para explorar eficazmente el gran espacio de soluciones creado por la consolidación combinatoria de múltiples combinaciones de genes/biblioteca de diseño genético (p. ej., biblioteca de intercambio de PRO x biblioteca de SNP o combinaciones dentro de una biblioteca de intercambio de PRO).

En esta aplicación ilustrativa de los métodos de mejora de cepas de HTP de la presente descripción, los intercambios de promotores identificados por tener un efecto positivo en el rendimiento del hospedante en el Ejemplo 5 se consolidan en combinaciones de segundo orden con la biblioteca de intercambio de PRO original. La decisión de consolidar las mutaciones de intercambio de PRO se basaba en el efecto general de cada mutación en el rendimiento o la productividad, y la probabilidad de que la combinación de las dos mutaciones produjera un efecto aditivo o sinérgico.

- 5 Por ejemplo, los autores de la invención se refieren a su elección de combinar Pcg0007::zwf y Pcg0007::lysA, basándose en los resultados del mapeo de epistasis del Ejemplo 6.

A. Ronda de consolidación para ingeniería de cepas por intercambio de PRO

- 10 Las cepas se transformaron como se describe en el Ejemplo 1 anterior. Brevemente, las cepas que ya contenían una mutación de intercambio de PRO deseada se transformaron una vez más con la segunda mutación de intercambio de PRO deseada. En total, los 145 intercambios de PRO ensayados del Ejemplo 5 se consolidaron en 53 cepas de consolidación de segunda ronda, comprendiendo cada una dos mutaciones de intercambio de PRO que se espera que presenten efectos aditivos o sinérgicos beneficiosos.
- 15 Las cepas de la segunda ronda resultantes se cribaron una vez más como se describe en el Ejemplo 3. Los resultados de este experimento se resumen en la Tabla 14 a continuación y se muestran en la Figura 11.

Tabla 14 - Cribado de HTP de bibliotecas de intercambio de PRO para lisina consolidadas de la segunda ronda

ID cepa	Número	Intercambio de PRO 1	Intercambio de PRO 2	Rendimiento medio (A ₅₆₀)	Desviación estándar
7000008489	4	Pcg0007-lysA	Pcg3121-pgi	1,17333	0,020121
7000008530	8	Pcg1860-pyc	Pcg0007-zwf	1,13144	0,030023
7000008491	7	Pcg0007-lysA	Pcg0007-zwf	1,09836	0,028609
7000008504	8	Pcg3121-pck	Pcg0007-zwf	1,09832	0,021939
7000008517	8	Pcg000739-ppc	Pcg0007-zwf	1,09502	0,030777
7000008502	4	Pcg3121-pck	Pcg3121-pgi	1,09366	0,075854
7000008478	4	Pcg3381-ddh	Pcg0007-zwf	1,08893	0,025505
7000008465	4	Pcg0007_265-dapB	Pcg0007-zwf	1,08617	0,025231
7000008535	8	Pcg0007-zwf	Pcg3121-pgi	1,06261	0,019757
7000008476	6	Pcg3381-ddh	Pcg3121-pgi	1,04808	0,084307
7000008510	8	Pcg3121-pgi	Pcg 1860-pyc	1,04112	0,021087
7000008525	8	Pcg1860-pyc	Pcg0007_265-dapB	1,0319	0,034045
7000008527	8	Pcg1860-pyc	Pcg0007-lysA	1,02278	0,043549
7000008452	5	Pcg1860-asd	Pcg0007-zwf	1,02029	0,051663
7000008463	4	Pcg0007_265-dapB	Pcg3121-pgi	1,00511	0,031604
7000008524	8	Pcg1860-pyc	Pcg1860-asd	1,00092	0,026355
7000008458	4	Pcg3381-aspB	Pcg1860-pyc	1,00043	0,020083
7000008484	8	Pcg3381-fbp	Pcg1860-pyc	0,99686	0,061364
7000008474	8	Pcg3381-ddh	Pcg3381-fbp	0,99628	0,019733
7000008522	8	Pcg0755-ptsG	Pcg3121-pgi	0,99298	0,066021
7000008528	8	Pcg1860-pyc	Pcg3121-pck	0,99129	0,021561
7000008450	4	Pcg1860-asd	Pcg3121-pgi	0,98262	0,003107

ES 2 875 579 T3

ID cepa	Número	Intercambio de PRO 1	Intercambio de PRO 2	Rendimiento medio (A ₅₆₀)	Desviación estándar
7000008448	8	Pcg1860-asd	Pcg3381-fbp	0,97814	0,022285
7000008494	8	Pcg000739-lysE	Pcg3381-fbp	0,97407	0,027018
7000008481	8	Pcg3381-fbp	Pcg0007-lysA	0,9694	0,029315
7000008497	8	Pcg0007_39-lysE	Pcg 1860-pyc	0,9678	0,028569
7000008507	8	Pcg3121-pgi	Pcg3381-fbp	0,96358	0,035078
7000008501	8	Pcg3121-pck	Pcg0007-lysA	0,96144	0,018665
7000008486	8	Pcg0007-lysA	Pcg0007_265-dapB	0,94523	0,017578
7000008459	8	Pcg0007_265-dapB	Pcg1860-asd	0,94462	0,023847
7000008506	2	Pcg3121-pgi	Pcg0007_265-dapD	0,94345	0,014014
7000008487	8	Pcg0007-lysA	Pcg3381-ddh	0,94249	0,009684
7000008498	8	Pcg3121-pck	Pcg1860-asd	0,94154	0,016802
7000008485	8	Pcg0007-lysA	Pcg1860-asd	0,94135	0,013578
7000008499	8	Pcg3121-pck	Pcg0007_265-dapB	0,93805	0,013317
7000008472	8	Pcg3381-ddh	Pcg1860-asd	0,93716	0,012472
7000008511	8	Pcg000739-ppc	Pcg1860-asd	0,93673	0,015697
7000008514	8	Pcg000739-ppc	Pcg0007-lysA	0,93668	0,027204
7000008473	8	Pcg3381-ddh	Pcg0007_265-dapB	0,93582	0,030377
7000008461	7	Pcg0007_265-dapB	Pcg3381-fbp	0,93498	0,037862
7000008512	8	Pcg0007_39-ppc	Pcg0007_265-dapB	0,93033	0,017521
7000008456	8	Pcg3381-aspB	Pcg3121-pck	0,92544	0,020075
7000008460	8	Pcg0007_265-dapB	Pcg0007_265-dapD	0,91723	0,009508
7000008492	8	Pcg0007_39-lysE	Pcg3381-aspB	0,91165	0,012988
7000008493	8	Pcg0007_39-lysE	Pcg0007_265-dapD	0,90609	0,031968
7000008453	8	Pcg3381-aspB	Pcg0007_265-dapB	0,90338	0,013228
7000008447	8	Pcg1860-asd	Pcg0007_265-dapD	0,89886	0,028896
7000008455	8	Pcg3381-aspB	Pcg0007-lysA	0,89531	0,027108
7000008454	6	Pcg3381-aspB	Pcg3381-ddh	0,87816	0,025807
7000008523	8	Pcg0755-ptsG	Pcg 1860-pyc	0,87693	0,030322
7000008520	8	Pcg0755-ptsG	Pcg3381-fbp	0,87656	0,018452
7000008533	4	Pcg0007-zwf	Pcg3381-fbp	0,84584	0,017012
7000008519	8	Pcg0755-ptsG	Pcg0007_265-dapD	0,84196	0,025747

Como predecía el modelo de epistasis, la cepa de intercambio de PRO de la segunda ronda que comprende las mutaciones Pcg0007::zwf y Pcg0007::lysA presentaba una de las mejoras de rendimiento más altas, con una mejora

de cerca del 30% en el rendimiento frente a Pcg0007::lysA solo, y 35,5% de mejora frente a la cepa base (véase los datos en un círculo en la Figura 11).

Los métodos de HTP para explorar el espacio de soluciones de mutaciones consolidadas simples y dobles también se pueden aplicar a la tercera, cuarta y posteriores consolidaciones de mutaciones. También se llama la atención, por ejemplo, sobre la cepa de consolidación de 3 cambios descrita que corresponden a zwf, pyc y lysA que se hizo entre los aciertos principales de identificados en las consolidaciones de 2 cambios como se muestra en la Tabla 14 anterior, y como se identifica por los métodos epistáticos de la presente descripción. Esta cepa de consolidación de 3 cambios se validó adicionalmente en tanques como mejorada significativamente en comparación con la progenitora o progenitora + zwf (véase la Tabla 10 más arriba y Figura 40).

10 **Ejemplo 8: Ingeniería genómica de HTP - implementación de una biblioteca de terminadores para mejorar una cepa hospedante industrial**

El presente ejemplo aplica los métodos de HTP de la presente descripción a bibliotecas de diseño genético de HTP adicionales, que incluyen intercambio de STOP. El ejemplo ilustra además la capacidad de la presente descripción para combinar elementos de bibliotecas de diseño genético básico (p. ej., intercambio de PRO, intercambio de SNP, intercambio de STOP, etc.) para crear bibliotecas de diseño genético más complejas (p. ej., bibliotecas de intercambio de PRO-STOP, que incorporan tanto un promotor como un terminador). En algunos ejemplos, la presente descripción enseña todas y cada una de las posibles bibliotecas de diseño genético, incluyendo las derivadas de combinar de cualquiera de las bibliotecas de diseño genético descritas previamente.

En este ejemplo, se llevó a cabo un experimento a pequeña escala para demostrar el efecto de los métodos de intercambio de STOP de la presente descripción en la expresión génica. Los terminadores T1-T8 de la presente descripción se emparejaron con uno de dos promotores nativos de *Corynebacterium glutamicum* como se describe a continuación, y se analizaron para determinar su capacidad para influir en la expresión de una proteína fluorescente.

A. Ensamblaje de construcciones de ADN

Los terminadores T1-T8 se emparejaron con uno de dos promotores nativos de *Corynebacterium glutamicum* (p. ej., Pcg0007 o Pcg0047) que expresaban una proteína de fluorescencia amarilla (YFP). Para facilitar la amplificación y el ensamblaje del ADN, la secuencia final del promotor-YFP-terminador se sintetizó en dos partes; la primera parte codificaba (de 5' a 3') **i**) el brazo de homología del vector, **ii**) el promotor seleccionado, **iii**) y 2/3 del gen de YFP. La segunda parte codificaba (de 5' a 3') **iv**) los siguientes 2/3 del gen de YFP, **v**) el terminador seleccionado, y **vi**) el segundo brazo de homología del vector. Cada parte se amplificó usando oligonucleótidos sintéticos y se purificó en gel. Los amplicones purificados en gel se ensamblaron con una cadena principal del vector usando recombinación homóloga de levaduras.

B. Transformación de clones ensamblados en *E. coli*

Los vectores que contenían las secuencias del promotor-YFP-terminador se transformó cada una individualmente en *E. coli* con el fin de identificar clones ensamblados correctamente y para amplificar el ADN del vector para la transformación de *Corynebacterium*. Los vectores ensamblados correctamente se confirmaron por digestión con enzimas de restricción y secuenciación de Sanger. Los clones positivos se almacenaron a -20°C para uso futuro.

C. Transformación de clones ensamblados en *Corynebacterium*

Los clones de vector verificados se transformaron individualmente en células hospedantes de *Corynebacterium glutamicum* por electroporación. Cada vector se diseñó para integrarse en un sitio de integración neutro dentro del genoma de *Corynebacterium glutamicum* que se determinó empíricamente que permitía la expresión de la proteína de fluorescencia amarilla heteróloga pero no era perjudicial para la célula hospedante. Para facilitar la integración, el vector de expresión comprendía además aproximadamente 2 kpb de secuencia homóloga (es decir, brazo de homología) al sitio de integración deseado, por lo que cada casete génico descrito antes se insertó secuencia abajo del brazo de homología. La integración en el genoma se produjo por integración de un solo cruce. A continuación, se analizó la correcta integración de *Corynebacterium* transformadas por PCR. Este procedimiento se repitió para cada una de las transformaciones realizadas para cada construcción génica.

D. Evaluación de construcciones de terminadores individuales en *Corynebacterium*

El fenotipo de cada transformante de *Corynebacterium* que contenía construcciones de promotor-YFP-terminador después se ensayó en dos tipos de medios (medios de ensayo de infusión de cerebro-corazón-BHI y HTP) en dos puntos de tiempo con el fin de evaluar la expresión. Brevemente, se eligieron entre cuatro y seis transformantes confirmados por PCR y se cultivaron en medio selectivo en un formato de 96 pocillos. Después, los cultivos iniciales se dividieron en medio BHI selectivo o medio de siembra selectivos. A las 48 horas, los cultivos en medio de siembra se inocularon en medio de ensayo selectivo de HTP o medio BHI y se analizaron en dos puntos de tiempo que representan diferentes partes de la curva de crecimiento. Los puntos de tiempo para los cultivos en medio de ensayo HTP eran 48 y 96 horas después de la inoculación. Los cultivos en el medio BHI selectivo se analizaron a las 48 y 72 horas después de la inoculación.

El análisis de los cultivos se llevó a cabo usando un citómetro de flujo de sobremesa. Brevemente, los cultivos se diluyeron 1:100 en 200 µl de solución salina tamponada con fosfato (PBS). Para cada cultivo, se analizaron entre 3000 y 5000 sucesos individuales (es decir, células) para determinar la fluorescencia amarilla. El citómetro de flujo de sobremesa traza un histograma de fluorescencia amarilla de cada "suceso" y calcula la fluorescencia mediana dentro de cada pocillo. La Figura 36 representa la media de la fluorescencia mediana para cada construcción (en las 4-6 repeticiones biológicas). Las barras de error indican el intervalo de confianza del 95% de cada punto de datos. Las condiciones A-D se refieren cada una a un solo medio y a un solo punto de tiempo. Por lo tanto, las condiciones A y B representan los dos puntos de tiempo para el medio BHI, mientras que los puntos C y D representan los dos puntos de tiempo para el medio de ensayo HTP. Obsérvese que las unidades arbitrarias (p. ej., AU) representan la fluorescencia mediana registrada por el citómetro de flujo de sobremesa.

Los resultados muestran que los terminadores 1-8 de la biblioteca de diseño genético de intercambio de STOP dan como resultado un intervalo continuo de expresión de YFP. Por lo tanto, estos terminadores forman una escalera de terminadores que se puede implementar en futuras bibliotecas de diseño genético, de acuerdo con los métodos de HTP de la presente descripción.

15 **Ejemplo 9: Comparación de conjuntos de herramientas de HTP frente a mutaciones por UV tradicionales.**

Este ejemplo demuestra los beneficios de las bibliotecas de diseño genético de HTP de la presente descripción en los programas tradicionales de mejora de cepas mutacionales. Los experimentos en esta parte de la memoria descriptiva cuantifican la magnitud y la velocidad mejoradas de las mejoras fenotípicas logradas a través de los métodos de HTP de la presente descripción frente a la mutagénesis por UV tradicional.

20 La presente descripción enseña nuevos métodos para acelerar los programas de mejora de cepas de las células hospedantes. En algunos ejemplos, el programa de mejora de cepas de HTP de la presente descripción se basa en la capacidad de los conjuntos de herramientas de HTP para generar e identificar perturbaciones genéticas. Los autores de la presente invención intentaron cuantificar los beneficios de los conjuntos de herramientas de HTP llevando a cabo un pequeño programa de mejora de cepas de ruta paralela comparando las técnicas de intercambio de promotores de la presente descripción con los enfoques tradicionales de mutaciones por UV.

Se eligió una cepa de referencia base que produjera un metabolito bioquímico de interés como punto de partida para las perturbaciones genéticas tanto de intercambio de promotores como por UV.

A. Mutaciones por UV

30 Los cultivos de la cepa base se cultivaron en medio BHI en cultivos que se normalizaron respecto a la DO a la DO600 de 10. Este cultivo se dividió en alícuotas en una placa de Petri estéril y se agitó usando una pequeña barra agitadora magnética. A continuación, un transiluminador UV a una longitud de onda de 254 nm se invirtió sobre el cultivo y se tomaron partes alícuotas a los 5 y 9 minutos de exposición al UV. Estas muestras se diluyeron de forma seriada 10 veces y cada dilución se sembró en bandejas Q de medio BHI. De estas bandejas Q, se recogieron aproximadamente 2500 colonias de cada punto de exposición a UV usando un aparato de selección de colonias automatizado y se evaluó el rendimiento como se indica a continuación.

B. Intercambio de promotores

40 Se generaron construcciones de intercambio de PRO en la cepa base para 15 dianas de genes usando todos o un subconjunto de promotores seleccionados de P1, P3, P4 y P8 descritos en la Tabla 1. La etapa final en la biosíntesis del producto de interés es catalizada por una enzima O-metiltransferasa que usa el cofactor potencialmente limitante de velocidad S-adenosilmetionina. Por lo tanto, las dianas genéticas para los intercambios de PRO se seleccionaron basándose en que están directamente implicadas en la biosíntesis de este cofactor o metabolitos secuencia arriba.

C. Evaluación de la biblioteca de intercambio de promotores y UV

45 Se analizó en el fenotipo de cada cepa de *Corynebacterium* desarrollada para este ejemplo, su capacidad para producir una biomolécula seleccionada. Brevemente, se eligieron entre cuatro y seis colonias de secuencia confirmada de cada cepa de intercambio de PRO, y se eligieron colonias individuales para cada cepa de UV y se propagaron en medio selectivos en un formato de 96 pocillos en medio líquido de producción.

50 Después de la propagación de la biomasa en placas de micropocillos de 96 pocillos, la masa celular se añadió al medio de fermentación que contenía sustrato en placas de micropocillos de 96 pocillos y se dejó que la bioconversión prosiguiera durante 24 horas. Los títulos de producto se determinaron para cada cepa usando cromatografía líquida de alta resolución a partir de muestras tomadas a las 24 horas. Se analizaron los resultados de los títulos para cada perturbación genética (UV e intercambio de PRO). Los resultados de cada repetición se promediaron y se asignaron para representar el rendimiento general de dicha cepa. A continuación, las cepas se agruparon en categorías basadas en el efecto de cada mutación en el rendimiento medido expresado como una relación frente al rendimiento de la cepa base.

La Figura 37 resume los resultados de este experimento, que se presentan como el número de cepas para cada técnica de mejora de cepas que proporcionaba: i) ningún cambio en el rendimiento, ii) una mejora de 1,2 a 1,4 veces del rendimiento, iii) una mejora de 1,4 a 1,6 veces del rendimiento, iv) una mejora de 1,6 a 1,8 veces del rendimiento, o, v) una mejora de 1,8 a 2 veces del rendimiento.

5 Los resultados son ilustrativos de los beneficios de los conjuntos de herramientas de HTP de la presente descripción frente a los enfoques tradicionales de mutagénesis por UV. Por ejemplo, los resultados de la Figura 37 demuestran que las cepas de intercambio de PRO presentaban una mayor tasa de cambios positivos en el rendimiento y, por lo tanto, tenían más probabilidades de proporcionar mutaciones que podrían mejorar significativamente la cepa. Lo más sorprendente era la alta incidencia de cepas de mejora alta que mostraban aumentos de 1,6, 1,8 y 2 veces en la biblioteca de intercambio de PRO, con poca o ninguna mejora identificada en la biblioteca de UV.

10 Los resultados también son importantes porque destacan la tasa acelerada de mejora de los métodos de intercambio de PRO de la presente descripción. De hecho, los resultados de la biblioteca de intercambio de PRO se basaron en menos de 100 perturbaciones promotor::gen, mientras que los resultados de las mutaciones por UV incluyeron el cribado de más de 4.000 cepas mutantes distintas. Por lo tanto, los métodos de la presente descripción reducen drásticamente el número de mutantes que se deben seleccionar antes de identificar perturbaciones genéticas capaces de conferir cepas con altas ganancias en el rendimiento.

Ejemplo 10 - Ingeniería genómica de HTP - Implementación de una biblioteca de mutagénesis de transposones para mejorar el rendimiento de cepas de *Escherichia coli*

20 Los ejemplos previos ilustran aplicaciones de los programas de mejora de cepas de HTP en *Corynebacterium*. Este ejemplo demuestra la aplicabilidad de las mismas técnicas a las células de *E. coli*.

Este ejemplo describe la aplicación de mutagénesis de transposones para generar bibliotecas de cepas aleatorias de *Escherichia coli* con el fin de mejorar las cepas. Estas bibliotecas de cepas se pueden cribar frente a un fenotipo deseado, tal como el rendimiento de triptófano, para identificar variantes con rendimiento mejorado.

25 La presente descripción describe un método para generar una biblioteca de mutantes mediante la aplicación del sistema de transposones EZ-Tn5 (Epicenter Bio) en *Escherichia coli*. La transposasa EZ-Tn5 se incuba con ADN de carga útil flanqueado por secuencias de elementos de mosaico. Tras la incubación, la transposasa Ez-Tn5 forma complejo con el ADN para formar un transposoma. El complejo de ADN/transposoma de proteína después se transforma en *Escherichia coli* por electroporación. La transposasa EZ-Tn5 cataliza la integración aleatoria del ADN de carga útil en el genoma de *Escherichia coli* dando lugar así a una biblioteca aleatoria de variantes de cepas.

30 La secuencia específica del ADN de carga útil se puede variar además para sesgar hacia los efectos de pérdida de función (LoF) o ganancia de función (GoF) de la inserción del transposón en el genoma diana.

35 La LoF se puede lograr mediante la inclusión de un marcador de selección de antibióticos en la carga útil del ADN. El marcador de antibiótico permite la selección de células con una inserción de transposón productivo. La inserción de carga útil del ADN puede alterar la función del ADN en el que se ha insertado de diversas formas, incluyendo, pero no limitado a la alteración de un marco de lectura abierto que evita la traducción del gen alterado.

La GoF se puede lograr mediante la inclusión de un marcador de antibiótico y un promotor fuerte en la carga útil de ADN. El marcador de antibiótico permite la selección de células con una inserción de transposón productivo. La inserción de carga útil del ADN puede aumentar la expresión en genes próximos al sitio de inserción a través de la acción del promotor fuerte.

40 Cualquiera de las cargas útiles de ADN de LoF o de GoF pueden contener además un marcador de contraselección además de un marcador de selección para permitir el reciclaje del marcador y, por lo tanto, más rondas de transformación.

45 La biblioteca de variantes de cepas generadas mediante el método de mutagénesis de transposones descrito antes se puede cribar frente a un fenotipo deseado. Las cepas se pueden cultivar y ensayar en alta capacidad para identificar cepas con un fenotipo deseado mejorado en relación con la cepa parental.

50 Las variantes de cepas mejoradas se pueden someter a rondas adicionales de transformación cíclica para mejorar aún más el fenotipo deseado, tal como el rendimiento de triptófano. Las rondas adicionales de transformación pueden ser compatibles con mutagénesis de transposones u otros tipos de bibliotecas tales como SNPSWP, PROSWP o mutagénesis aleatoria. Las cepas mejoradas también se pueden consolidar con otras variantes de cepas que presentan un fenotipo mejorado para producir una cepa mejorada adicional a través del efecto aditivo de distintas mutaciones beneficiosas.

55 Los métodos descritos en el presente documento reducen el coste de construcción de bibliotecas de alta calidad para cribado en ingeniería cíclica. La aplicación de mutagénesis de transposones a *Escherichia coli* permite la producción de miles de mutantes de LoF o GoF en todo el genoma en una sola reacción. Un método alternativo es construir miles de plásmidos asignados para transformar cepas mediante recombinación homóloga de cruzamiento simple (SCHR).

Otro método alternativo es construir miles de fragmentos lineales asignados para transformar cepas por recombinogénica con lambda Red. Ambos métodos son costosos porque requieren generar fragmentos de ADN únicos para cada mutante que contiene el ADN de carga útil previsto y la homología de secuencia que dirige la recombinación a una ubicación específica en el genoma diana. La mutagénesis de transposones usa una carga útil de ADN única y la diversidad se genera mediante la integración aleatoria en el genoma diana.

Ejemplo 11- Ingeniería genómica de HTP - Generación de cadenas principales de vectores para usar en ingeniería genómica de HTP en *Escherichia coli*

Este ejemplo describe la generación de vectores para usar en la ingeniería genómica de HTP para la ingeniería recombinogénica en *Escherichia coli* de manera que dichos vectores confieren transformación e integración de plásmidos eficientes.

Se generó el vector 1 (ácido nucleico de SEQ ID NO. 214) y comprende un origen de replicación R6K, un gen SacB, un gen PheS como marcador de contraselección y un marcador auxotrófico de levadura URA3. Con el fin de mejorar la eficiencia y la recombinación homóloga de cruzamiento simple, se modificó la cadena principal del vector 1 para que contuviera los elementos de la Tabla 15, lo que dio como resultado el vector 2 (ácido nucleico de SEQ ID NO. 215). El mapa del plásmido que se muestra en la Figura 55 muestra los componentes generales del vector 1. En el vector 2, se añadieron secuencias de aislantes aleatorias Aislante1 (SEQ ID NO. 218) y Aislante2 (SEQ ID NO. 219) flanqueando los brazos de homología, y las secuencias de terminadores T1 (ácido nucleico de SEQ ID NO. 220; véase Orosz et al., *Eur J Biochem.* 1 de noviembre de 1991;201 (3): 653-9) y B0015 (ácido nucleico de SEQ ID NO. 221) se añadieron para eliminar la lectura transcripcional en el sitio de inserción genómica. El mapa del plásmido que se muestra en la Figura 56 muestra los componentes generales del vector 2.

La utilidad de los vectores o plásmidos se ensayó en experimentos de inactivación. En resumen, se inoculó *E. coli* en caldo LB y se cultivó durante 8 horas a 37°C con agitación. Posteriormente, se usó entonces una parte alícuota del cultivo de la noche para inocular un volumen mayor de caldo LB y se cultivó durante 16 horas a 18°C con agitación. Para las transformaciones, se añadieron 100-400 ng de plásmido de ensayo a células competentes y la transformación se llevó a cabo mediante electroporación. Las células se recuperaron en medio SOC con incubación a 37°C durante 3 horas antes de sembrar en placa en LB-agar con kanamicina. La placa se incubó a 37°C para desarrollar las colonias con el plásmido de ensayo.

El gen diana que se inactivó era el gen *aroA* de *E. coli*. Así pues, los plásmidos de ensayo de "aroA-KO en la versión 1" (es decir, vector 1) y "aroA-KO en la versión 2" (es decir, vector 2) se construyeron por la inserción de brazos de homología en el gen *aroA* de *E. coli*. en las cadenas principales del vector 1 (versión 1) y del vector 2 (versión 2), respectivamente, de modo que los brazos de homología flanquearan un gen de resistencia a la kanamicina para permitir la recombinación de homología de cruzamiento simple en la célula hospedante de *E. coli*. La transformación de estos plásmidos de ensayo y la selección en kanamicina verificó que "aroA-KO en la versión 2" mostraba una eficiencia mejorada de transformación e integración del plásmido (Figura 53).

La modificación adicional de la cadena principal del vector 1 en el vector 2 permitió una contraselección eficaz en medios que contenían sacarosa y 4-clorofenilalanina añadiendo la secuencia PheS (Tabla 15). Cabe señalar que la secuencia del promotor PheS en el vector 2 consiste en el promotor PL del fago λ identificado por Kincade y deHaseth (véase Kinacade y deHaseth *Gene.* 2 de enero de 1991; 97(1): 7-12) seguida inmediatamente por una secuencia RBS llamada B0032 que procedía de iGEM. Además, en el vector 2, la secuencia promotora del gen *sacB* se reemplazó con un promotor que contenía el promotor P5-MCD2 (Mutalik et al, *Nat Methods.* Abril 2013; 10 (4): 354-60) y un ATG adicional. Esta modificación permitió una contraselección eficiente en sacarosa con cepas integradas con la cadena principal de la versión 2. Para generar el vector 3, la secuencia de promotor y el CDS del gen *pheS** de *C. glutamicum* en la cadena principal del vector 1 se reemplazaron con una nueva secuencia de promotor y CDS, específicamente una versión de codón optimizado de la *pheS* de *E. coli* nativa que contiene las mutaciones requeridas (T251A/A294G, véase Miyazaki, K. *Biotechniques.* 1 de febrero de 2015; 58 (2): 86-8) (Tabla 15). Esta modificación permitía una contraselección mejorada en 4 clorofenilalanina con cepas integradas con la cadena principal del vector 3. El mapa de plásmido que se muestra en la Figura 57 muestra los componentes generales del vector 3.

La cadena principal usada para la ingeniería genómica de HTP puede contener varios marcadores de selección de levaduras para usar en el ensamblaje de plásmidos. En la presente descripción, la modificación de la cadena principal del vector 3 reemplazó el marcador de selección de levadura URA3 con un marcador TRP1 para dar el vector 4. El mapa de plásmido mostrado en la Figura 58 muestra los componentes generales del vector 4.

Ejemplo 12 - Ingeniería genómica de HTP - Generación y ensayo de una biblioteca de intercambio de promotores adicional para su uso en la mejora de una cepa microbiana industrial

Este ejemplo describe la generación de una biblioteca de PROSWP adicional para su uso posterior en los métodos de ingeniería genómica de HTP proporcionados en el presente documento para transformar genéticamente células hospedantes microbianas (p. ej., *Escherichia coli*) en un esfuerzo por mejorar el rendimiento de cepas industriales.

En este ejemplo, se compiló una serie de promotores de *E. coli* nativos y promotores sintéticos para generar la biblioteca de intercambio de promotores que se encuentra en la Tabla 1.4. Para los promotores nativos, se seleccionó

un conjunto de secuencias de promotores de 60-90 pb de longitud del genoma de una cepa de *E. coli* K-12 (es decir, *E. coli* W3110). En particular, se seleccionaron promotores que mostraban una variación mínima en la expresión del gen asociado, de acuerdo con los datos de expresión basados en micromatrices en múltiples condiciones de crecimiento (Lewis et al., *Mol Syst Biol.* 2010; 6: 390). Las secuencias de promotores nativos consistían en 50 pb delante de los sitios de inicio de la transcripción putativos, así como la secuencia hasta, pero sin incluir, el codón de inicio putativo (véase Tabla 1.4). Además, se creó un conjunto de secuencias de promotores sintéticas quiméricas que consisten en porciones de los promotores conocidos del fago lambda p_L y p_R , el promotor delante del gen *acs* de *E. coli* y secuencias variables de 6 pb que constituyen las regiones -35 y -10 (Figura 54, Tabla 1.5). Cada uno de los promotores sintéticos tenía una longitud de 60 a 90 pb.

Con el fin de ensayar la capacidad de cada uno de los promotores que se encuentran en la Tabla 1.4 para dirigir la expresión de un gen operativamente unido al mismo, se construyó un conjunto de plásmidos de replicación de bajo número de copias, cada uno de los cuales contiene el gen RFP dirigido por uno de los promotores dados en la Tabla 1.4. El plásmido de replicación de bajo número de copias de elección fue un plásmido llamado Ori_Plsmid27, que tiene la secuencia nucleica de SEQ ID NO. 213. El vector se eligió porque se deseaba un plásmido de replicación con el fin de construir y evaluar la biblioteca de promotores lo más rápidamente posible y un plásmido de bajo número de copias tal como Ori_Plsmid27 se aproximaría más al escenario en el que solo se integra una sola copia en el genoma. Ori_Plsmid27 es de bajo número de copias porque posee el origen de replicación p15A de *E. coli*. El origen de replicación p15A típicamente da como resultado aproximadamente 10 copias del plásmido en cada célula. Esto es de "bajo número de copias" en comparación con otros plásmidos comunes que pueden mantener >20 o incluso varios cientos de copias de plásmido por célula.

Los plásmidos se construyeron usando técnicas convencionales de biología molecular. Específicamente, se adquirieron cebadores de PCR directos que consistían en una secuencia para reasociación con el gen RFP; la secuencia de promotor a introducir; y una secuencia que se solapaba con Ori_Plsmid27. Se obtuvo un solo cebador de PCR inversa que consistía en una secuencia para reasociarse con el terminador ECK120033737 (un terminador nativo de *E. coli*) y una secuencia que se solapaba con Ori_Plsmid27. El gen RFP se amplificó por PCR con los cebadores directos y el cebador inverso para generar un conjunto de amplicones de PCR, cada uno de los cuales contiene el gen RFP y uno de los promotores dados en la Tabla 1.4. Los plásmidos se construyeron digiriendo Ori_Plsmid27 con la enzima de restricción XhoI e insertando el amplicón de PCR correspondiente usando una mezcla de enzimas de ensamblaje de ADN comercial. Como control negativo, se generó una construcción que comprendía el terminador Tsod de *C. glutamicum* (ácido nucleico de SEQ ID NO. 224 en la Tabla 16) situado secuencia arriba del gen RFP.

Los plásmidos se transformaron mediante electroporación en *E. coli* W3110. Para cada promotor a evaluar, se seleccionaron cuatro colonias y se inocularon en 1 ml de caldo LB que contenía 25 µg/ml de kanamicina en una placa de cultivo de 96 pocillos. Los cultivos se desarrollaron a 37°C durante la noche con agitación a 1000 rpm. Se usaron 10 µl de cultivo para inocular 1 ml de Medio 1 (un medio rico que contiene glucosa, extracto de levadura, sal y tampón de fosfato) que contiene 25 µg/ml de kanamicina en una placa de cultivo de 96 pocillos. Los cultivos se desarrollaron a 37°C durante 24 horas con agitación a 1000 rpm. Los cultivos se diluyeron en agua en una placa de 96 pocillos de fondo transparente y pared negra y se tomaron dos medidas en un espectrofotómetro: DO₆₀₀ (densidad óptica a 600 nm) y fluorescencia (excitación a 554 nm, emisión a 590 nm). Se usaron 10 µl de los cultivos en el Medio 1 para inocular 1 ml del Medio 2 (un medio rico que contiene más glucosa que el Medio 1 pero solo una pequeña cantidad de extracto de levadura, que en su lugar contenía sulfato de amonio como fuente de nitrógeno junto con oligoelementos) que contenía 25 µg/ml de kanamicina en una placa de cultivo de 96 pocillos. Los cultivos en el Medio 2 también se desarrollaron a 37°C durante la noche y se midieron después de 24 horas.

Las mediciones de DO₆₀₀ se corrigieron restando el valor de los pocillos en blanco (pocillos que contenían solo medio). La Tabla 16 muestra las medidas de fluorescencia normalizadas para la DO₆₀₀ corregida. Como puede verse en la Tabla 16, las cepas resultantes se cultivaron eficazmente en dos medios diferentes y permitieron la expresión de proteínas fluorescentes en un intervalo de ~5000 veces.

Tabla 16. Niveles de expresión de RFP para construcciones de promotor-RFP en medios diferentes (2).

Nombre del promotor	SEQ ID NO.	Tipo*	Fluorescencia en medio 1	Fluorescencia en medio 2
Vector vacío		Control	0,59310262	1,038945766
Sin promotor		Control	0,372576311	0,759290158
Terminador	224	Control	0,41480568	0,638974345
b0904_promotor	71	Nativo	797,7465067	5344,268555
b2405_promotor	72	Nativo	181,125065	918,6698873
b0096_promotor	73	Nativo	124,2236508	404,6536222

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*	Fluorescencia en medio 1	Fluorescencia en medio 2
b0576_promotor	74	Nativo	22,74750166	231,1578591
b2017_promotor	75	Nativo	43,07006732	214,1483559
b1278_promotor	76	Nativo	24,51965879	167,5502652
b4255_promotor	77	Nativo	35,05859344	159,6253961
b0786_promotor	78	Nativo	34,42969151	136,7168488
b0605_promotor	79	Nativo	19,80356197	116,9009802
b1824_promotor	80	Nativo	24,81069692	111,2950744
b1061_promotor	81	Nativo	23,595542	104,971453
b0313_promotor	82	Nativo	20,68806067	75,34101875
b0814_promotor	83	Nativo	16,46706854	70,36934251
b4133_promotor	84	Nativo	18,1034447	67,11707926
b4268_promotor	85	Nativo	12,80314528	60,91103948
b0345_promotor	86	Nativo	9,57124889	56,38701501
b2096_promotor	87	Nativo	9,852378505	44,58545057
b1277_promotor	88	Nativo	7,821796935	43,50459312
b1646_promotor	89	Nativo	18,43768693	33,66381279
b4177_promotor	90	Nativo	12,51676981	30,71734482
b0369_promotor	91	Nativo	6,614285367	28,3955664
b1920_promotor	92	Nativo	6,196134846	28,12617522
b3742_promotor	93	Nativo	5,646106239	27,85277232
b3929_promotor	94	Nativo	9,120314333	27,26180409
b3743_promotor	95	Nativo	5,416134854	26,11889953
b1613_promotor	96	Nativo	6,653798388	24,49553744
b1749_promotor	97	Nativo	4,14048349	24,04592462
b2478_promotor	98	Nativo	5,387025634	20,49940077
b0031_promotor	99	Nativo	3,227415659	18,8898903
b2414_promotor	100	Nativo	5,054709408	17,84254467
b1183_promotor	101	Nativo	5,109245812	14,42155472
b0159_promotor	102	Nativo	3,662900763	14,16361794
b2837_promotor	103	Nativo	3,427953441	13,22830567
b3237_promotor	104	Nativo	3,238274954	11,27938911
b3778_promotor	105	Nativo	1,778132048	8,89448301
b2349_promotor	106	Nativo	1,918547188	8,854115649
b1434_promotor	107	Nativo	3,05979593	7,982575199

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*	Fluorescencia en medio 1	Fluorescencia en medio 2
b3617_promotor	108	Nativo	1,631428047	7,872487968
b0237_promotor	109	Nativo	2,524629463	6,00416496
b4063_promotor	110	Nativo	1,75938278	5,338598994
b0564_promotor	111	Nativo	1,099810487	5,27271929
b0019_promotor	112	Nativo	1,407515599	5,174702317
b2375_promotor	113	Nativo	0,758754127	4,329319889
b1187_promotor	114	Nativo	1,042830698	4,119645144
b2388_promotor	115	Nativo	1,068375672	4,008318722
b1051_promotor	116	Nativo	2,595570861	3,770122087
b4241_promotor	117	Nativo	0,86986247	3,520447288
b4054_promotor	118	Nativo	1,166071321	3,009684331
b2425_promotor	119	Nativo	0,684355198	2,847357334
b0995_promotor	120	Nativo	0,783460564	2,701492023
b1399_promotor	121	Nativo	0,67920166	2,675743061
b3298_promotor	122	Nativo	0,693842895	2,248914103
b2114_promotor	123	Nativo	0,680005978	2,166569672
b2779_promotor	124	Nativo	0,516911046	1,708793442
b1114_promotor	125	Nativo	0,477678475	1,423398279
b3730_promotor	126	Nativo	0,527510171	1,40174991
b3025_promotor	127	Nativo	0,528069061	1,375386866
b0850_promotor	128	Nativo	0,580500872	1,365244168
b2365_promotor	129	Nativo	0,600099606	1,293746086
b4117_promotor	130	Nativo	0,465598493	1,292314074
b2213_promotor	131	Nativo	0,502863501	1,066458143
pMB029_promotor	132	Sintético	3244,159758	6397,50355
pMB023_promotor	133	Sintético	2716,103558	5324,355721
pMB025_promotor	134	Sintético	2529,260134	5315,845766
pMB019_promotor	135	Sintético	2632,199487	4730,881793
pMB008_promotor	136	Sintético	3183,229982	4685,405472
pMB020_promotor	137	Sintético	1719,381656	4331,834596
pMB022_promotor	138	Sintético	1432,727416	4224,994617
pMB089_promotor	139	Sintético	1588,496061	4173,815385
pMB001_promotor	140	Sintético	1344,15431	3710,844718
pMB051_promotor	141	Sintético	1485,624649	3614,713536

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*	Fluorescencia en medio 1	Fluorescencia en medio 2
pMB070_promotor	142	Sintético	1148,371596	3494,100223
pMB074_promotor	143	Sintético	1372,63931	3423,518954
pMB046_promotor	144	Sintético	1070,100822	3325,635875
pMB071_promotor	145	Sintético	829,6337908	2753,919683
pMB013_promotor	146	Sintético	649,3261186	2255,268459
pMB080_promotor	147	Sintético	584,6548742	2064,645666
pMB03_8_promotor	148	Sintético	580,7846278	2043,19218
pMB060_promotor	149	Sintético	575,1167059	1916,349295
pMB064_promotor	150	Sintético	376,0013091	1390,767111
pMB05_8_promotor	151	Sintético	292,4875916	1093,376645
pMB085_promotor	152	Sintético	288,8178401	1085,653214
pMB081_promotor	153	Sintético	278,7856475	670,8190328
pMB091_promotor	154	Sintético	296,1836771	655,9453316
pMB027_promotor	155	Sintético	262,5866089	601,1559107
pMB048_promotor	156	Sintético	235,0853436	567,7578783
pMB055_promotor	157	Sintético	189,6983907	499,3358783
pMB006_promotor	158	Sintético	205,2252242	497,2257006
pMB012_promotor	159	Sintético	118,144338	480,5859474
pMB014_promotor	160	Sintético	119,9861606	474,4330596
pMB028_promotor	161	Sintético	204,8853226	454,9894988
pMB059_promotor	162	Sintético	101,7399612	387,0993978
pMB061_promotor	163	Sintético	128,6404482	381,8073111
pMB043_promotor	164	Sintético	139,0031576	380,0980602
pMB066_promotor	165	Sintético	95,64407537	328,5708555
pMB079_promotor	166	Sintético	105,9768914	312,5115498
pMB032_promotor	167	Sintético	89,71910699	311,1818243
pMB068_promotor	168	Sintético	97,56292371	276,8593245
pMB082_promotor	169	Sintético	84,20594634	262,0127742
pMB030_promotor	170	Sintético	92,28874802	255,1052708
pMB067_promotor	171	Sintético	49,74078463	243,9642233
pMB050_promotor	172	Sintético	69,03704925	223,6129199
pMB069_promotor	173	Sintético	66,09043115	213,5258967
pMB017_promotor	174	Sintético	50,43740198	213,0316985
pMB039_promotor	175	Sintético	52,58639745	209,9432421

ES 2 875 579 T3

Nombre del promotor	SEQ ID NO.	Tipo*	Fluorescencia en medio 1	Fluorescencia en medio 2
pMB011_promotor	176	Sintético	63,05192023	194,8104082
pMB072_promotor	177	Sintético	47,21948722	165,6730417
pMB016_promotor	178	Sintético	47,20169015	163,6018648
pMB077_promotor	179	Sintético	43,98192292	158,5866563
pMB047_promotor	180	Sintético	30,55918215	121,2753356
pMB052_promotor	181	Sintético	34,86596074	114,4169675
pMB090_promotor	182	Sintético	14,01229264	55,03412277
pMB035_promotor	183	Sintético	14,7011564	50,95895734
pMB073_promotor	184	Sintético	11,79654115	47,47608915
pMB004_promotor	185	Sintético	8,701914436	33,9085268
pMB054_promotor	186	Sintético	8,934971498	32,06510507
pMB024_promotor	187	Sintético	6,760860897	27,59246488
pMB007_promotor	188	Sintético	11,5557009	25,41738039
pMB005_promotor	189	Sintético	5,354276069	23,77742451
pMB003_promotor	190	Sintético	6,389447374	22,7246451
pMB088_promotor	191	Sintético	5,873205603	22,42214302
pMB065_promotor	192	Sintético	5,44173707	20,60442307
pMB037_promotor	193	Sintético	4,809680789	19,26108105
pMB009_promotor	194	Sintético	3,591881545	13,6498423
pMB041_promotor	195	Sintético	4,78150005	13,03957332
pMB036_promotor	196	Sintético	2,508122856	11,98560161
pMB049_promotor	197	Sintético	2,916151498	11,02944113
pMB044_promotor	198	Sintético	2,719080697	10,47251128
pMB042_promotor	199	Sintético	2,661753833	8,883733663
pMB086_promotor	200	Sintético	2,33212712	8,316649305
pMB053_promotor	201	Sintético	2,336160735	7,262981543
pMB057_promotor	202	Sintético	1,518801989	6,27420435
pMB018_promotor	203	Sintético	1,261000991	4,714509042
pMB002_promotor	204	Sintético	1,290587949	4,473711828
pMB015_promotor	205	Sintético	0,994888783	4,328541658
pMB087_promotor	206	Sintético	0,94621781	2,900396821
pMB063_promotor	207	Sintético	0,425085793	1,459419227
*Promotores nativos de Escherichia coli				

Ejemplo 13 - Ingeniería genómica de HTP - Ensayo de integración de la biblioteca de intercambio de promotores de la Tabla 1.4 en el genoma de *E. coli* usando la cadena principal del vector 2

5 Este ejemplo describe una prueba de concepto del uso de la cadena principal del vector 2 del Ejemplo 11 en combinación con un subconjunto de promotores que forman la biblioteca de promotores de la Tabla 1.4 para dirigir la integración de una sola copia de una construcción heteróloga de promotor-gen en el genoma de *E. coli*.

Para este ejemplo, se construirá un conjunto de plásmidos para insertar genes fluorescentes RFP y GFP en dos locus (nupG y asl) en *E. coli* con un subconjunto de 14 promotores del conjunto de la Tabla 1.4. Esto permitirá que esos promotores se evalúen como una única copia integrada en el genoma, en lugar de en plásmidos de replicación de bajo número de copias (véase el Ejemplo 12).

10 Los plásmidos comprenderán brazos de homología que flanquean los genes RFP o GFP con el fin de facilitar la integración en el genoma de *E. coli* mediante "entrada de bucle" como se proporciona a lo largo de esta descripción. Se analizará en las cepas resultantes la fluorescencia, lo que demostrará que este subconjunto de 14 promotores de la Tabla 1.4 se ha ensayado usando una cadena principal de vector descrita en el Ejemplo 11 y se puede usar para insertar un gen heterólogo en el genoma de *E. coli* usando el métodos descritos en esta descripción.

15 Ejemplo 14 - Ingeniería genómica de HTP - Implementación de métodos de PROSWP usando una biblioteca de promotores derivada de la Tabla 1.4.

20 La siguiente sección proporciona una implementación ilustrativa de las herramientas del programa de mejora de cepas de diseño de HTP de intercambio de PRO de la presente descripción, como se describe en los Ejemplos 4 y 5. En este ejemplo, una cepa de *E. coli* se sometió a los métodos de intercambio de PRO de la presente descripción con el fin de modular la expresión de genes en el genoma de *E. coli*. Este ejemplo se basa en los resultados de los Ejemplos 12 y 13 en cuanto que este ejemplo ilustra el uso de una biblioteca de promotores que comprende los promotores de la Tabla 1.4 en los métodos PROSWP de la presente descripción.

A. Intercambio de promotores

25 Los intercambios de promotores se llevarán a cabo como se describe en el Ejemplo 4. Los genes del genoma de *E. coli* se someterán a intercambios de promotores usando la biblioteca de promotores descrita en el Ejemplo 13, que comprende un subconjunto de 14 promotores de la Tabla 1.4. El subconjunto de 14 promotores que se usará en este ejemplo se seleccionará basándose en su efecto en la expresión génica según se determina en los ejemplos 12 y 13.

B. Ingeniería de HTP y cribado de alta capacidad

30 La ingeniería de HTP de los intercambios de promotores se llevará a cabo como se describe en los Ejemplos 1 y 3. El cribado de HTP de las cepas de intercambio de promotores resultantes se llevará a cabo como se describe en el Ejemplo 3. En total, se llevarán a cabo 14 intercambios de PRO. Finalmente, se analizará el impacto de estas modificaciones en la producción de productos de interés.

Ejemplo 15 - Ingeniería genómica de HTP: implementación de una biblioteca de intercambio de TERMINADORES para mejorar el rendimiento de cepas para la producción de licopeno

35 La siguiente sección proporciona una implementación ilustrativa de las herramientas del programa de mejora de cepas de diseño de HTP de intercambio de TERMINADORES de la presente descripción. En este ejemplo, una cepa de *E. coli* se sometió a los métodos de intercambio de TERMINADORES de la presente descripción con el fin de afectar al rendimiento de licopeno de la célula hospedante.

40 Los intercambios de terminadores que se dirigen a genes en la ruta biosintética del licopeno que se muestra en la Figura 59 se llevaron a cabo usando los métodos de intercambio de terminadores presentes a lo largo de esta descripción. Las construcciones se diseñaron como se describe a continuación y la recombinación fue mediada con el sistema CRISPR/Cas9. Los terminadores usados para los intercambios de terminadores en este ejemplo eran los terminadores que se encuentran en la Tabla 19.

Tabla 19 - Terminadores usados en este ejemplo para dirigir genes en la ruta biosintética del licopeno

Nombre	Descripción	Longitud (pb)
Spy	Terminador (SEQ ID NO. 225)	90
pheA	Terminador (SEQ ID NO. 226)	51
osmE	Terminador (SEQ ID NO. 227)	42
rpoH	Terminador (SEQ ID NO. 228)	41

vibE	Terminador (SEQ ID NO. 229)	71
Thr1_ABC	Terminador (SEQ ID NO. 230)	57

Construcción del diseño

5 Se identificó un ARN guía de 20 bases cerca de la secuencia de inserción de la diana y adyacente a una secuencia NGG PAM para cortar el genoma en la posición deseada. La secuencia que se pretendía insertar en el genoma estaba flanqueada en ambos extremos por 90 bases de homología, de modo que la homología dirigiría la secuencia nativa para ser eliminada o retenida según se desee. Debe indicarse que aunque la recombinación se facilitó por el sistema CRISPR/Cas9 en este Ejemplo, todas las cepas también pueden construirse por métodos tradicionales de recombinación homóloga de cruzamiento simple y doble así como el sistema Lambda Red como se describe a lo largo de la presente descripción. Así pues, cada uno de los tipos de bibliotecas de intercambio de terminadores es independiente del método de construcción/recombinación.

Inoculación del cultivo de semillas

Se seleccionó una colonia de cepa base de edición (W3110 pKD46-cas9 pLYC4) de una placa de Petri y se inoculó en un volumen mayor de LB clin100 cmp25 y se cultivó a 30°C con agitación durante ~16 horas.

Preparación de células competentes y transformación

15 Se preparó una dilución 1:10 del cultivo de una noche y se midió la DO600. Se inoculó LB clin100 cmp25 a una DO600 de 0,05 y se cultivó a 30°C con agitación durante ~2 horas.

Después de 2 horas, se midió periódicamente la DO600 hasta que se alcanzó la DO objetivo de inducción y, una vez alcanzada la DO objetivo, se añadió 20% de arabinosa hasta una concentración final de 0,2%.

20 Se centrifuga el cultivo durante 5 minutos a 5.000 x G a 4°C. Se vierte el líquido sobrenadante y se resuspende hasta un volumen final equivalente al volumen de cultivo original.

Se repite la etapa 7 para lavar las células por tercera vez.

Después de 3 lavados, las células se sedimentaron y se resuspendieron en glicerol al 10% hasta ~1/250 del volumen de cultivo original. Se prepara una dilución 1:500 de las células resuspendidas, se prepararon y se resuspendieron a una DO600 deseada con un volumen adecuado para 40 ul de células por transformación.

25 En una placa de PCR Framestar (o tubos de microcentrífuga) se mezclaron 40 ul de células con 100 ng de plásmido de ARN guía y ~4 ul de molde de reparación de producto de PCR purificado. Si se usan oligos para el molde de reparación, los oligos se añadieron en una concentración final de 2 uM.

Las células se electroporaron y se resuspendieron inmediatamente en LB y se recuperaron en una placa de pocillos profundos durante 1 hora a 30°C con agitación.

30 Las células recuperadas se diluyeron y se sembraron en placas de LB agar clin100 kan50 cmp25 y a 30°C de 24 a 36 horas. Las colonias se cribaron por PCR de colonias, se secuenciaron o se cribaron por fenotipos. El plásmido pKD46-cas9 se puede curar por crecimiento a 37°C o más, y pCRISPR2 se puede curar por crecimiento en sacarosa al 10%.

Ingeniería de HTP y cribado de alta capacidad

35 La ingeniería de HTP de las combinaciones de intercambios se llevó a cabo como se describe en los Ejemplos 1 y 3, con la excepción de que se usó CRISPR/Cas 9 para facilitar la recombinación homóloga de construcciones en el genoma de *E. coli*. El cribado de HTP de las cepas de intercambio de promotores/intercambio de terminadores, cepas de intercambio de promotores/intercambio de marcadores de degradación, cepas de intercambio de promotores/intercambio de marcadores de solubilidad y cepas intercambio de promotores/intercambio de terminadores/intercambio de marcadores de degradación/intercambio de marcadores de solubilidad resultantes se llevó a cabo como se describe en el Ejemplo 3. Los resultados de los experimentos se muestran en las Figuras 60 y 61.

45 Como se muestra en la Figura 60, los intercambios de terminadores en las dianas de la ruta del licopeno *idi* y *ymgA* usando el terminador TyjbE demostraban un rendimiento de las cepas reducido con respecto al control, lo que destaca la utilidad de estos tipos de bibliotecas para identificar las dianas de rutas críticas. Esta conclusión estaba respaldada además por los resultados que se muestran en la Figura 61, donde se realizaron intercambios de terminadores en múltiples dianas de la ruta del licopeno.

Ejemplo 16 - Ingeniería genómica de HTP - implementación de una biblioteca de intercambio de TERMINADORES o una biblioteca de intercambio de PRO en combinación con una biblioteca de intercambio de MARCADORES de SOLUBILIDAD o biblioteca de intercambio de marcadores de DEGRADACIÓN para mejorar el rendimiento de las cepas para la producción de licopeno

La siguiente sección proporciona una implementación ilustrativa de las herramientas del programa de mejora de cepas de diseño de HTP intercambio de MARCADORES DE SOLUBILIDAD e intercambio de TERMINADORES de la presente descripción, así como las herramientas del programa de mejora de cepas de diseño de intercambio de PRO e intercambio de MARCADORES DE DEGRADACIÓN de la presente descripción. En este ejemplo, una cepa de *E. coli* se sometió al intercambio de PRO en combinación con métodos de intercambio de MARCADORES DE DEGRADACIÓN de la presente descripción, así como los métodos de intercambio de MARCADORES DE SOLUBILIDAD y TERMINADORES de la presente descripción con el fin de afectar al rendimiento de licopeno de la célula hospedante.

Intercambio de promotores/Intercambio de terminadores/Intercambio de marcadores de solubilidad/Intercambio de marcadores de degradación

El intercambio de terminadores se llevó a cabo como se describe en el ejemplo 15, mientras que el intercambio de marcador de solubilidad y el intercambio de promotores en combinación con los intercambios de marcadores de degradación se llevaron a cabo esencialmente como se describe en los ejemplos 4 y 5. Las construcciones de diseñaron como se describe a continuación y la recombinación fue mediada por el sistema CRISPR/Cas9. Los promotores bicistrónicos usados para los intercambios de promotores en este ejemplo eran de Mutalik et al., *Nat Methods*. Abril 2013; 10(4):354-60 y se puede encontrar en la Tabla 20. Debe indicarse que cualquiera de los promotores proporcionados en el presente documento podría usarse en los métodos descritos a continuación.

Tabla 20 - Promotores usados para combinaciones de intercambios de promotores en este ejemplo.

Nombre	Longitud	SEQ ID NO.
P3_BCD1	133	255
P4_BCD22	121	256
P7_BCD19	132	257
P8_BCD15	121	258
P11_BCD17	121	259
P13_BCD8	121	260

Construcción del diseño

Se identificó un ARN guía de 20 bases cerca de la secuencia de inserción de la diana y adyacente a una secuencia NGG PAM para cortar el genoma en la posición deseada. La secuencia que se pretendía insertar en el genoma estaba flanqueada en ambos extremos por 90 bases de homología, de modo que la homología dirigiría la secuencia nativa para ser eliminada o retenida según se desee. Debe indicarse que aunque la recombinación se facilitó por el sistema CRISPR/Cas9 en este Ejemplo, todas las cepas también pueden construirse por métodos tradicionales de recombinación homóloga de cruzamiento simple y doble así como el sistema Lambda Red como se describe a lo largo de la presente descripción. Así pues, cada uno de los tipos de bibliotecas (intercambio de promotores, intercambio de marcadores de solubilidad de proteínas, intercambio de marcadores de degradación de proteínas e intercambio de terminadores) solos o en combinación son independientes del método de construcción/recombinación.

Inoculación del cultivo de semillas

Se seleccionó una colonia de cepa base de edición (W3110 pKD46-cas9 pLYC4) de una placa de Petri y se inoculó en un volumen mayor de LB clin100 cmp25 y se cultivó a 30°C con agitación durante ~16 horas.

Preparación de células competentes y transformación

Se preparó una dilución 1:10 del cultivo de una noche y se midió la DO600. Se inoculó LB clin100 cmp25 a una DO600 de 0,05 y se cultivó a 30°C con agitación durante ~2 horas.

Después de 2 horas, se midió periódicamente la DO600 hasta que se alcanzó la DO objetivo de inducción y, una vez alcanzado el objetivo, se añadió 20% de arabinosa hasta una concentración final de 0,2%.

Se centrifuga el cultivo durante 5 minutos a 5.000 x G a 4°C. Se vierte el líquido sobrenadante y se resuspende hasta un volumen final equivalente al volumen de cultivo original.

Se repite la etapa 7 para lavar las células por tercera vez.

Después de 3 lavados, las células se sedimentaron y se resuspendieron en glicerol al 10% hasta ~1/250 del volumen de cultivo original. Se prepara una dilución 1:500 de las células resuspendidas, se prepararon y se resuspendieron a una DO600 deseada con un volumen adecuado para 40 ul de células por transformación.

5 En una placa de PCR Framestar (o tubos de microcentrífuga) se mezclaron 40 ul de células con 100 ng de plásmido de ARN guía y ~4 ul de molde de reparación de producto de PCR purificado. Si se usan oligos para el molde de reparación, los oligos se añadieron en una concentración final de 2 uM.

Las células se electroporaron y se resuspendieron inmediatamente en LB y se recuperaron en una placa de pocillos profundos durante 1 hora a 30°C con agitación.

10 Las células recuperadas se diluyeron y se sembraron en placas de LB agar clin100 kan50 cmp25 y a 30°C de 24 a 36 horas. Las colonias se cribaron por PCR de colonias, se secuenciaron o se cribaron por fenotipos. El plásmido pKD46-cas9 se puede curar por crecimiento a 37°C o más, y pCRISPR2 se puede curar por crecimiento en sacarosa al 10%.

Ingeniería de HTP y cribado de alta capacidad

15 La ingeniería de HTP de las combinaciones de intercambios se llevó a cabo como se describe en los Ejemplos 1 y 3, con la excepción de que se usó CRISPR/Cas 9 para facilitar la recombinación homóloga de construcciones en el genoma de *E. coli*. El cribado de HTP de las cepas de intercambio de promotores/intercambio de terminadores, cepas de intercambio de promotores/intercambio de marcadores de degradación, cepas de intercambio de promotores/intercambio de marcadores de solubilidad y cepas intercambio de promotores/intercambio de terminadores/intercambio de marcadores de degradación/intercambio de marcadores de solubilidad resultantes se llevó a cabo como se describe en el Ejemplo 3.

20 Debe indicarse que el promotor P3_BCD1 se usó en todas las cepas en las que se estaba estudiando la modificación en el locus dxs, a menos que se indique lo contrario, tal como P4_BCD22. En cualquier locus distinto de dxs, se usó la secuencia de promotor nativo a menos que se indique lo contrario. Esto significa que una cepa descrita como ssrA_LAA en el locus dxs, por ejemplo, también contenía P3_BCD1, pero una cepa descrita como ssrA_LAA en el locus gdhA usaba la secuencia de promotor nativo. El contenido completo de las cepas ensayadas en la Tabla 21 a
25 continuación.

Tabla 21 - Contenido de las cepas generadas en el Ejemplo 16.

Modificaciones en el locus dxs						
	*Control -P	P4_BCD2 2	ssrA_AS V	ssrA_LA A	Tspy	TyjbE
Promotor	P3_BCD 1	P4_BCD2 2	P3_BCD1	P3_BCD1	P3_BCD 1	P3_BCD 1
Marcador de solubilidad	ninguno	ninguno	ninguno	ninguno	ninguno	ninguno
Marcador de degradación de	ninguno	ninguno	ssrA_AS V	ssrA_LAA	ninguno	ninguno
Terminador	secuencia nativa	secuencia nativa	secuencia nativa	secuencia nativa	Tspy	TyjbE
Modificaciones en el locus gdhA						
	*Control-P	FH8	GB1	P4_BCD2 2	Tspy	TyjbE
Promotor	secuencia nativa	secuencia nativa	secuencia nativa	P4_BCD2 2	secuencia nativa	secuencia nativa
Marcador de solubilidad	ninguno	FH8	GB1	ninguno	ninguno	ninguno
Marcador de degradación de	ninguno	ninguno	ninguno	ninguno	ninguno	ninguno
Terminador	secuencia nativa	secuencia nativa	secuencia nativa	secuencia nativa	Tspy	TyjbE

Los resultados de los experimento se resumen y se muestran en las Figuras 62 y 63.

5 Como se muestra en la Figura 62, el marcador de degradación *ssrA_LAA* demuestra un rendimiento de cepas mejorado en relación con el control. Esto es inesperado ya que esta cepa es una combinación de un PROSWP con una marcador de degradación en una única diana de la ruta. Se espera que el PROSWP inicial aumente la abundancia de proteínas, y se espera que el marcador de degradación disminuya la abundancia de proteínas, demostrando así la utilidad de las combinaciones de tipos de bibliotecas para ajustar el rendimiento óptimo de las cepas. Como se muestra en la Figura 63, el marcador de solubilidad FH8 demuestra un rendimiento de cepas mejorado en relación con el control, pero el marcador de solubilidad GB1 no lo hace, demostrando así la necesidad de evaluar bibliotecas de cada tipo de modificación.

15 En general, lo que demuestra este ejemplo es que, aunque los componentes de la presente descripción pueden ser útiles individualmente para la mejora sistemática de cepas, también se pueden usar sinérgicamente con otros enfoques. Por ejemplo, después de mejorar la estabilidad del ARNm por modificación del terminador, se puede insertar un promotor fuerte para aumentar más la producción de proteínas más allá del nivel de cualquiera de los enfoques por sí solo. Asimismo, este nuevo nivel elevado de producción de proteínas se puede mejorar más mediante una marcador de solubilidad de proteínas junto con otras modificaciones. Cuando se emplean juntos y con enfoques previos, los componentes de la presente descripción pueden permitir una mejora de la cepa más robusta y eficaz para la producción de una molécula diana.

20 SECUENCIAS DE LA DESCRIPCIÓN CON IDENTIFICADORES SEQ ID NO

NOMBRE ABREVIADO)	(NOMBRE FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
Pcg0007_lib_39 (P1)		1		
Pcg0007 (P2)		2		
Pcg1860 (P3)		3		
Pcg0755 (P4)		4		
Pcg0007_265 (P5)		5		

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
Pcg3381 (P6)			6		
Pcg0007_119 (P7)			7		
Pcg3121 (P8)			8		
cg0001 (T1)			9		
cg0007 (T2)			10		
cg0371 (T3)			11		
cg0480 (T4)			12		
cg0494 (T5)			13		
cg0564 (T6)			14		
cg0610 (T7)			15		
cg0695 (T8)			16		
Secuencia de nucleótidos Cisl			17		
Conector rico en serina			18		
Conector rico en serina			19		
Conector rico en serina			20		
Conector rico en alanina			21		
conector			22		
conector			23		
conector			24		
conector			25		
Zif268			26		Secuencia de dominio de unión de ADN
PBSII			27		Secuencia de dominio de unión de ADN
ZFa			28		Secuencia de dominio de unión de ADN
ZFb			29		Secuencia de dominio de unión de ADN
ZFc			30		Secuencia de dominio de unión de ADN
Tyr123			31		Secuencia de dominio de unión de ADN
Tyr456			32		Secuencia de dominio de unión de ADN
Dedo de zinc Blues			33		Secuencia de dominio de unión de ADN

ES 2 875 579 T3

NOMBRE (NOMBRE ABREVIADO)	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
Dedo de zinc Jazz		34		Secuencia de dominio de unión de ADN
Dedo de zinc Bagly		35		Secuencia de dominio de unión de ADN
Sitio de unión de dedo de zinc Bagly		36		Secuencia de unión de ADN (5'→3')
GliI		37		Secuencia de dominio de unión de ADN
Sitio de unión de GliI		38		Secuencia de unión de ADN (5'→3')
Dedo de zinc HIVC		39		Secuencia de dominio de unión de ADN
Dedos de zinc B3		40		Secuencia de dominio de unión de ADN
Dedo de zinc N1		41		Secuencia de dominio de unión de ADN
Primer dedo de zinc Sp-1		42		Secuencia de dominio de unión de ADN
Segundo dedo Sp-1		43		Secuencia de dominio de unión de ADN
Dominio de unión de proteína SH3 clase I			44	
Sitio de unión de proteína SH3 clase II			45	
Sitio de unión de proteína SH3			46	
Péptido de reclutamiento SH3			47	
Péptido de reclutamiento SH3			48	
Péptido de reclutamiento SH3			49	
Dominio de unión de proteína PDZ			50	
Dominio de unión de proteína PDZ			51	
Dominio de unión de proteína PDZ			52	
Dominio de unión de proteína PDZ			53	
Dominio de unión de proteína PDZ			54	
Péptido de reclutamiento PDZ			55	
Péptido de reclutamiento PDZ			56	
Péptido de reclutamiento PDZ			57	
Péptido de reclutamiento PDZ			58	
Péptido de reclutamiento PDZ			59	
Péptido de reclutamiento PDZ			60	

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
Péptido de reclutamiento PDZ				61	
Dominio de unión de proteína GBD				62	
Dominio de unión de proteína GBD				63	
Dominio de unión de proteína GBD				64	
Dominio de unión de proteína GBD				65	
Dominio de unión de cremallera de leucina				66	
Dominio de unión de cremallera de leucina				67	
Dominio de unión de cremallera de leucina				68	
Secuencia reguladora de diseño bicistrónico (BCD)				69	<i>véase la Figura 43</i>
Secuencia reguladora de diseño bicistrónico (BCD)				70	<i>véase la Figura 43</i>
b0904_promotor		<i>E. coli</i>	71		
b2405_promotor		<i>E. coli</i>	72		
b0096_promotor		<i>E. coli</i>	73		
b0576_promotor		<i>E. coli</i>	74		
b2017_promotor		<i>E. coli</i>	75		
b1278_promotor		<i>E. coli</i>	76		
b4255_promotor		<i>E. coli</i>	77		
b0786_promotor		<i>E. coli</i>	78		
b0605_promotor		<i>E. coli</i>	79		
b1824_promotor		<i>E. coli</i>	80		
b1061_promotor		<i>E. coli</i>	81		
b0313_promotor		<i>E. coli</i>	82		
b0814_promotor		<i>E. coli</i>	83		
b4133_promotor		<i>E. coli</i>	84		
b4268_promotor		<i>E. coli</i>	85		
b0345_promotor		<i>E. coli</i>	86		
b2096_promotor		<i>E. coli</i>	87		
b1277_promotor		<i>E. coli</i>	88		
b1646_promotor		<i>E. coli</i>	89		
b4177_promotor		<i>E. coli</i>	90		

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
b0369_promotor		<i>E. coli</i>	91		
b1920_promotor		<i>E. coli</i>	92		
b3742_promotor		<i>E. coli</i>	93		
b3929_promotor		<i>E. coli</i>	94		
b3743_promotor		<i>E. coli</i>	95		
b1613_promotor		<i>E. coli</i>	96		
b1749_promotor		<i>E. coli</i>	97		
b2478_promotor		<i>E. coli</i>	98		
b0031_promotor		<i>E. coli</i>	99		
b2414_promotor		<i>E. coli</i>	100		
b1183_promotor		<i>E. coli</i>	101		
b0159_promotor		<i>E. coli</i>	102		
b2837_promotor		<i>E. coli</i>	103		
b3237_promotor		<i>E. coli</i>	104		
b3778_promotor		<i>E. coli</i>	105		
b2349_promotor		<i>E. coli</i>	106		
b1434_promotor		<i>E. coli</i>	107		
b3617_promotor		<i>E. coli</i>	108		
b0237_promotor		<i>E. coli</i>	109		
b4063_promotor		<i>E. coli</i>	110		
b0564_promotor		<i>E. coli</i>	111		
b0019_promotor		<i>E. coli</i>	112		
b2375_promotor		<i>E. coli</i>	113		
b1187_promotor		<i>E. coli</i>	114		
b23 8 8_promotor		<i>E. coli</i>	115		
b1051_promotor		<i>E. coli</i>	116		
b4241_promotor		<i>E. coli</i>	117		
b4054_promotor		<i>E. coli</i>	118		
b2425_promotor		<i>E. coli</i>	119		
b0995_promotor		<i>E. coli</i>	120		
b1399_promotor		<i>E. coli</i>	121		
b3298_promotor		<i>E. coli</i>	122		
b2114_promotor		<i>E. coli</i>	123		

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
b2779_promotor		<i>E. coli</i>	124		
b1114_promotor		<i>E. coli</i>	125		
b3 73 0_promotor		<i>E. coli</i>	126		
b3025_promotor		<i>E. coli</i>	127		
b0850_promotor		<i>E. coli</i>	128		
b2365_promotor		<i>E. coli</i>	129		
b4117_promotor		<i>E. coli</i>	130		
b2213_promotor		<i>E. coli</i>	131		
pMB029_promotor		Sintética	132		
pMB023_promotor		Sintética	133		
pMB025_promotor		Sintética	134		
pMB019_promotor		Sintética	135		
pMB008_promotor		Sintética	136		
pMB020_promotor		Sintética	137		
pMB022_promotor		Sintética	138		
pMB089_promotor		Sintética	139		
pMB001_promotor		Sintética	140		
pMB051_promotor		Sintética	141		
pMB070_promotor		Sintética	142		
pMB074_promotor		Sintética	143		
pMB046_promotor		Sintética	144		
pMB071_promotor		Sintética	145		
pMB013_promotor		Sintética	146		
pMB080_promotor		Sintética	147		
pMB03 8_promotor		Sintética	148		
pMB060_promotor		Sintética	149		
pMB064_promotor		Sintética	150		
pMB05 8_promotor		Sintética	151		
pMB085_promotor		Sintética	152		
pMB081_promotor		Sintética	153		
pMB091_promotor		Sintética	154		
pMB027_promotor		Sintética	155		
pMB048_promotor		Sintética	156		

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
pMB055_promotor		Sintética	157		
pMB006_promotor		Sintética	158		
pMB012_promotor		Sintética	159		
pMB014_promotor		Sintética	160		
pMB028_promotor		Sintética	161		
pMB059_promotor		Sintética	162		
pMB061_promotor		Sintética	163		
pMB043_promotor		Sintética	164		
pMB066_promotor		Sintética	165		
pMB079_promotor		Sintética	166		
pMB032_promotor		Sintética	167		
pMB068_promotor		Sintética	168		
pMB082_promotor		Sintética	169		
pMB030_promotor		Sintética	170		
pMB067_promotor		Sintética	171		
pMB050_promotor		Sintética	172		
pMB069_promotor		Sintética	173		
pMB017_promotor		Sintética	174		
pMB039_promotor		Sintética	175		
pMB011_promotor		Sintética	176		
pMB072_promotor		Sintética	177		
pMB016_promotor		Sintética	178		
pMB077_promotor		Sintética	179		
pMB047_promotor		Sintética	180		
pMB052_promotor		Sintética	181		
pMB090_promotor		Sintética	182		
pMB035_promotor		Sintética	183		
pMB073_promotor		Sintética	184		
pMB004_promotor		Sintética	185		
pMB054_promotor		Sintética	186		
pMB024_promotor		Sintética	187		
pMB007_promotor		Sintética	188		
pMB005_promotor		Sintética	189		

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
pMB003_promotor		Sintética	190		
pMB088_promotor		Sintética	191		
pMB065_promotor		Sintética	192		
pMB037_promotor		Sintética	193		
pMB009_promotor		Sintética	194		
pMB041_promotor		Sintética	195		
pMB036_promotor		Sintética	196		
pMB049_promotor		Sintética	197		
pMB044_promotor		Sintética	198		
pMB042_promotor		Sintética	199		
pMB086_promotor		Sintética	200		
pMB053_promotor		Sintética	201		
pMB057_promotor		Sintética	202		
pMB018_promotor		Sintética	203		
pMB002_promotor		Sintética	204		
pMB015_promotor		Sintética	205		
pMB087_promotor		Sintética	206		
pMB063_promotor		Sintética	207		
Parte distal del promotor sintético		Fago λ	208		promotor P_R
Parte central del promotor sintético		Fago λ	209		promotor P_R
Parte distal del promotor sintético		Fago λ	210		promotor P_L
parte 5'UTR/RBS del promotor sintético		Fago λ	211		promotor P_R
parte 5'UTR/RBS del promotor sintético		<i>E. coli</i>	212		promotor del gen <i>acs</i>
Ori_Plsmd27			213		
cadena principal de vector 1			214		
cadena principal de vector 2			215		
cadena principal de vector 3			216		
cadena principal de vector 4			217		
Aislante 1			218		véase Tabla 15
Aislante 2			219		véase Tabla 15
T1			220		de Orosz et al., Eur J Biochem. 1991 Nov

ES 2 875 579 T3

NOMBRE ABREVIADO)	(NOMBRE	FUENTE	ÁCIDO NUCLEICO DE SEQ ID NO.	AMINOÁCIDO DE SEQ ID NO.	DESCRIPCIÓN
					1;201(3):653-9; véase Tabla 15
B0015			221		véase Tabla 15
promotor de SacB			222		véase Tabla 15
promotor y secuencia de PheS			223		véase Tabla 15
Terminador Tsod	<i>C. glutamicum</i>		224		
Spy	<i>E. coli</i>		225		Secuencia de terminador
pheA	<i>E. coli</i>		226		Secuencia de terminador
osmE	<i>E. coli</i>		227		Secuencia de terminador
rpoH	<i>E. coli</i>		228		Secuencia de terminador
vibE	<i>E. coli</i>		229		Secuencia de terminador
Thr1_ABC	<i>E. coli</i>		230		Secuencia de terminador
GB1 (PST1)	Streptococcus sp.		231	235	dominio B1 de IgG de proteína G
FH8 (PST2)	F. hepatica		232	236	Antígeno de Fasciola hepatica 8-kDa
Ubiquitina (PST3)			233	237	
SUMO (PST4)	Homo sapiens		234	238	Ubiquitina modificada pequeña
ssrA_LAA (PDT1)	<i>E. coli</i>		239	247	nativo
ssrA_LVA (PDT2)	<i>E. coli</i>		240	248	mutante
ssrA_AAV (PDT3)	<i>E. coli</i>		241	249	mutante
ssrA_ASV (PDT4)	<i>E. coli</i>		242	250	mutante
ftsH-cll89-97 (PDT5)	<i>E. coli</i>		243	251	nativo
cl108 (PDT6)	<i>E. coli</i>		244	252	nativo
su120 (PDT7)	<i>E. coli</i>		245	253	nativo
β20 (PDT8)	<i>E. coli</i>		246	254	nativo
P3_BCD1	Artificial		255		
P4_BCD22	Artificial		256		
P7_BCD19	Artificial		257		
P8_BCD15	Artificial		258		
P11_BCD17	Artificial		259		
P13_BCD8	Artificial		260		

REFERENCIAS

5 Los siguientes se refieren a: solicitud de EE.UU. nº 15/396,230 (Pub. de EE.UU. nº US 2017/0159045 A1) presentada el 30 de diciembre, 2016; PCT/US2016/065465 (WO 2017/100377 A1) presentada el 7 de diciembre, 2016; solicitud de EE.UU. nº 15/140.296 (US 2017/0316353 A1) presentada el 27 de abril, 2016; PCT/US2017/029725 (WO 2017/189784 A1) presentada el 26 de abril, 2017; PCT/US2016/065464 (WO 2017/100376 A2) presentada el 7 de diciembre, 2016; solicitud prov. de EE.UU. nº 62/431.409 presentada el 7 de diciembre, 2016; solicitud prov. de EE.UU. nº 62/264.232 presentada el 7 de diciembre, 2015; y solicitud prov. de EE.UU. nº 62/368.786 presentada el 29 de julio, 2016.

10 **LISTA DE SECUENCIAS**

<110> Zymergen Inc.

Kimball, Aaron

Szyjka, Shawn

Frewen, Barbara

15 Treynor, Thomas

Serber, Zach

Dean, Erich Jedediah

Manchester, Shawn

Gora, Katherine

20 Flashman, Michael

Shellman, Erin

Haushalter, Robert

Morgan, Stacy-Anne

Blaisse, Michael

25 Ramakrishnan, Prabha

Rothschild-Mancinelli, Kyle

Kim, Youngnyun

Davis, Matthew

Wisnewski, Christy

30 Westfall, Patrick

<120> PLATAFORMA DE INGENIERÍA GENÓMICA DE HTP PARA MEJORAR ESCHERICHIA COLI

<130> ZYMR-012/01WO 327574-2058

35

<150> US 62/515,870

<151> 2017-06-06

<160> 260

<170> PatentIn versión 3.5

<210> 1

5 <211> 97

<212> ADN

<213> Desconocido

<220>

10 <223> Promotor de expresión derivado de Pcg0007_lib_39

<400> 1

tgccgtttct cgcgtttgtt gtggtactac gtggggacct aagcgtgtat tatggaaacg 60

tctgtatcgg ataagtagcg aggagtgttc gttaaaa 97

15 <210> 2

<211> 97

<212> ADN

<213> Desconocido

20 <220>

<223> Promotor de expresión derivado de Pcg0007

<400> 2

tgccgtttct cgcgtttgtt gtggtactac gtggggacct aagcgtgtaa gatggaaacg 60

tctgtatcgg ataagtagcg aggagtgttc gttaaaa 97

25

<210> 3

<211> 93

<212> ADN

<213> Desconocido

30

<220>

<223> Promotor de expresión derivado de Pcg1860

<400> 3

ES 2 875 579 T3

	cttagctttg acctgcacaa atagttgcaa attgtccac atacacataa agtagcttgc	60
	gtatttaaaa ttatgaacct aaggggttta gca	93
	<210> 4	
	<211> 98	
5	<212> ADN	
	<213> Desconocido	
	<220>	
	<223> Promotor de expresión derivado de Pcg0755	
10	<400> 4	
	aataaattta taccacacag tctattgcaa tagaccaagc tgttcagtag ggtgcatggg	60
	agaagaattt cctaataaaa actcttaagg acctccaa	98
	<210> 5	
15	<211> 97	
	<212> ADN	
	<213> Desconocido	
	<220>	
20	<223> Promotor de expresión derivado de Pcg0007_265	
	<400> 5	
	tgccgtttct cgcgttggt gtggtactac gtggggacct aagcgtgtac gctggaaacg	60
	tctgtatcgg ataagtagcg aggagtgttc gttaaaa	97
25	<210> 6	
	<211> 86	
	<212> ADN	
	<213> Desconocido	
30	<220>	
	<223> Promotor de expresión derivado de Pcg3381	
	<400> 6	

ES 2 875 579 T3

	cgccggataa atgaattgat tatttttaggc tcccagggat taagtctagg gtggaatgca	60
	gaaatatttc ctacggaagg tccggtt	86
	<210> 7	
	<211> 97	
5	<212> ADN	
	<213> Desconocido	
	<220>	
	<223> Promotor de expresión derivado de Pcg0007_119	
10	<400> 7	
	tgccgtttct cgcgttgtgt gtggtactac gtggggacct aagcgtgttg catggaacg	60
	tctgtatcgg ataagtagcg aggagtgttc gttaaaa	97
	<210> 8	
15	<211> 87	
	<212> ADN	
	<213> Desconocido	
	<220>	
20	<223> Promotor de expresión derivado de Pcg3121	
	<400> 8	
	gtggctaataa cttttggaaa cttaagttac cttaaatcgg aaacttattg aattcgggtg	60
	aggcaactgc aactctggac ttaaagc	87
25	<210> 9	
	<211> 25	
	<212> ADN	
	<213> Desconocido	
30	<220>	
	<223> Terminador cg0001	
	<400> 9	

ES 2 875 579 T3

gacctatctt cggatgggtc tttt 25

<210> 10

<211> 30

5 <212> ADN

<213> Desconocido

<220>

<223> Terminador cg0007

10

<400> 10

ccgcccctg gaattctggg ggcgggttt 30

<210> 11

15 <211> 24

<212> ADN

<213> Desconocido

<220>

20 <223> Terminador cg0371

<400> 11

ccggaactt ttgtaagtg ccgg 24

25 <210> 12

<211> 27

<212> ADN

<213> Desconocido

30 <220>

<223> Terminador cg0480

<400> 12

cccctcagaa gcgattctga ggggttt 27

35

<210> 13

<211> 28

<212> ADN

<213> Desconocido

<220>

5 <223> Terminador cg0494

<400> 13

gcaccgcctt tcggggcggg gcttttt 28

10 <210> 14

<211> 28

<212> ADN

<213> Desconocido

15 <220>

<223> Terminador cg0564

<400> 14

ggcccatgc ttgcatggg gtctttt 28

20

<210> 15

<211> 30

<212> ADN

<213> Desconocido

25

<220>

<223> Terminador cg0610

<400> 15

30 gcacttacct taactgtag gtgcttttt 30

<210> 16

<211> 24

<212> ADN

35 <213> Desconocido

<220>

<223> Terminador cg0695

<400> 16

accgggtcac cagaccgggt cttt 24

5

<210> 17

<211> 54

<212> ADN

<213> Desconocido

10

<220>

<223> Secuencia codificante de Cis1

<400> 17

15 atgaaagcaa tttcgtact gaaacatctt aatcatgcac aggagacttt ctaa 54

<210> 18

<211> 4

<212> PRT

20 <213> Desconocido

<220>

<223> Péptido conector rico en serina

25 <400> 18

Gly Ser Gly Ser
1

<210> 19

<211> 5

30 <212> PRT

<213> Desconocido

<220>

<223> Secuencia de conector rico en serina

35

<400> 19

ES 2 875 579 T3

Gly Ser Gly Ser Gly
1 5

<210> 20

<211> 4

5 <212> PRT

<213> Desconocido

<220>

<223> Secuencia de conector rico en serina

10

<400> 20

Gly Ser Asn Gly
1

<210> 21

15 <211> 5

<212> PRT

<213> Desconocido

<220>

20 <223> Secuencia de conector rico en alanina

<400> 21

Thr Ser Ala Ala Ala
1 5

25 <210> 22

<211> 6

<212> PRT

<213> Desconocido

30 <220>

<223> Secuencia de conector

<400> 22

Ala Ala Ala Gly Gly Met
1 5

35

ES 2 875 579 T3

<210> 23

<211> 14

<212> PRT

<213> Desconocido

5

<220>

<223> Secuencia de conector

<400> 23

10 Ala Ala Ala Gly Gly Met Pro Pro Ala Ala Ala Gly Gly Met
1 5 10

<210> 24

<211> 6

<212> PRT

15 <213> Desconocido

<220>

<223> Secuencia de conector

20 <400> 24

Ala Ala Ala Gly Gly Met
1 5

<210> 25

<211> 9

25 <212> PRT

<213> Desconocido

<220>

<223> Secuencia de conector

30

<400> 25

Pro Pro Ala Ala Ala Gly Gly Met Met
1 5

<210> 26

35 <211> 87

ES 2 875 579 T3

<212> PRT

<213> Desconocido

<220>

5 <223> Zif268

<400> 26

Pro Gly Glu Lys Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg
1 5 10 15

Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr Gly
20 25 30

Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser
35 40 45

Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro Phe
50 55 60

Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Glu Arg Lys
65 70 75 80

10 Arg His Thr Lys Ile His Thr
85

<210> 27

<211> 85

<212> PRT

15 <213> Desconocido

<220>

<223> PBSII

20 <400> 27

ES 2 875 579 T3

Pro Gly Glu Lys Pro Tyr Ala Cys Pro Glu Cys Gly Lys Ser Phe Ser
1 5 10 15

Gln Arg Ala Asn Leu Arg Ala His Gln Arg Thr His Thr Gly Glu Lys
20 25 30

Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His
35 40 45

Leu Thr Thr His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys Cys
50 55 60

Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp Val Leu Val Arg His
65 70 75 80

Gln Arg Thr His Thr
85

<210> 28

<211> 84

5 <212> PRT

<213> Desconocido

<220>

<223> ZFa

10

<400> 28

Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Asp Ser Pro Thr Leu Arg Arg His Thr Arg Thr His Thr Gly Glu Lys
20 25 30

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Val Arg His Asn
35 40 45

Leu Thr Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
50 55 60

Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Thr Ser Leu Ala Arg His
65 70 75 80

Leu Lys Thr His

<210> 29

15 <211> 84

<212> PRT

ES 2 875 579 T3

<213> Desconocido

<220>

<223> ZFb

5

<400> 29

Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Lys Lys Asp His Leu His Arg His Thr Arg Thr His Thr Gly Glu Lys
20 25 30

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr
35 40 45

Leu Lys Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
50 55 60

Arg Ile Cys Met Arg Asn Phe Ser Arg Leu Asp Met Leu Ala Arg His
65 70 75 80

Leu Lys Thr His

<210> 30

10 <211> 84

<212> PRT

<213> Desconocido

<220>

15 <223> ZFc

<400> 30

ES 2 875 579 T3

Pro Gly Glu Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser
1 5 10 15

Ser Pro Ser Lys Leu Ile Arg His Thr Arg Thr His Thr Gly Glu Lys
20 25 30

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Gly Ser Asn
35 40 45

Leu Ala Arg His Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
50 55 60

Arg Ile Cys Met Arg Asn Phe Ser Arg Val Asp Asn Leu Pro Arg His
65 70 75 80

Leu Lys Thr His

<210> 31

<211> 83

5 <212> PRT

<213> Desconocido

<220>

<223> Tyr123

10

<400> 31

Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Asp Arg
1 5 10 15

Ser Asn Leu Thr Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr
20 25 30

Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Thr Thr Ser Asn Leu Ala
35 40 45

Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Lys Cys Pro Glu
50 55 60

Cys Gly Lys Ser Phe Ser Arg Ser Asp Ala Leu Thr Arg His Gln Arg
65 70 75 80

Thr His Thr

<210> 32

15 <211> 83

<212> PRT

ES 2 875 579 T3

<213> Desconocido

<220>

<223> Tyr456

5

<400> 32

Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser
1 5 10 15

Ser Asn Leu Ala Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr
20 25 30

Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His Leu Thr
35 40 45

Lys His Gln Arg Thr His Thr Gly Glu Lys Pro Phe Lys Cys Pro Glu
50 55 60

Cys Gly Lys Ser Phe Ser Gln Ser Ser Asn Leu Ala Arg His Gln Arg
65 70 75 80

Thr His Thr

<210> 33

10

<211> 94

<212> PRT

<213> Desconocido

<220>

15

<223> Dedo de zinc Blues

<400> 33

Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15

Arg Phe Ser Arg Arg Asp Val Leu Met Asn His Ile Arg Ile His Thr
20 25 30

Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg
35 40 45

Ser Asp His Leu Thr Thr His Ile Arg Thr His Thr Gly Glu Lys Pro
50 55 60

ES 2 875 579 T3

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asn Arg Asp Thr Leu
65 70 75 80

Thr Arg His Ser Lys Ile His Leu Arg Gln Asn Asp Leu Glu
85 90

<210> 34

<211> 94

5 <212> PRT

<213> Desconocido

<220>

<223> Dedo de zinc Jazz

10

<400> 34

Ala Ser Asp Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15

Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile His Thr
20 25 30

Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Ser
35 40 45

Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu Lys Pro
50 55 60

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ser Arg Asp Val Leu
65 70 75 80

Arg Arg His Asn Arg Ile His Leu Arg Gln Asn Asp Leu Glu
85 90

<210> 35

15 <211> 127

<212> PRT

<213> Desconocido

<220>

20 <223> Dedo de zinc Bagly

<400> 35

ES 2 875 579 T3

Glu Phe Met Thr Gly Asp Arg Pro Tyr Ala Cys Pro Val Glu Ser Cys
 1 5 10 15

Asp Arg Arg Phe Ser Arg Ser Asp Glu Leu Thr Arg His Ile Arg Ile
 20 25 30

His Thr Gly Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe
 35 40 45

Ser Ser Arg Asp Val Leu Arg Arg His Asn Arg Thr His Thr Gly Glu
 50 55 60

Lys Pro Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Ser Arg Asp
 65 70 75 80

Val Leu Arg Arg His Asn Arg Ile His Leu Arg Gln Gly Arg Ser His
 85 90 95

Val Cys Ala Glu Cys Gly Lys Ala Phe Val Glu Ser Ser Lys Leu Lys
 100 105 110

Arg His Gln Leu Val His Thr Gly Glu Lys Pro Phe Gln Leu Glu
 115 120 125

- 5 <210> 36
- <211> 11
- <212> ADN
- <213> Desconocido

- 10 <220>
- <223> Sitio de unión de dedo de zinc Bagly

<400> 36
 cgggctgctg c 11

- 15 <210> 37
- <211> 175
- <212> PRT
- <213> Desconocido

- 20 <220>
- <223> Gli1

ES 2 875 579 T3

<400> 37

Lys Arg Glu Pro Glu Ser Val Tyr Glu Thr Asp Cys Arg Trp Asp Gly
 1 5 10 15

Cys Ser Gln Glu Phe Asp Ser Gln Glu Gln Leu Val His His Ile Asn
 20 25 30

Ser Glu His Ile His Gly Glu Arg Lys Glu Phe Val Cys His Trp Gly
 35 40 45

Gly Cys Ser Arg Glu Leu Arg Pro Phe Lys Ala Gln Tyr Met Leu Val
 50 55 60

Val His Met Arg Arg His Thr Gly Glu Lys Pro His Lys Cys Thr Phe
 65 70 75 80

Glu Gly Cys Arg Lys Ser Tyr Ser Arg Leu Glu Asn Leu Lys Thr His
 85 90 95

Leu Arg Ser His Thr Gly Glu Lys Pro Tyr Met Cys Glu His Glu Gly
 100 105 110

Cys Ser Lys Ala Phe Ser Asn Ala Ser Asp Arg Ala Lys His Gln Asn
 115 120 125

Arg Thr His Ser Asn Glu Lys Pro Tyr Val Cys Lys Leu Pro Gly Cys
 130 135 140

Thr Lys Arg Tyr Thr Asp Pro Ser Ser Leu Arg Lys His Val Lys Thr
 145 150 155 160

Val His Gly Pro Asp Ala His Val Thr Lys Arg His Arg Gly Asp
 165 170 175

5

<210> 38

<211> 15

<212> ADN

<213> Desconocido

10

<220>

<223> Sitio de unión Gli1

<400> 38

15 gaccaccaa gacga 15

<210> 39

ES 2 875 579 T3

<211> 84

<212> PRT

<213> Desconocido

5 <220>

<223> Dedo de zinc HIVC

<400> 39

Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Leu Arg Thr Asp
1 5 10 15

Leu Asp Arg His Thr Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys
20 25 30

Arg Ile Cys Met Arg Asn Phe Ser Leu Ser Gln Thr Leu Arg Arg His
35 40 45

Leu Arg Thr His Thr Gly Glu Lys Pro Phe Gln Cys Arg Ile Cys Met
50 55 60

Arg Asn Phe Ser Leu Arg Ser Asn Leu Gly Arg His Leu Lys Thr His

10

65 70 75 80

Thr Gly Glu Lys

<210> 40

<211> 100

15 <212> PRT

<213> desconocido

<220>

<223> Dedo de zinc B3

20

<400> 40

ES 2 875 579 T3

Ala Gln Ala Ala Leu Glu Pro Lys Glu Lys Pro Tyr Ala Cys Pro Glu
1 5 10 15

Cys Gly Lys Ser Phe Ser Asp Pro Gly Asn Leu Val Arg His Gln Arg
20 25 30

Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser
35 40 45

Phe Ser Arg Ser Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly
50 55 60

Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Gln Ser
65 70 75 80

Ser His Leu Val Arg His Gln Arg Thr His Thr Gly Lys Lys Thr Ser
85 90 95

Gly Gln Ala Gly
100

<210> 41

<211> 100

5 <212> PRT

<213> desconocido

<220>

<223> Dedo de zinc N1

10

<400> 41

Ala Gln Ala Ala Leu Glu Pro Lys Glu Lys Pro Tyr Ala Cys Pro Glu
1 5 10 15

Cys Gly Lys Ser Phe Ser Gln Ser Ser Ser Leu Val Arg His Gln Arg
20 25 30

ES 2 875 579 T3

Thr His Thr Gly Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser
 35 40 45

Phe Ser Gln Ser Ser Asn Leu Val Arg His Gln Arg Thr His Thr Gly
 50 55 60

Glu Lys Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser
 65 70 75 80

Asp Lys Leu Val Arg His Gln Arg Thr His Thr Gly Lys Lys Thr Ser
 85 90 95

Gly Gln Ala Gly
 100

<210> 42

<211> 94

5 <212> PRT

<213> Desconocido

<220>

<223> Dedo de zinc primero Sp-1

10

<400> 42

Pro Gly Lys Lys Lys Gln His Ile Cys His Ile Gln Gly Cys Gly Lys
 1 5 10 15

Val Tyr Gly Lys Thr Ser His Leu Arg Ala His Leu Arg Trp His Thr
 20 25 30

Gly Glu Arg Pro Phe Met Cys Thr Trp Ser Tyr Cys Gly Lys Arg Phe
 35 40 45

Thr Arg Ser Asp Glu Leu Gln Arg His Lys Arg Thr His Thr Gly Glu
 50 55 60

Lys Lys Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp
 65 70 75 80

His Leu Ser Lys His Ile Lys Thr His Gln Asn Lys Lys Gly
 85 90

<210> 43

15 <211> 90

<212> PRT

<213> Desconocido

ES 2 875 579 T3

<220>

<223> Dedo segundo Sp-1

5 <400> 43

Pro Gly Lys Lys Lys Gln His Ala Cys Pro Glu Cys Gly Lys Ser Phe
 1 5 10 15
 Ser Lys Ser Ser His Leu Arg Ala His Gln Arg Thr His Thr Gly Glu
 20 25 30
 Arg Pro Tyr Lys Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp
 35 40 45
 Glu Leu Gln Arg His Gln Arg Thr His Thr Gly Glu Lys Pro Tyr Lys
 50 55 60
 Cys Pro Glu Cys Gly Lys Ser Phe Ser Arg Ser Asp His Leu Ser Lys
 65 70 75 80
 His Gln Arg Thr His Gln Asn Lys Lys Gly
 85 90

<210> 44

<211> 74

10 <212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína SH3 clase I

15

<400> 44

Glu Gly Tyr Gln Tyr Arg Ala Leu Tyr Asp Tyr Lys Lys Glu Arg Glu
 1 5 10 15
 Glu Asp Ile Asp Leu His Leu Gly Asp Ile Leu Thr Val Asn Lys Gly
 20 25 30
 Ser Leu Val Ala Leu Gly Phe Ser Asp Gly Gln Glu Ala Arg Pro Glu
 35 40 45
 Glu Ile Gly Trp Leu Asn Gly Tyr Asn Glu Thr Thr Gly Glu Arg Gly
 50 55 60
 Asp Phe Pro Gly Thr Tyr Val Glu Tyr Ile
 65 70

ES 2 875 579 T3

<210> 45

<211> 54

<212> PRT

5 <213> Desconocido

<220>

<223> Sitio de unión de proteína SH3 clase II

10 <400> 45

Tyr Val Arg Ala Leu Phe Asp Phe Asn Gly Asn Asp Glu Glu Asp Leu
1 5 10 15

Pro Phe Lys Lys Gly Asp Ile Leu Arg Ile Arg Asp Lys Pro Glu Glu
20 25 30

Gln Trp Trp Asn Ala Glu Asp Ser Glu Gly Lys Arg Gly Met Ile Pro
35 40 45

Val Pro Tyr Val Glu Lys
50

<210> 46

<211> 58

15 <212> PRT

<213> Desconocido

<220>

<223> Sitio de unión de proteína SH3

20

<400> 46

Met Ala Glu Tyr Val Arg Ala Leu Phe Asp Phe Asn Gly Asn Asp Glu
1 5 10 15

Glu Asp Leu Pro Phe Lys Lys Gly Asp Ile Leu Arg Ile Arg Asp Lys
20 25 30

Pro Glu Glu Gln Trp Trp Asn Ala Glu Asp Ser Glu Gly Lys Arg Gly
35 40 45

Met Ile Pro Val Pro Tyr Val Glu Lys Tyr
50 55

<210> 47

ES 2 875 579 T3

<211> 7

<212> PRT

<213> Desconocido

5 <220>

<223> Péptido de reclutamiento SH3

<400> 47

Arg Pro Leu Pro Val Ala Pro
1 5

10

<210> 48

<211> 13

<212> PRT

<213> Desconocido

15

<220>

<223> Péptido de reclutamiento SH3

<400> 48

Pro Pro Pro Ala Leu Pro Pro Lys Arg Arg Arg Pro Gly
1 5 10

20

<210> 49

<211> 10

<212> PRT

25 <213> Desconocido

<220>

<223> Péptido de reclutamiento SH3

30 <400> 49

Pro Pro Pro Ala Leu Pro Pro Lys Lys Arg
1 5 10

<210> 50

<211> 84

35 <212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína PDZ

5

<400> 50

Glu Ile Thr Leu Glu Arg Gly Asn Ser Gly Leu Gly Phe Ser Ile Ala
 1 5 10 15

Gly Gly Thr Asp Asn Pro His Ile Gly Asp Asp Pro Ser Ile Phe Ile
 20 25 30

Thr Lys Ile Ile Pro Gly Gly Ala Ala Ala Gln Asp Gly Arg Leu Arg
 35 40 45

Val Asn Asp Ser Ile Leu Phe Val Asn Glu Val Asp Val Arg Glu Val
 50 55 60

Thr His Ser Ala Ala Val Glu Ala Leu Lys Glu Ala Gly Ser Ile Val
 65 70 75 80

Arg Leu Tyr Val

<210> 51

10 <211> 87

<212> PRT

<213> Desconocido

<220>

15 <223> Dominio de unión de proteína PDZ

<400> 51

ES 2 875 579 T3

Val Met Glu Ile Lys Leu Ile Lys Gly Pro Lys Gly Leu Gly Phe Ser
1 5 10 15

Ile Ala Gly Gly Val Gly Asn Gln His Ile Pro Gly Asp Asn Ser Ile
20 25 30

Tyr Val Thr Lys Ile Ile Glu Gly Gly Ala Ala His Lys Asp Gly Arg
35 40 45

Leu Gln Ile Gly Asp Lys Ile Leu Ala Val Asn Ser Val Gly Leu Glu
50 55 60

Asp Val Met His Glu Asp Ala Val Ala Ala Leu Lys Asn Thr Tyr Asp
65 70 75 80

Val Val Tyr Leu Lys Val Ala
85

<210> 52

<211> 79

5 <212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína PDZ

10

<400> 52

Arg Ile Val Ile His Arg Gly Ser Thr Gly Leu Gly Phe Asn Ile Val
1 5 10 15

Gly Gly Glu Asp Gly Glu Gly Ile Phe Ile Ser Phe Ile Leu Ala Gly
20 25 30

Gly Pro Ala Asp Leu Ser Gly Glu Leu Arg Lys Gly Asp Gln Ile Leu
35 40 45

Ser Val Asn Gly Val Asp Leu Arg Asn Ala Ser His Glu Gln Ala Ala
50 55 60

Ile Ala Leu Lys Asn Ala Gly Gln Thr Val Thr Ile Ile Ala Gln
65 70 75

<210> 53

15 <211> 82

<212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína PDZ

5 <400> 53

Arg Arg Val Thr Val Arg Lys Ala Asp Ala Gly Gly Leu Gly Ile Ser
1 5 10 15

Ile Lys Gly Gly Arg Glu Asn Lys Met Pro Ile Leu Ile Ser Lys Ile
20 25 30

Phe Lys Gly Leu Ala Ala Asp Gln Thr Glu Ala Leu Phe Val Gly Asp
35 40 45

Ala Ile Leu Ser Val Asn Gly Glu Asp Leu Ser Ser Ala Thr His Asp
50 55 60

Glu Ala Val Gln Ala Leu Lys Lys Thr Gly Lys Glu Val Val Leu Glu
65 70 75 80

Val Lys

<210> 54

<211> 98

10 <212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína PDZ

15

<400> 54

ES 2 875 579 T3

Met Leu Gln Arg Arg Arg Val Thr Val Arg Lys Ala Asp Ala Gly Gly
 1 5 10 15

Leu Gly Ile Ser Ile Lys Gly Gly Arg Glu Asn Lys Met Pro Ile Leu
 20 25 30

Ile Ser Lys Ile Phe Lys Gly Leu Ala Ala Asp Gln Thr Glu Ala Leu
 35 40 45

Phe Val Gly Asp Ala Ile Leu Ser Val Asn Gly Glu Asp Leu Ser Ser
 50 55 60

Ala Thr His Asp Glu Ala Val Gln Ala Leu Lys Lys Thr Gly Lys Glu
 65 70 75 80

Val Val Leu Glu Val Lys Tyr Met Lys Glu Val Ser Pro Tyr Phe Lys
 85 90 95

Gly Ser

- 5 <210> 55
- <211> 5
- <212> PRT
- <213> Desconocido

- 10 <220>
- <223> Péptido de reclutamiento PDZ

<400> 55

Ile Glu Ser Asp Val
 1 5

- 15 <210> 56
- <211> 6
- <212> PRT
- <213> Desconocido

- 20 <220>
- <223> Péptido de reclutamiento PDZ

<400> 56

Val Lys Glu Ser Leu Val
 1 5

25

<210> 57

<211> 7

<212> PRT

5 <213> Desconocido

<220>

<223> Péptido de reclutamiento PDZ

10 <400> 57

Gly Val Lys Glu Ser Leu Val
1 5

<210> 58

<211> 7

15 <212> PRT

<213> Desconocido

<220>

<223> Péptido de reclutamiento PDZ

20

<400> 58

Gly Val Lys Gln Ser Leu Leu
1 5

<210> 59

25 <211> 7

<212> PRT

<213> Desconocido

<220>

30 <223> Péptido de reclutamiento PDZ

<400> 59

Gly Val Lys Glu Ser Gly Ala
1 5

35 <210> 60

ES 2 875 579 T3

<211> 7

<212> PRT

<213> Desconocido

5 <220>

<223> Péptido de reclutamiento PDZ

<400> 60

Tyr Val Lys Glu Ser Leu Val
1 5

10

<210> 61

<211> 5

<212> PRT

<213> Desconocido

15

<220>

<223> Péptido de reclutamiento PDZ

<400> 61

Val Glu Thr Asp Val
1 5

20

<210> 62

<211> 43

<212> PRT

25 <213> Desconocido

<220>

<223> Dominio de unión de proteína GBD

30 <400> 62

Ala Asp Ile Gly Thr Pro Ser Asn Phe Gln His Ile Gly His Val Gly
1 5 10 15

Trp Asp Pro Asn Thr Gly Phe Asp Leu Asn Asn Leu Asp Pro Glu Leu
20 25 30

Lys Asn Leu Phe Asp Met Cys Gly Ile Ser Glu
35 40

<210> 63

<211> 36

<212> PRT

5 <213> Desconocido

<220>

<223> Dominio de unión de proteína GBD

10 <400> 63

Lys Glu Arg Pro Glu Ile Ser Leu Pro Ser Asp Phe Glu His Thr Ile
1 5 10 15

His Val Gly Phe Asp Ala Val Thr Gly Glu Phe Thr Gly Met Pro Glu
20 25 30

Gln Trp Ala Arg
35

<210> 64

<211> 80

15 <212> PRT

<213> Desconocido

<220>

<223> Dominio de unión de proteína GBD

20

<400> 64

Met Thr Lys Ala Asp Ile Gly Thr Pro Ser Asn Phe Gln His Ile Gly
1 5 10 15

His Val Gly Trp Asp Pro Asn Thr Gly Phe Asp Leu Asn Asn Leu Asp
20 25 30

Pro Glu Leu Lys Asn Leu Phe Asp Met Cys Gly Ile Ser Glu Ala Gln
35 40 45

Leu Lys Asp Arg Glu Thr Ser Lys Val Ile Tyr Asp Phe Ile Glu Lys
50 55 60

Thr Gly Gly Val Glu Ala Val Lys Asn Glu Leu Arg Arg Gln Ala Pro
65 70 75 80

<210> 65

<211> 32

<212> PRT

<213> Desconocido

5

<220>

<223> Péptido de reclutamiento GBD

<400> 65

10 Leu Val Gly Ala Leu Met His Val Met Gln Lys Arg Ser Arg Ala Ile
1 5 10 15

His Ser Ser Asp Glu Gly Glu Asp Gln Ala Gly Asp Glu Asp Glu Asp
20 25 30

<210> 66

15 <211> 47

<212> PRT

<213> Desconocido

<220>

20 <223> Dominio de unión de cremallera de leucina

<400> 66

Leu Glu Ile Glu Ala Ala Phe Leu Glu Arg Glu Asn Thr Ala Leu Glu
1 5 10 15

Thr Arg Val Ala Glu Leu Arg Gln Arg Val Gln Arg Leu Arg Asn Arg
20 25 30

Val Ser Gln Tyr Arg Thr Arg Tyr Gly Pro Leu Gly Gly Gly Lys
35 40 45

25 <210> 67

<211> 47

<212> PRT

<213> Desconocido

30 <220>

<223> Dominio de unión de cremallera de leucina

ES 2 875 579 T3

<400> 67

Leu Glu Ile Glu Ala Ala Phe Leu Glu Arg Glu Asn Thr Ala Leu Glu
 1 5 10 15

Thr Arg Val Ala Glu Leu Arg Gln Arg Val Gln Arg Leu Arg Asn Arg
 20 25 30

Val Ser Gln Tyr Arg Thr Arg Tyr Gly Pro Leu Gly Gly Gly Lys
 35 40 45

5 <210> 68

<211> 47

<212> PRT

<213> Desconocido

10 <220>

<223> Dominio de unión de cremallera de leucina

<400> 68

Leu Glu Ile Arg Ala Ala Phe Leu Arg Gln Arg Asn Thr Ala Leu Arg
 1 5 10 15

15

Thr Glu Val Ala Glu Leu Glu Gln Glu Val Gln Arg Leu Glu Asn Glu
 20 25 30

Val Ser Gln Tyr Glu Thr Arg Tyr Gly Pro Leu Gly Gly Gly Lys
 35 40 45

<210> 69

<211> 88

20 <212> ADN

<213> Desconocido

<220>

<223> Secuencia reguladora de diseño bicistrónico (BCD) Figura 43

25

<400> 69

gggcccaagt tcaacttaaaa aggagatcaa caatgaaagc aattttcgta ctgaaacatc 60

ttaatcatgc acaggagact ttctaatag 88

ES 2 875 579 T3

<210> 70
 <211> 88
 <212> ADN
 <213> Desconocido
 5
 <220>
 <223> Secuencia reguladora de diseño bicistrónico (BCD) Figura 43

 <220>
 10 <221> misc_feature
 <222> (36)..(70)
 <223> n es a, c, g, o t

 <400> 70
 gggcccaagt tcacttaaaa aggagatcaa caatgnnnnn nnnnnnnnnn nnnnnnnnnn 60
 15 nnnnnnnnnn acaggagact ttctaatg 88

 <210> 71
 <211> 77
 <212> ADN
 20 <213> Escherichia coli

 <400> 71
 gatatgatct atatcaattt ctcatctata atgctttggt agtatctcgt cgccgactta 60
 ataaagagag agttagt 77

 25 <210> 72
 <211> 78
 <212> ADN
 <213> Escherichia coli

 30 <400> 72
 atgtcggata tctggtggtg aaatacttta tgccatgata atttaatacg atgtatttat 60
 tatatggagc acttaatt 78

 <210> 73

ES 2 875 579 T3

<211> 77

<212> ADN

<213> Escherichia coli

5 <400> 73

gaggctcttt gtgctaaact ggcccgccga atgtatagta cacttcggtt ggataggtaa 60

tttggcgaga taatacg 77

<210> 74

<211> 80

10 <212> ADN

<213> Escherichia coli

<400> 74

cgtgtggatt gtgtcttgcg acgatgggca ctaaagtta aaaggtgccc ctcaacaaaa 60

aagacacaca ggggaaaggc 80

15

<210> 75

<211> 70

<212> ADN

<213> Escherichia coli

20

<400> 75

caaaactaatt aataaatagt taattaacgc tcatcattgt acaatgaact gtacaaaaga 60

ggagattgac 70

<210> 76

25 <211> 71

<212> ADN

<213> Escherichia coli

<400> 76

caaatctgtc aatttttcct ggaactggcg ttttcagtta tgattgtggg acttatcaaa 60

30 aaggagaggc c 71

<210> 77

ES 2 875 579 T3

<211> 78

<212> ADN

<213> Escherichia coli

5 <400> 77

tctttgtcac aaaggtggag gcaatgtcag tgggtgtgta caataagagt atcggcagga 60

cattaagagg aatgagcc 78

<210> 78

<211> 76

10 <212> ADN

<213> Escherichia coli

<400> 78

tttacgctgc gtcgatgcac agcctcatca ttttgcagta tccttaagat attccttata 60

15

tcttcaggag atcgtc 76

<210> 79

<211> 74

20 <212> ADN

<213> Escherichia coli

<400> 79

taatgaaaac gcattagccg aatcggcaaa aattggttac cttacatctc atcgaaaaca 60

cggaggaagt atag 74

25

<210> 80

<211> 76

<212> ADN

<213> Escherichia coli

30

<400> 80

tgtttaaaaa tggcttgcca taattaacgt tgtatgtgat aacagatttc gggttaaacg 60

aggtacagtt ctgttt 76

ES 2 875 579 T3

	<210> 81	
	<211> 72	
	<212> ADN	
	<213> Escherichia coli	
5	<400> 81	
	aatcgtagct tcctggtgtc attaggttat tttacctgta taaataacca gtatattcaa	60
	cagggggcta tt	72
	<210> 82	
10	<211> 77	
	<212> ADN	
	<213> Escherichia coli	
	<400> 82	
	aagacacatt ttatattgaa cgtccaatca ataaccgctt taatagataa acaccgctga	60
15	tgaatggagt ggcgaaa	77
	<210> 83	
	<211> 81	
	<212> ADN	
20	<213> Escherichia coli	
	<400> 83	
	atcacaatc gcgaagagtt tcccattaat tttgatata tttaaaactt aggacttatt	60
	tgaatcacat ttgaggtggt t	81
25	<210> 84	
	<211> 79	
	<212> ADN	
	<213> Escherichia coli	
30	<400> 84	
	aataaaccca tctatagatg gtaaaaatag gttgtggcaa ttatcattgc atcattccct	60
	tttcgaatga gtttctatt	79

ES 2 875 579 T3

<210> 85
 <211> 76
 <212> ADN
 <213> Escherichia coli
 5
 <400> 85
 agaaataaaa aaacgtgaaa ttattatgcc gccagggcgtg gtatcgcagc aggtaagatg 60
 attcaggaga ttttaa 76

 <210> 86
 10 <211> 78
 <212> ADN
 <213> Escherichia coli

 <400> 86
 gacaccatcg aatggcgcaa aacctttcgc ggtatggcat gatagcgccc ggaagagagt 60
 15 caattcaggg tggatgaat 78

 <210> 87
 <211> 80
 <212> ADN
 20 <213> Escherichia coli

 <400> 87
 tttgtgatcg ttatctcgat atttaaaaac aaataatttc attatatttt gaaatcgaaa 60
 acaaacgaca ggatatgaaa 80

 25 <210> 88
 <211> 80
 <212> ADN
 <213> Escherichia coli

 30 <400> 88
 cgacgcgtcg tgattaggtg aacccttct cgttatggca aaataagcca atacagaacc 60
 agcattatct ggagaatttc 80

ES 2 875 579 T3

<210> 89
 <211> 74
 <212> ADN
 <213> Escherichia coli

5

<400> 89
 ttaggaatag cgcgccgttca aaaatgtgtc actggtttac acttattcag gaatgcacaa 60
 tgaacggagg tcct 74

<210> 90
 10 <211> 73
 <212> ADN
 <213> Escherichia coli

<400> 90
 15 gaggcagcaaaa agtgctgtaa ctctgaaaaa gcgatggtag aatccatttt taagcaaacg 60
 gtgattttga aaa 73

<210> 91
 <211> 77
 <212> ADN
 20 <213> Escherichia coli

<400> 91
 gcgacaactt tcgtaaaaca tocctaccct gcttcaggta tactatgcc ctcgattcca 60
 caaacatcag gcagacc 77

25 <210> 92
 <211> 75
 <212> ADN
 <213> Escherichia coli

30 <400> 92
 tgtagcggag ttgtttttgt gtttacaaac aatggctcta cactgcaaac agacataaca 60
 acattcgggg tgaat 75

ES 2 875 579 T3

<210> 93
 <211> 70
 <212> ADN
 <213> Escherichia coli
 5
 <400> 93
 acttttccac aggtagatcc caacgcgttc acagcgtaca atacgccact cttataaag 60
 gtggcggttt 70

 <210> 94
 10 <211> 78
 <212> ADN
 <213> Escherichia coli

 <400> 94
 ttaacaattg atgattttgc caacagccca catagcgcca tatactgaaa attctcgag 60
 15 caactgaatg ttaagcct 78

 <210> 95
 <211> 78
 <212> ADN
 20 <213> Escherichia coli

 <400> 95
 aaaagcctaa ctcaacttttg attcattaaa aaagaaggct aaaatagaat gaatcatcaa 60
 tccgcataag aaaatcct 78

 25 <210> 96
 <211> 75
 <212> ADN
 <213> Escherichia coli

 30 <400> 96
 cggctccagg ttacttcccg taggattctt gctttaatag tgggattaat ttccacatta 60
 aaacagggat tgatc 75

ES 2 875 579 T3

	<210> 97		
	<211> 79		
	<212> ADN		
	<213> Escherichia coli		
5	<400> 97		
	tcacttaaca acagggcgga agcaacgcga aattctgcta ccatccacgc actctttatc	60	
	tgaataaatg gcagcgact	79	
	<210> 98		
10	<211> 74		
	<212> ADN		
	<213> Escherichia coli		
	<400> 98		
	acggttctgt ctgcttgctt ttaatgccat accaaacgta ccattgagac acttgtttgc	60	
15	acagaggatg gccc	74	
	<210> 99		
	<211> 81		
	<212> ADN		
20	<213> Escherichia coli		
	<400> 99		
	atcagcgttt ttggctggcg gcgtagcgat gcgctgggta ctctgaaaac ggtctatgca	60	
	aattaacaaa agagaatagc t	81	
25	<210> 100		
	<211> 82		
	<212> ADN		
	<213> Escherichia coli		
30	<400> 100		
	atgctaaatc cttacttccg catattctct gagcgggtat gctacctgtt gtatcccaat	60	
	ttcatacagt taaggacagg cc	82	

ES 2 875 579 T3

<210> 101
 <211> 79
 <212> ADN
 <213> Escherichia coli

5

<400> 101
 gatagcctga atcagtattg atctgctggc aagaacagac tactgtatat aaaaacagta 60
 taacttcagg cagattatt 79

<210> 102
 10 <211> 80
 <212> ADN
 <213> Escherichia coli

<400> 102
 ctctccccgg tgagaaatac gcttccccgt aagcgcattg taaactatgc cttcaaatcg 60
 15 ggcttatcgc gagtaaatct 80

<210> 103
 <211> 80
 <212> ADN
 20 <213> Escherichia coli

<400> 103
 aaaacacgcc accccttgaa ccaacgggcy ttttccgtaa cactgaaaga atgtaagcgt 60
 ttaccacta aggtattttc 80

25 <210> 104
 <211> 76
 <212> ADN
 <213> Escherichia coli

30 <400> 104
 ttcagggctg actgtttgca taaaaattca tctgtatgca caataatgtt gtatcaacca 60
 ccatatcggg tgactt 76

ES 2 875 579 T3

<210> 105
 <211> 80
 <212> ADN
 <213> Escherichia coli
 5
 <400> 105
 gttagcggct tacacgcggt cacattcaaa tgcgattctg ctacaatcct cccccgttc 60
 gaagattgag caatacacct 80

 <210> 106
 10 <211> 80
 <212> ADN
 <213> Escherichia coli

 <400> 106
 ttcgcttcag atcgttgaca gccgcactcc atgacgggta aaaagtggat aaaataattt 60
 15 taccacccgg atttttaccc 80

 <210> 107
 <211> 72
 <212> ADN
 20 <213> Escherichia coli

 <400> 107
 gcgtacaggc accgtagcat ttgtccgta taacgcacaa gtgataaact tccgttttgc 60
 cggaggagtc gc 72

 25 <210> 108
 <211> 76
 <212> ADN
 <213> Escherichia coli

 30 <400> 108
 tgttaacgcg ttatctcgtc gcgacctata agtttgggta atatgtgctg gaatttgccc 60
 tgtctggaga atcgca 76

ES 2 875 579 T3

<210> 109
 <211> 71
 <212> ADN
 <213> Escherichia coli

5

<400> 109
 gttgacaaca ttttctgcta accctgtgac ctgcaatact gttttgcggg tgatcgacaa 60
 ggagacttaa c 71

<210> 110
 10 <211> 72
 <212> ADN
 <213> Escherichia coli

<400> 110
 ttcagtgttc agttcgtaa ttcattctgtt ggggagtata attcctcaag ttaacttgag 60
 15 gtaaagcgat tt 72

<210> 111
 <211> 75
 <212> ADN
 20 <213> Escherichia coli

<400> 111
 ctgattatga ttgtgtatth aattggttgt tatttgacta ctatcaactt gttttaatth 60
 tatgataggt gcaag 75

25 <210> 112
 <211> 81
 <212> ADN
 <213> Escherichia coli

30 <400> 112
 aactataat ctgattthaa cgatgattcg tgcggggtaa aatagtaaaa acgatctatt 60
 cacctgaaag agaaataaaa a 81

ES 2 875 579 T3

<210> 113
 <211> 81
 <212> ADN
 <213> Escherichia coli
 5
 <400> 113
 ccgttgctct ctgatttctc atttcatgct cacccaatat gatggcggcg ttttctaaaa 60
 ctgttaaaga atgaggttaag t 81

 <210> 114
 10 <211> 72
 <212> ADN
 <213> Escherichia coli

 <400> 114
 ctgctatcag cgtagttagc cctctggtat gatgagtcca actttgttt gctgtgttat 60
 15 ggaaatctca ct 72

 <210> 115
 <211> 71
 <212> ADN
 20 <213> Escherichia coli

 <400> 115
 gttgttgta tgccccagg tatttacagt gtgagaaaga attatattga ctttagcgga 60
 gcagttgaag a 71

 25 <210> 116
 <211> 82
 <212> ADN
 <213> Escherichia coli

 30 <400> 116
 tttttcgcct ttcatacttg caaaagcgga gaatcagcta tccttttccc tgaaacctca 60
 tcaactcaaa gggagaatcg tg 82

ES 2 875 579 T3

	<210> 117	
	<211> 75	
	<212> ADN	
	<213> Escherichia coli	
5	<400> 117	
	cagactttac ccattgctga atgcacgggt aacgttaggc tcaaataatt aaacaacacg	60
	ttacaggaca acagg	75
	<210> 118	
10	<211> 82	
	<212> ADN	
	<213> Escherichia coli	
	<400> 118	
15	ctcctgaaca tccactcgat cttcgccttc ttccggttta ttgtgtttta accacctgcc	60
	cgtaaacctg gagaaccatc gc	82
	<210> 119	
20	<211> 79	
	<212> ADN	
	<213> Escherichia coli	
	<400> 119	
	atataaaacc gtactcctt tcacgtccgt tataaatatg atggctatta gaaagtcatt	60
25	aaattataa ggtgcgca	79
	<210> 120	
	<211> 80	
	<212> ADN	
30	<213> Escherichia coli	
	<400> 120	
	tatgaacagc ggcaactggtc aggatgaacg gcttacggca gaatatgaac agatatgaac	60
	agaatgagta aaaccctctg	80

ES 2 875 579 T3

<210> 121
 <211> 80
 <212> ADN
 5 <213> Escherichia coli

<400> 121
 gggtaaagcg ccagggccag aagtcgatac gacctgtgct atgattcata aatcacaaca 60
 ataacaacag actgaagcga 80

10 <210> 122
 <211> 74
 <212> ADN
 <213> Escherichia coli

15 <400> 122
 tgcgtttcca ttgagtatc ctgaaaacgg gcttttcagc atggaacgta catattaaat 60
 agtaggagtg cata 74

<210> 123
 <211> 81
 20 <212> ADN
 <213> Escherichia coli

<400> 123
 gaatatagag aagtacttac ttaacatfff cccatttggg actatctaac cccttttcac 60
 tattaagaag taatgcctac t 81

25
 <210> 124
 <211> 77
 <212> ADN
 <213> Escherichia coli

30
 <400> 124
 tagagcggca acgcgtaccc tgggtacgcg ttgtttgtct ggagtttcag ttaactagt 60
 gacttgagga aaaccta 77

ES 2 875 579 T3

	<210> 125		
	<211> 78		
	<212> ADN		
5	<213> Escherichia coli		
	<400> 125		
	ggataatgaa cagcattgct tadcagagat atgccccat atgttgaggc atatcctaac	60	
	gagaatctga caaccggt	78	
10	<210> 126		
	<211> 73		
	<212> ADN		
	<213> Escherichia coli		
15	<400> 126		
	cacgatgaaa aaaatgtagt ttttcaagg tgaagcgggt taaattcgtt ctcaaattac	60	
	agtcaggacg cgt	73	
	<210> 127		
	<211> 77		
20	<212> ADN		
	<213> Escherichia coli		
	<400> 127		
	ataaaattta gtgctgtaca gagcgcgtta caacacgggt tactggcagc aaatacgggt	60	
	atcgcagggg tgaaaaa	77	
25	<210> 128		
	<211> 71		
	<212> ADN		
	<213> Escherichia coli		
30	<400> 128		
	aaagctataa ctgttaaaca caatacagtg aaaagtttta gactgaaggc tcactttgca	60	
	gagggaagcg t	71	

ES 2 875 579 T3

	<210> 129	
	<211> 76	
	<212> ADN	
5	<213> Escherichia coli	
	<400> 129	
	gcaaatttta tcgtttgtca gcctgcgttg tttttttgtc caatatcatc aggttaatca	60
	caggggaagg tgagat	76
10	<210> 130	
	<211> 82	
	<212> ADN	
	<213> Escherichia coli	
15	<400> 130	
	tttgtttttc acgcgcttta cagcccgaaa aggccggaag atacttgccc gcaacgaaga	60
	ttccttcata accgggtaag ca	82
	<210> 131	
	<211> 72	
20	<212> ADN	
	<213> Escherichia coli	
	<400> 131	
	gcgcaagatt gttggttttt gcgtgatggt gaccgggcag cctaaaggct atccttaacc	60
	agggagctga tt	72
25	<210> 132	
	<211> 63	
	<212> ADN	
	<213> Secuencia artificial	
30	<220>	
	<223> promotor pMB029	

ES 2 875 579 T3

	<400> 132		
	accgtgcggtg ttgactatth tacctctggc ggttataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 133		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB023		
	<400> 133		
	accgtgcggtg ttgacaatth tacctctggc ggttataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 134		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB025		
	<400> 134		
	accgtgcggtg ttgacaatth tacctctggc ggtagagtg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 135		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB019		

ES 2 875 579 T3

	<400> 135		
	accgtgcggtg ttgacttaaa taccactggc gggtataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 136		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB008		
	<400> 136		
	accgtgcggtg ttgacataaa taccactggc gggtataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 137		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB020		
	<400> 137		
	accgtgcggtg ttgacataaa taccactggc gggtagagtg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 138		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB022		

ES 2 875 579 T3

	<400> 138		
	accgtgctg ttgacataaa taccactggc ggtgataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 139		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB089		
	<400> 139		
	accgtgctg ttgactatatt tacctctggc ggtagagtg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 140		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB001		
	<400> 140		
	accgtgctg ttgacataaa taccactggc ggttatattg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 141		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB051		

ES 2 875 579 T3

	<400> 141		
	accgtgctg ttgactat tt tacctctggc ggttat tttg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 142		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB070		
	<400> 142		
	accgtgctg ttgacaat tt tacctctggc ggtgatactg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 143		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB074		
	<400> 143		
	accgtgctg tttacataaa taccactggc ggttataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 144		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB046		

ES 2 875 579 T3

<400> 144
accgtgcggtg ttgactatatt tacctctggc ggtgataatg gttgcatgta ctaaggaggt 60
tgt 63

<210> 145
5 <211> 63
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB071

<400> 145
accgtgcggtg ttgacttaaa taccactggc ggtgataatg gttgcatgta ctaaggaggt 60
tgt 63

15 <210> 146
<211> 63
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB013

<400> 146
accgtgcggtg ttgacataaa taccactggc ggtgataactg gttgcatgta ctaaggaggt 60
tgt 63

25 <210> 147
<211> 63
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB080

ES 2 875 579 T3

	<400> 147		
	accgtgctg tagacttaaa taccactggc ggttataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 148		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB038		
	<400> 148		
	accgtgctg ttgactatth tacctctggc ggtgatactg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 149		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB060		
	<400> 149		
	accgtgctg ttacaatth tacctctggc ggttagagtg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 150		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB064		

ES 2 875 579 T3

	<400> 150		
	accgtgctg ttgacttaaa taccactggc ggtgatactg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 151		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB058		
	<400> 151		
	accgtgctg tttacaattt tacctctggc ggttattttg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 152		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB085		
	<400> 152		
	accgtgctg tttacaattt tacctctggc ggtgataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 153		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB081		

ES 2 875 579 T3

<400> 153
accgtgcg⁵tg ttgacttaaa taccactggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

<210> 154
<211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB091

<400> 154
accgtgcg¹⁵tg ttgacaattt tacctctggc gg¹⁵ttagagtt aacatcctac aaggagaaca 60
aaagc 65

<210> 155
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB027

<400> 155
accgtgcg²⁵tg ttgacataaa taccactggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

<210> 156
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB048

ES 2 875 579 T3

<400> 156
accgtgctg ttgacataaa taccactggc ggtagagtt aacatcctac aaggagaaca 60
aaagc 65

<210> 157
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB055

<400> 157
accgtgctg ttgacaattt tacctctggc ggttattttt aacatcctac aaggagaaca 60
aaagc 65

15 <210> 158
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB006

<400> 158
accgtgctg ttgactattt tacctctggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

25 <210> 159
<211> 63
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB012

ES 2 875 579 T3

<400> 159
 accgtgcg⁶⁰tg tttacataaa taccactggc ggtgataatg gttgcatgta ctaaggaggt
 tgt ⁶³

<210> 160
 5 <211> 63
 <212> ADN
 <213> Secuencia artificial

<220>
 10 <223> promotor pMB014

<400> 160
 accgtgcg⁶⁰tg tttacaattt tacctctggc ggtgataactg gttgcatgta ctaaggaggt
 tgt ⁶³

15 <210> 161
 <211> 65
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> promotor pMB028

<400> 161
 accgtgcg⁶⁰tg ttgacaattt tacctctggc ggtgataatt aacatcctac aaggagaaca
 aaagc ⁶⁵

25
 <210> 162
 <211> 63
 <212> ADN
 <213> Secuencia artificial

30
 <220>
 <223> pMB059 promotor

ES 2 875 579 T3

<400> 162
 accgtgcggtg ttacataaa taccactggc ggttatcttg gttgcatgta ctaaggaggt 60
 tgt 63

<210> 163
 5 <211> 65
 <212> ADN
 <213> Secuencia artificial

<220>
 10 <223> promotor pMB061

<400> 163
 accgtgcggtg ttgacaattt tacctctggc ggtgatactt aacatcctac aaggagaaca 60
 aaagc 65

15 <210> 164
 <211> 65
 <212> ADN
 <213> Secuencia artificial

20 <220>
 <223> promotor pMB043

<400> 164
 accgtgcggtg ttgactatctt tacctctggc ggttagagtt aacatcctac aaggagaaca 60
 aaagc 65

25 <210> 165
 <211> 63
 <212> ADN
 <213> Secuencia artificial

30 <220>
 <223> promotor pMB066

ES 2 875 579 T3

<400> 165
accgtgctg tagactatth tacctctggc ggttatthtg gttgcatgta ctaaggaggt 60
tgt 63

<210> 166
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB079

<400> 166
accgtgctg ttgacataaa taccactggc ggtgataatt aacatcctac aaggagaaca 60
aaagc 65

15 <210> 167
<211> 63
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB032

<400> 167
accgtgctg tagacttaaa taccactggc ggttagagtg gttgcatgta ctaaggaggt 60
tgt 63

25 <210> 168
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB068

ES 2 875 579 T3

<400> 168
accgtgcg⁶⁰tg tttacaattt tacctctggc ggttataatt aacatcctac aaggagaaca
aaagc ⁶⁵

<210> 169
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB082

<400> 169
accgtgcg⁶⁰tg ttgactat⁶⁰ttt tacctctggc ggttat⁶⁰ttttt aacatcctac aaggagaaca
aaagc ⁶⁵

15 <210> 170
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB030

<400> 170
accgtgcg⁶⁰tg ttacataaa taccactggc ggttataatt aacatcctac aaggagaaca
aaagc ⁶⁵

25 <210> 171
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB067

ES 2 875 579 T3

<400> 171
accgtgcggtg ttgacttaaa taccactggc ggtagagtt aacatcctac aaggagaaca 60
aaagc 65

<210> 172
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB050

<400> 172
accgtgcggtg ttgacttaaa taccactggc gggtattttt aacatcctac aaggagaaca 60
aaagc 65

15 <210> 173
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB069

<400> 173
accgtgcggtg ttgactatth tacctctggc ggtgataatt aacatcctac aaggagaaca 60
aaagc 65

25 <210> 174
<211> 63
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB017

ES 2 875 579 T3

	<400> 174		
	accgtgcggtg taggctatatt tacctctggc ggttataatg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 175		
5	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB039		
	<400> 175		
	accgtgcggtg tagacttaaa taccactggc ggttatatttg gttgcatgta ctaaggaggt	60	
	tgt	63	
15	<210> 176		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB011		
	<400> 176		
	accgtgcggtg ttgacttaaa taccactggc ggtgataatt aacatcctac aaggagaaca	60	
	aaagc	65	
25	<210> 177		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB072		

ES 2 875 579 T3

<400> 177
accgtgcg⁵tg tagacttaaa taccactggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

<210> 178
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB016

<400> 178
accgtgcg⁵tg ttgactat¹⁵ttt tacctctggc ggtgatactt aacatcctac aaggagaaca 60
aaagc 65

15 <210> 179
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB077

<400> 179
accgtgcg⁵tg tagactat¹⁵ttt tacctctggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

25 <210> 180
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB047

ES 2 875 579 T3

<400> 180
accgtgcg⁵tg ttgacttaaa taccactggc ggtgatactt aacatcctac aaggagaaca 60
aaagc 65

<210> 181
<211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB052

<400> 181
accgtgcg¹⁵tg ttacaattt tacctctggc gg¹⁵ttagagtt aacatcctac aaggagaaca 60
aaagc 65

<210> 182
<211> 65
<212> ADN
<213> Secuencia artificial

<220>
20 <223> promotor pMB090

<400> 182
accgtgcg²⁵tg ttacaattt tacctctggc ggtgataatt aacatcctac aaggagaaca 60
aaagc 65

<210> 183
<211> 65
<212> ADN
<213> Secuencia artificial

<220>
30 <223> promotor pMB035

ES 2 875 579 T3

<400> 183
accgtgcg⁶⁰tg ttacataaa taccactggc gg⁶⁰ttagagtt aacatcctac aaggagaaca
aaagc ⁶⁵

<210> 184
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB073

<400> 184
accgtgcg⁶⁰tg tagactat⁶⁰ttt tacctctggc gg⁶⁰ttagagtt aacatcctac aaggagaaca
aaagc ⁶⁵

15 <210> 185
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB004

<400> 185
accgtgcg⁶⁰tg ttacaat⁶⁰ttt tacctctggc ggtgatactt aacatcctac aaggagaaca
aaagc ⁶⁵

25 <210> 186
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB054

ES 2 875 579 T3

	<400> 186		
	accgtgcg ^{tg} tagacttaaa taccactggc ggttagagtt aacatcctac aaggagaaca	60	
	aaagc	65	
	<210> 187		
5	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB024		
	<400> 187		
	accgtgcg ^{tg} tagactat ^{tt} tacctctggc ggtgataatt aacatcctac aaggagaaca	60	
	aaagc	65	
15	<210> 188		
	<211> 63		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB007		
	<400> 188		
	accgtgcg ^{tg} taggctat ^{tt} tacctctggc ggtgatactg gttgcatgta ctaaggaggt	60	
	tgt	63	
25	<210> 189		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB005		

ES 2 875 579 T3

<400> 189
accgtgcg^{tg} taggctattt tacctctggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

<210> 190
5 <211> 63
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB003

<400> 190
accgtgcg^{tg} taggctattt tacctctggc ggtgataatg gttgcatgta ctaaggaggt 60
tgt 63

15 <210> 191
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB088

<400> 191
accgtgcg^{tg} ttacataaa taccactggc ggtgataatt aacatcctac aaggagaaca 60
aaagc 65

25 <210> 192
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> pMB065
promotor

ES 2 875 579 T3

<400> 192
accgtgcg⁵tg tagacttaaa taccactggc ggttat⁶⁰ttttt aacatcctac aaggagaaca
aaagc 65

<210> 193
<211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB037

<400> 193
accgtgcg⁵tg tagacttaaa taccactggc ggtgata⁶⁰aatt aacatcctac aaggagaaca
aaagc 65

15 <210> 194
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB009

<400> 194
accgtgcg⁵tg tagacttaaa taccactggc ggtgata⁶⁰ctt aacatcctac aaggagaaca
aaagc 65

25 <210> 195
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB041

ES 2 875 579 T3

<400> 195
accgtgcg⁵tg taggctattt tacctctggc ggttattttt aacatcctac aaggagaaca 60
aaagc 65

<210> 196
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB036

<400> 196
accgtgcg⁵tg taggctattt tacctctggc ggtgataatt aacatcctac aaggagaaca 60
aaagc 65

15 <210> 197
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB049

<400> 197
accgtgcg⁵tg taggcttaaa taccactggc ggttataatt aacatcctac aaggagaaca 60
aaagc 65

25 <210> 198
<211> 65
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB044

ES 2 875 579 T3

	<400> 198	
	accgtgcg ^{tg} g ttacataaa taccactggc ggtgatactt aacatcctac aaggagaaca	60
	aaagc	65
	<210> 199	
5	<211> 63	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
10	<223> promotor pMB042	
	<400> 199	
	accgtgcg ^{tg} g taggcttaaa taccactggc ggtgataatg gttgcatgta ctaaggaggt	60
	tgt	63
15	<210> 200	
	<211> 63	
	<212> ADN	
	<213> Secuencia artificial	
20	<220>	
	<223> promotor pMB086	
	<400> 200	
	accgtgcg ^{tg} g taggcttaaa taccactggc gg ^{tt} agagtg gttgcatgta ctaaggaggt	60
	tgt	63
25	<210> 201	
	<211> 63	
	<212> ADN	
	<213> Secuencia artificial	
30	<220>	
	<223> promotor pMB053	

ES 2 875 579 T3

	<400> 201		
	accgtgcg tg taggcttaaa taccactggc ggttattttg gttgcatgta ctaaggaggt	60	
	tgt	63	
	<210> 202		
5	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
10	<223> promotor pMB057		
	<400> 202		
	accgtgcg tg taggctattt tacctctggc ggtgatactt aacatcctac aaggagaaca	60	
	aaagc	65	
15	<210> 203		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
20	<220>		
	<223> promotor pMB018		
	<400> 203		
	accgtgcg tg taggcttaaa taccactggc ggttattttt aacatcctac aaggagaaca	60	
	aaagc	65	
25	<210> 204		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
30	<220>		
	<223> promotor pMB002		

ES 2 875 579 T3

<400> 204
accgtgcg⁶⁰tg taggcttaaa taccactggc ggttagagtt aacatcctac aaggagaaca
aaagc ⁶⁵

<210> 205
5 <211> 65
<212> ADN
<213> Secuencia artificial

<220>
10 <223> promotor pMB015

<400> 205
accgtgcg⁶⁰tg taggcttaaa taccactggc ggtgataatt aacatcctac aaggagaaca
aaagc ⁶⁵

15 <210> 206
<211> 65
<212> ADN
<213> Secuencia artificial

20 <220>
<223> promotor pMB087

<400> 206
accgtgcg⁶⁰tg taggcttaaa taccactggc ggtgatactt aacatcctac aaggagaaca
aaagc ⁶⁵

25 <210> 207
<211> 63
<212> ADN
<213> Secuencia artificial

30 <220>
<223> promotor pMB063

ES 2 875 579 T3

<400> 207

accgtgcggtg taggcttaaa taccactggc ggtgatactg gttgcatgta ctaaggaggt 60

tgt 63

<210> 208

5 <211> 10

<212> ADN

<213> Fago Lambda

<400> 208

10 accgtgcggtg 10

<210> 209

<211> 17

<212> ADN

15 <213> Fago Lambda

<400> 209

atttacctc tggcgggt 17

20 <210> 210

<211> 17

<212> ADN

<213> Fago Lambda

25 <400> 210

taaataccac tggcgggt 17

<210> 211

<211> 24

30 <212> ADN

<213> Fago Lambda

<400> 211

ggtgcatgt actaaggagg ttgt 24

35

<210> 212

<211> 26

<212> ADN

<213> Escherichia coli

5

<400> 212

taacatccta caaggagaac aaaagc 26

<210> 213

10 <211> 6357

<212> ADN

<213> Secuencia artificial

<220>

15 <223> vector Ori_Plsmid27

<400> 213

ES 2 875 579 T3

actacctagc tgcattttca ggaggaagcg atggggcgcc gcacaccttc ttaataagat 60
 gatcttcttg agatcgtttt ggtctgcgcg taatctcttg ctctgaaaac gaaaaaacgg 120
 ccttgacagg cggtttttcg aaggttctct gagctaccaa ctctttgaac cgaggttaact 180
 ggcttgagg agcgcagtca ccaaaacttg tcctttcagt ttagccttaa cgggcgcatg 240
 acttcaagac taactcctct aatcaatta ccagtggctg ctgccagtgg tgcttttgca 300
 tgtctttccg ggttgactc aagacgatag ttaccggata aggcgcagcg gtcggactga 360
 acggggggtt cgtgcataca gtccagcttg gagcgaactg cctaccggga actgagtgtc 420
 aggcgtggaa tgagacaaac gcggccataa cagcggaatg acaccggtaa accgaaaggc 480
 agaacacgga gagcgcacga gggagccgcc aggggaaacg cctggtatct ttatagtctt 540
 gtcgggtttc gccaccactg atttgagcgt cagatttctg gatgcttgtc agggggcgcg 600
 agcctatgga aaaacggctt tgccgcggcc ctctcacttc cctgttaagt atcttcctgg 660
 catcttccag gaaatctccg ccccgttcgt aagccatttc cgctcgccgc agtcgaacga 720
 ccgagcgtag cgagtcagt agcgaggaag cggaaatat cctgtatcac atattctgct 780
 gacgcaccgg tgcagccttt tttctcctgc cacatgaagc acttactga caccctcatc 840
 agtgccaaca tagtaagcca gtatacactc cgctatgata atgggtgagt gagtgtgtgc 900
 gtgtggggcg cgccagatgg gaacaaactt aatcgccttg cagcacatcc cccttcgcc 960
 agctggcgta atagcgaaga ggcccgacc gatcgcctt cccaacagt ggcgagcctg 1020
 aatggcgaat ggcgataagc tagcttcacg ctgccgaag cactcagggc gcaagggctg 1080
 ctaaaggaag cgaacacgt agaaagccag tccgcagaaa cggtgctgac cccgatgaa 1140
 tgtcagctac tgggtatct ggacaagga aaacgcaagc gcaaagagaa agcaggtagc 1200
 ttgcagtggg cttacatggc gatagctaga ctgggcgggt ttatggacag caagcgaacc 1260
 ggaattgcca gctggggcgc cctctggtaa ggttgggaag ccctgcaaag taaactggat 1320
 ggctttcttg ccgccaagga tctgatggcg caggggatca agatctgatc aagagacagg 1380
 atgaggatcg tttcgcatga ttgaacaaga tggattgcac gcaggttctc cggccgcttg 1440
 ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct ctgatgccgc 1500
 cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg acctgtccgg 1560
 tgccctgaat gaactccaag acgaggcagc gcggctatcg tggtggcca cgacgggctg 1620
 tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc tgctattggg 1680
 cgaagtgccg gggcaggatc tctgtcatc tcacctgct cctgccgaga aagtatccat 1740
 catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc cattcgacca 1800

ES 2 875 579 T3

ccaagcgaag catcgcatcg agcagacacg tactcggatg gaagccggtc ttgtcgatca 1860
 ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg ccaggctcaa 1920
 ggcgcggatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa 1980
 tatcatgggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc tgggtgtggc 2040
 ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc ttggcggcga 2100
 atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc agcgcacgac 2160
 cttctatcgc cttcttgacg agttcttctg agcgggactc tggggttcgc tagaggatcg 2220
 atccttttta acccatcaca tatacctgcc gttcactatt atttagtgaa atgagatatt 2280
 atgatatttt ctgaattgtg attaaaaagg caactttatg cccatgcaac agaaactata 2340
 aaaaatacag agaatgaaaa gaaacagata gattttttag ttcttttaggc ccgtagtctg 2400
 caaatccttt tatgattttc tatcaaaaca aagaggaaaa tagaccagtt gcaatccaaa 2460
 cgagagtcta atagaatgag gtcgaaaagt aaatcgcgcg ggtttgttac tgataaagca 2520
 ggcaagacct aaaatgtgta aagggcaaag tgtatacttt gccgtcacc cttacatatt 2580
 ttaggtcttt ttttattgtg cgtaactaac ttgccatctt caaacaggag ggctggaaga 2640
 agcagaccgc taacacagta cataaaaaag gagacatgaa cgactccagt ctttctagaa 2700
 gatggcaaac agctattatg ggtattatgg gtgctccccg aaaagtgcc cctgggtcct 2760
 tttcatcacg tgctataaaa ataattataa tttaaatfff ttaatataaa tatataaatt 2820
 aaaaatagaa agtaaaaaaa gaaattaaag aaaaaatagt tttgttttc cgaagatgta 2880
 aaagactcta gggggatcgc caacaatac taccttttat cttgctcttc ctgctctcag 2940
 gtattaatgc cgaattgttt catcttgtct gtgtagaaga ccacacacga aaatcctgtg 3000
 attttacatt ttacttatcg ttaatogaat gtatatctat ttaatctgct tttcttgtct 3060
 aataaatata tatgtaaagt acgctttttg ttgaaatfff ttaaaccttt gtttattfff 3120
 tttcttcat tccgtaactc ttctacctc tttatttact ttctaaaatc caaatacaaa 3180
 acataaaaat aaataaacac agagtaaatt cccaaattat tccatcatta aaagatacga 3240
 ggcgcgtgta agttacagc aagcgcgctg tcctaagaaa ccattattat catgacatta 3300
 acctataaaa ataggcgtat cacgagccc tttcgtctcg cgcgtttcgg tgatgacggt 3360
 gaaaacctct gacacatgca gctcccggag acggtcacag cttgtctgta agcggatgcc 3420
 gggagcagac aagcccgtca gggcgcgtca gcgggtgtg gcgggtgtcg gggctggctt 3480
 aactatgagg catcagagca gattgtactg agagtgcacc ataccacttt ttcgtgacgc 3540
 gcggttttga aaacatagac aagtttttgg cgtcgttgtt aatttcgaag aggatgtcca 3600
 atattttttt taaggaataa ggataactca agactagatt cccccctgca ttcccatcag 3660

ES 2 875 579 T3

aaccgtaaac	cttggcgctt	tccttgggaa	gtattcaaga	agtgccttgt	ccggtttctg	3720
tggctcacia	accagcgcg	ccgatatggc	tttcttttca	cttatgaatg	taccagtacg	3780
ggacaattag	aacgctcctg	taacaatctc	tttgcaaatg	tggggttaca	ttctaaccat	3840
gtcacactgc	tgacgaaatt	caaagtaaaa	aaaaatggga	ccacgtcttg	agaacgatag	3900
atthctthta	ttttacattg	aacagtcggt	gtctcagcgc	gctttatggt	ttcattcata	3960
cttcatatta	taaaataaca	aaagaagaat	ttcatattca	cgccaagaa	atcaggctgc	4020
tttccaaatg	caattgacac	ttcattagcc	atcacacaaa	actctttctt	getggagctt	4080
cttttaaaaa	agacctcagt	acaccaaaaa	cgttaccgga	cctcgttatt	ttacgacaac	4140
tatgataaaa	ttctgaagaa	aaaataaaaa	aatthtcata	cttcttgctt	ttatthtaac	4200
cattgaatga	tttctthtga	acaaaactac	ctgthtcacc	aaaggaaata	gaaagaaaaa	4260
atcaattaga	agaaaacaaa	aaacaaaatg	tctgttatta	atthcacagg	tagthctggt	4320
ccattgggtga	aagthtgcg	cttgcaagc	acagagccg	cagaatgtgc	tctagattcc	4380
gatgctgact	tgctgggtat	tatatgtgtg	cccaatagaa	agagaacaat	tgacctgggt	4440
attgcaagga	aaatthcaag	tctgttaaaa	gcatataaaa	atagthcagg	cactccgaaa	4500
tacttggttg	gcgtgtthcg	taatcaacct	aaggaggatg	thttggctct	ggtcaatgat	4560
tacggcattg	atctcgtcca	actgcatgga	gatgagctgt	ggcaagaata	ccaagagthc	4620
ctcggthtgc	cagthtatta	aagactcgta	thtccaaaag	actgcaacat	actactcagt	4680
gcagcttcac	agaaacctca	thcgtthtatt	cccttgthtg	atthcagaagc	aggtgggaca	4740
ggtgaactth	tggattggaa	ctcagatthct	gactgggttg	gaaggcaaga	gagccccgaa	4800
agcttacatt	thtatgttagc	tgggtgactg	acgccagaaa	atgthtggga	tgctgcttaga	4860
thaaatggcg	thattgggtgt	tgatgtaagc	ggagggtgtg	agacaaatgg	tgtaaaagac	4920
tctaacaaaa	tagcaaatht	cgtcaaaaat	gctaagaat	aggtthattac	tgagtagtat	4980
thattthaagt	atgthtthtg	cactthcctg	caggccttht	gaaaagcaag	cataaaagat	5040
ctaaacataa	aatctgtaaa	ataacaagat	gtaaagataa	tgctaaatca	thtggcttht	5100
tgattgattg	tacaggactg	ggtggaatcc	ctthctgcagc	acctggatta	ccctgthtacc	5160
cctagthggc	aaaggcacgg	thcgtctcgca	ccggaagggt	tctthcaggg	aacctctga	5220
aacctccgaa	acatctgact	tggthtacgat	ggactthgaa	cacgccgagg	aatgaacctc	5280
gacctcgcac	agthcatatga	gaaagthcagc	thcatgagac	catccacctg	ccctgagggga	5340
cgctthgagc	agctctggct	acctctatgg	ccactagcaa	gcgacaacct	ccgtgacggc	5400
atctaccgca	cctcacggca	taaagcccta	gacaagcgt	atctcgaaac	caacctcgat	5460
gccatctcga	atctctgggt	cgtcgacatc	gacaagcaag	acgcactctt	gcgtthcagtg	5520
tgggacagag	acggctggcg	gcctaaccgc	gtggthgaaa	acctthaac	gggcacgcac	5580

ES 2 875 579 T3

acgccgtgtg ggcgctcgcg gagcattta cccgcaccga gtacgccagg cgcaagccac 5640
ctgctacgcg gccgctgtca ccgaaggact acgccgctct gtcgacggcg acaaaggcta 5700
ctccgggctg atcaccaaaa accccgagca cacagcctgg gacagccact gggtaaccga 5760
taagctctac aactcogatg aactgcgctt ttggctggaa gaaaccggct tcatgccgcc 5820
agagtccctg aagaaaaccc gcagaaaatc gcccatcggg ctaggccgaa actgcgcgct 5880
cttcgagagc gcacgcagct gggcttatcg cgaaatacgc caccacttcg gcgaccccga 5940
cgggctagga cgctctatcc aagccaccgc ccaggcgctc aaccaggagc tgttttctga 6000
gccgctaccc gtggcagaag ttgaccaa atccaggat ggatcatcac 6060
caaatcacgc atgtggacag acggcccagc cgtctacgaa gccacattca ccacaatcca 6120
agccacacgc ggaaaacgaa gcgccgagca ccgctggggc accaccaacg cggaacgaat 6180
cgaaagatt atcaatgact aagcgcactc gtattccccg aaacggtaaa actatccgcg 6240
aagtagccga aggaactggc ctatcaactg cgacaatcga acgctggaca tctgcccctc 6300
gtaatgagta gtccctcatct ccctcaagca ggcgcggcg gtactgccat cctcgag 6357

<210> 214

<211> 6697

5 <212> ADN

<213> Secuencia artificial

<220>

<223> cadena principal de vector 1

10

<400> 214

ctggggcgcc ctctggtaag gttgggaagc cctgcaaagt aaactggatg gctttcttgc 60
cgccaaggat ctgatggcgc aggggatcaa gatctgatca agagacagga tgaggatcgt 120
ttcgcgatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg gtggagagggc 180
tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc gtgttccggc 240
tgtcagcgca gggcgccccg gttctttttg tcaagaccga cctgtccggt gccctgaatg 300
aactccaaga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt ccttgccgag 360
ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc gaagtgccgg 420
ggcaggatct cctgtcatct caccttgctc ctgccagaaa agtatccatc atggctgatg 480
caatgcggcg gctgcatacg cttgatccgg ctacctgcc attcgaccac caagcgaaac 540
atcgcatcga gcgagcacgt actcggatgg aagccggctc tctcgatcag gatgatctgg 600
acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag gcgcggatgc 660
ccgacggcga ggatctcgtc gtgacctatg gcgatgcctg cttgccgaat atcatggtgg 720
aaaatggccg cttttctgga ttcatcgact gtggccggct ggggtgtggcg gaccgctatc 780

ES 2 875 579 T3

aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa tgggctgacc 840
 gcttcctcgt gctttacggg atcgccgctc ccgattcgca gcgcatcgcc ttctatcgcc 900
 ttcttgacga gttcttctga ggggactct ggggttcgct agaggatcga tcctttttaa 960
 cccatcacat atacctgccg ttcactatta tttagtghaa tgagatatta tgatattttc 1020
 tgaattgtga ttaaaaaggc aactttatgc ccatgcaaca gaaactataa aaaatacaga 1080
 gaatgaaaag aaacagatag attttttagt tctttaggcc cgtagtctgc aaatcctttt 1140
 atgattttct atcaaaacaa agaggaaaat agaccagttg caatcctaac gagagtctaa 1200
 tagaatgagg tcgaaaagta aatcgcgagg gtttgttact gataaagcag gcaagaccta 1260
 aatgtgtaa agggcaaatg gtatactttg gcgtcaccoc ttacatattt taggtccttt 1320
 tttattgtgc gtaactaact tgccatcttc aaacaggagg gctggaagaa gcagaccgct 1380
 aacacagtac ataaaaagg agacatgaac gtgagctggt tacaattaat catcgtgtgg 1440
 taccatgtgt ggaattggaa aggacatgaa catcaaaaag tttgcaaac aagcaacagt 1500
 attaaccttt actaccgcac tgctggcagg aggcgcaact caagcgtttg cgaagaaac 1560
 gaaccaaag ccatataagg aaacatacgg catttcccat attacacgcc atgatatgct 1620
 gcaaatccct gaacagcaaa aaaatgaaaa atatcaagtt tctgaatttg attcgtccac 1680
 aattaaat atctcttctg caaaaggcct ggacgtttgg gacagctggc cattacaaaa 1740
 cgctgacggc actgtcgcaa actatcacgg ctaccacatc gtctttgcat tagccggaga 1800
 tcctaaaaat gcggatgaca catcgattta catgttctat caaaaagtcg gcgaaacttc 1860
 tattgacagc tggaaaaacg ctggccgctt ctttaagac agcgacaaat tcgatgcaaa 1920
 tgattctatc ttaaaaagacc aaacacaaga atgggtcaggt tcagccacat ttacatctga 1980
 cggaaaaatc cgtttattct aactgattt ctccggtaaa cattacggca acaaacact 2040
 gacaactgca caagttaacg tatcagcatc agacagctct ttgaacatca acggtgtaga 2100
 ggattataaa tcaatctttg acggtgacgg aaaaacgtat caaaatgtac agcagttcat 2160
 cgatgaaggc aactacagct caggcgacaa ccatacctg agagatcctc actacgtaga 2220
 agataaaggc cacaaatact tagtatttga agcaaacact ggaactgaag atggctacca 2280
 aggcgaagaa tctttattta acaaagcata ctatggcaaa agcacatcat tcttccgtca 2340
 agaaagtcaa aaacttctgc aaagcgataa aaaacgcacg gctgagttag caaacggcgc 2400
 tctcggtatg attgagctaa acgatgatta cacactghaa aaagtgatga aaccgctgat 2460
 tgcatctaac acagtaacag atgaaattga acgcgcgaac gtctttaaaa tgaacggcaa 2520
 atggtacctg ttcactgact cccgcggatc aaaaatgacg attgacggca ttacgtctaa 2580
 cgatatttac atgcttggtt atgtttctaa ttctttaact ggcccataca agccgctgaa 2640
 caaaactggc cttgtgtaa aaatggatct tgatcctaac gatgtaacct ttacttactc 2700

ES 2 875 579 T3

acacttcgct gtacctcaag cgaaaggaaa caatgtcgtg attacaagct atatgacaaa 2760
 cagaggattc tacgcagaca aacaatcaac gtttgcccg agcttcctgc tgaacatcaa 2820
 aggcaagaaa acatctgttg tcaaagacag catccttgaa caaggacaat taacagttaa 2880
 caaataaaaa cgcaaaagaa aatgccgatg ggtaccgagc gaaatgaccg accaagcgac 2940
 gcccaacctg ccatcacgag atttcgattc caccgccgcc ttctatgaaa ggttgggctt 3000
 cggaatcgtt ttccgggacg ccctcgcgga cgtgctcata gtccaataac tacattgagc 3060
 gaaatgccaa ccacatgtcc catgctttta ctaatgtggg gtcttagaag aaagcgacca 3120
 atttaaggag agttgaatat gtctgaaatc cagctgacgg aggcattcatt gaacgaagca 3180
 gccgatgctg caattaaagc gttcogatga gcacaaaacc tcgatgaatt ggctgctctg 3240
 cgacgagatc acctgggtga tgcggcacca atccctcagg cacgccgctc gcttggaaac 3300
 attccaaaag atcagcgtaa ggatgccgga cgattcgtaa acatggcgct gggccgcgcy 3360
 gaaaagcact tcgcccaggt taagtggtt cttgaagaaa agcgaacgc agaagtcctg 3420
 gagctggagc gcgtggatgt taccgtccct accacacgtg aacaagtcgg cgactgcac 3480
 ccaattacga ttctcaacga acagatcgcg gacatctttg ttggcatggg ctgggagatc 3540
 gcagagggcc cggaagttga agccgaatac ttcaatttcg atgcacttaa ctttctccca 3600
 gaccaccag cccgcaccct gcaggatacc ttccacatcg cacctgaagg atcgcgcaa 3660
 gtgttgcgca ccatacctc tcctgtccag gttcgcacga tgctgaatcg agaggtacct 3720
 atctatatcg cctgtcctgg tcgctcttc cgactgacg aattggatgc taccacacc 3780
 cctgtctttc accagatcga gggcctggct gtcgacaaaag gcctgacaat ggctcacctt 3840
 cgcggaactc tggatcactt ggctaaagaa ctgttcggac ctgagactaa aaccgcatg 3900
 cgttcaaaact acttcccatt ttctgagccc agcgcggaag ttgatgtctg gttcccaaat 3960
 aagaaggcg gtgccggctg gatcgaatgg ggcggctcg gcattggtcaa cccaaacgtg 4020
 ctccgcgctg taggcgtcga cccggaagaa tacactggat tcggcttcg gatgggtatt 4080
 gaacgcacct tgcaattccg aaatggactc tcagatatgc gcgatatggt agaaggcgac 4140
 attcgcttta ccctcccttt cggcattcag gcttagcat ttttagtacg tgcaataacc 4200
 actctggttt ttccagggtg gttttttgat gccctttttg gactcttcaa ctgagcctcg 4260
 cagagcagga ttcccgttga gcaccgccag gtgcgaataa gggacagtga agaaggaaca 4320
 cccgctcgcg ggtgggcta cttcacctat cctgcccggc tgacgccgtt ggatacacca 4380
 aggaaagtct acacgaacct tttggcaaaa tcctgtatat cgtgcgaaaa aggatggata 4440
 taccgaaaaa atcgctataa tgaccccgaa gcagggttat gcagcggaaa agctccccga 4500
 aaagtgccac ctgggtcctt ttcatcacgt gctataaaaa taattataat ttaaattttt 4560

ES 2 875 579 T3

taatataaat atataaatta aaaatagaaa gtaaaaaaag aaattaaaga aaaaatagtt	4620
tttgttttcc gaagatgtaa aagactctag ggggatcgcc aacaaatact accttttctc	4680
ttgctcttcc tgctctcagg tattaatgcc gaattgtttc atcttgtctg tgtagaagac	4740
cacacacgaa aatcctgtga ttttacatth tacttatcgt taatcgaatg tatatctatt	4800
taatctgctt ttcttgtcta ataaatatat atgtaaagta cgctttttgt tgaattttt	4860
taaacctttg tttatthttt ttttcttcat tccgtaactc ttctaccttc tttatttact	4920
ttctaaaaat caaatacaaaa acataaaaaat aaataaacac agagtaaatt cccaaattat	4980
tccatcatta aaagatacga ggcgcgtgta agttacaggc aagcgatccg tctaagaaac	5040
cattattatc atgacattaa cctataaaaa taggcgtatc acgaggccct ttcgtctcgc	5100
gcgtttcggg gatgacgggtg aaaacctctg acacatgcag ctcccggaga cggtcacagc	5160
ttgtctgtaa gcggatgccg ggagcagaca agcccgtcag ggcgcgtcag cgggtgttgg	5220
cgggtgtcgg ggctggctta actatgcggc atcagagcag attgtactga gagtgcacca	5280
taccaccttt tcaattcctc atthtttttt tattcttttt tttgatttct gtttctttga	5340
cattthtttg attcggtaat ctccgaacag aaggaagaac gaaggaagga gcacagactt	5400
agattgggat atatacgcct atgtagtgtt gaagaaacat gaaattgccc agtattctta	5460
acccaactgc acagaacaaa aacctgcagg aaacgaagat aaatcatgtc gaaagctaca	5520
tataaggaac gtgctgctac tcatcctagt cctgttctg ccaagctatt taatatcatg	5580
cacgaaaagc aaacaaactt gtgtgcttca ttggatgttc gtaccaccaa ggaattactg	5640
gagtagttg aagcattagg tcccaaaatt tgtttactaa aaacacatgt ggatatcttg	5700
actgatthtt ccatggaggg cacagttaag ccgctaaagg cattatccgc caagtacaat	5760
thtttactct tcgaagacag aaaatthtct gacattggta atacagtcaa attgcagtac	5820
tctgcgggtg tatacagaat agcagaatgg gcagacatta cgaatgcaca cgggtgtgtg	5880
ggcccaggta ttgttagcgg tttgaagcag gcggcagaag aagtaacaaa ggaacctaga	5940
ggcctthtga tgttagcaga attgtcatgc aagggctccc tatctactgg agaataact	6000
aaggtactg ttgacattgc gaagagcgc aaagattht ttatcggctt tattgctcaa	6060
agagacatgg gtggaagaga tgaagttac gattggttga ttatgacacc cgggtgtggg	6120
ttagatgaca agggagacgc attgggtcaa cagtatagaa ccgtggatga tgtgtctct	6180
acaggatctg acattattat tgttgaaga ggactatthg caaagggag ggatgctaag	6240
gtagagggg aacgttacag aaaagcaggc tgggaagcat atthgagaag atgcggccag	6300
caaaactaaa aaactgtatt ataagtaaat gcatgtatac taaactcaca aattagagct	6360
tcaatttaat tatatcagtt attaccccc tggcttgttg tccacaaccg ttaaacctta	6420
aaagctthaa aagccttata tattctthtt tttcttataa aactthaaac cttagaggct	6480
atthaaagtt ctgatttata ttaatttht tgtthaaaca tgagagctta gtacgtgaaa	6540
catgagagct tagtacgtta gccatgaggg thtagttcgt tagccatgag ggtthtagttc	6600
gtthaaacatg agagcttagt acgtthaaaca tgagagctta gtacgtgaaa catgagagct	6660
tagtacgtac tatcaacagc ttgaactgct gatcttc	6697

ES 2 875 579 T3

<210> 215

<211> 7810

<212> ADN

<213> Secuencia artificial

5

<220>

<223> cadena principal de vector 2

<400> 215

gtttaaaciaa	ttttgtgtcg	cccttaattg	tgagcgctca	caattccaca	acggtttccc	60
tctagaata	atthttgttta	acttttcgag	accttaggag	gtaaacaat	ggcgagcttg	120
gttaagaaag	atatgtgtat	taagatgacg	atggaaggta	ctgtgaacgg	ttatcacttt	180
aagtgcgttg	gcgaggggtga	aggcaagccg	ttcgagggca	cgcagaacat	gcgactcgt	240
gtcaccgagg	gcggtccgct	gccttttgca	ttcgacatcc	tggccccgtg	ctgtatgtac	300
ggctctaaga	ccttcattaa	acacgtgagc	ggatccccgg	attactttaa	agagtccttt	360
ccagagggct	tcaactggga	acgtaccag	atthttgagg	acggtgggtg	tctgaccgag	420
caccaagaca	ccagcctgga	aggtaattgc	ctgatctata	aagtgaagg	tctgggtacc	480
aatthccccg	cgaatgggtcc	ggatgatgcaa	aagaaaaccg	cgggttgga	gccgtgcgtc	540
gagatgctgt	atccgcgtga	cggcgtcttg	tgtggtcaga	gcttgatggc	gctgaagtgc	600
accgatggca	atcatctgac	cagccacctg	cgcacgacgt	atcgtagccg	taaaccgagc	660
aacgccgtta	acatgccgga	gttccatttt	ggatgaccatc	gcatcgaaat	cctgaaagct	720
gagcagggca	aattctacga	acaatacga	tcggctgtcg	cacgttacag	cgatgtgccg	780
gaaaaagcga	cgtaatgaga	attctgtaca	ctcgagggtc	tcacccaag	ggcgacaccc	840
cctaattagc	ccgggcgaaa	ggcccagtct	ttcgactgag	cctttcgttt	tatttgagtt	900
taaacctggg	gcgccctctg	gtaagggttg	gaagccctgc	aaagtaact	ggatgccagg	960
catcaaataa	aacgaaaggc	tcagtcgaaa	gactgggcct	ttcgttttat	ctggtgtttg	1020
tcggtgaacg	ctctctacta	gagtcacact	ggctcacctt	cgggtgggcc	tttctgcggt	1080
tatagcacta	ggacttgccg	cggatactgc	cccattacat	gaattgcagc	ctcagggacg	1140
tcagtagatc	atggaggtag	ggcatatgtc	ctctgttggt	aaaatgtgag	ttctcaacga	1200
agcacgaatc	ggtcagaacc	tacactaagg	agatthggta	ggatgcacgg	ttctgtcgca	1260
10 tagaccagtt	catttcagat	gtctggcacg	taagaggttc	caactttcac	cataatgaaa	1320

ES 2 875 579 T3

taagatcact accggcgta tttttgagt tatcgagatt ttcaggagct aaggaagcta 1380
 aatgattga acaagatgga ttgcacgcag gttctccggc cgcttgggtg gagaggctat 1440
 tcggctatga ctgggcacaa cagacaatcg gctgctctga tgccgccgtg ttcgggctgt 1500
 cagcgcaggg gcgcccgggt cttttgtca agaccgacct gtccggtgcc ctgaatgaac 1560
 tccaagacga ggcagcgcgg ctatcgtggc tggccacgac gggcgttcct tgcgcagctg 1620
 tgctcgacgt tgtcactgaa gcgggaaggg actggctgct attgggcgaa gtgccggggc 1680
 aggatctcct gtcactcac ettgctcctg ccgagaaagt atccatcatg gctgatgcaa 1740
 tgcggggct gcatacgctt gatccggcta cctgccatt cgaccaccaa gcgaaacatc 1800
 gcacgcagcg agcacgtact cggatggaag ccggtcttgt cgatcaggat gatctggacg 1860
 aagagcatca ggggctcgcg ccagccgaac tgttcgccag gctcaaggcg cggatgcccg 1920
 acggcgagga tctcgtcgtg acccatggcg atgcctgctt gccgaatatac atggtggaac 1980
 atggccgctt ttctggattc atcgactgtg gccggctggg tgtggcggac cgctatcagg 2040
 acatagcgtt ggctaccctg gatattgctg aagagcttgg cggcgaatgg gctgaccgct 2100
 tcctcgtgct ttacggatc gccgctcccg attcgcagcg catcgccttc taticgccttc 2160
 ttgacgagtt cttctgatga tgaatggcaa gggggcgcgt agcccccaa ccgaagtga 2220
 ggggattttt ttgacaatta atcatccggc tcgtaattta tgtggatctt aatcatgcta 2280
 aggaggtttt ctaatgatga acatcaaaaa gtttgcaaaa caagcaacag tattaacctt 2340
 tactaccgca ctgctggcag gaggcgaac tcaagcgtt gccgaaagaaa cgaacaaaa 2400
 gccatataag gaaacatacg gcatttccca tattacacgc catgatatgc tgcaaatccc 2460
 tgaacagcaa aaaaatgaaa aatatcaagt ttctgaattt gattcgtcca caattaaaaa 2520
 tatctcttct gcaaaaaggcc tggacgtttg ggacagctgg ccattacaaa acgctgacgg 2580
 cactgtcgca aactatcacg gctaccacat cgtctttgca ttagccggag atcctaataaa 2640
 tgcggatgac acatcgattt acatgttcta tcaaaaagtc ggcgaaactt ctattgacag 2700
 ctgaaaaaac gctggccgcg totttaaga cagcgacaaa ttcgatgcaa atgattctat 2760
 cctaaaagac caaacacaag aatggtcagg ttcagccaca ttacatctg acggaaaaat 2820
 ccgtttattc tacactgatt tctccgtaa acattacggc aaacaaacac tgacaactgc 2880
 acaagttaac gtatcagcat cagacagctc tttgaacatc aacgggtgtag aggattataa 2940
 atcaatcttt gacggtgacg gaaaaacgta tcaaaatgta cagcagttca tcgatgaagg 3000
 caactacagc tcagggcaca accatacgct gagagatcct cactacgtag aagataaagg 3060
 ccacaaatac ttagtatttg aagcaaacac tggaaactgaa gatggctacc aaggcgaaga 3120
 atctttattt aacaaagcat actatggcaa aagcacatca ttcttcgctc aagaaagtca 3180
 aaaacttctg caaagcgata aaaaacgcac ggctgagtta gcaaacggcg ctctcggtat 3240

ES 2 875 579 T3

gattgagcta aacgatgatt acacactgaa aaaagtgatg aaaccgctga ttgcatctaa 3300
cacagtaaca gatgaaattg aacgcgcgaa cgtctttaa atgaacggca aatggtacct 3360
gttactgac tcccgcgat caaaaatgac gattgacggc attacgtcta acgatattta 3420
catgcttggg tatgtttcta attctttaac tggccatac aagccgctga acaaaactgg 3480
ccttgtgtta aaaatggatc ttgatcctaa cgatgtaacc ttacttact cacacttcgc 3540
tgtacctcaa gcgaaaggaa acaatgtcgt gattacaagc tatatgacaa acagaggatt 3600
ctacgcagac aaacaatcaa cgtttgcgcc gagcttcctg ctgaacatca aaggcaagaa 3660
aacatctggt gtcaaagaca gcatccttga acaaggacaa ttaacagtta acaataaaa 3720
acgcaaaaga aaatgccgat gggtaaccgag cgaaatgacc gaccaagcga cgcccaacct 3780
gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct tcggaatcgt 3840
tttccgggac gccctcgcg acgtgctcat agtccaggcg gtgttgacat aaataccact 3900
ggcggtgata ctgagcacat cagcagggtca cacaggaaag tactagatgt ctgaaatcca 3960
gctgacggag gcatcattga acgaagcagc cgatgctgca attaaagcgt tcgatggagc 4020
acaaaacctc gatgaattgg ctgctctgcg acgagatcac ctgggtgatg cggcaccaat 4080
ccctcaggca cgccgctcgc ttggaacct tccaaaagat cagcgttaagg atgccggacg 4140
attcgtaaac atggcgctgg gccgcgcgga aaagcacttc gcccaggta aggtggttct 4200
tgaagaaaag gcgaaacgcag aagtccctgga gctggagcgc gtggatgtta ccgtccctac 4260
cacacgtgaa caagtcggcg cactgcaccc aattacgatt ctcaacgaac agatcgcgga 4320
catctttggt ggcattggct gggagatcgc agagggcccg gaagttgaag ccgaataact 4380
caatttcgat gcacttaact ttctcccaga ccaccagcc cgcaccctgc aggatacctt 4440
ccacatcgca cctgaaggat cgcgcgcaagt gttgcgcacc catacctctc ctgtccagggt 4500
tcgcacgatg ctgaatcgag aggtacctat ctatatgcc tgcctgggtc gcgtcttccg 4560
cactgacgaa ttggatgcta cccacacccc tgtctttcac cagatcgagg gcctggctgt 4620
cgacaaaggc ctgacaatgg ctcaccttcg cggaactctg gatcacttgg ctaaagaact 4680
gttcggacct gagactaaaa cccgcatgcg ttcaaacctac ttcccatttt ctgagcccag 4740
cgcggaagtt gatgtctggt tcccaaataa gaaggcggt gccggctgga tcgaatgggg 4800
cggctcggc atggtcaacc caaacgtgct ccgcgctgta ggcgtcgacc cggaagaata 4860
cactggattc ggcttcggtg tgggtattga acgcaccttg caattccgaa atggactctc 4920
agatatcgc gatatggtag aaggcgacat tcgctttacc ctccctttcg gcattcaggc 4980
ttaggcattt ttagtacgtg caataaccac tctggttttt ccagggtggt tttttgatgc 5040
cctttttgga gtcttcaact gagcctcgca gagcaggatt cccgttgagc accgccagggt 5100

ES 2 875 579 T3

gcgaataagg gacagtgaag aaggaacacc cgctcgcggg tgggcctact tcacctatcc 5160
 tgcccggctg acgccgttgg atacaccaag gaaagtctac acgaaccctt tggcaaaatc 5220
 ctgtatatcg tgcgaaaaag gatggatata ccgaaaaaat cgctataatg accccgaagc 5280
 agggttatgc agcggaaaag ctccccgaaa agtgccacct gggtcctttt catcacgtgc 5340
 tataaaaata attataatth aaatthttta atataaatat ataaattaaa aatagaaagt 5400
 aaaaaagaa attaaagaaa aatagthttt tgttttccga agatgtaaaa gactctaggg 5460
 ggatcgccaa caaatactac cttttatctt gctcttctcg ctctcaggtt ttaatgccga 5520
 attgtttcat cttgtctgtg tagaagacca cacacgaaaa tcctgtgatt ttacatttta 5580
 cttatcgtha atcgaatgta tatctattha atctgctttt cttgtctaat aaatatatat 5640
 gtaaagtacg ctttttgttg aaatthttta aacctttgtt tathththth ttcttcattc 5700
 cgtaactctt ctacctctt taththcttt ctaaaatcca aatacaaaac ataaaaataa 5760
 ataaacacag agtaaatthc caaaththt cactathaaa agatacgagg cgcgtgtaag 5820
 ttacaggcaa gcgatccgtc taagaaacca ttattatcat gacattaacc tataaaaata 5880
 ggcgtatcac gaggcccttt cgtctcgcgc gtttcgggtg tgacgggtgaa aacctctgac 5940
 acatgcagct cccggagacg gtcacagctt gctgtgaagc ggatgccggg agcagacaag 6000
 cccgtcaggg cgcgtcagcg ggtgttggcg ggtgtcgggg ctggcttaac tatgcccgat 6060
 cagagcagat tgtactgaga gtgcaccata ccaccttht aaththctat ththththth 6120
 thctththth tgaththggg thctthgaca ththththth tchgtaatct ccgaacagaa 6180
 ggaagaacga aggaaggagc acagacttag atthggtatat atacgcatat gtagtgthga 6240
 agaaacatga aaththcccag ththctthaac ccaactgcac agaacaaaaa cctgcaggaa 6300
 acgaagataa atcatgtcga aagctacata taaggaacgt gctgctactc atcctagtcc 6360
 tgttgctgcc aagctatthta atatcatgca cgaaaagcaa acaactthgt gtgctthatt 6420
 ggatgttctg accaccaagg aaththctgga gthtagthgaa gcaththgth ccaaaaththg 6480
 ththctthaaa acacatgtgg atatctthgac tgathththct atggagggca cagththagcc 6540
 gctaaaggca ththctccca agthacaath ththctthct gaagacagaa aaththgctga 6600
 caththgthaat acagthcaaat thcagthactc thcgggtgta thcagaatag cagaaththggc 6660
 agacaththac aaththcacag ththgthgggg cccagthatt gththagcgtt thgaagcaggc 6720
 ggcagaagaa gthacaagg aacctagagg cthththgath ththagcaaat ththctthgcaa 6780
 gggctcccta ththctthggag aaththactha gggthctthgt gacaththcga agagcagaaa 6840
 agathththgt atcggctthta ththctthaaa agacaththggg ggaagagath aagththacga 6900
 ththgthgath atgacacccg ththgthggthh agaththacaag gththagcath thggthcaaca 6960
 gththagaacc ththgathgath ththgthctthac agththctthgac atththththg ththgaagagg 7020

ES 2 875 579 T3

actatattgca aagggaaggg atgctaaggt agagggtgaa cgttacagaa aagcaggctg 7080
 ggaagcatat ttgagaagat gcggccagca aaactaaaaa actgtattat aagtaaatgc 7140
 atgtatacta aactcacaaa ttagagcttc aatttaatta tatcagttat tacccccctg 7200
 gcttgtgtgc cacaaccggt aaaccttaaa agctttaaaa gccttatata ttcttttttt 7260
 tcttataaaa cttaaacct tagaggctat ttaagttgct gatttatatt aattttattg 7320
 ttcaaacatg agagcttagt acgtgaaaca tgagagctta gtacgtagc catgagggtt 7380
 tagttcgta gccatgaggg tttagttcgt taaacatgag agcttagtac gttaaacatg 7440
 agagcttagt acgtgaaaca tgagagctta gtacgtacta tcaacaggtt gaactgctga 7500
 tcttcggcat caaataaaac gaaaggctca gtcgaaagac tgggcctttc gttttatctg 7560
 ttgtttgtcg gtgaacgctc tcttgagtag gacaaatccg ccgccctaga tattcacacg 7620
 caatcaacag gcaggataat cgctggttaag gtcagtgtt tcttcaggta gtagagatac 7680
 aatagttccc aacgataggt ggcagatttc actttacaga ccgactggtt cagaagcgtg 7740
 gataatagcc cgtgttttcc aataagggat agtgtaggta agtcaactcc tccgtcagag 7800
 ccaaccgttt 7810

<210> 216

<211> 7756

5 <212> ADN

<213> Secuencia artificial

<220>

<223> cadena principal de vector 3

10

<400> 216

gtttaaacia ttttgtgtcg cccttaattg tgagcgctca caattccaca acggtttccc 60
 tctagaata attttgttta acttttcgag accttaggag gtaaacadat ggcgagcttg 120
 gttaagaaag atatgtgtat taagatgacg atggaaggta ctgtgaacgg ttatcacttt 180
 aagtgcgttg gcgaggggtga aggcaagccg ttcgagggca cgcagaacat gcgcactcgt 240
 gtcaccgagg gcggtccgct gccttttgca ttcgacatcc tggccccgtg ctgtatgtac 300
 ggctctaaga ccttcattaa acacgtgagc ggtatcccgg attactttaa agagtccttt 360
 ccagagggct tcacttggga acgtaccag atttttgagg acggtggtgt tctgaccgcg 420
 caccaagaca ccagcctgga aggtaattgc ctgatctata aagtgaaggt tctgggtacc 480
 aatttcccgg cgaatggtcc ggtgatgcaa aagaaaaccg cgggttgga gccgtgcgtc 540
 gagatgctgt atccgcgtga cggcgtcttg tgtggtcaga gcttgatggc gctgaagtgc 600
 accgatggca atcatctgac cagccacctg cgcacgacgt atcgtagccg taaaccgagc 660
 aacgccgtta acatgccgga gttccatttt ggtgaccatc gcatcgaaat cctgaaagct 720

ES 2 875 579 T3

gagcagggca aattctacga acaatacga tccgctgtcg cacgttacag cgatgtgccg 780
gaaaaagcga cgtaatgaga attctgtaca ctcgagggtc tcacccaag gccgacaccc 840
cctaattagc ccgggcgaaa ggcccagtct ttcgactgag cctttcgttt tatttgagtt 900
taaacctggg gcgccctctg gtaaggttgg gaagccctgc aaagtaaact ggatgccagg 960
catcaaataa aacgaaaggc tcagtcgaaa gactgggcct ttcgttttat ctgttgtttg 1020
tcggtgaacg ctcttacta gagtcacact ggctcacctt cgggtgggcc tttctgcgtt 1080
tatagcacta ggacttgccg cggatactgc cccattacat gaattgcagc ctcagggacg 1140
tcagtagatc atggaggtag ggcatatgtc ctctgttgtt aaaatgtgag ttctcaacga 1200
agcacgaatc ggtcagaacc tacactaagg agatttgta ggtgcacggt ttctgtcgca 1260
tagaccagtt catttcagat gtctggcacg taagaggttc caactttcac cataatgaaa 1320
taagatcact accgggcgta ttttttgagt tatcgagatt ttcaggagct aaggaagcta 1380
aatgattga acaagatgga ttgcacgacg gttctccggc cgtttgggtg gagaggctat 1440
tcggctatga ctgggcacaa cagacaatcg gctgctctga tgccgccgtg ttcgggctgt 1500
cagcgcaggg gcgcccggtt ctttttgta agaccgacct gtccgggtgcc ctgaatgaac 1560
tccaagacga ggcagcgcgg ctatcgtggc tggccacgac gggcgttcct tgcgcagctg 1620
tgctcgacgt tgtcactgaa gcgggaaggg actggctgct attgggcgaa gtgccggggc 1680
aggatctcct gtcatctcac cttgctcctg ccgagaaagt atccatcatg gctgatgcaa 1740
tgccggcgct gcatacgctt gatccgcta cctgccatt cgaccaccaa gcgaaacatc 1800
gcacgcagcg agcacgtact cggatggaag ccggctctgt cgatcaggat gatctggacg 1860
aagagcatca ggggctcgcg ccagccgaac tgttcgccag gctcaaggcg cggatgcccg 1920
acggcgagga tctcgtcgtg acccatggcg atgcctgctt gccgaatc atggtggaaa 1980
atggccgctt ttctggattc atcgactgtg gccggctggg tgtggcgac cgctatcagg 2040
acatagcgtt ggctaccgct gatattgctg aagagcttgg cggcgaatgg gctgaccgct 2100
tcctcgtgct ttacggtatc gccgctcccg attcgcagcg catgccttc tatgccttc 2160
ttgacgagtt cttctgatga tgaatggcaa ggcggcgcgt agcccccaa ccgaagtga 2220
ggggatTTTT ttgacaatta atcatccggc tcgtaattta tgtggatctt aatcatgcta 2280
aggaggtttt ctaatgatga acatcaaaaa gtttgcaaaa caagcaacag tattaacctt 2340
tactaccgca ctgctggcag gaggcgcaac tcaagcgttt gcgaaagaaa cgaacaaaa 2400
gccatataag gaaacatacg gcatttccca tattacacgc catgatatgc tgcaaatccc 2460
tgaacagcaa aaaaatgaaa aatatcaagt ttctgaattt gattcgtcca caattaataa 2520
tatctcttct gcaaaaaggcc tggacgtttg ggacagctgg ccattacaaa acgctgacgg 2580
cactgtcgca aactatcacg gctaccacat cgtctttgca ttagccggag atcctaataa 2640

ES 2 875 579 T3

tgcggatgac	acatcgattt	acatgttcta	tcaaaaagtc	ggcgaaactt	ctattgacag	2700
ctggaaaaac	gctggccgcg	tctttaaaga	cagcgacaaa	ttc gatgcaa	atgattctat	2760
cctaaaagac	caaacacaag	aatggtcagg	ttcagccaca	ttt acatctg	acggaaaaat	2820
ccgtttattc	tacactgatt	tctccggtaa	acattacggc	aaacaaacac	tgacaactgc	2880
acaagttaac	gtatcagcat	cagacagctc	tttgaacatc	aacggtgtag	aggattataa	2940
atcaatcttt	gacggtgacg	gaaaaacgta	tcaaaatgta	cagcagttca	tcgatgaagg	3000
caactacagc	tcaggcgaca	accatacgct	gagagatcct	cactacgtag	aagataaagg	3060
ccacaaatac	ttagtatttg	aagcaaacac	tggaaactgaa	gatggctacc	aaggcgaaga	3120
atctttatth	aacaaagcat	actatggcaa	aagcacatca	ttcttccgtc	aagaaagtca	3180
aaaacttctg	caaagcgata	aaaaacgcac	ggctgagtta	gcaaacggcg	ctctcggtat	3240
gattgagcta	aacgatgatt	acacactgaa	aaaagtgatg	aaaccgctga	ttgcatctaa	3300
cacagtaaca	gatgaaattg	aacgcgcgaa	cgtctttaa	atgaacggca	aatggtacct	3360
gttcaactgac	tcccgcggat	caaaaatgac	gattgacggc	attacgtcta	acgatattta	3420
catgcttggt	tatgtttcta	attctttaa	tggcccatac	aagccgctga	acaaaactgg	3480
ccttgtgtta	aaaatggatc	ttgatcctaa	cgatgtaacc	tttacttact	cacacttgc	3540
tgtacctcaa	gcgaaaggaa	acaatgtcgt	gattacaagc	tatatgacaa	acagaggatt	3600
ctacgcagac	aaacaatcaa	cgtttgcgcc	gagcttccctg	ctgaacatca	aaggcaagaa	3660
aacatctggt	gtcaaagaca	gcatccttga	acaaggacaa	ttaacagtta	acaaataaaa	3720
acgcaaaaaga	aaatgccgat	gggtaccgag	cgaaatgacc	gaccaagcga	cgcccaacct	3780
gccatcacga	gatttcgatt	ccaccgccgc	cttctatgaa	aggttgggct	tcggaatcgt	3840
tttccgggac	gccctcgcgg	acgtgctcat	agtccaggcg	gtgttgacat	aaataccact	3900
ggcggtgata	ctgagcacat	cagcagggtca	cacaggaaa	tactagatgt	cgcatcttgc	3960
agaattagta	gcttcagcga	aggcccgcat	ttctcaggcg	agtgacgtcg	cagcactgga	4020
taatgtacgt	gttgagtacc	tgggaaagaa	gggacacctt	actcttcaa	tgacaaccct	4080
gcgcgaaactg	ccgccggagg	aacgccccgc	agcaggagcg	gtaatcaatg	aggcaaagga	4140
gcaagtacaa	caggcactga	acgcccgtaa	ggctgagttg	gaatccgccg	cattaaacgc	4200
gcgccttgct	gcggaacca	ttgatgtctc	gctgcccggg	cgccgcattg	agaatggagg	4260
cttacacca	gtgactcgta	ccatcgaccg	tatcgaatct	ttctttggcg	aacttggett	4320
cactgtggca	actggaccgg	agattgagga	cgactaccac	aatttcgatg	ccttgaacat	4380
tcccggatcat	catcctgcac	gcgccgatca	tgatacatct	tggtttgata	ccacccttt	4440
gcttcgtacc	cagacaagcg	gtgtccaaat	ccgtacgatg	aaggctcagc	aaccaccgat	4500

ES 2 875 579 T3

ccgtatcatt gctccagggc gcgtgtaccg taacgattat gaccagacac atacaccgat 4560
 gtttcaccaa atggaagggg tgattgtgga tacgaatata tctttcacga atctgaaggg 4620
 caccttacct gatttcttac gcaacttttt cgaggaggac cttcaaattc gctttcgtcc 4680
 atcgtacttc ccttttgtag aaccttcggc tgaagtggat gtaatgggga aaaacggtaa 4740
 gtggctggag gttttagggt gcgggatggg tcatccaaat gtgcttcgca acgtcggcat 4800
 cgaccccgaa gtctacagtg gattcggatt cgggatggga atggaacgtc tgactatgct 4860
 tcgttacggc gtaacggatt tgcgtcctt ttttgagaac gatcttcggt tcttgaagca 4920
 attcaaataa gcatttttag tacgtgcaat aaccactctg gttttccag ggtggttttt 4980
 tgatgccctt tttggagtct tcaactgagc ctgcgagagc aggattcccg ttgagcaccg 5040
 ccaggtgcga ataagggaca gtgaagaagg aacaccgct cgcgggtggg cctacttcac 5100
 ctatcctgcc cggctgagc cgttggatac accaaggaaa gtctacacga accctttggc 5160
 aaaatcctgt atatcgtgcg aaaaaggatg gatataccga aaaaatcgct ataagacc 5220
 cgaagcaggg ttatgcagcg gaaaagctcc ccgaaaagt ccacctgggt ccttttcac 5280
 acgtgctata aaaataatta taatttaaat tttttaatat aaatatataa attaaaaata 5340
 gaaagtaaaa aaagaaatta aagaaaaaat agttttgtt ttccgaagat gtaaaagact 5400
 ctagggggat cgccaacaaa tactaccttt tatcttctc ttctctctct caggtattaa 5460
 tgccgaattg tttcatctt tctgtgtaga agaccacaca cgaaaatcct gtgattttac 5520
 attttactta tcgttaatcg aatgtatata tatttaatct gcttttcttg tctaataaat 5580
 atatatgtaa agtacgcttt ttgttgaat ttttaaacct tttgtttatt tttttttct 5640
 tcattccgta actctctac cttctttatt tactttctaa aatocaaata caaaacataa 5700
 aaataaataa acacagagta aattcccaa ttattccatc attaaaagat acgaggcgcg 5760
 tgtaagttac aggcaagcga tccgtctaag aaaccattat tatcatgaca ttaacctata 5820
 aaaataggcg tatcacgagg ccctttcgtc tcgcgcttt cggatgatgac ggtgaaaacc 5880
 tctgaacacat gcagctcccg gagacggtca cagcttctct gtaagcggat gccgggagca 5940
 gacaagcccc tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttaactatg 6000
 cggcatcaga gcagattgta ctgagagtgc accataccac cttttcaatt catcattttt 6060
 tttttattct ttttttgat ttcggtttct ttgacatttt tttgattcgg taatctccga 6120
 acagaaggaa gaacgaagga aggagcacag acttagattg gtatatatac gcatatgtag 6180
 tgttgaagaa acatgaaatt gccagatatt cttaacccaa ctgcacagaa caaaaacctg 6240
 caggaaacga agataaatca tgtcgaagc tacatataag gaacgtgctg ctactcatcc 6300
 tagtctggtt gctgccaagc ttttaatat catgcacgaa aagcaaaca acttgtgtgc 6360
 ttcattggat gttcgtacca ccaaggaatt actggagtta gttgaagcat taggtcccaa 6420

ES 2 875 579 T3

aatttgttta ctaaaaacac atgtggatat cttgactgat ttttccatgg agggcacagt 6480
 taagccgcta aaggcattat cgcceaagta caatttttta ctcttcgaag acagaaaatt 6540
 tgctgacatt ggttaatacag tcaaattgca gtactctgcg ggtgtataca gaatagcaga 6600
 atgggcagac attacgaatg cacacggtgt ggtgggcca ggtattgta gcggttgaa 6660
 gcaggcggca gaagaagtaa caaaggaacc tagaggcctt ttgatgtag cagaattgtc 6720
 atgcaagggc tccctatcta ctggagaata tactaagggt actggtgaca ttgcaagag 6780
 cgacaaagat tttgttatcg gctttattgc tcaaagagac atgggtggaa gagatgaagg 6840
 ttacgattgg ttgattatga caccggtgt ggttttagat gacaaggag acgattggg 6900
 tcaacagtat agaaccgtgg atgatgtggt ctctacagga tctgacatta ttattgttg 6960
 aagaggacta tttgcaaagg gaagggatgc taaggtagag ggtgaacgtt acagaaaagc 7020
 aggctgggaa gcatatttga gaagatgcgg ccagcaaac taaaaactg tattataagt 7080
 aatgcatgt atactaaact cacaattag agcttcaatt taattatatac agttattacc 7140
 cccctggctt gttgtccaca accgttaaac cttaaaagct ttaaagcct tatatattct 7200
 ttttttctt ataaaactta aaaccttaga ggctatntaa gttgctgatt tatattaatt 7260
 ttattgttca aacatgagag cttagtacgt gaaacatgag agcttagtac gttagccatg 7320
 aggttttagt tcgttagcca tgagggttta gttcgttaa catgagagct tagtacgtta 7380
 aacatgagag cttagtacgt gaaacatgag agcttagtac gtactatcaa caggttgaac 7440
 tgctgatctt cggcatcaaa taaaacgaaa ggctcagtcg aaagactggg cctttcgttt 7500
 tatctgttgt ttgtcgggta acgctctcct gagtaggaca aatccgccgc cctagatatt 7560
 cacacgcaat caacagggcag gataatcgct ggtaagggtca gtgctttctt caggtagtag 7620
 agatacaata gttccaacg ataggtggca gatttcaact tacagaccga ctggttcaga 7680
 agcgtagata atagcccgtg tttccaata agggatagtg taggtaagtc aactcctccg 7740
 tcagagccaa ccgttt 7756

<210> 217

<211> 8067

5 <212> ADN

<213> Secuencia artificial

<220>

<223> cadena principal de vector 4

10

<400> 217

gtttaaacia ttttgtgctg cccttaattg tgagcgtca caattccaca acggtttccc 60
 tctagaata attttgttta acttttcgag accttaggag gtaaacatat ggcgagcttg 120
 gttaagaaag atatgtgtat taagatgacg atggaaggta ctgtgaacgg ttatcacttt 180

ES 2 875 579 T3

aagtgcgttg gcgaggggtga aggcaagccg ttcgagggca cgcagaacat gcgcactcgt 240
gtcacccgagg gcgggccgct gccttttgca ttcgacatcc tggccccgtg ctgtatgtac 300
ggctctaaga ccttcattaa acacgtgagc ggtatcccgg attactttaa agagtccttt 360
ccagagggct tcacttggga acgtaccag atttttgagg acggtggtgt tctgaccgcg 420
caccaagaca ccagcctgga aggtaattgc ctgatctata aagtgaaggt tctgggtacc 480
aatttcccgg cgaatggtcc ggtgatgcaa aagaaaaccg cgggttggga gccgtgcgtc 540
gagatgctgt atccgcgtga cggcgtcttg tgtggtcaga gcttgatggc gctgaagtgc 600
accgatggca atcatctgac cagccacctg cgcacgacgt atcgtagccg taaaccgagc 660
aacgccgtta acatgccgga gttccatfff ggtgaccatc gcatcgaat cctgaaagct 720
gagcagggca aattctacga acaatacгаа tcggctgtcg cacgttacag cgatgtgccg 780
gaaaaagcga cgtaatgaga attctgtaca ctcgagggtc tcacccaag gccgacaccc 840
cctaattagc ccgggcgaaa ggcccagtct ttcgactgag cctttcgttt tatttgagtt 900
taaacctggg gcgccctctg gtaaggttgg gaagccctgc aaagtaaact ggatgccagg 960
catcaataa aacgaaaggc tcagtcgaaa gactgggcct ttcgttttat ctgttgttg 1020
tcggtgaacg ctcttacta gagtcacact ggctcacctt cgggtgggcc tttctgcgtt 1080
tatagcacta ggacttgccg cggatactgc cccattacat gaattgcagc ctcagggacg 1140
tcagtagatc atggaggtag ggcataatgc ctctgttgtt aaaatgtgag ttctcaacga 1200
agcacgaatc ggtcagaacc tacactaagg agatttgta ggtgcacggt ttctgtcgca 1260
tagaccagtt catttcagat gtctggcacg taagaggttc caactttcac cataatgaaa 1320
taagatcact accgggcgta ttttttgagt tatcgagatt ttcaggagct aaggaagcta 1380
aatgattga acaagatgga ttgcacgagc gttctccggc cgcttgggtg gagaggctat 1440
tcggctatga ctgggcacaa cagacaatcg gctgctctga tgcgcgctg ttcggctgt 1500
cagcgcaggg gcgcccggtt ctttttgca agaccgacct gtccggtgcc ctgaatgaa 1560
tccaagacga ggcagcggc ctatcgtggc tggccacgac gggcgttcct tgcgcagctg 1620
tgctcgacgt tgtcactgaa gcgggaaggg actggctgct attgggcgaa gtgccggggc 1680
aggatctcct gtcactcac cttgctcctg ccgagaaagt atccatcatg gctgatgcaa 1740
tgccgggct gcatacgtt gatccggcta cctgccatt cgaccacaa gcgaaacatc 1800
gcacgcagc agcacgtact cggatggaag ccggtcttgt cgatcaggat gatctggacg 1860
aagagcatca ggggctcgcg ccagccgaac tgttcgccag gctcaaggcg cggatgcccg 1920
acggcgagga tctcgtcgtg acccatggcg atgcctgctt gccgaatatc atggtggaaa 1980
atggccgctt ttctggatc atcactgtg gccggctggg tgtggcggac cgetatcagg 2040
acatagcgtt ggctaccctg gatattgctg aagagcttgg cggcgaatgg gctgaccgct 2100

ES 2 875 579 T3

tcctcgtgct ttacgggatc gccgctcccg attcgcagcg catcgccttc tatcgccttc	2160
ttgacgagtt cttctgatga tgaatggcaa ggcggcgcgt agcccccaa ccgaagttga	2220
ggggatTTTT ttgacaatta atcatccggc tcgtaattta tgtggatcct aatcatgcta	2280
aggaggtttt ctaatgatga acatcaaaaa gtttgcaaaa caagcaacag tattaacctt	2340
tactaccgca ctgctggcag gaggcgcaac tcaagcgttt gcgaaagaaa cgaaccaaaa	2400
gccatataag gaaacatacg gcatttccca tattacacgc catgatatgc tgcaaatccc	2460
tgaacagcaa aaaaatgaaa aatatcaagt ttctgaattt gattcgtcca caattaaaaa	2520
tatctcttct gcaaaaggcc tggacgtttg ggacagctgg ccattacaaa acgctgacgg	2580
cactgtcgca aactatcacg gctaccacat cgtctttgca ttagccggag atcctaataa	2640
tgcggatgac acatcgattt acatgttcta tcaaaaagtc ggcgaaactt ctattgacag	2700
ctggaaaaac gctggccgcg tctttaaaga cagcgacaaa ttcgatgcaa atgattctat	2760
cctaaaagac caaacacaag aatggtcagc ttcagccaca tttacatctg acggaaaaat	2820
ccgtttattc tacactgatt tctccggtaa acattacggc aaacaaacac tgacaactgc	2880
acaagttaac gtatcagcat cagacagctc tttgaacatc aacggtgtag aggattataa	2940
atcaatcttt gacggtgacg gaaaaacgta tcaaaatgta cagcagttca tcgatgaagg	3000
caactacagc tcaggcgaca accatacgcct gagagatcct cactacgtag aagataaagg	3060
ccacaatac ttagtatttg aagcaaacac tggaactgaa gatggctacc aaggcgaaga	3120
atctttattt aacaaagcat actatggcaa aagcacatca ttcttccgtc aagaaagtca	3180
aaaacttctg caaagcgata aaaaacgcac ggctgagtta gcaaacggcg ctctcggtat	3240
gattgagcta aacgatgatt acacactgaa aaaagtgatg aaaccgctga ttgcatctaa	3300
cacagtaaca gatgaaattg aacgcgcgaa cgtctttaa atgaacggca aatggtacct	3360
gttactgac tcccgcggat caaaaatgac gattgacggc attacgtcta acgatattta	3420
catgcttggc tatgtttcta attctttaac tggcccatac aagccgctga acaaaactgg	3480
ccttgtgta aaaatggatc ttgatcctaa cgatgtaacc tttacttact cacacttgc	3540
tgtacctcaa gcgaaaggaa acaatgtcgt gattacaagc tatatgaaa acagaggatt	3600
ctacgcagac aaacaatcaa cgtttgcgcc gagcttctctg ctgaacatca aaggcaagaa	3660
aacatctggt gtcaaagaca gcatccttga acaaggacaa ttaacagtta acaataaaa	3720
acgcaaaaga aaatgccgat gggtaaccgag cgaaatgacc gaccaagcga cgcccaacct	3780
gccatcacga gatttcgatt ccaccgccgc cttctatgaa aggttgggct tcggaatcgt	3840
tttccgggac gccctcgcgg acgtgctcat agtcaggcg gtgttgacat aaataccact	3900
ggcggtgata ctgagcacat cagcaggtca cacaggaaag tactagatgt cgcaccttgc	3960

ES 2 875 579 T3

agaattagta gcttcagcga aggccgcgat ttctcaggcg agtgacgtcg cagcactgga 4020
 taatgtacgt gttgagtacc tgggaaagaa gggacacctt actcttcaaa tgacaaccct 4080
 ggcggaactg ccgccggagg aacgccccgc agcaggagcg gtaatcaatg aggcaaagga 4140
 gcaagtacaa caggcactga acgcccgtaa ggctgagttg gaatccgccg cattaaacgc 4200
 gcgcccttgct gcggaaacca ttgatgtctc gctgccgggg cgccgcattg agaatggagg 4260
 cttacacca gtgactcgtg ccatcgaccg tatcgaatct ttctttggcg aacttggctt 4320
 cactgtggca actggaccgg agattgagga cgactaccac aatttcgatg cettgaacat 4380
 tcccggctcat catcctgcac gcgccgatca tgatacattc tggtttgata ccaccgttt 4440
 gcttcgtacc cagacaagcg gtgtcctaat ccgtacgatg aaggctcagc aaccaccgat 4500
 ccgtatcatt gctccagggc gcgtgtaccg taacgattat gaccagacac atacaccgat 4560
 gtttcaccaa atggaagggt tgattgtgga tacgaatctc tctttcacga atctgaaggg 4620
 caccttacat gatttcttac gcaacttttt cgaggaggac cttcaaattc gctttcgtcc 4680
 atcgtacttc ccttttgagc aaccttcggc tgaagtggat gtaatgggga aaaacggtaa 4740
 gtggctggag gttttagggt gcgggatggt tcatcctaat gtgcttcgca acgtcggcat 4800
 cgaccccgaa gtctacagtg gattcggatt cgggatggga atggaacgtc tgactatgct 4860
 tcggtacggc gtaacggatt tgcgctcctt ttttgagaac gatcttcggt ttctgaagca 4920
 attcaataa gcatttttag tacgtgcaat aaccactctg gttttccag ggtggttttt 4980
 tgatgccctt tttggagtct tcaactgagc ctgcgagagc aggattcccg ttgagcaccg 5040
 ccagtgcgga ataagggaca gtgaagaagg aacaccgctc cgccgggtggg cctacttcac 5100
 ctatcctgcc cggctgacgc cgttgatac accaaggaaa gtctacacga accctttggc 5160
 aaaatcctgt atatcgtgcg aaaaaggatg gatataccga aaaaatcgtc ataatagccc 5220
 cgaagcaggg ttatgcagcg gaaaagctcc ccgaaaagtg ccacctgggt ccttttcac 5280
 acgtgctata aaaataatta taatttaaat tttttaatat aaatatataa attaaaaata 5340
 gaaagtaaaa aaagaaatta aagaaaaaat agtttttgtt ttccgaagat gtaaaagact 5400
 ctagggggat cgccaacaaa tactaccttt tatcttgctc ttctgctct caggtattaa 5460
 tgccgaattg tttcatcttg tctgtgtaga agaccacaca cgaaaatcct gtgattttac 5520
 attttactta tcgttaatcg aatgtatata tatttaatct gcttttcttg tctaataaat 5580
 atatatgtaa agtacgcttt ttggtgaaat tttttaaac tttgtttatt tttttttct 5640
 tcattccgta actctcttac cttctttatt tactttctaa aatccaaata caaaacataa 5700
 aaataaataa acacagagta aattcccaaa ttattccatc attaaaagat acgaggcgcg 5760
 tgtaagttac aggcaagcga tccgtctaag aaaccattat tatcatgaca ttaacctata 5820
 aaaataggcg tatcacgagg ccctttcgtc tcgcgcgctt cggatgatgac ggtgaaaacc 5880

ES 2 875 579 T3

tctgacacat gcagctcccg gagacggtca cagcttgtct gtaagcggat gccgggagca 5940
gacaagcccc tcagggcgcg tcagcgggtg ttggcgggtg tcggggctgg cttactatg 6000
cggcatcaga gcagattgta ctgagagtgc accataccac agccggaaga ggagtaggga 6060
atattactgg ctgaaaataa gtcttgaatg aacgtatagc cgtatatttc taccaatctc 6120
tcaacactga gtaatgtag ttataagaaa gagaccgagt tagggacagt tagaggcggc 6180
ggagatattc cttatggcat gtctggcgat gataaaactt ttcaaacggc agccccgatc 6240
taaaagagct gacagggaaa tggtcagaaa aagaaacgtg caccgcgccg tctggacgcg 6300
ccgctcaccg gcacggcaga gaccaatcag taaaaatcaa cggttaacga cttactata 6360
tatataatat aggaagcatt taatagaaca gcatcgtaat atatgtgtac tttgcagtta 6420
tgacgccaga tggcagtagt ggaagatatt ctttattgaa aaatagcttg tcaccttagc 6480
tacaatcttg atccggagct tttctttttt tgccgattaa gaattcggtc gaaaaaagaa 6540
aaggagaggg ccaagagggg gggcattggt gactattgag cacgtgagta tacgtgatta 6600
agcacacaaa ggcagcttgg agtatgtctg ttattaattt cacaggtagt tctggtccat 6660
tggtgaaagt ttgccccttg cagagcacag aggccgcaga atgtgctcta gattccgatg 6720
ctgacttgct gggattata tgtgtgccca atagaaagag aacaattgac ccggttattg 6780
caaggaaaat ttcaagtctt gtaaaagcat ataaaaatag ttcaggcact ccgaaatact 6840
tggttggcgt gtttcgtaat caacctaagg aggatgtttt ggctctggtc aatgattacg 6900
gcattgatat cgtccaactg catggagatg agtcgtggca agaataccaa gagttcctcg 6960
gtttgccagt tattaanaaga ctctgttttc caaaagactg caacatacta ctcagtgcag 7020
cttcacagaa acctcattcg tttattccct tgtttgattc agaagcaggt gggacaggtg 7080
aacttttggg ttggaactcg atttctgact gggttggaag gcaagagagc cccgaaagct 7140
tacatthtat gttagctggt ggactgacgc cagaaaatgt tggatgatgcg cttagattaa 7200
atggcgttat tgggtttgat gtaagcggag gtgtggagac aaatggtgta aaagactcta 7260
acaaaatagc aaatttcgtc aaaaatgcta agaaataggt tattactgag tagtatttat 7320
ttaagtattg tttgtgcaact tgcctgcagg ccttttgaaa agcaagcata aaagatctaa 7380
acataaaatc tgtaaaataa caagatgtaa agataatgct aaatcatttg gctttttgat 7440
tgattgtaca ggccttgct tgggtgccac aaccgttaa ccttaaaagc tttaaaagcc 7500
ttatatattc ttttttttct tataaaactt aaaaccttag aggctattta agttgctgat 7560
ttatattaat tttattgttc aaacatgaga gcttagtacg tgaacatga gagcttagta 7620
cgtagccat gagggtttag ttcgtagcc atgagggttt agttcgtaa acatgagagc 7680
ttagtacggt aaacatgaga gcttagtacg tgaacatga gagcttagta cgtactatca 7740

ES 2 875 579 T3

acaggttgaa ctgctgatct tcggcatcaa ataaaacgaa aggctcagtc gaaagactgg 7800
gcctttcggtt ttatctggtg tttgtcgggtg aacgctctcc tgagtaggac aaatccgccg 7860
ccctagatat tcacacgcaa tcaacaggca ggataatcgc tggttaaggtc agtgctttct 7920
tcaggtagta gagatacaat agttcccaac gataggtggc agatttcact ttacagaccg 7980
actggttcag aagcgtagat aatagcccgt gttttccaat aagggatagt gtaggtaagt 8040
caactcctcc gtcagagcca accgttt 8067

<210> 218

<211> 200

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Elemento de vector aislante 1

10

<400> 218

tattcacacg caatcaacag gcaggataat cgctggtaag gtcagtgcctt tcttcaggtta 60
gtagagatac aatagttccc aacgataggt ggcagatttc actttacaga ccgactgggtt 120
cagaagcgta gataatagcc cgtgttttcc aataagggat agtgtaggtta agtcaactcc 180
tccgtcagag ccaaccgttt 200

<210> 219

15 <211> 200

<212> ADN

<213> Secuencia artificial

<220>

20 <223> Elemento de vector aislante 2

<400> 219

gcactaggac ttgccgcgga tactgcccga ttacatgaat tgcagcctca gggacgtcag 60
tagatcatgg aggtagggca tatgtcctct gttgttaaaa tgtgagttct caacgaagca 120
cgaatcggtc agaacctaca ctaaggagat ttggtaggtg cacggtttct gtcgcataga 180
ccagttcatt tcagatgtct 200

25 <210> 220

ES 2 875 579 T3

<211> 105

<212> ADN

<213> Secuencia artificial

5 <220>

<223> Elemento de vector T1

<400> 220

ggcatcaaat aaaacgaaag gctcagtcga aagactgggc ctttcgtttt atctgttggt 60

tgtcggtgaa cgctctcctg agtaggacaa atccgccgcc ctaga 105

10

<210> 221

<211> 129

<212> ADN

<213> Secuencia artificial

15

<220>

<223> Elemento de vector B0015

<400> 221

ccaggcatca aataaaacga aaggctcagt cgaaagactg ggcctttcgt tttatctggt 60

gtttgtcggg gaacgctctc tactagagtc aactcgctc accttcgggt gggcctttct 120

20

gcgtttata 129

<210> 222

<211> 66

<212> ADN

25 <213> Secuencia artificial

<220>

<223> promotor de SacB

30 <400> 222

ttgacaatta atcatccggc tcgtaattta tgtggatctt aatcatgcta aggaggtttt 60

ctaatg 66

ES 2 875 579 T3

<210> 223

<211> 1054

<212> ADN

<213> Secuencia artificial

5

<220>

<223> promotor y secuencia de PheS

<400> 223

```

ggcgggtggtg acataaatac cactggcgggt gatactgagc acatcagcag gtcacacagg      60
aaagtactag atgtcgcatc ttgcagaatt agtagcttca gcgaaggccg cgatttctca      120
ggcgagtgac gtcgcagcac tggataatgt acgtgttgag tacctgggaa agaagggaca      180
ccttactctt caaatgacaa ccctgcgcga actgccgccg gaggaacgcc cgcgagcagg      240
agcggtaatc aatgaggcaa aggagcaagt acaacaggca ctgaacgccc gtaaggctga      300
gttggaatcc gccgcattaa acgcgcgcct tgctgcggaa accattgatg tctcgtgccc      360
cgggcgccgc attgagaatg gaggcttaca cccagtgact cgtaccatcg accgtatcga      420
atctttcttt ggcgaacttg gcttcaactg ggcaactgga ccggagattg aggacgacta      480
ccacaatttc gatgccttga acattcccgg tcatcatcct gcacgcgccg atcatgatac      540
attctggttt gataccaccg gtttgcttcg taccagaca agcgggtgtcc aaatccgtac      600
gatgaaggct cagcaaccac cgatccgatc cattgctcca gggcgcgtgt accgtaacga      660
10 ttatgaccag acacatacac cgatgtttca ccaaattgaa gggttgattg tggatacгаа      720

tatctctttc acgaatctga agggcacctt acatgatttc ttacgcaact ttttcgagga      780
ggaccttcaa attcgctttc gtccatcgta cttccctttt gcagaacctt cggctgaagt      840
ggatgtaatg gggaaaaacg gtaagtggct ggaggtttta ggttgcggga tggttcatcc      900
aatgtgctt cgcaacgtcg gcatcgacct cgaagtctac agtggattcg gattcgggat      960
gggaatggaa cgtctgacta tgcttcgtta cggcgtaacg gatttgcgct ccttttttga     1020
gaacgatctt cgttttctga agcaattcaa ataa                                     1054

```

<210> 224

15 <211> 77

<212> ADN

<213> Corynebacterium glutamicum

<400> 224

ES 2 875 579 T3

gcatttttag tacgtgcaat aaccactctg gttttccag ggtggttttt tgatgccctt 60
tttgagtct tcaactg 77

<210> 225

<211> 90

5 <212> ADN

<213> Escherichia coli

<400> 225

ttcagccaaa aaacttaaga cgcgctct tgtccactac cttgcagtaa tgcggtggac 60
aggatcggcg gttttctttt ctcttctcaa 90

10

<210> 226

<211> 52

<212> ADN

<213> Escherichia coli

15

<400> 226

gacgaacaat aaggcctccc aaatcggggg gccctttta ttgataacaa aa 52

<210> 227

20 <211> 42

<212> ADN

<213> Escherichia coli

<400> 227

25 gtaacaacgg aaaccggcca ttgcgccggt tttttggc ct 42

<210> 228

<211> 41

<212> ADN

30 <213> Escherichia coli

<400> 228

aagcagagaa ccctggatga gagtccgggg tttttgttt t 41

ES 2 875 579 T3

<210> 229
 <211> 71
 <212> ADN
 <213> Escherichia coli
 5
 <400> 229
 taaagtatgt atccccaaaa taattcgagt cattgcatct gtggctagaa gtatgaaggg 60
 attaaccata a 71

 <210> 230
 10 <211> 57
 <212> ADN
 <213> Escherichia coli

 <400> 230
 15 ggaaacacag aaaaaagccc gcacctgaca gtgcgggctt tttttcga ccaaagg 57

 <210> 231
 <211> 168
 <212> ADN
 20 <213> Streptococcus sp.

 <400> 231
 atgacgtata agttgatttt aatggaag accctgaaag gggagacgac cactgaagct 60
 gtagatgagg cgactgctga aaaagtcttt aagcaatatg caaacgacaa cggagttgat 120
 ggtgaatgga cttatgatga cgccactaaa acgttcacag taaccgaa 168

 25 <210> 232
 <211> 207
 <212> ADN
 <213> Fasciola hepatica

 30 <400> 232

ES 2 875 579 T3

	atgccaagcg tacaggaagt ggagaagtta cttcatgtgt tggatcggaa cggagacgga	60
	aaagtttcag ccgaagagtt aaaggccttt gcggacgact ctaagtgcc tctggattcc	120
	aacaagataa aagcgtttat taaggagcat gataagaaca aggatggtaa acttgacctg	180
	aaggagctgg tctccatact gtcgagc	207
	<210> 233	
	<211> 228	
5	<212> ADN	
	<213> Desconocido	
	<220>	
	<223> Ubiquitina marcador de solubilidad (PST3)	
10		
	<400> 233	
	atgcagatct ttgtaagac tttaacagga aaaacgataa cgctggaagt tgagtctagc	60
	gacacaatag acaatgtcaa atcaaagatt caagataaag agggatatacc accggaccag	120
	caacggttga ttttcgcagg taagcaatta gaggatggga gaacactgag tgactacaat	180
	atccagaaag aatcgacggt gcatttggtg cttagattgc gtggaggt	228
	<210> 234	
15	<211> 303	
	<212> ADN	
	<213> Homo sapiens	
	<400> 234	
	atgagtgatt ctgaggttaa tcaagaggcc aaaccggaag taaaaccaga agtgaagcct	60
	gaaacgcata tcaatttgaa agtctcggac ggctcgtctg aatatattctt caagataaaa	120
	aagacgacac cacttcggcg tctgatggag gcgtttgcc aaagacaagg aaaagaaatg	180
	gactcgcttc gtttcttata cgacggtatc agaatacaag cggaccagac tccagaagat	240
	cttgatatgg aggataatga cattatcga gcccatcgcg agcagatagg tggcgcgaca	300
20	tac	303
	<210> 235	
	<211> 56	
	<212> PRT	
25	<213> Streptococcus sp.	

ES 2 875 579 T3

<400> 235

Met Thr Tyr Lys Leu Ile Leu Asn Gly Lys Thr Leu Lys Gly Glu Thr
1 5 10 15

Thr Thr Glu Ala Val Asp Ala Ala Thr Ala Glu Lys Val Phe Lys Gln
20 25 30

Tyr Ala Asn Asp Asn Gly Val Asp Gly Glu Trp Thr Tyr Asp Asp Ala
35 40 45

Thr Lys Thr Phe Thr Val Thr Glu
50 55

5 <210> 236

<211> 69

<212> PRT

<213> Fasciola hepatica

10 <400> 236

Met Pro Ser Val Gln Glu Val Glu Lys Leu Leu His Val Leu Asp Arg
1 5 10 15

Asn Gly Asp Gly Lys Val Ser Ala Glu Glu Leu Lys Ala Phe Ala Asp
20 25 30

Asp Ser Lys Cys Pro Leu Asp Ser Asn Lys Ile Lys Ala Phe Ile Lys
35 40 45

Glu His Asp Lys Asn Lys Asp Gly Lys Leu Asp Leu Lys Glu Leu Val
50 55 60

Ser Ile Leu Ser Ser
65

15 <210> 237

<211> 76

<212> PRT

<213> Desconocido

20 <220>

<223> Ubiquitina marcador de solubilidad (PST3)

<400> 237

ES 2 875 579 T3

Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15

Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ser Lys Ile Gln Asp
20 25 30

Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45

Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60

Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly
65 70 75

<210> 238

<211> 101

5 <212> PRT

<213> Homo sapiens

<400> 238

Met Ser Asp Ser Glu Val Asn Gln Glu Ala Lys Pro Glu Val Lys Pro
1 5 10 15

Glu Val Lys Pro Glu Thr His Ile Asn Leu Lys Val Ser Asp Gly Ser
20 25 30

Ser Glu Ile Phe Phe Lys Ile Lys Lys Thr Thr Pro Leu Arg Arg Leu
35 40 45

Met Glu Ala Phe Ala Lys Arg Gln Gly Lys Glu Met Asp Ser Leu Arg
50 55 60

Phe Leu Tyr Asp Gly Ile Arg Ile Gln Ala Asp Gln Thr Pro Glu Asp
65 70 75 80

10

Leu Asp Met Glu Asp Asn Asp Ile Ile Glu Ala His Arg Glu Gln Ile
85 90 95

Gly Gly Ala Thr Tyr
100

<210> 239

<211> 36

15 <212> ADN

<213> Escherichia coli

ES 2 875 579 T3

<400> 239

gcagcaaagc acgagaatta tgccttgccg gcttaa 36

<210> 240

5 <211> 36

<212> ADN

<213> Escherichia coli

<400> 240

10 gctgcaaacg atgagaatta tgccttgcta gcataa 36

<210> 241

<211> 36

<212> ADN

15 <213> Escherichia coli

<400> 241

gctgctaagc atgagaacta tgcagccgag gtataa 36

20 <210> 242

<211> 36

<212> ADN

<213> Escherichia coli

25 <400> 242

gccgccaatg atgagaatta tgcagcgagc gtgtaa 36

<210> 243

<211> 30

30 <212> ADN

<213> Escherichia coli

<400> 243

cgcagtgagc agatttcgat ggagttttaa 30

35

<210> 244

<211> 18

ES 2 875 579 T3

<212> ADN

<213> Escherichia coli

<400> 244

5 tcattgttat ggtcctaa 18

<210> 245

<211> 63

<212> ADN

10 <213> Escherichia coli

<400> 245

gcttcctcgc atgcgacccg tcaactgtca ggtctgaaaa ttcactcaaa cctttaccat 60

taa 63

15 <210> 246

<211> 66

<212> ADN

<213> Escherichia coli

20 <400> 246

atgggacagc tgcgctcatt aaatggggag tggcgcttcg cgtggttccc agctcccga 60

gctgtg 66

<210> 247

<211> 11

25 <212> PRT

<213> Escherichia coli

<400> 247

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Ala Ala
1 5 10

30

<210> 248

<211> 11

<212> PRT

<213> Escherichia coli

<400> 248

Ala Ala Asn Asp Glu Asn Tyr Ala Leu Val Ala
1 5 10

5 <210> 249

<211> 11

<212> PRT

<213> Escherichia coli

10 <400> 249

Ala Ala Asn Asp Glu Asn Tyr Ala Ala Ala Val
1 5 10

<210> 250

<211> 11

15 <212> PRT

<213> Escherichia coli

<400> 250

Ala Ala Asn Asp Glu Asn Tyr Ala Ala Ser Val
1 5 10

20

<210> 251

<211> 9

<212> PRT

<213> Escherichia coli

25

<400> 251

Arg Ser Glu Gln Ile Ser Met Glu Phe
1 5

<210> 252

30 <211> 5

<212> PRT

<213> Escherichia coli

<400> 252

ES 2 875 579 T3

Ser Leu Leu Trp Ser
1 5

<210> 253

<211> 20

5 <212> PRT

<213> Escherichia coli

<400> 253

Ala Ser Ser His Ala Thr Arg Gln Leu Ser Gly Leu Lys Ile His Ser
1 5 10 15

Asn Leu Tyr His
20

10

<210> 254

<211> 22

<212> PRT

<213> Escherichia coli

15

<400> 254

Met Gly Gln Leu Arg Ser Leu Asn Gly Glu Trp Arg Phe Ala Trp Phe
1 5 10 15

Pro Ala Pro Glu Ala Val
20

<210> 255

20 <211> 133

<212> ADN

<213> Secuencia artificial

<220>

25 <223> P3_BCD1

<400> 255

aaaaaattta ttgcttatt aattcatccg gctcgtataa tgtgtggagg gcccaagttc 60
acttaaaaag gagatcaaca atgaaagcaa ttttcgtact gaaacatctt aatcatgcac 120
aggagacttt cta 133

ES 2 875 579 T3

<210> 256

<211> 121

<212> ADN

<213> Secuencia artificial

5

<220>

<223> P4_BCD22

<400> 256

ttgacatcag gaaaatTTTT ctgtataatg tgtggagggc ccaagttcac ttaaaaagga 60

gatcaacaat gaaagcaatt ttcgtactga aacatcttaa tcatgcctag gaagtttct 120

10 a 121

<210> 257

<211> 132

<212> ADN

15 <213> Secuencia artificial

<220>

<223> P7_BCD19

20 <400> 257

aaaaaattta ttgctttcg catcttttg tacctataat gtgtggaggg cccaagttca 60

cttaaaaagg agatcaacaa tgaaagcaat ttcgtactg aaacatctta atcatgctat 120

ggaggttttc ta 132

<210> 258

<211> 121

25 <212> ADN

<213> Secuencia artificial

<220>

<223> P8_BCD15

30

<400> 258

ES 2 875 579 T3

ttcactttta atcatccggc tegtataatg tgtggagggc ccaagttcac ttaaaaagga 60
gatcaacaat gaaagcaatt ttcgtactga aacatcttaa tcatgcgggg gagtctttct 120
a 121

<210> 259

<211> 121

5 <212> ADN

<213> Secuencia artificial

<220>

<223> P11_BCD17

10

<400> 259

ttgacaatta atcatccggc tcttagtggt tgtggagggc ccaagttcac ttaaaaagga 60
gatcaacaat gaaagcaatt ttcgtactga aacatcttaa tcatgcgggg gagggtttct 120
a 121

<210> 260

15 <211> 121

<212> ADN

<213> Secuencia artificial

<220>

20 <223> P13_BCD8

<400> 260

ttccctatta atcatccggc tegtataatg tgtggagggc ccaagttcac ttaaaaagga 60
gatcaacaat gaaagcaatt ttcgtactga aacatcttaa tcatgcatcg gaccgtttct 120
a 121

25

REIVINDICACIONES

1. Un método para generar una biblioteca de cepas de *E. coli* de intercambio de promotores, que comprende las etapas de:

5 a. proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*, en donde al menos uno de la pluralidad de promotores es un promotor quimérico sintético, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que son cada uno de seis nucleótidos de longitud, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) que es la parte 5' UTR/RBS del promotor p_R del fago lambda o la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli*; y

10 b. transformar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los promotores de la escalera de promotores operativamente unidos a uno de los genes diana endógenos para la cepa base de *E. coli*.

2. El método para generar una biblioteca de cepas de *E. coli* de intercambio de promotores según la reivindicación 1, en donde al menos uno de la pluralidad de promotores comprende una secuencia reguladora de diseño bicistrónico (BCD).

20 3. Un método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción, que comprende las etapas de:

25 a. proporcionar una pluralidad de genes diana endógenos para una cepa base de *E. coli*, y una escalera de promotores, en donde dicha escalera de promotores comprende una pluralidad de promotores que presentan diferentes perfiles de expresión en la cepa base de *E. coli*, en donde al menos uno de la pluralidad de promotores es un promotor quimérico sintético, en donde el promotor quimérico sintético es de 60-90 nucleótidos de longitud y consiste en una parte distal del promotor p_R del fago lambda, regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda que son cada uno de seis nucleótidos de longitud, partes centrales de los promotores p_L y p_R del fago lambda y una parte 5' UTR/sitio de unión al ribosoma (RBS) que es la parte 5' UTR/RBS del promotor p_R del fago lambda o la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli*;

30 b. transformar el genoma de la cepa base de *E. coli*, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores inicial que comprende una pluralidad de cepas de *E. coli* individuales con variaciones genéticas únicas encontradas dentro de cada cepa de dicha pluralidad de cepas de *E. coli* individuales, en donde cada una de dichas variaciones genéticas únicas comprende uno o más de los promotores de la escalera de promotores operativamente unidos a uno de los genes diana endógenos para la cepa base de *E. coli*;

35 c. cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores inicial según las mejoras del rendimiento fenotípico frente a una cepa de *E. coli* de referencia, identificando así variaciones genéticas únicas que confieren mejoras del rendimiento fenotípico;

40 d. proporcionar una pluralidad subsiguiente de microbios de *E. coli* que comprenden cada uno una combinación de variaciones genéticas únicas de las variaciones genéticas presentes en al menos dos cepas de *E. coli* individuales cribadas en la etapa precedente, para crear así una biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente;

45 e. cribar y seleccionar cepas de *E. coli* individuales de la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente según las mejoras del rendimiento fenotípico frente la cepa de *E. coli* de referencia, identificando así combinaciones únicas de variación genética que confieren mejoras de rendimiento fenotípico adicionales; y

50 f. repetir las etapas d)-e) una o más veces, de forma lineal o no lineal, hasta que una cepa de *E. coli* presente un nivel deseado de rendimiento fenotípico mejorado en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción, en donde cada iteración subsiguiente crea una nueva biblioteca de cepas microbianas de *E. coli* de intercambio de promotores, donde cada cepa de la nueva biblioteca comprende variaciones genéticas que son una combinación de variaciones genéticas seleccionadas entre al menos dos cepas de *E. coli* individuales de una biblioteca precedente.

4. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente es una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de promotores inicial.

5. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente es un subconjunto de una biblioteca combinatoria completa de la biblioteca de cepas de *E. coli* de intercambio de promotores inicial.
- 5 6. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente es una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de promotores precedente.
7. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde la biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente es un subconjunto de una biblioteca combinatoria completa de una biblioteca de cepas de *E. coli* de intercambio de promotores precedente.
- 10 8. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según una cualquiera de las reivindicaciones 3-7, en donde las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente presenta al menos un 10% de aumento en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.
- 15 9. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según una cualquiera de las reivindicaciones 3-7, en donde las etapas d)-e) se repiten hasta que el rendimiento fenotípico de una cepa de *E. coli* de una biblioteca de cepas de *E. coli* de intercambio de promotores subsiguiente presenta al menos un aumento de una vez en una variable fenotípica medida en comparación con el rendimiento fenotípico de la cepa de *E. coli* de producción.
- 20 10. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde el rendimiento fenotípico mejorado de la etapa f) se selecciona del grupo que consiste en: productividad volumétrica de un producto de interés, productividad específica de un producto de interés, rendimiento de un producto de interés, título de un producto de interés y combinaciones de los mismos.
- 25 11. El método para mejorar el rendimiento fenotípico de una cepa de *E. coli* de producción según la reivindicación 3, en donde el rendimiento fenotípico mejorado de la etapa f) es: producción mayor o más eficiente de un producto de interés, dicho producto de interés seleccionado del grupo que consiste en: una molécula pequeña, enzima, péptido, aminoácido, ácido orgánico, compuesto sintético, combustible, alcohol, metabolito extracelular primario, metabolito extracelular secundario, molécula de componente intracelular y combinaciones de los mismos.
- 30 12. El método de la reivindicación 1 o 3, en donde la parte 5' UTR/sitio de unión al ribosoma (RBS) es la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda y las secuencias de ácido nucleico de la parte distal del promotor p_R del fago lambda, las regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda, las partes centrales de los promotores p_L y p_R del fago lambda y la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor p_R del fago lambda se seleccionan de las secuencias de ácido nucleico que se encuentran en la Tabla 1.5.
- 35 13. El método de la reivindicación 1 o 3, en donde la parte 5' UTR/sitio de unión al ribosoma (RBS) es la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli* y las secuencias de ácido nucleico de la parte distal del promotor p_R del fago lambda, las regiones variables -35 y -10 de los promotores p_L y p_R del fago lambda, las partes centrales de los promotores p_L y p_R del fago lambda y la parte 5' UTR/sitio de unión al ribosoma (RBS) del promotor del gen *acs* de *E. coli* se seleccionan de las secuencias de ácido nucleico que se encuentran en la Tabla 1.5.
- 40

FIGURA 1

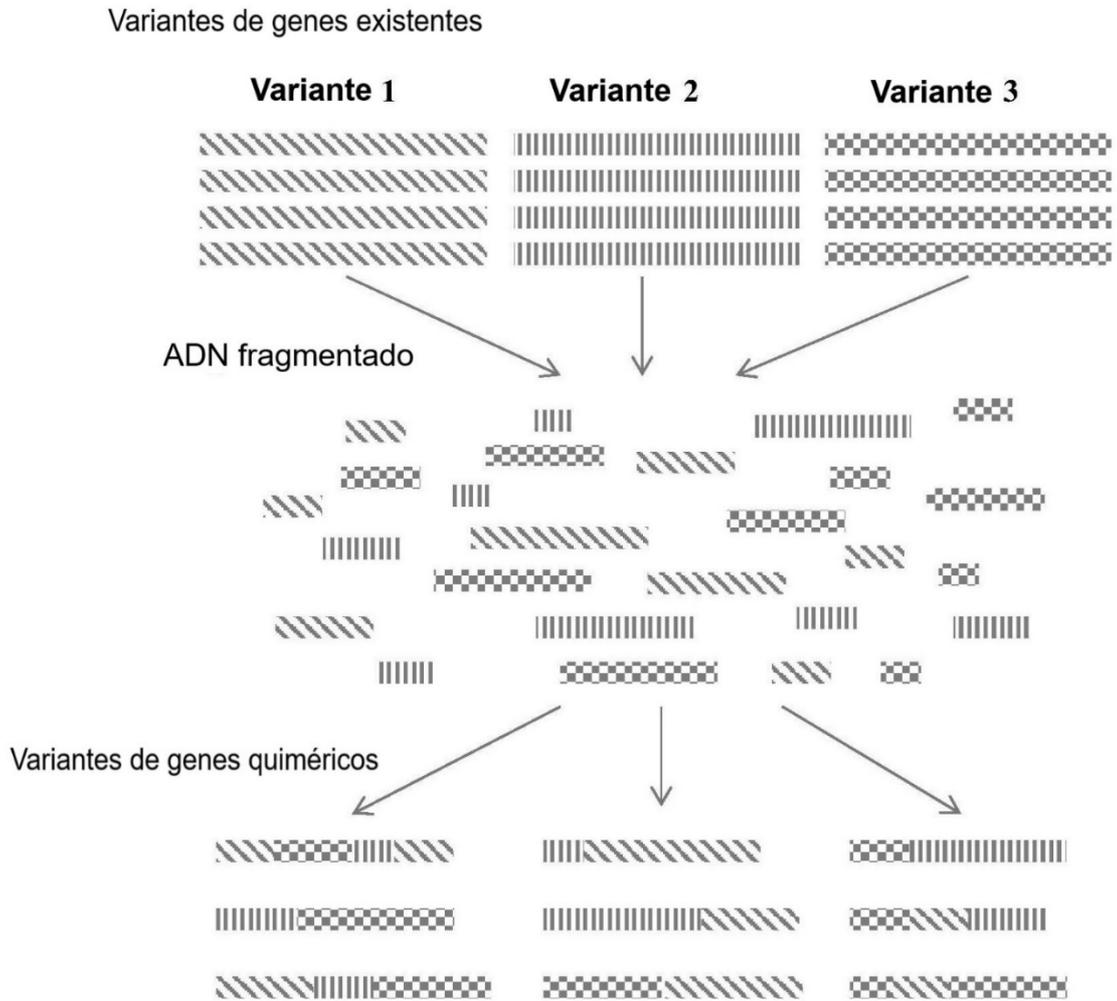


FIGURA 2



FIGURA 3

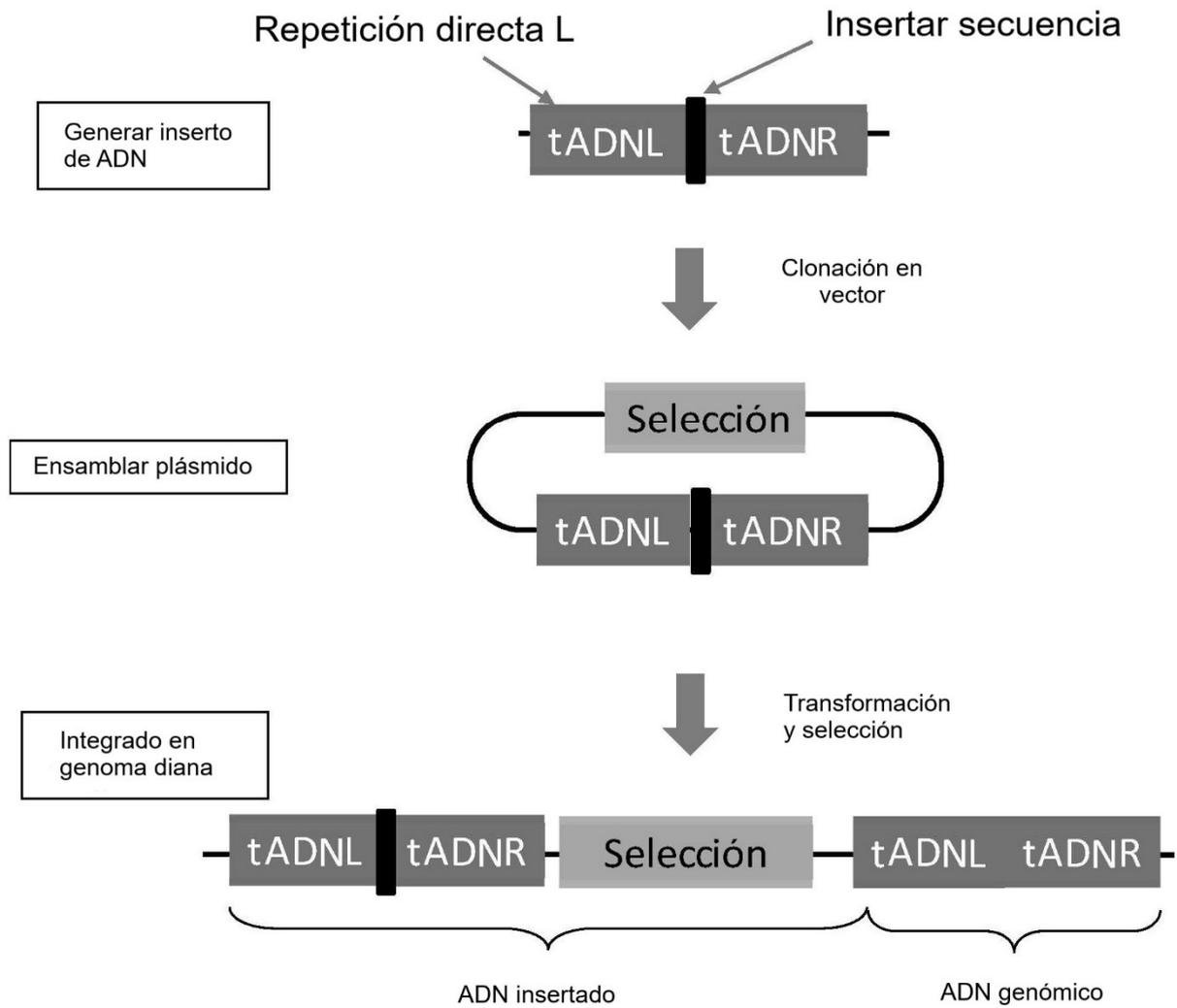


FIGURA 4

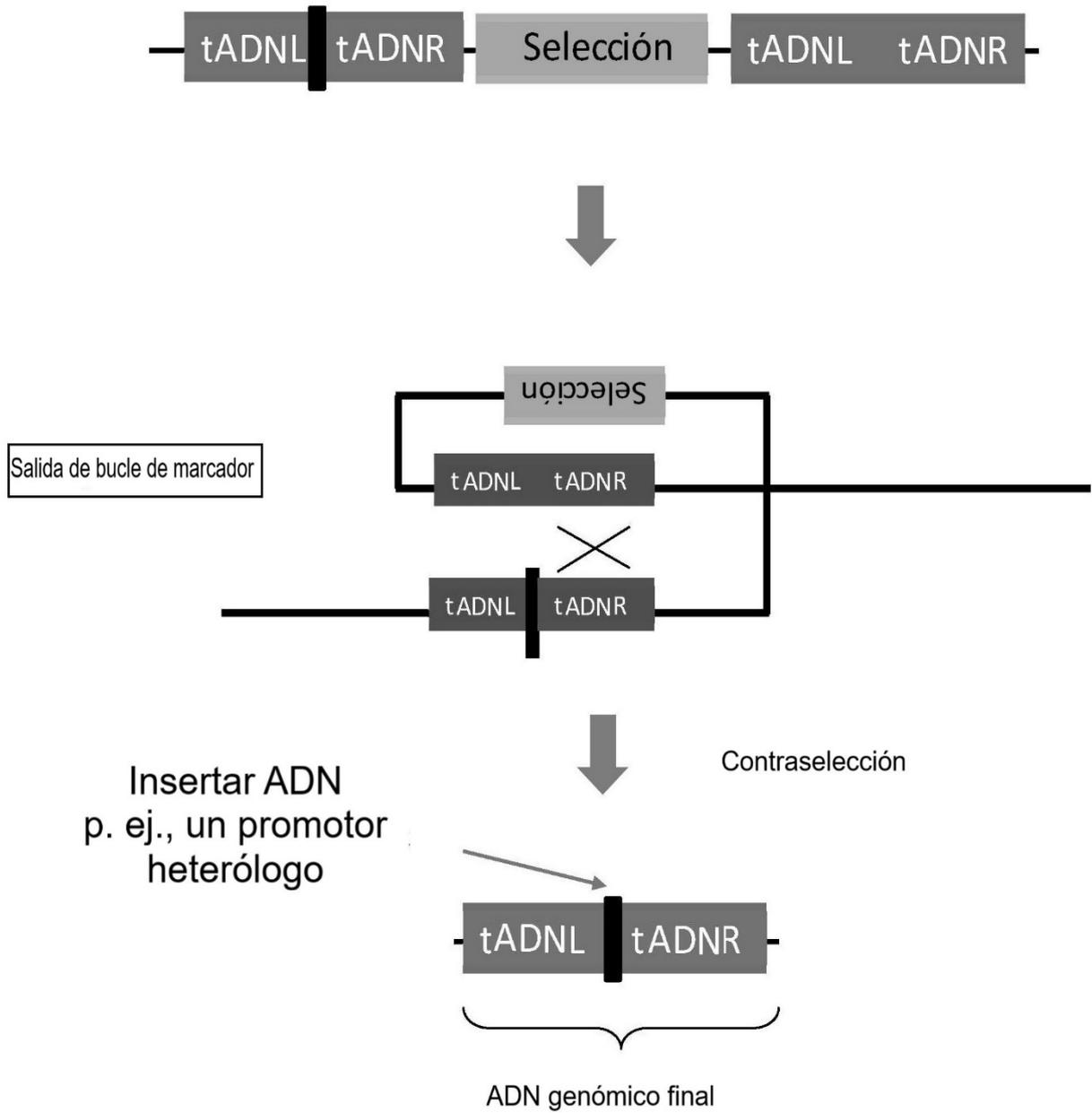


FIGURA 5

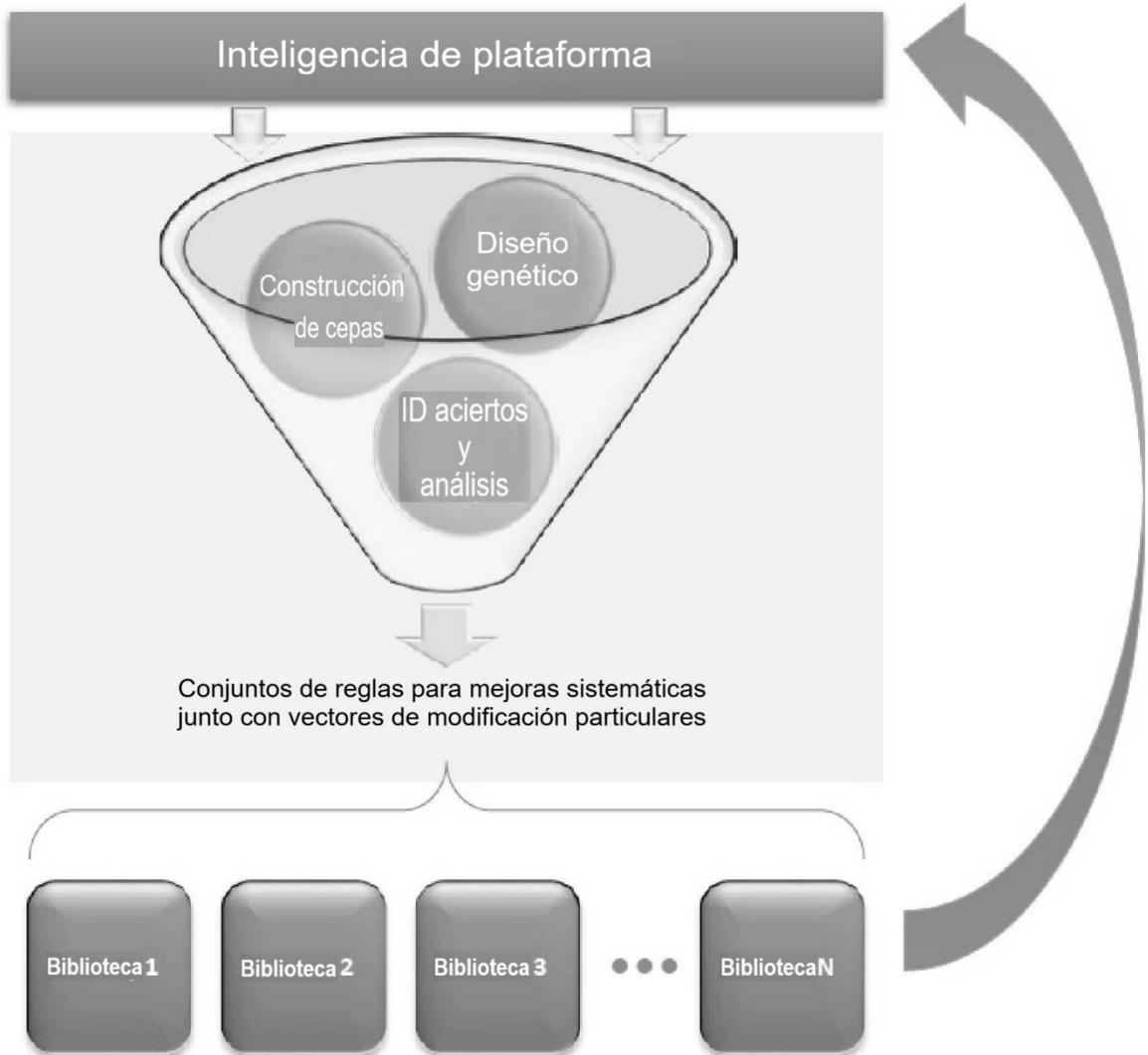


FIGURA 6A

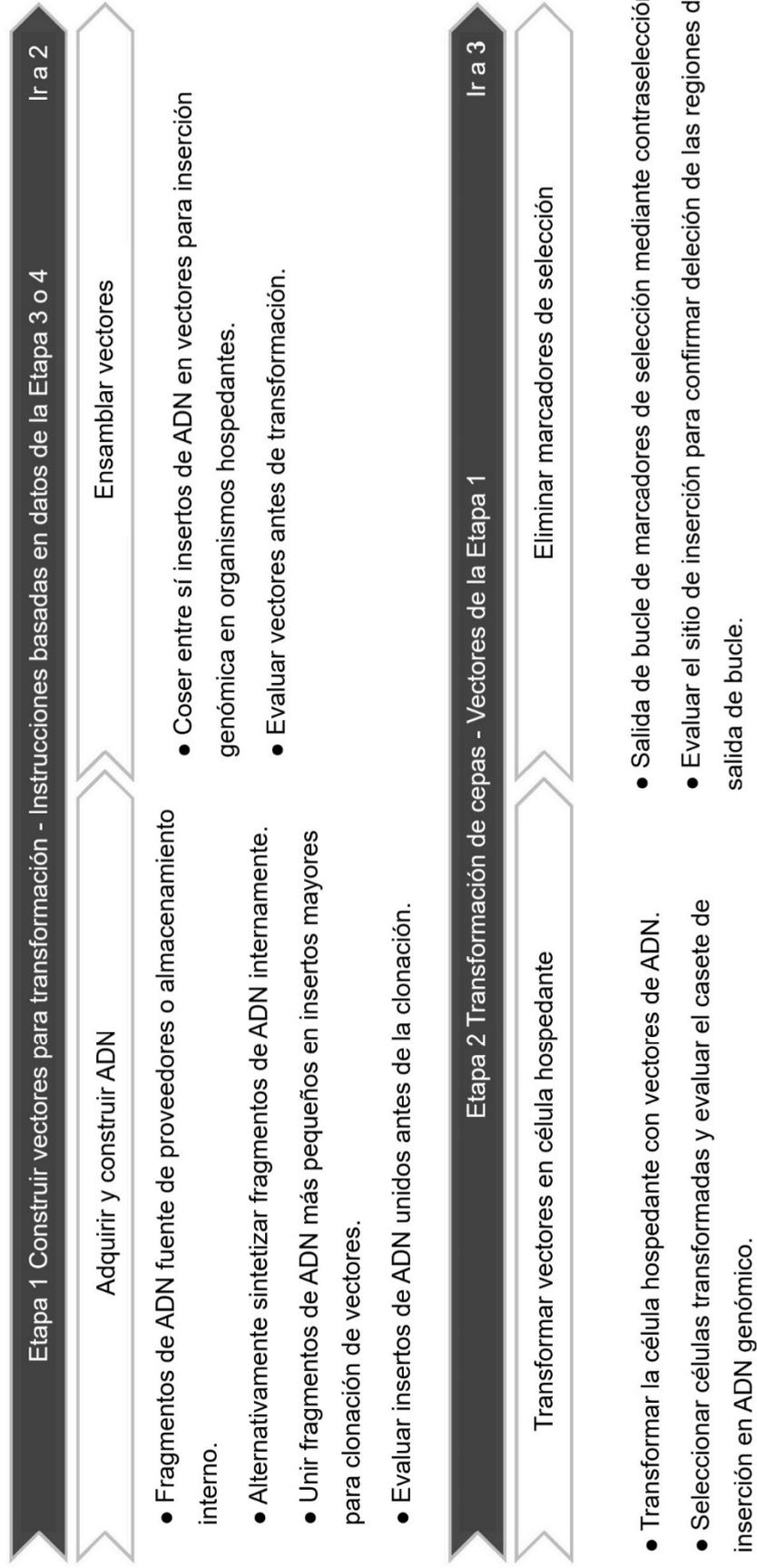


FIGURA 6B

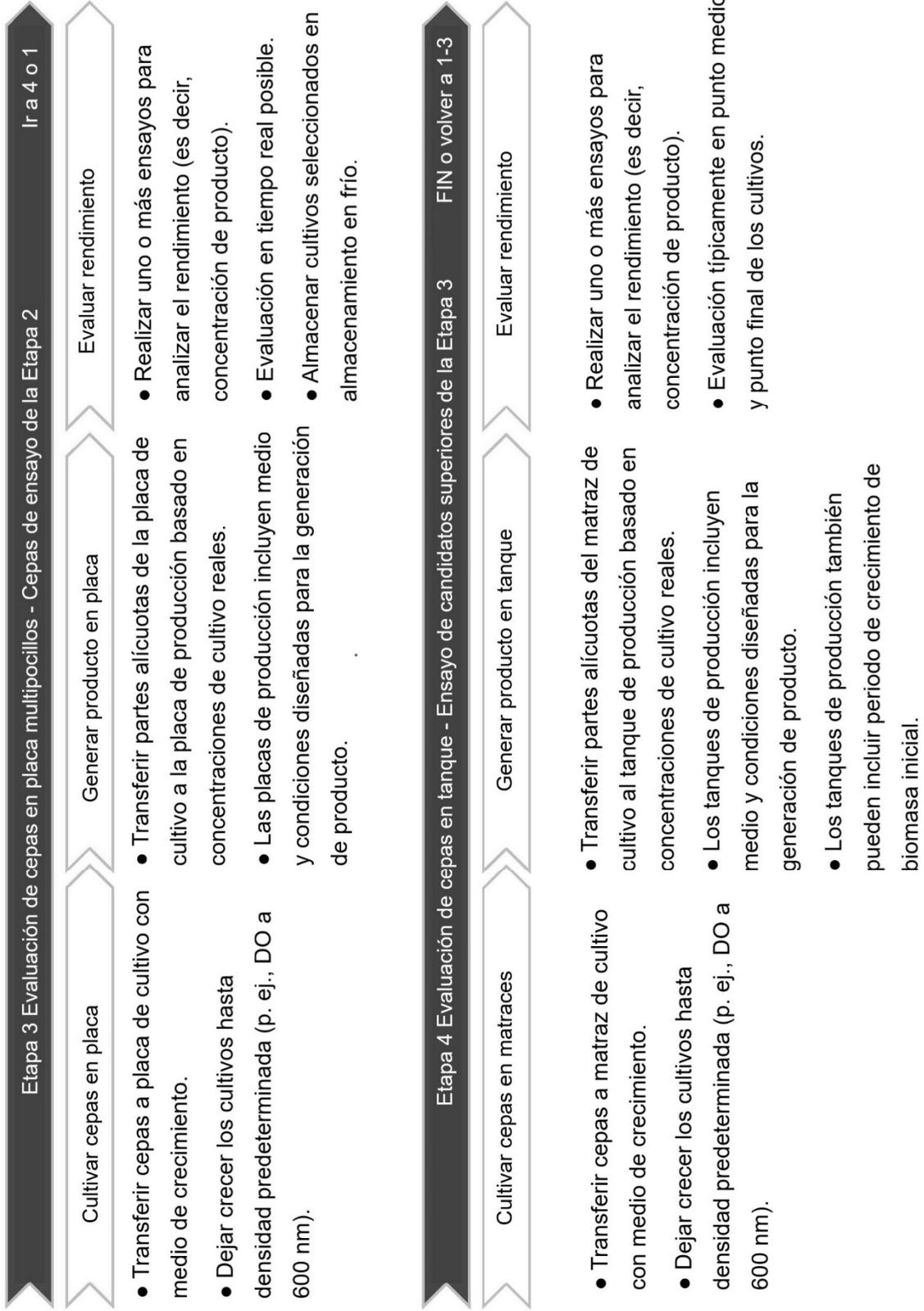
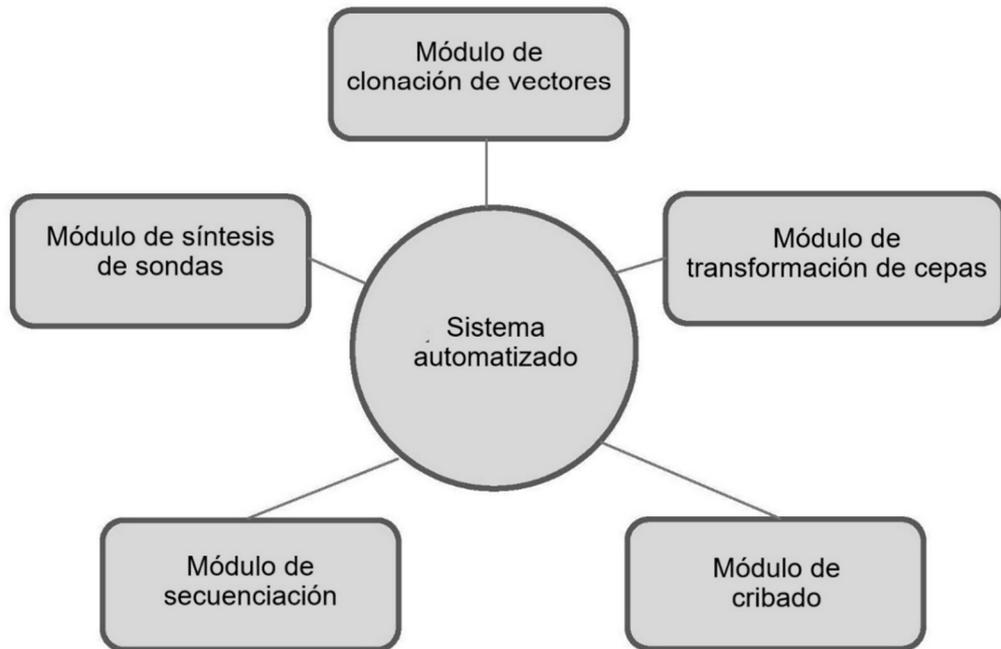


FIGURA 7



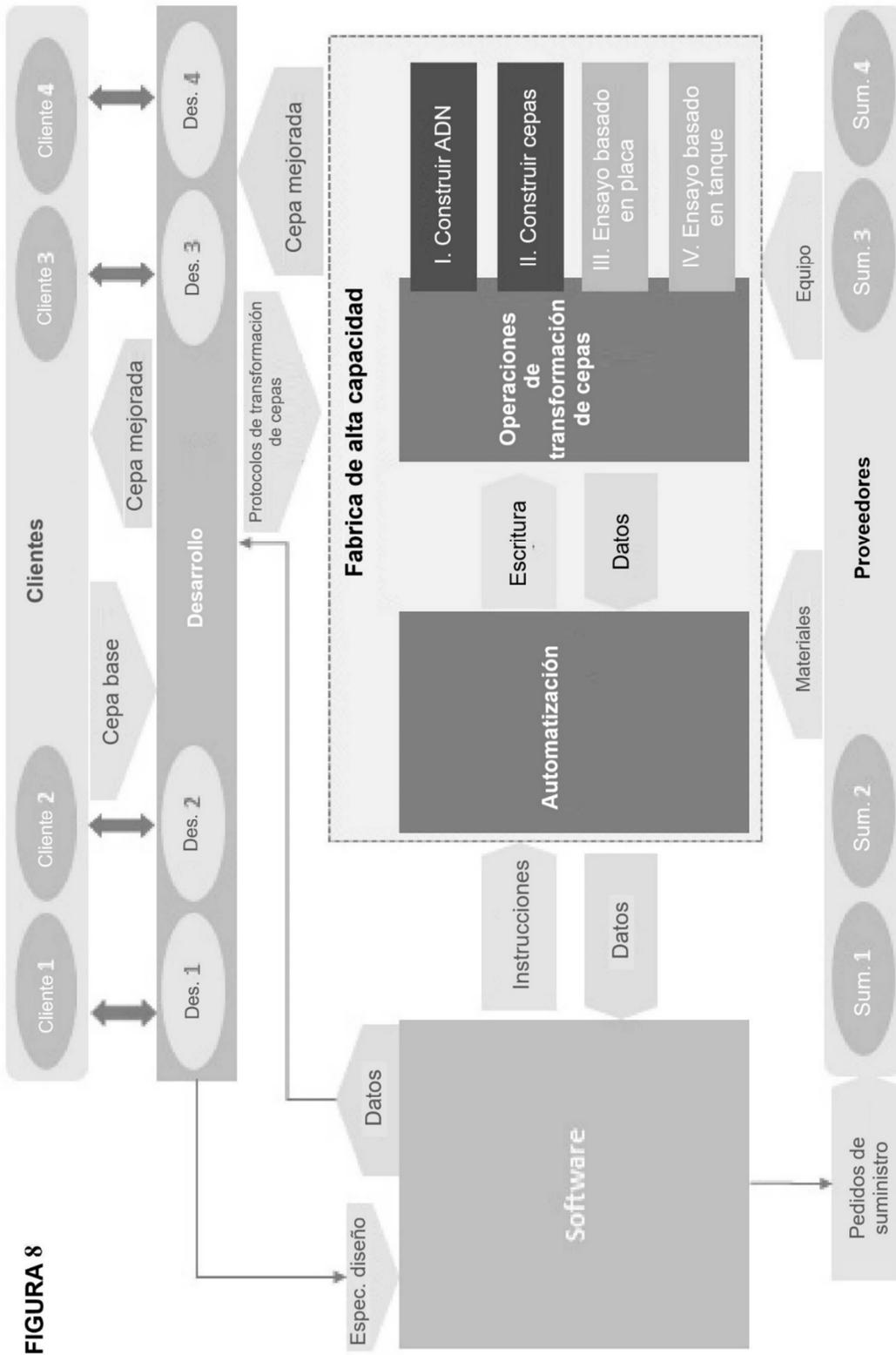


FIGURA 9

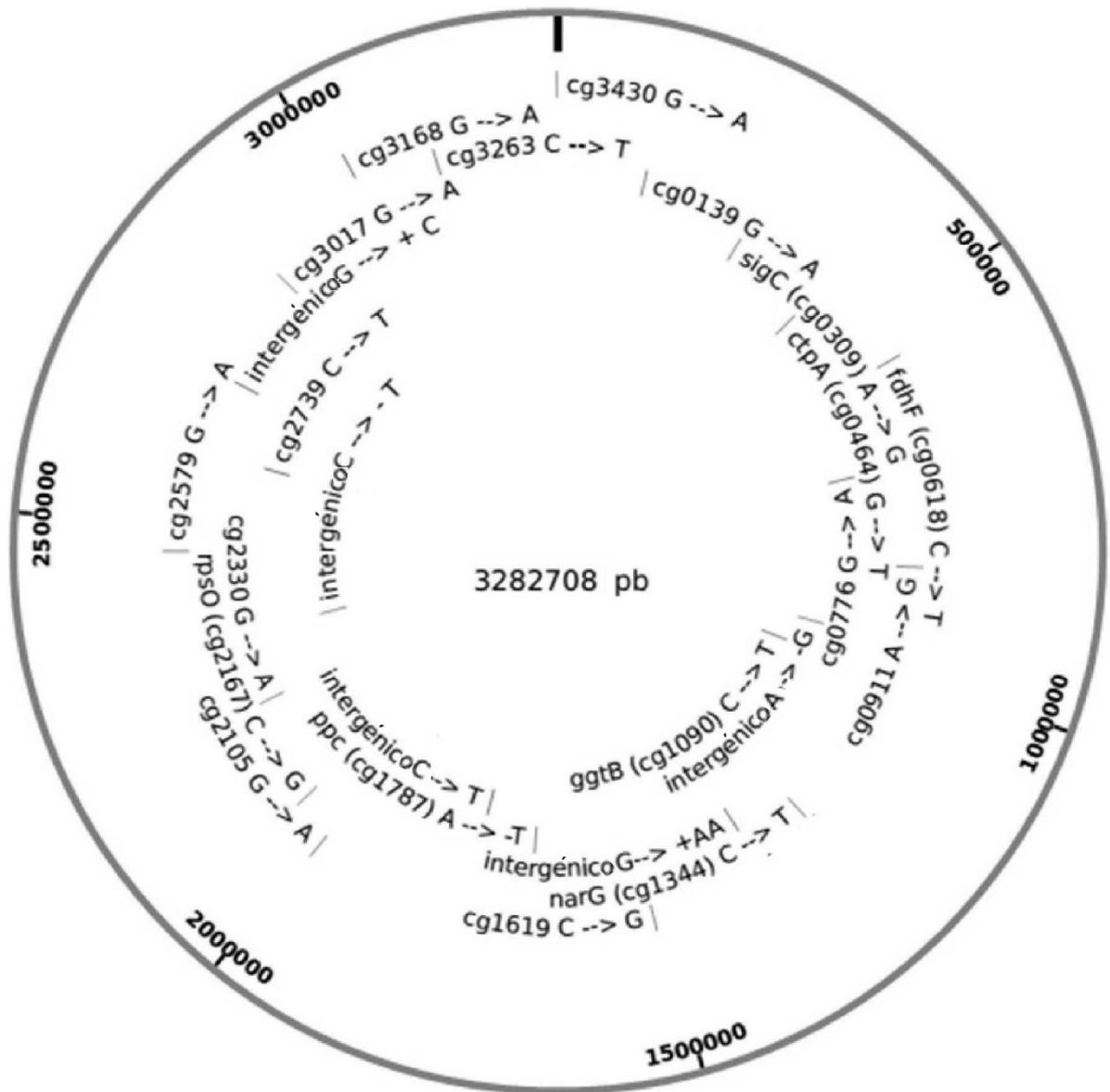


FIGURA 10

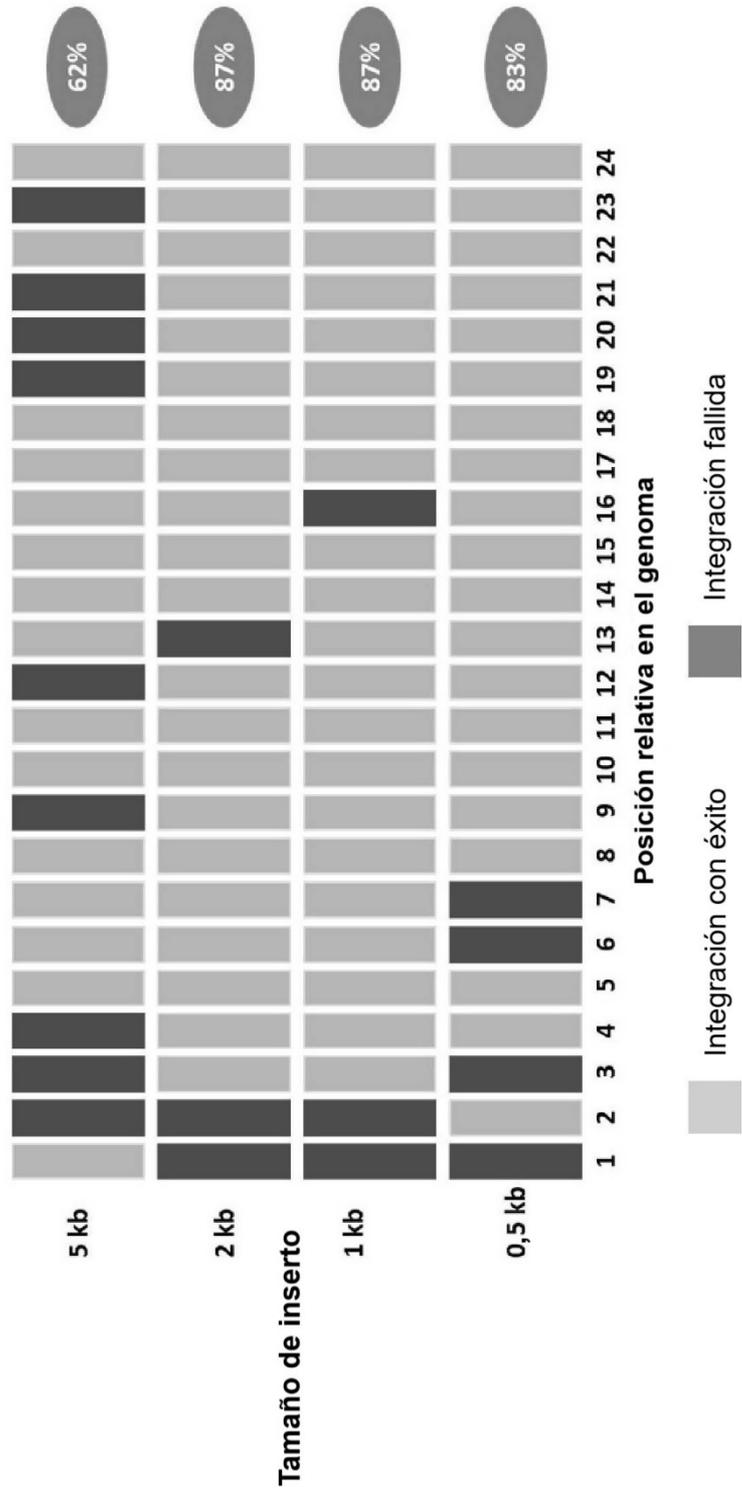


FIGURA 11

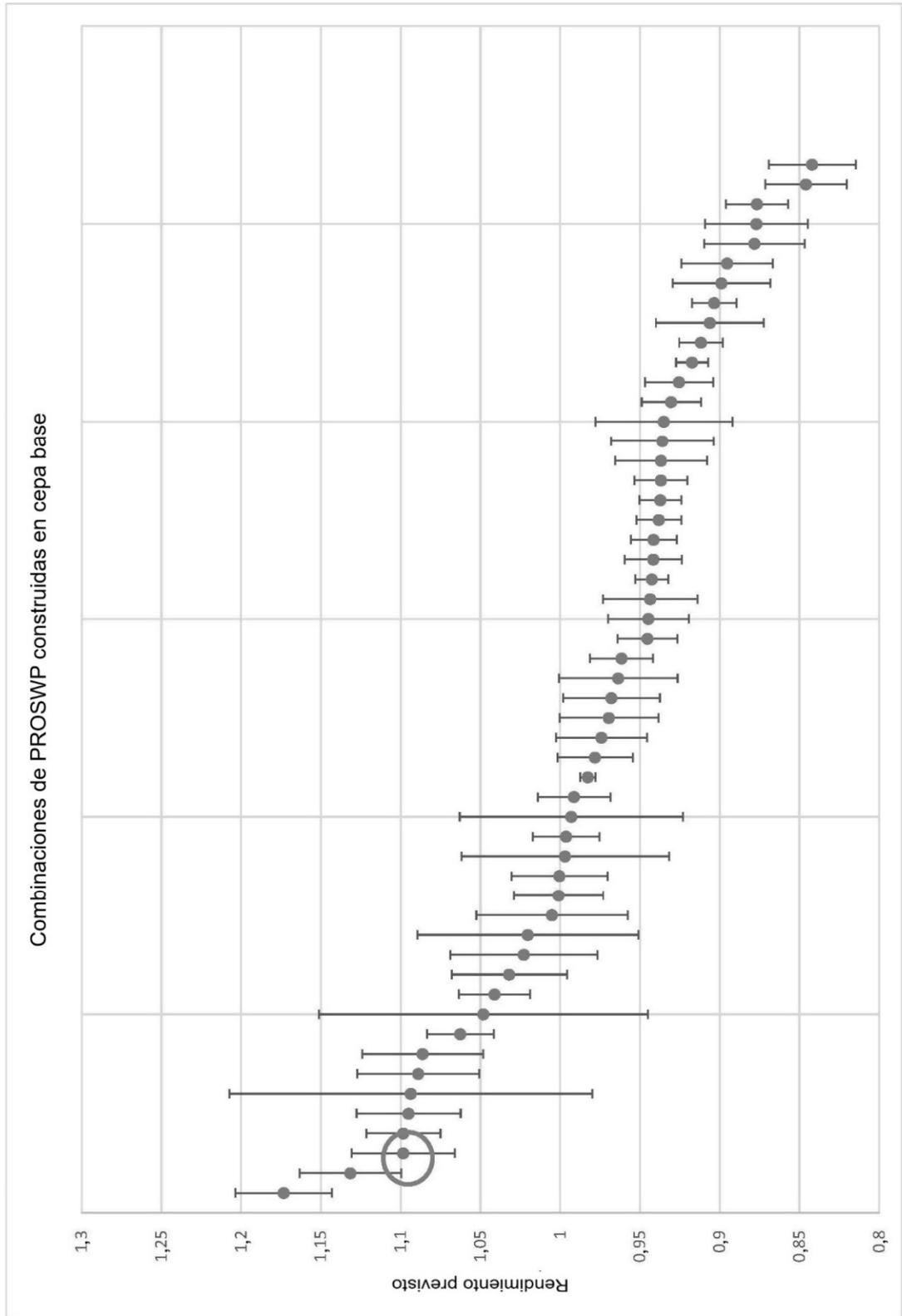


FIGURA 12

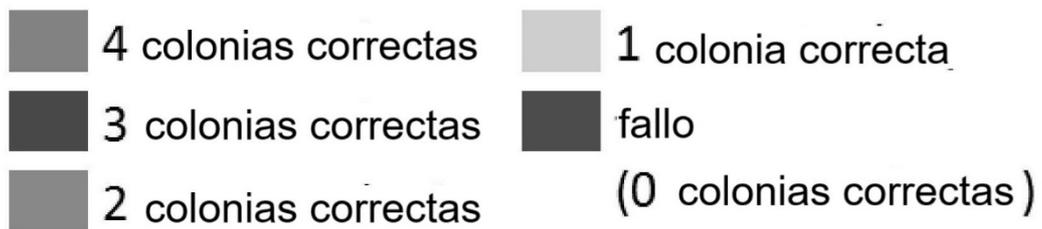
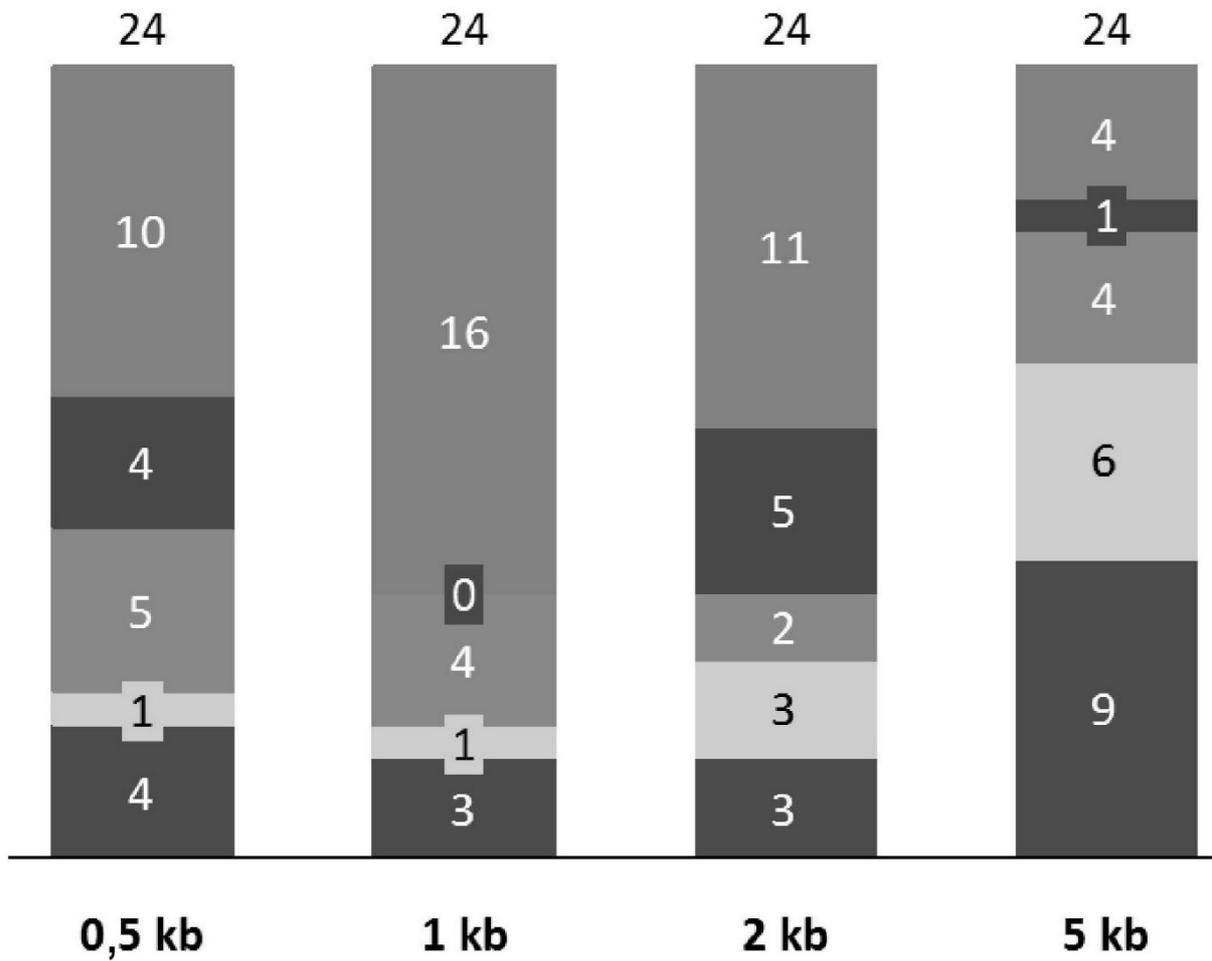


FIGURA 13

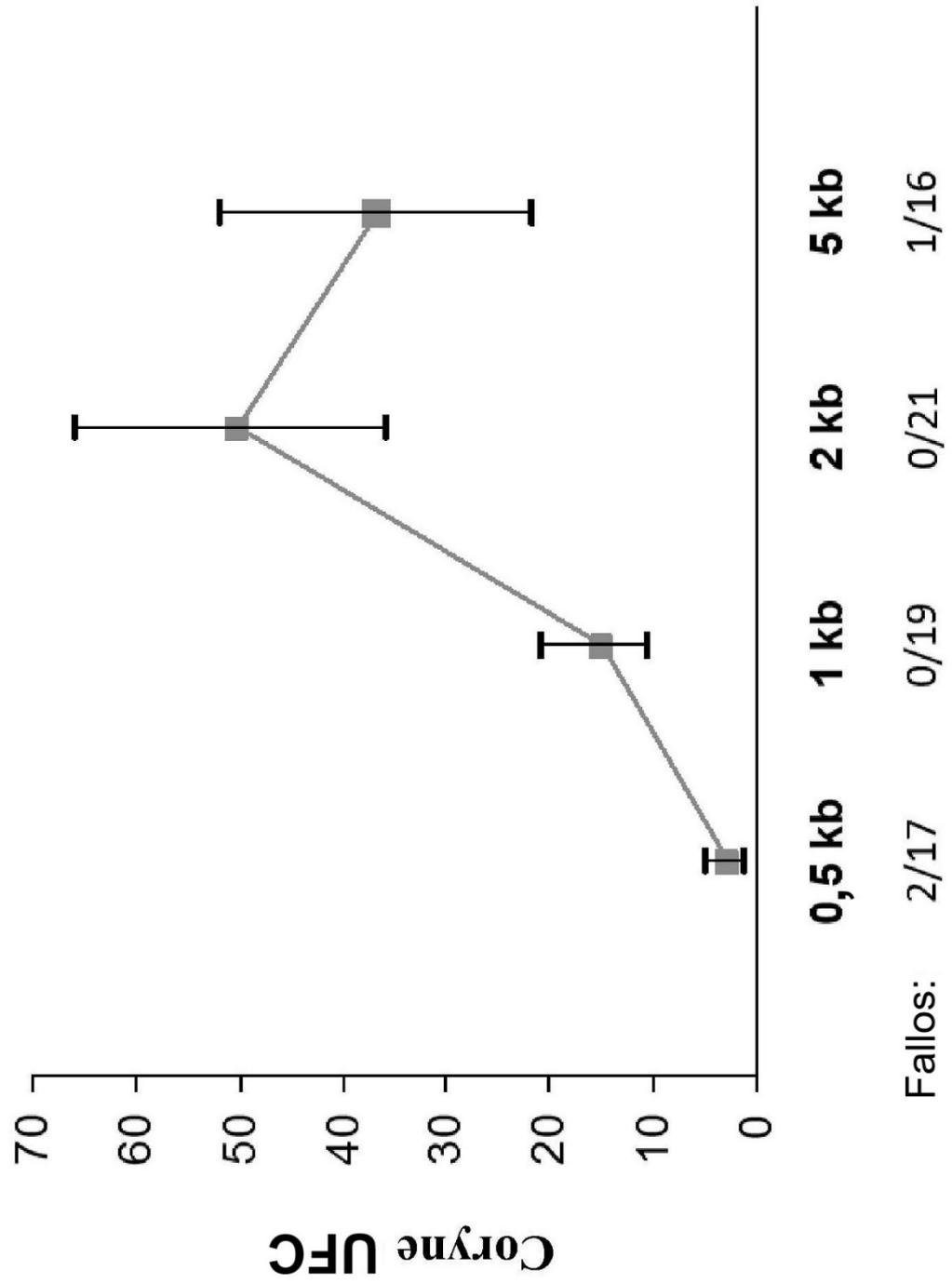
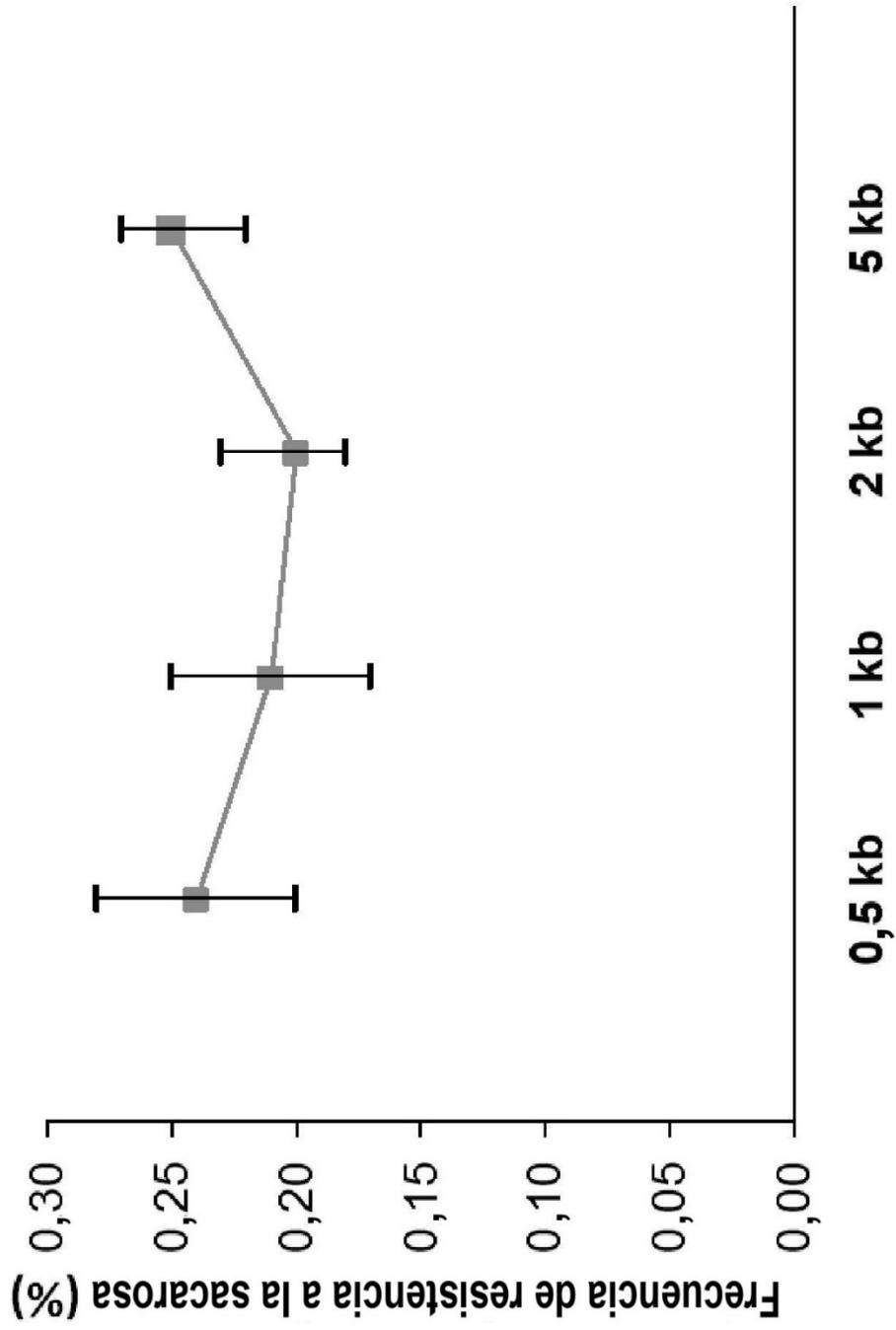
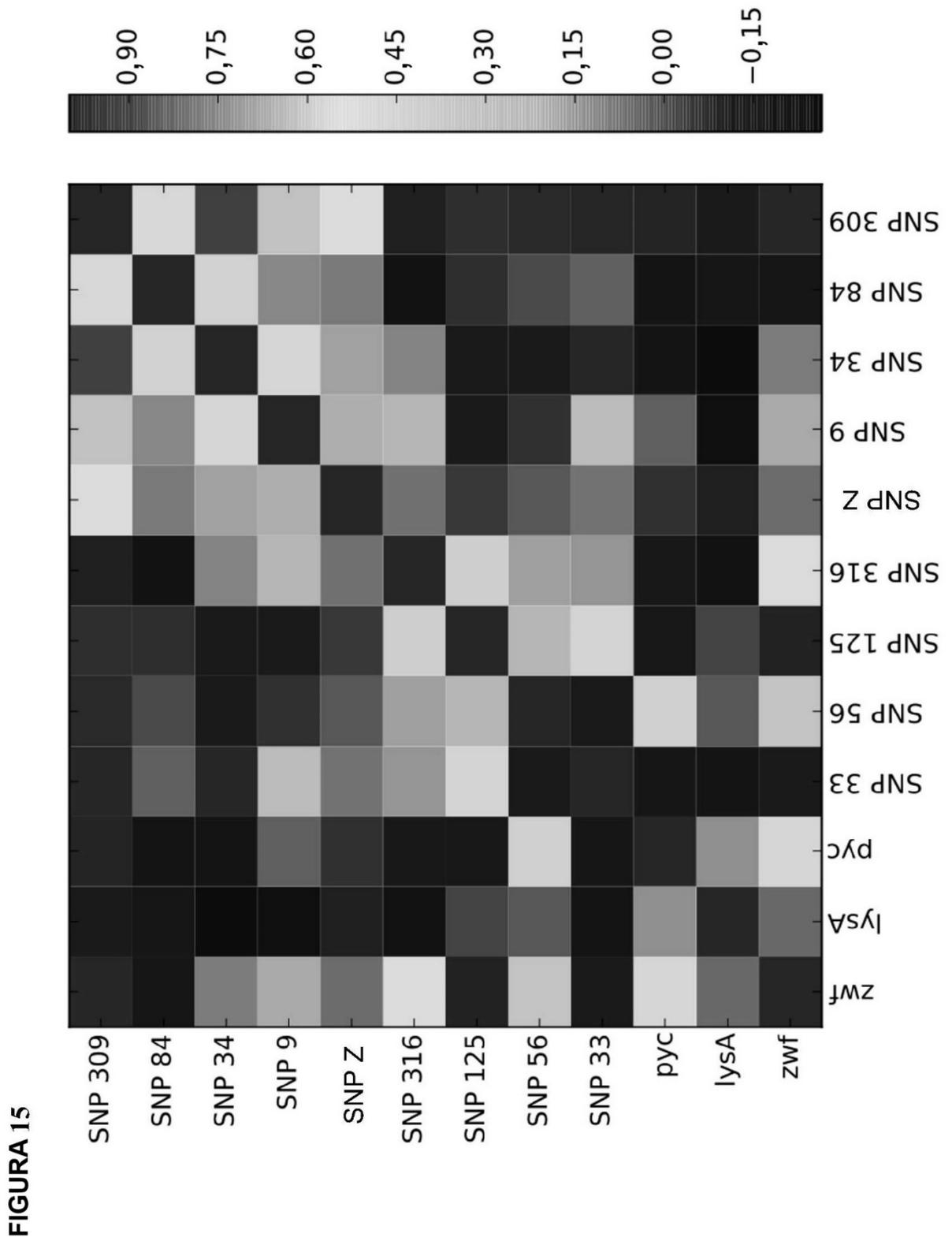


FIGURA 14





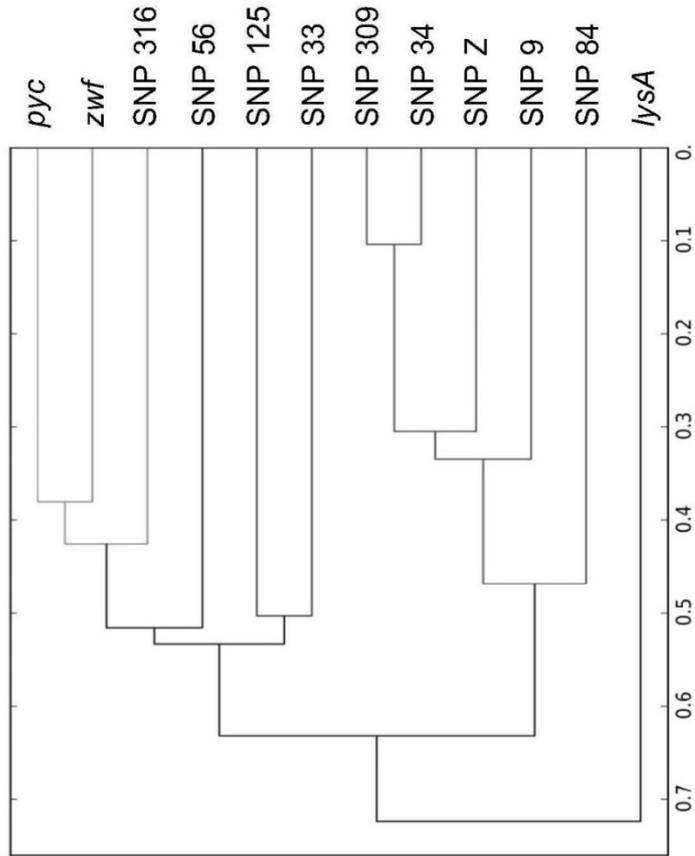


FIGURA 16 A

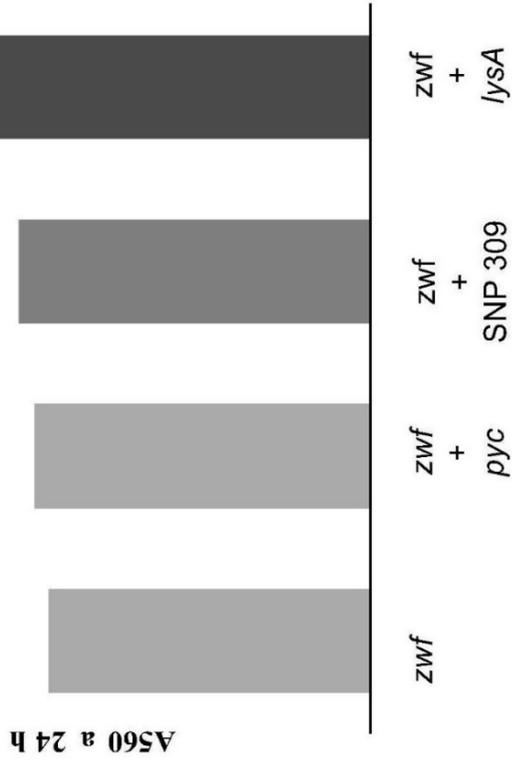


FIGURA 16 B

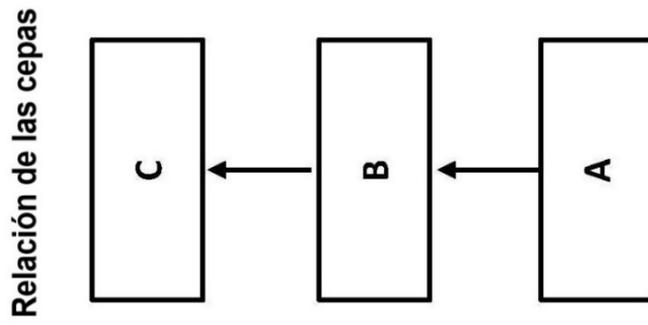


FIGURA 17 A

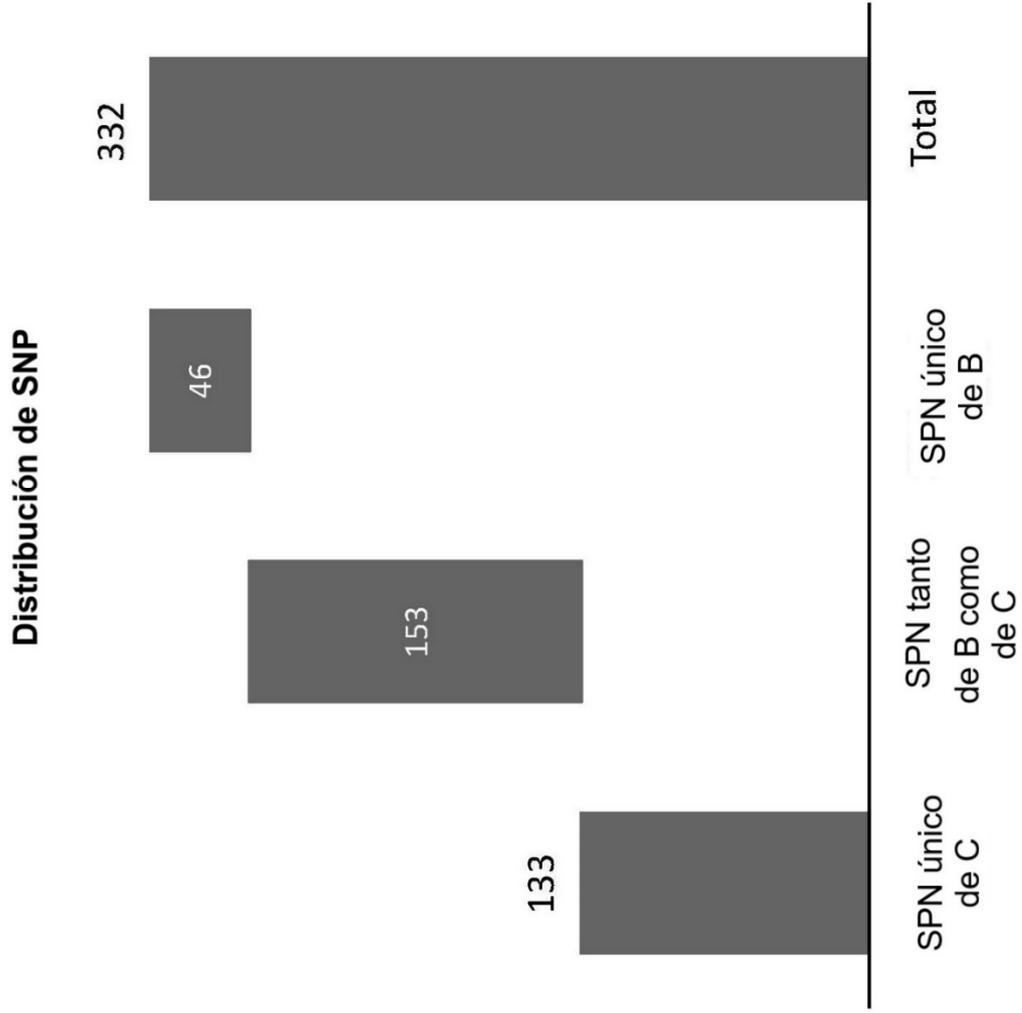


FIGURA 17 B

FIGURA 18

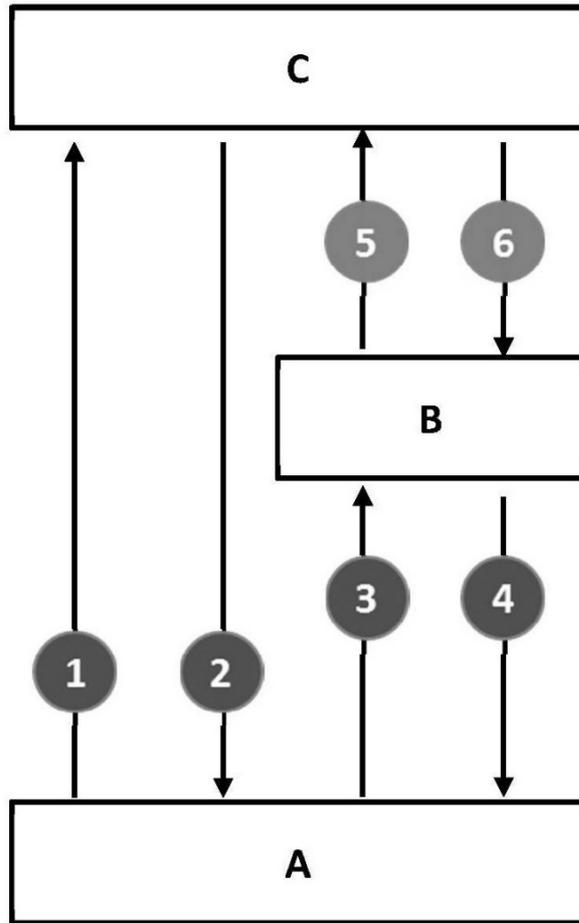


FIGURA 19

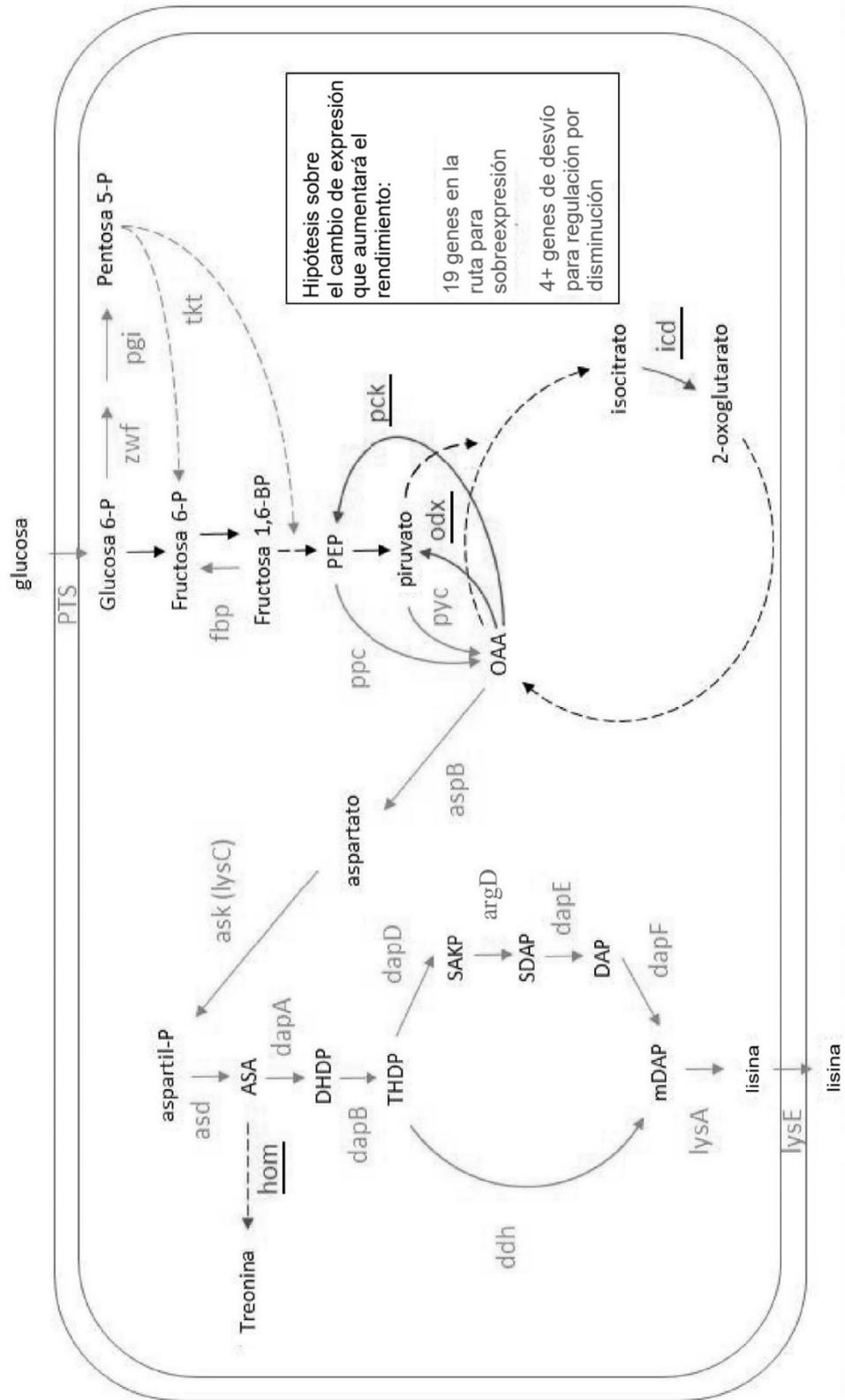


FIGURA 20

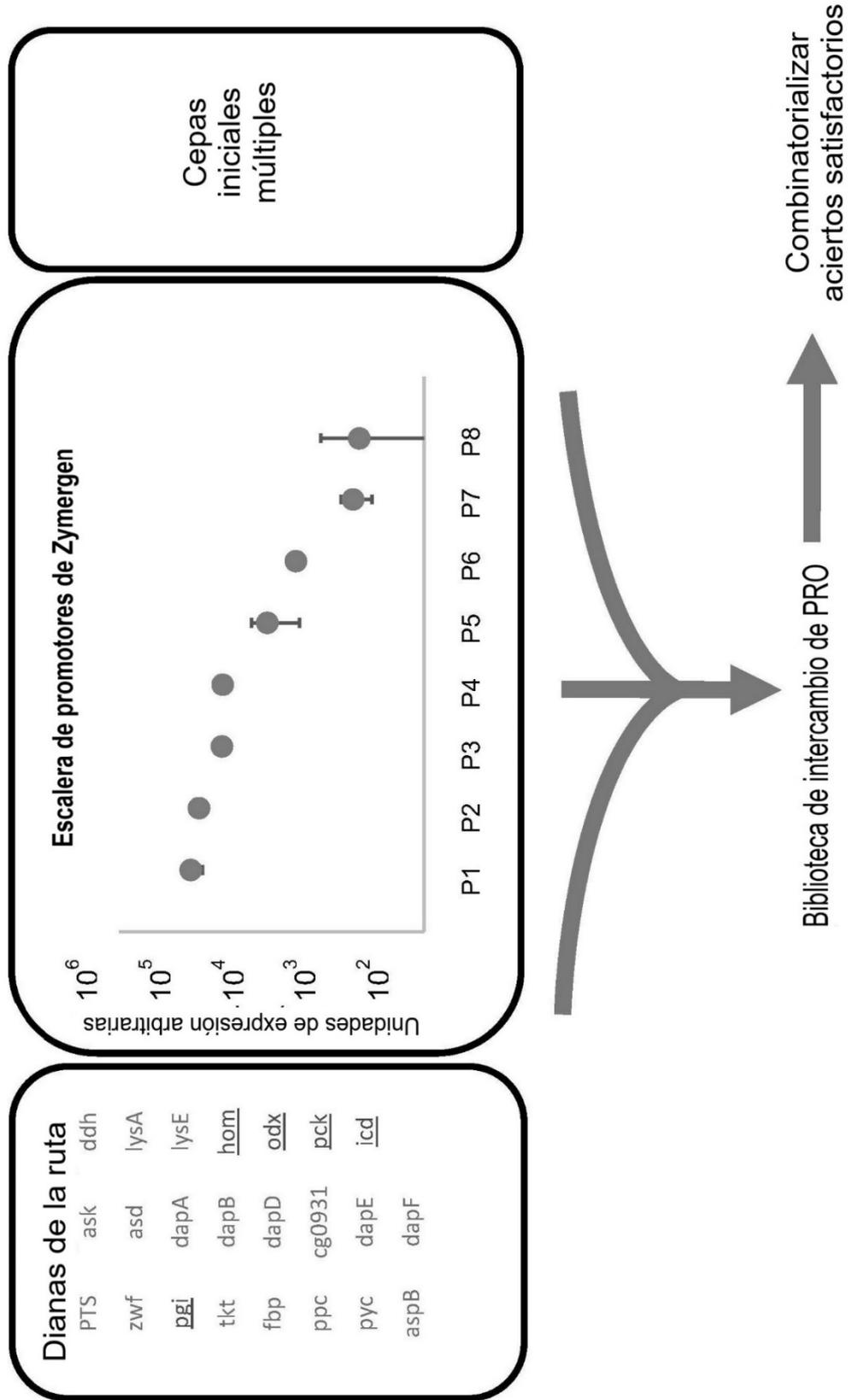


FIGURA 21

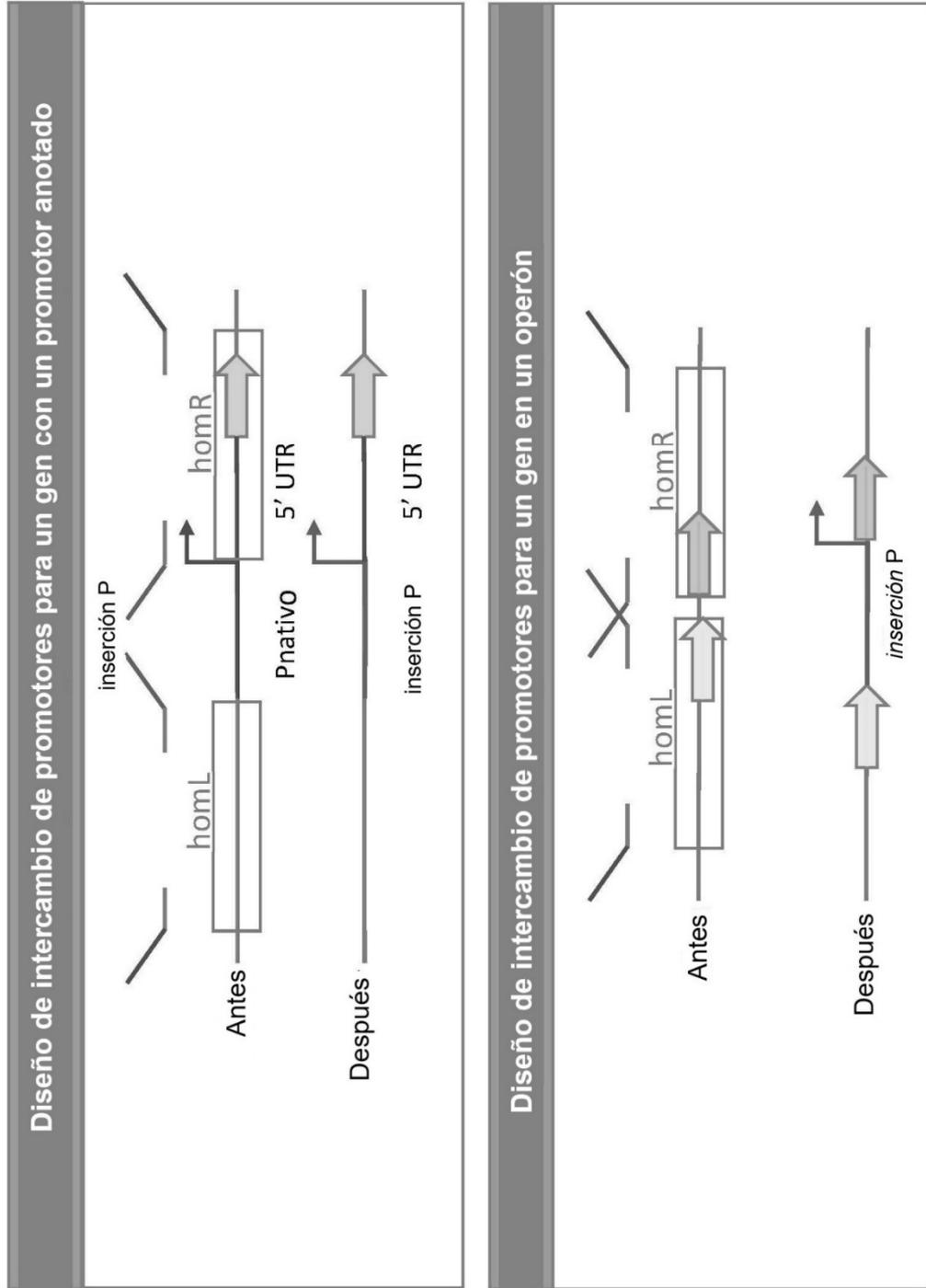


FIGURA 22

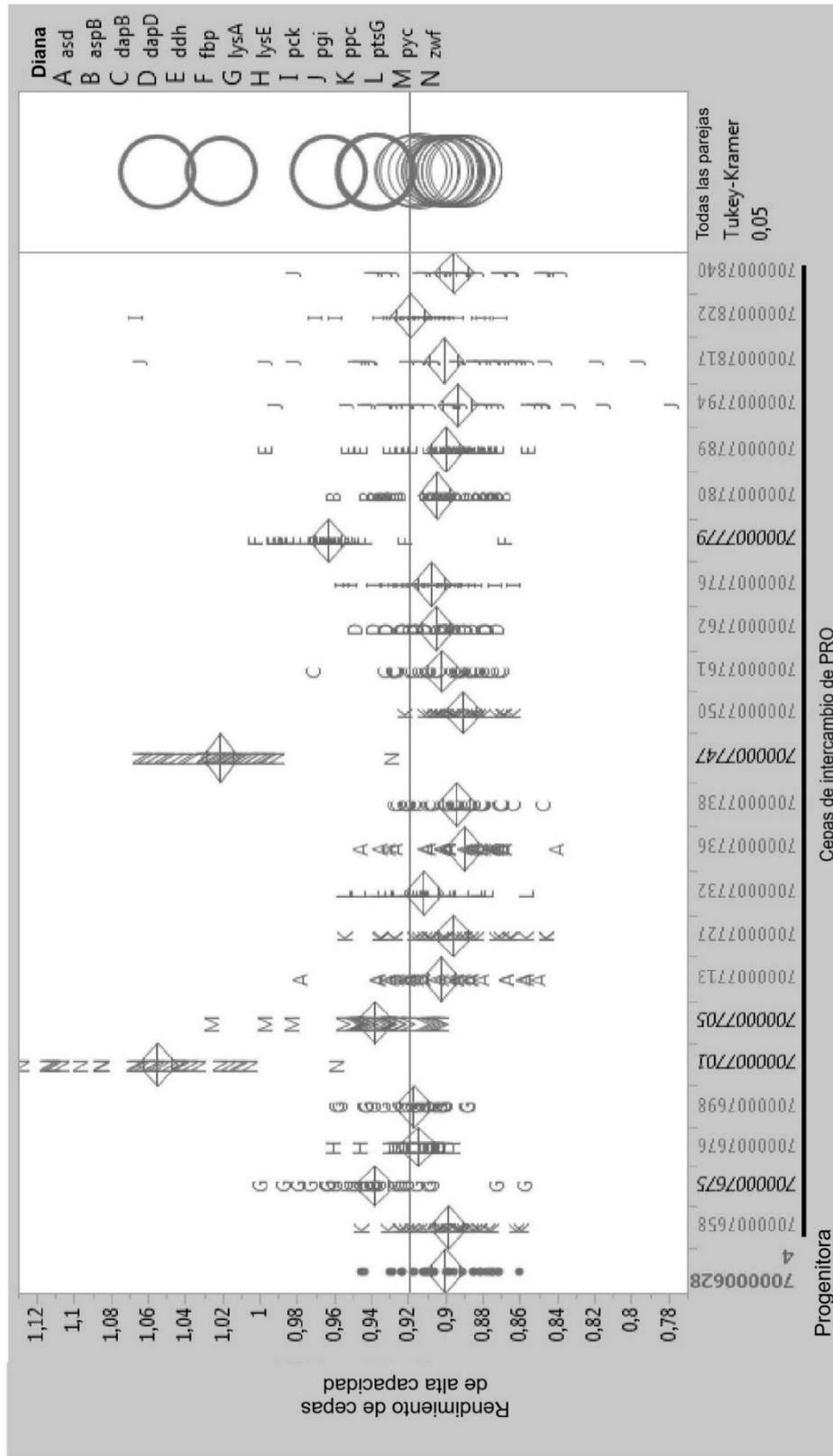


FIGURA 23

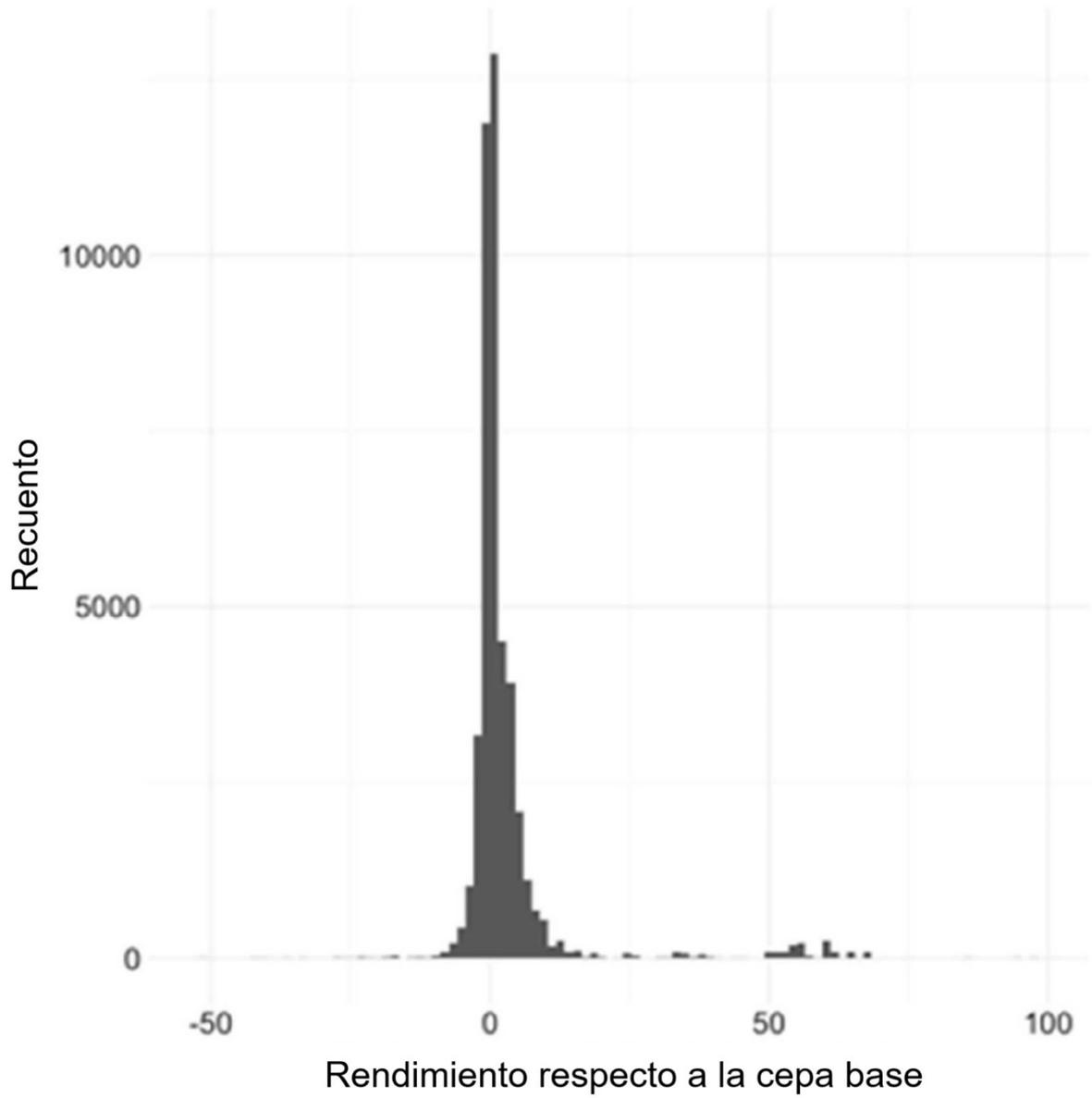


FIGURA 24

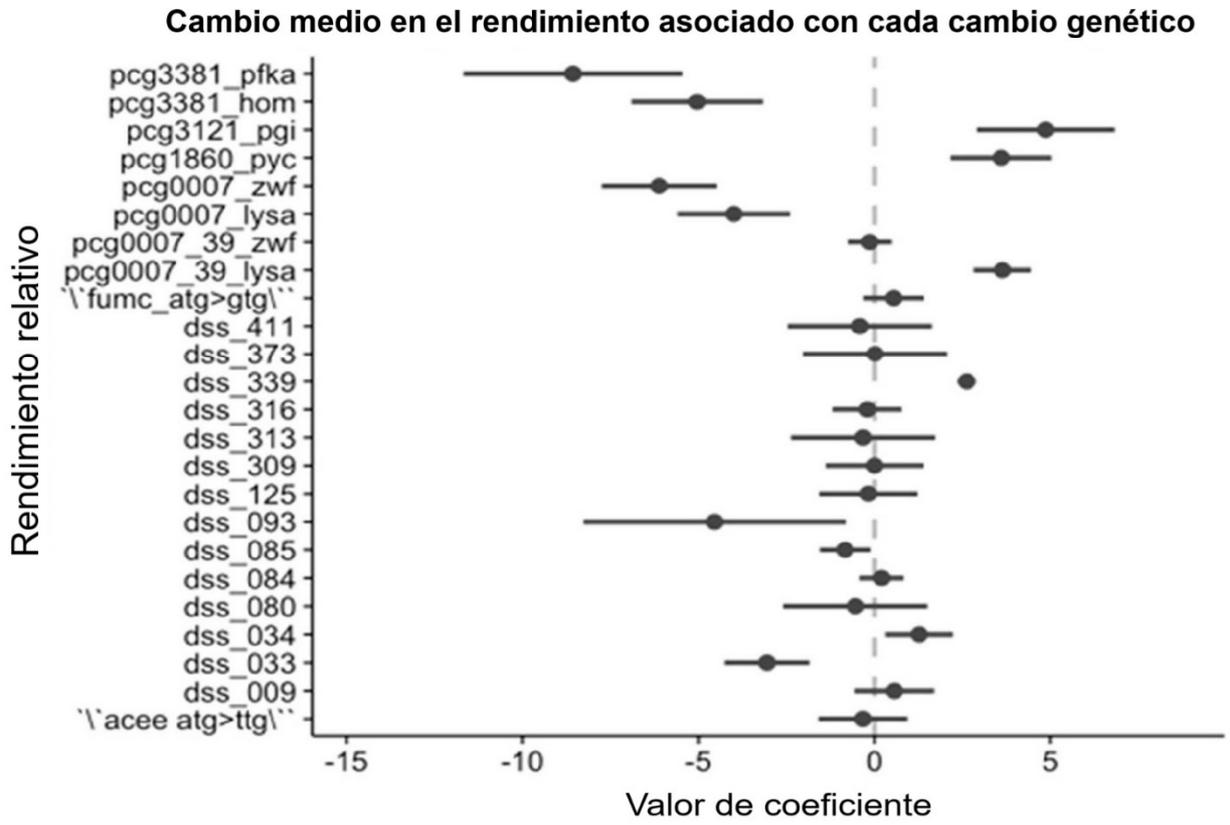


FIGURA 25

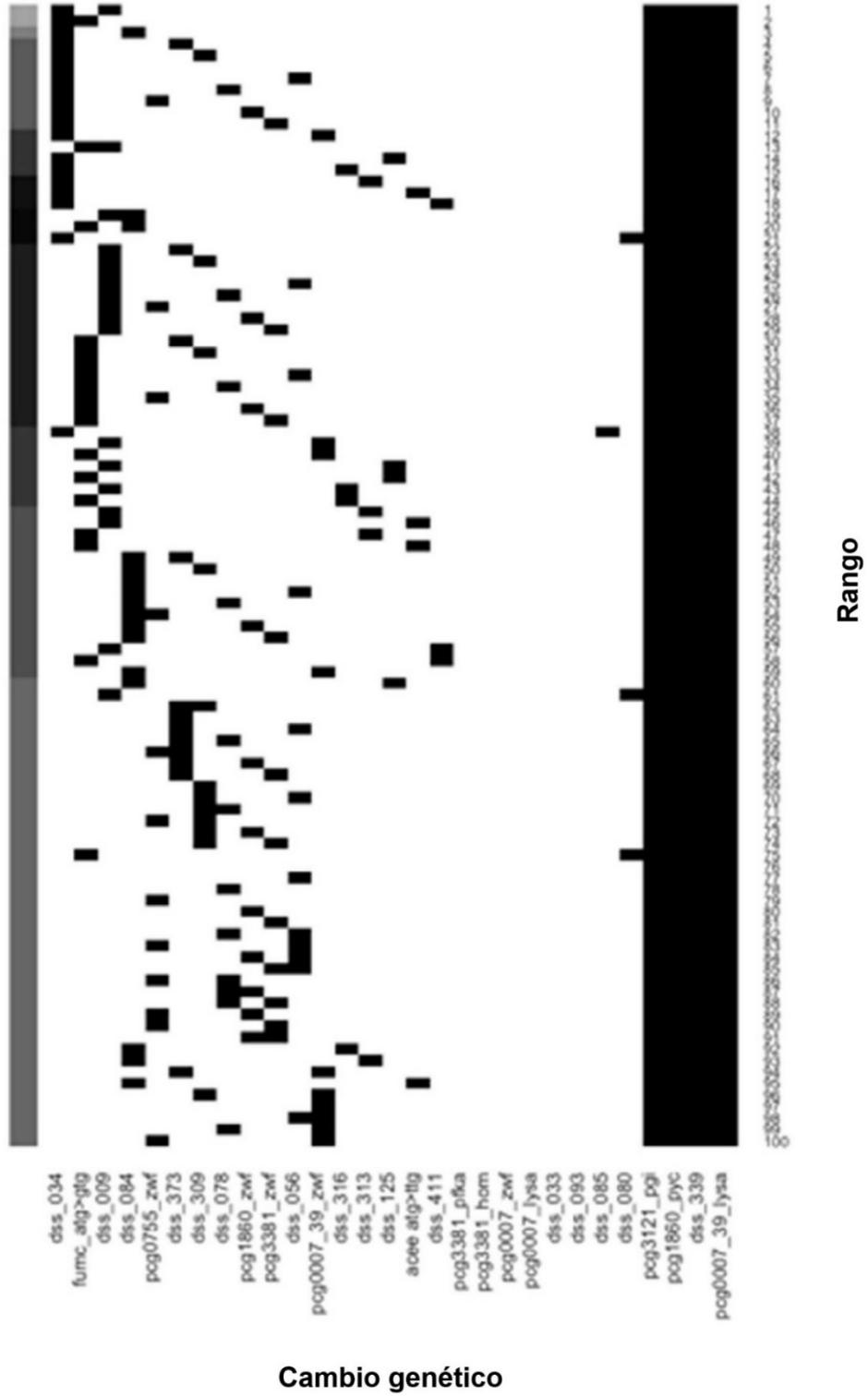


FIGURA 26

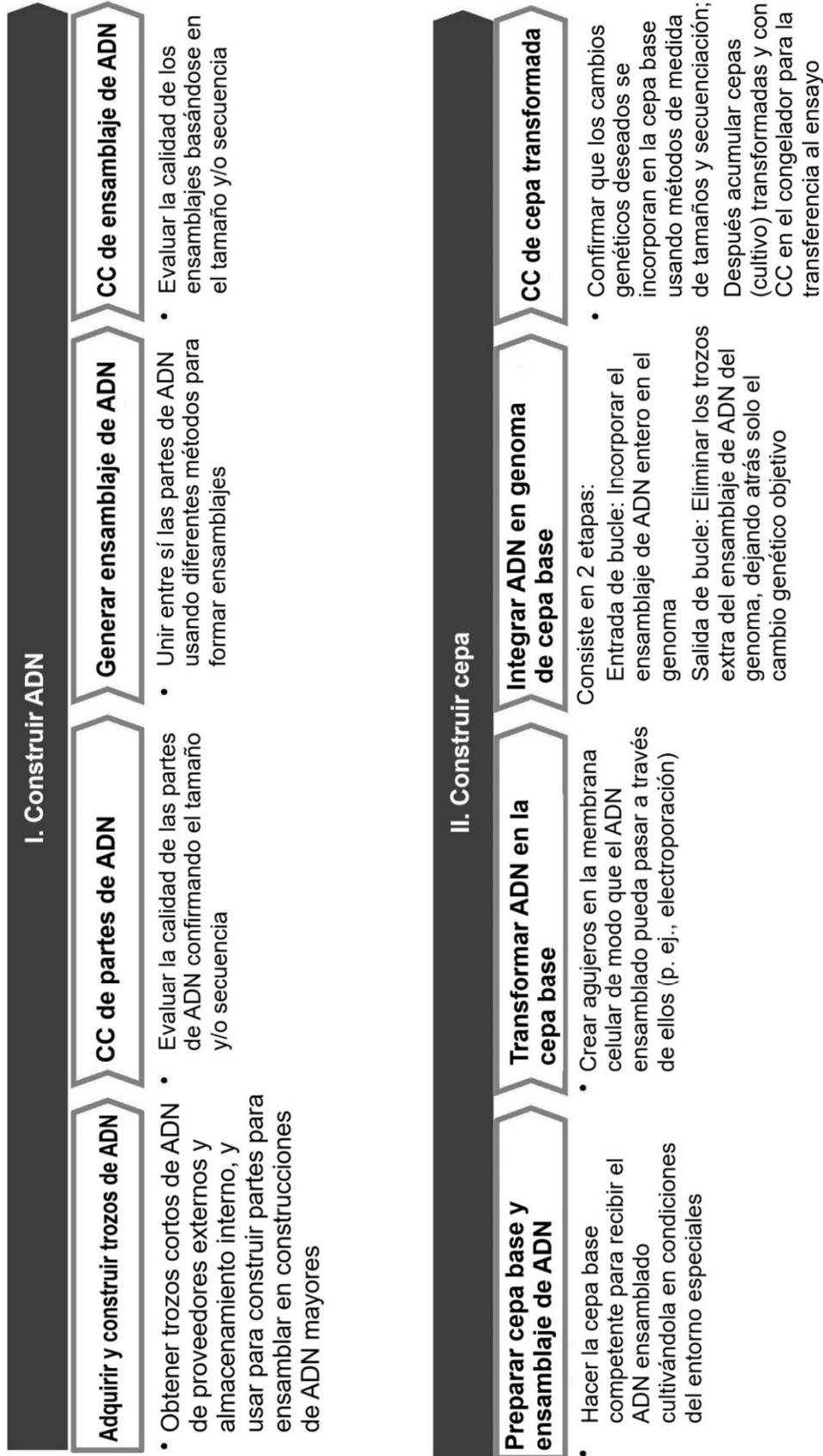


FIGURA 27

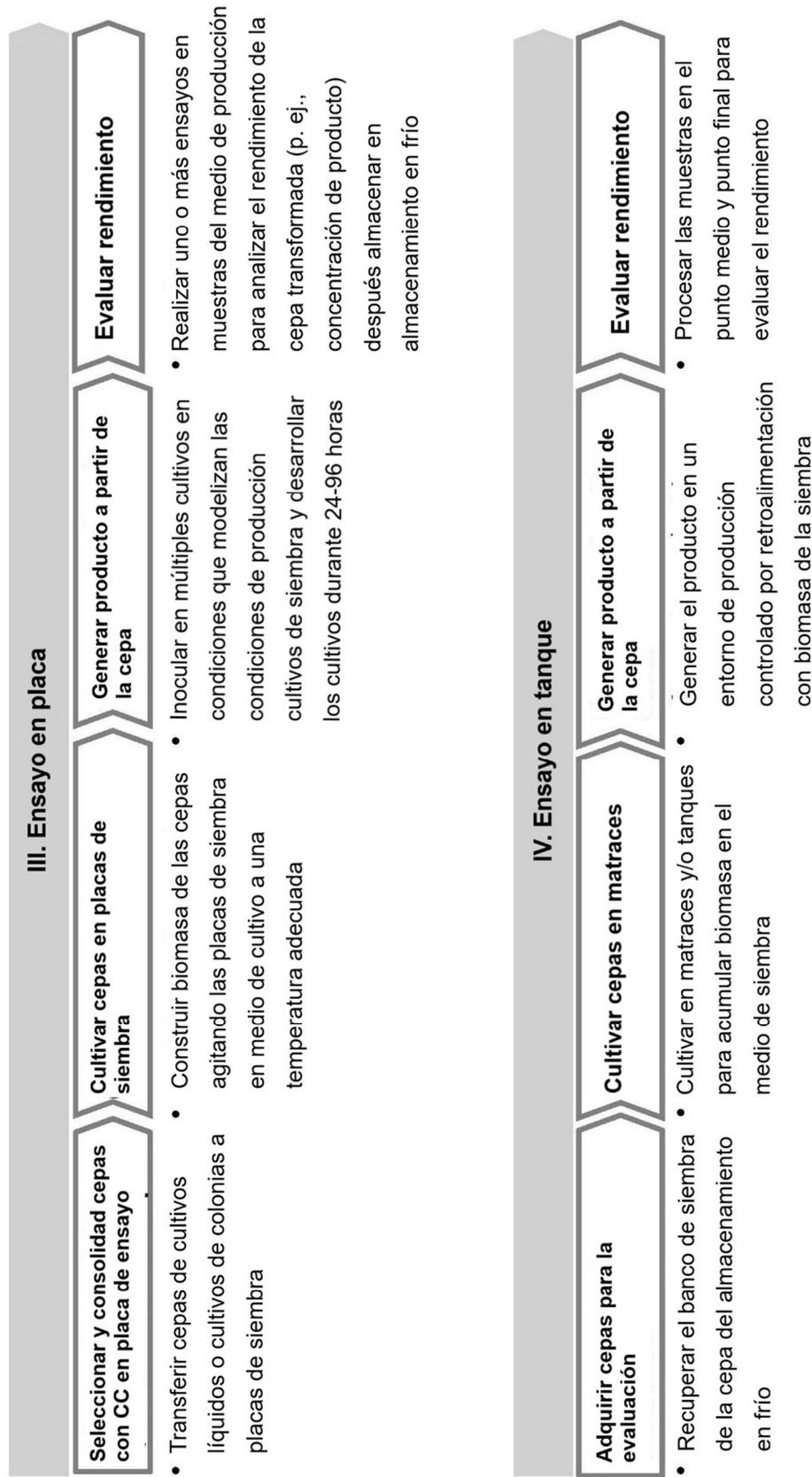


FIGURA 28

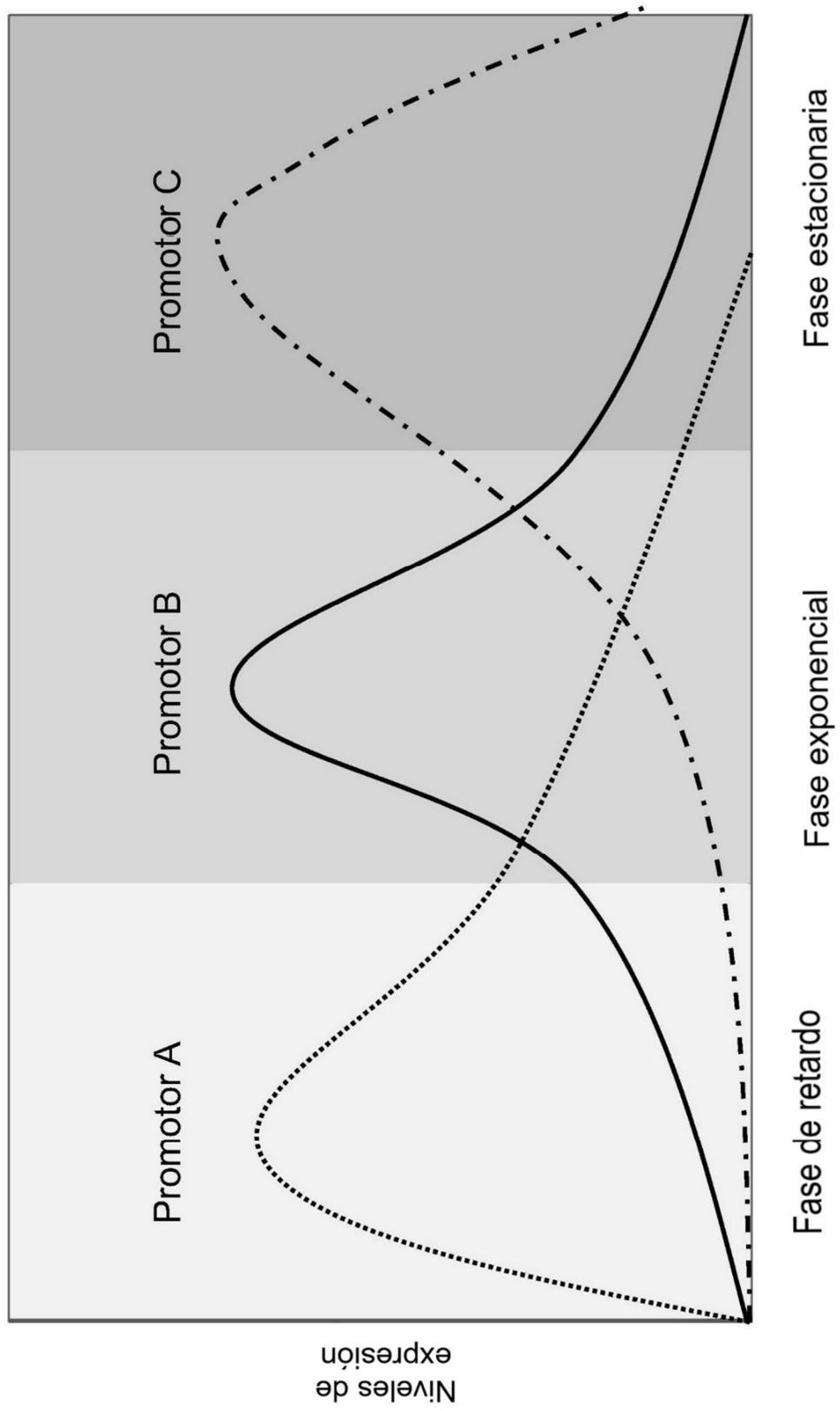


FIGURA 29

Concentración de sustrato a lo largo del tiempo

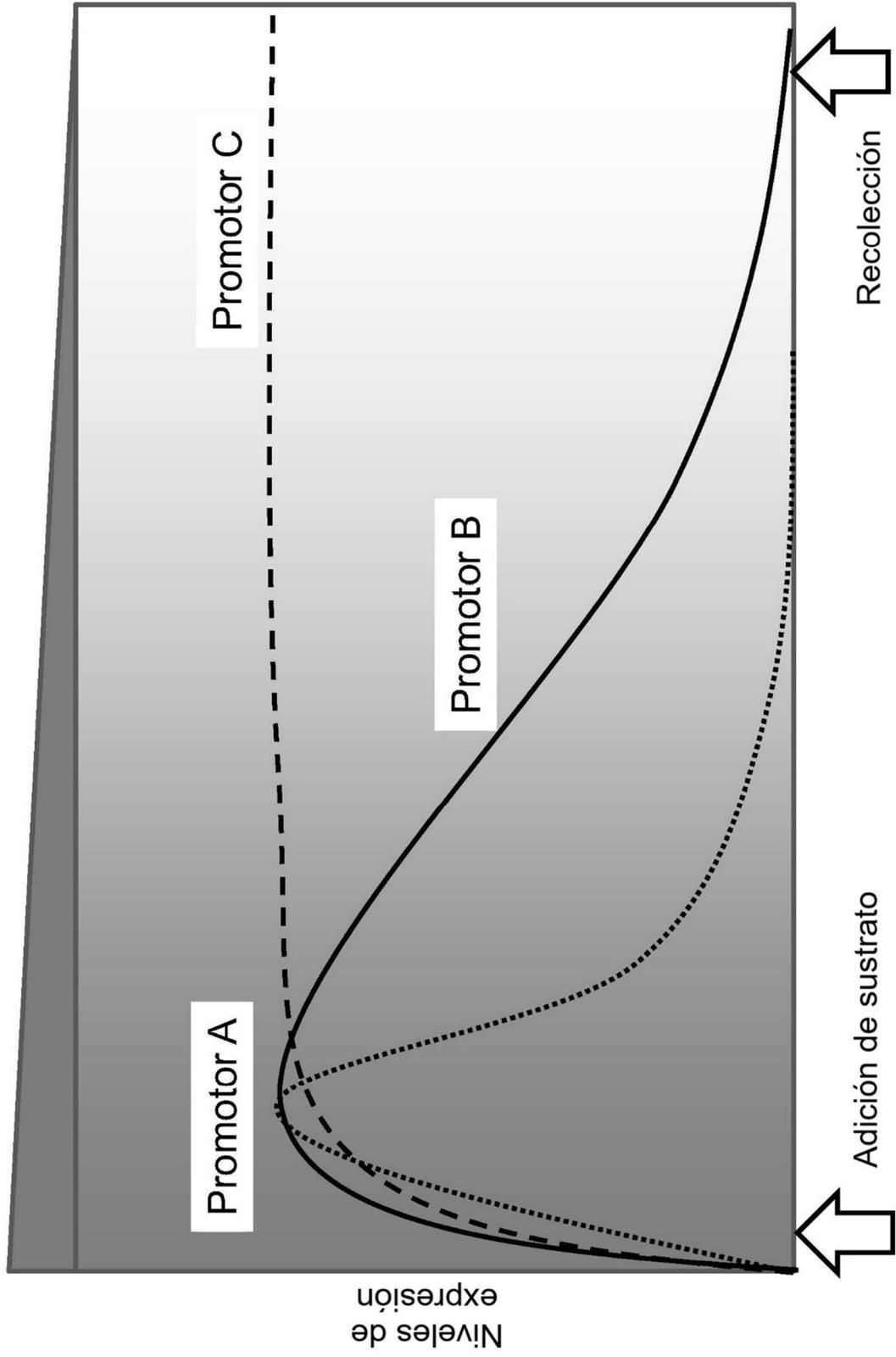


FIGURA 30

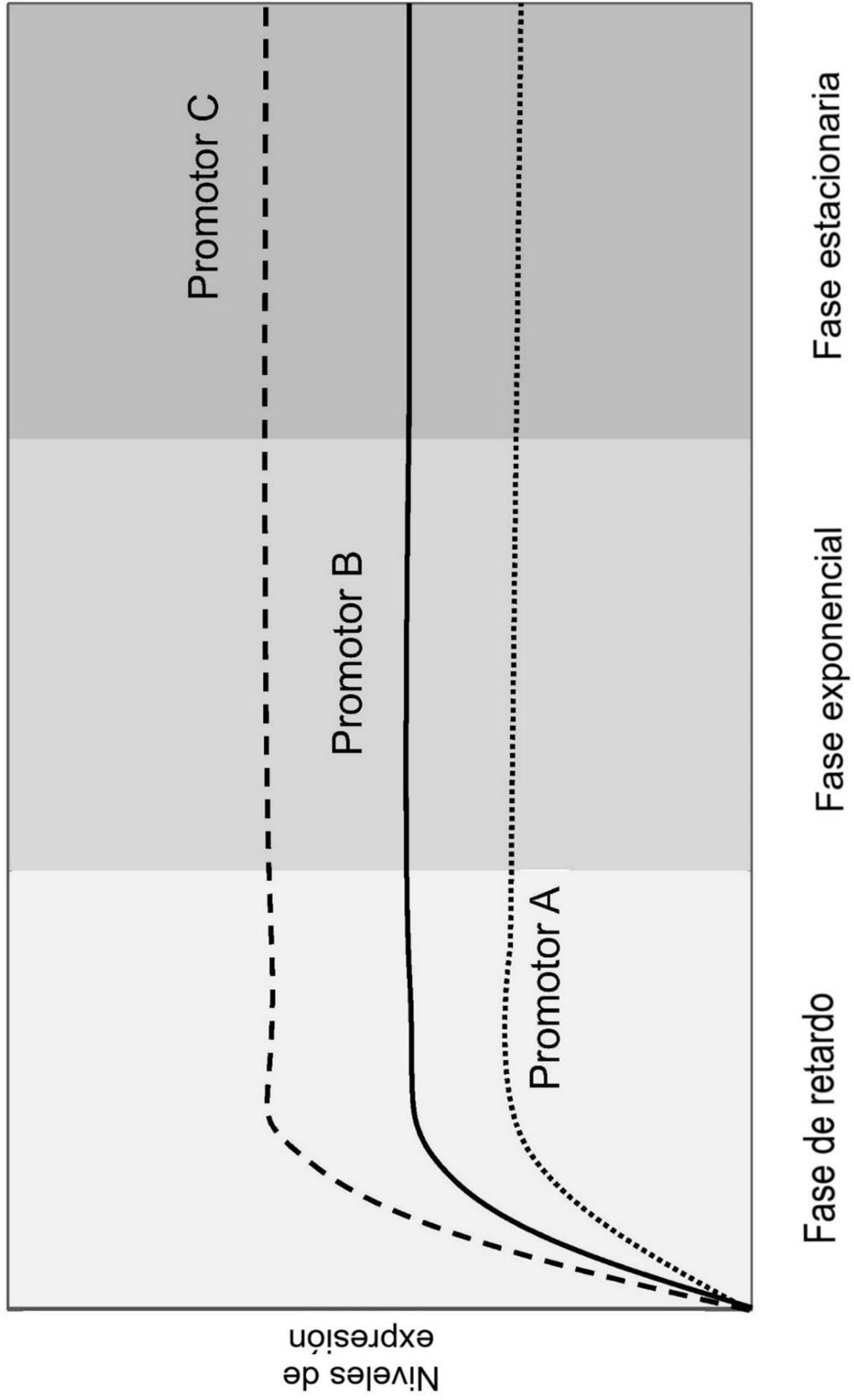


FIGURA 31

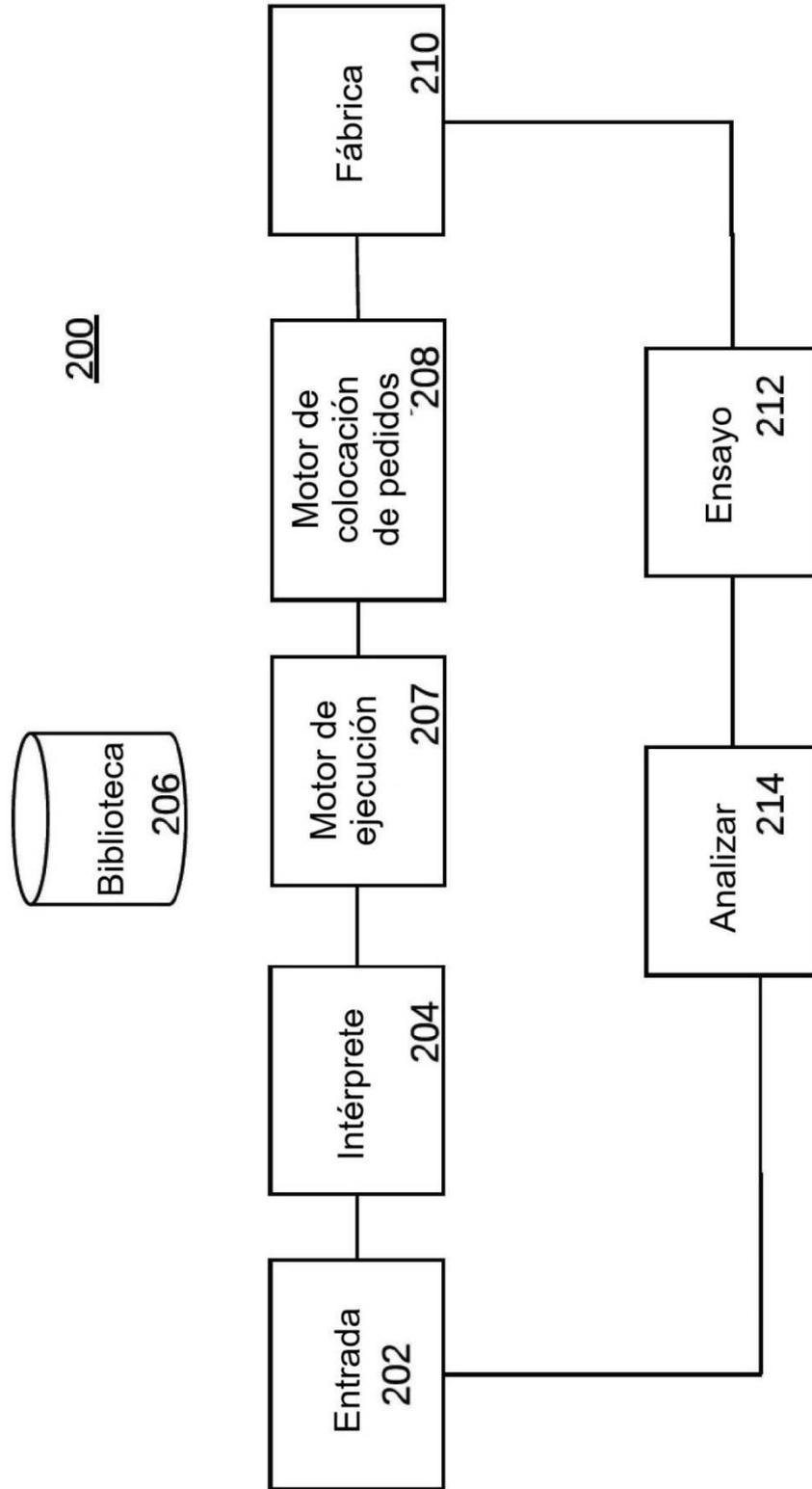


FIGURA 32

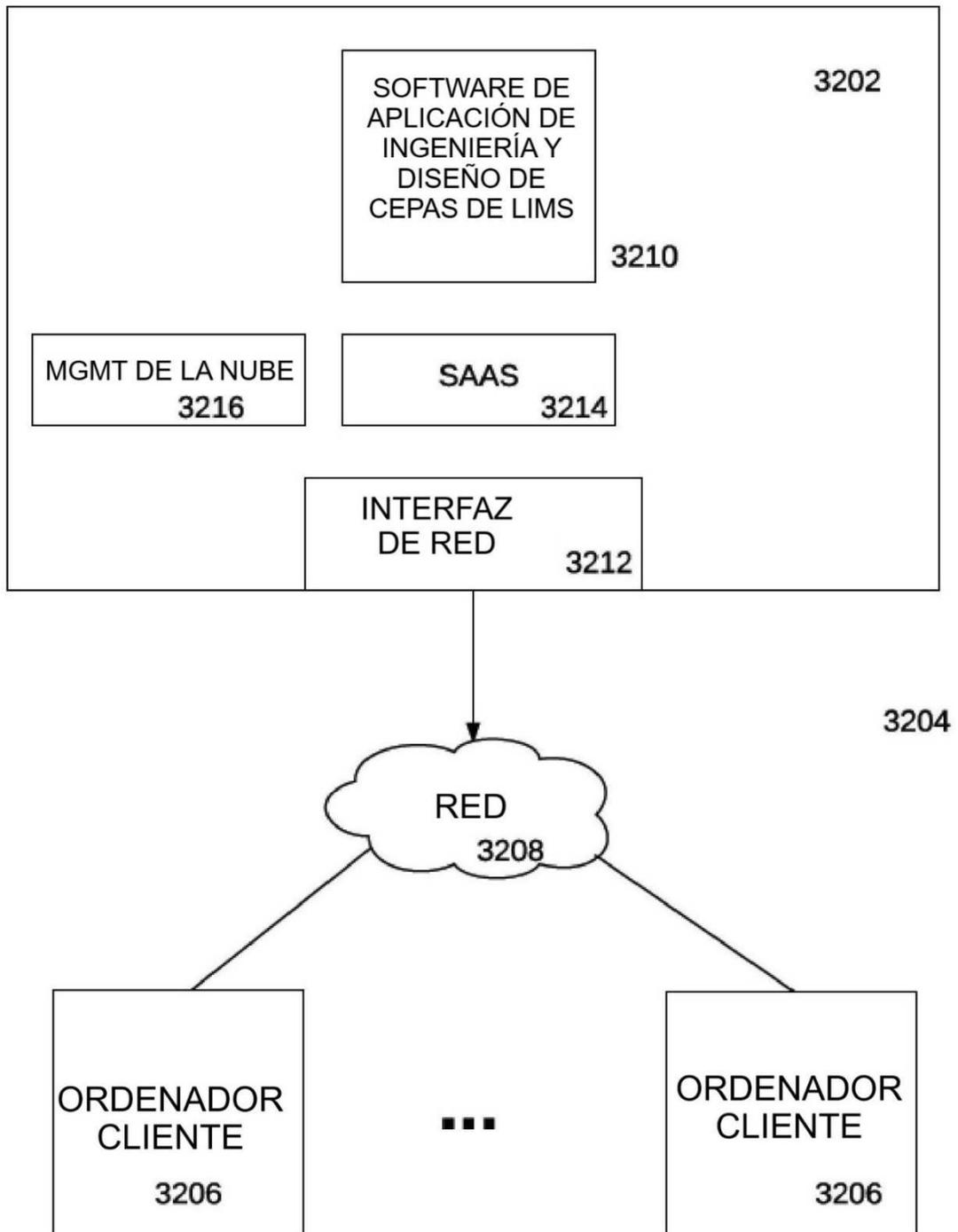


FIGURA 33

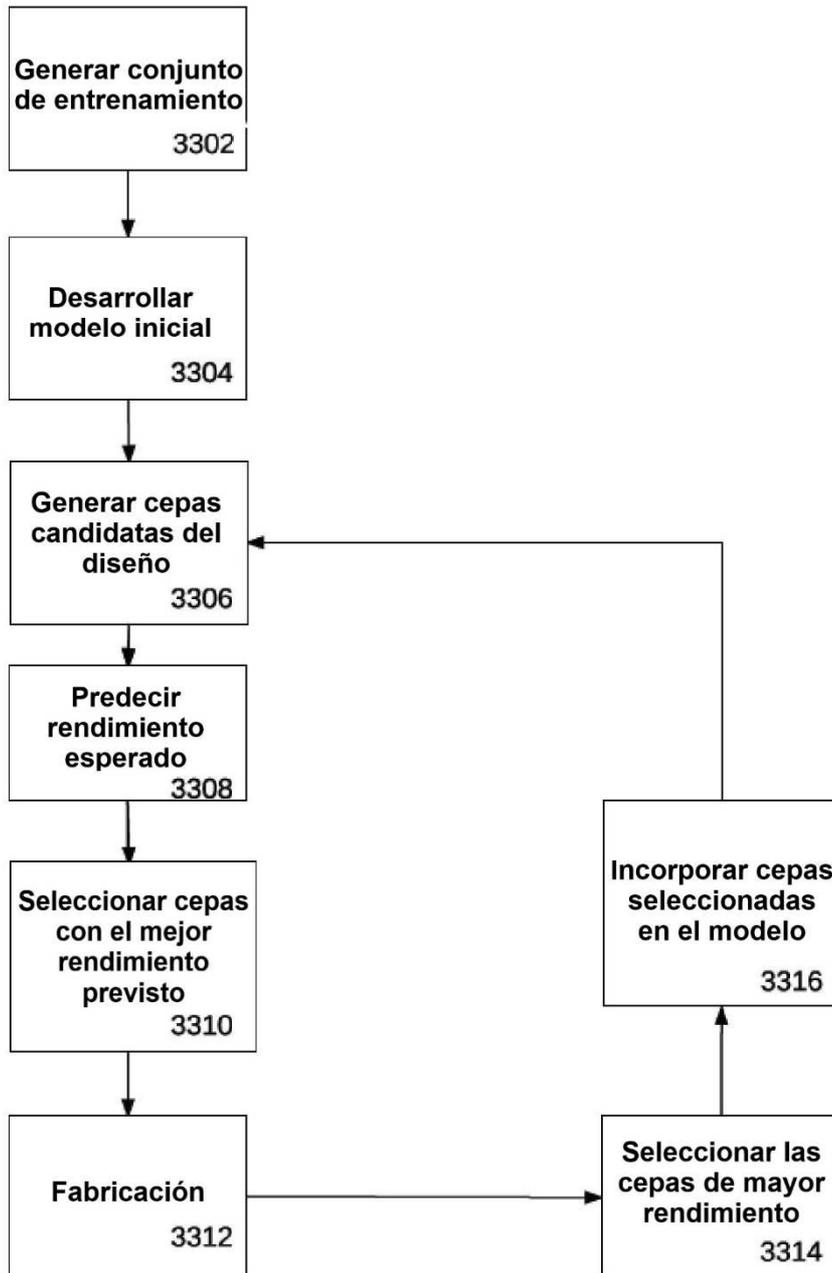


FIGURA 34

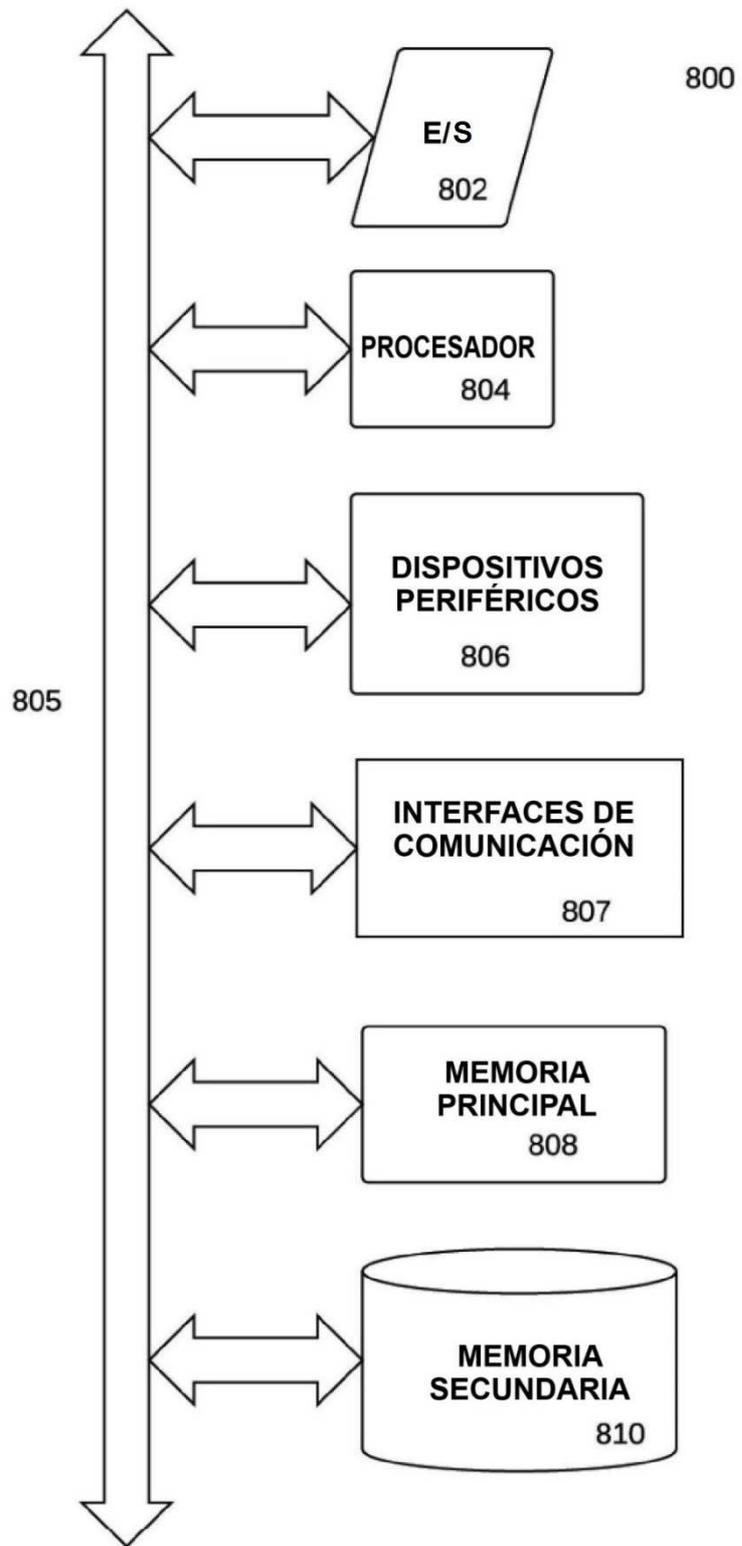


FIGURA 35

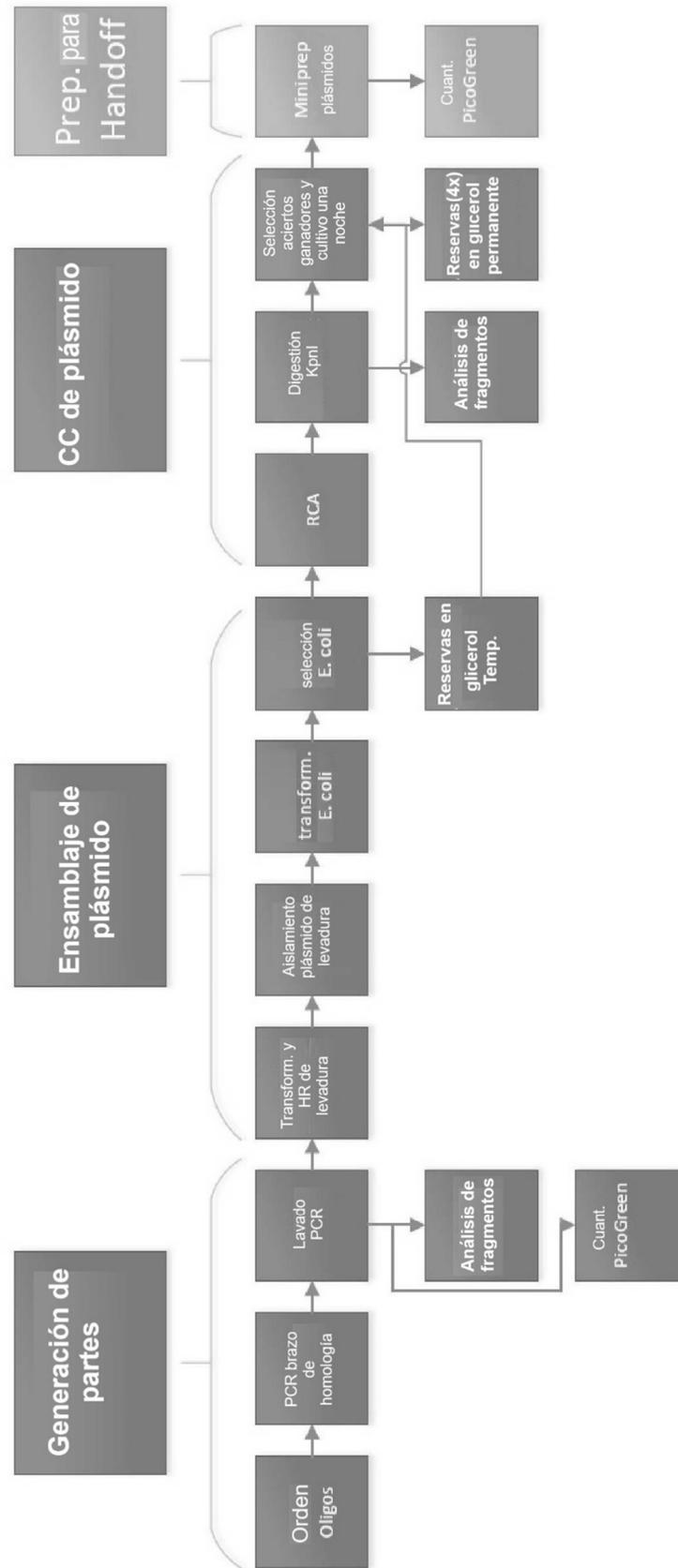


FIGURA 36

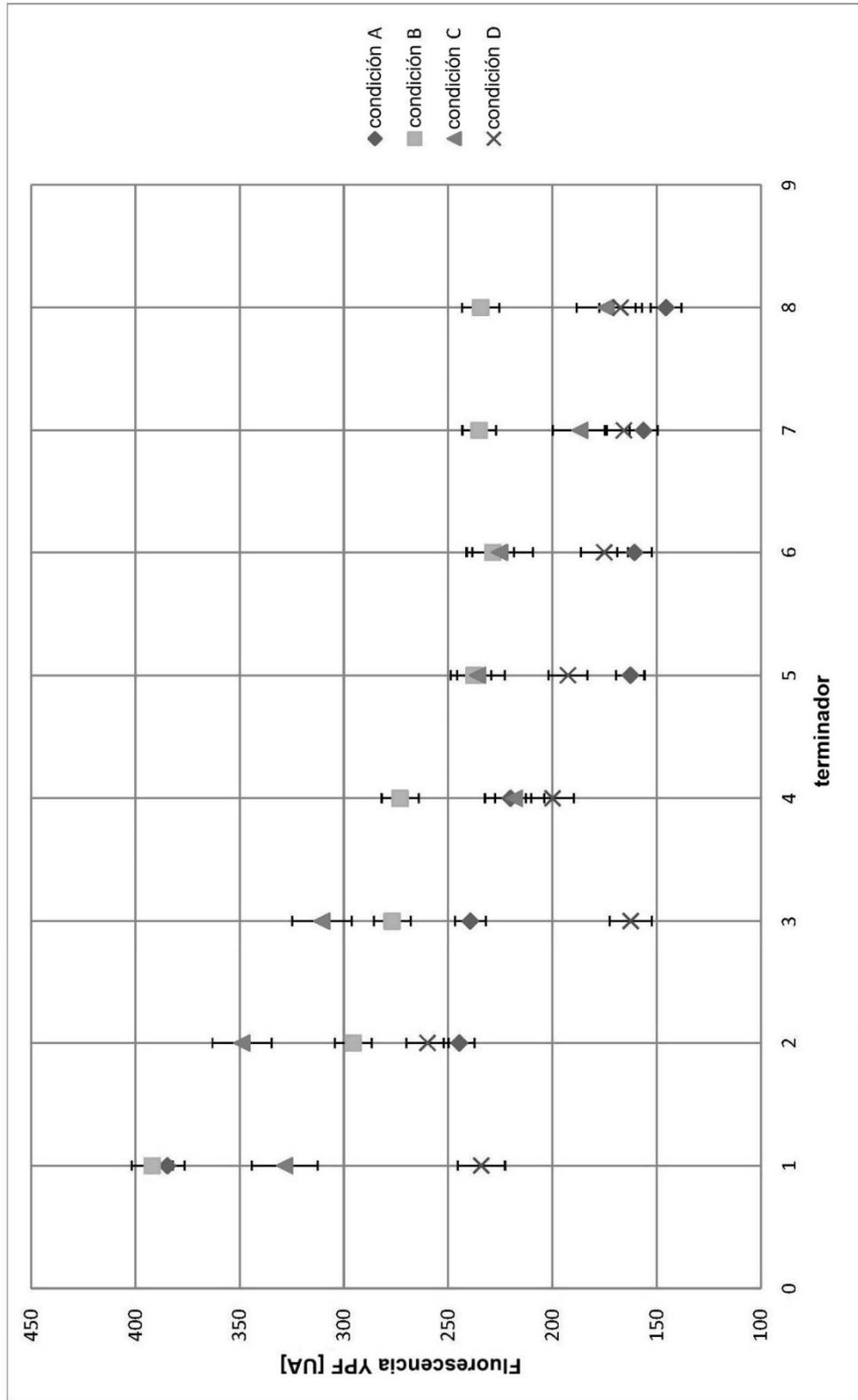


FIGURA 37

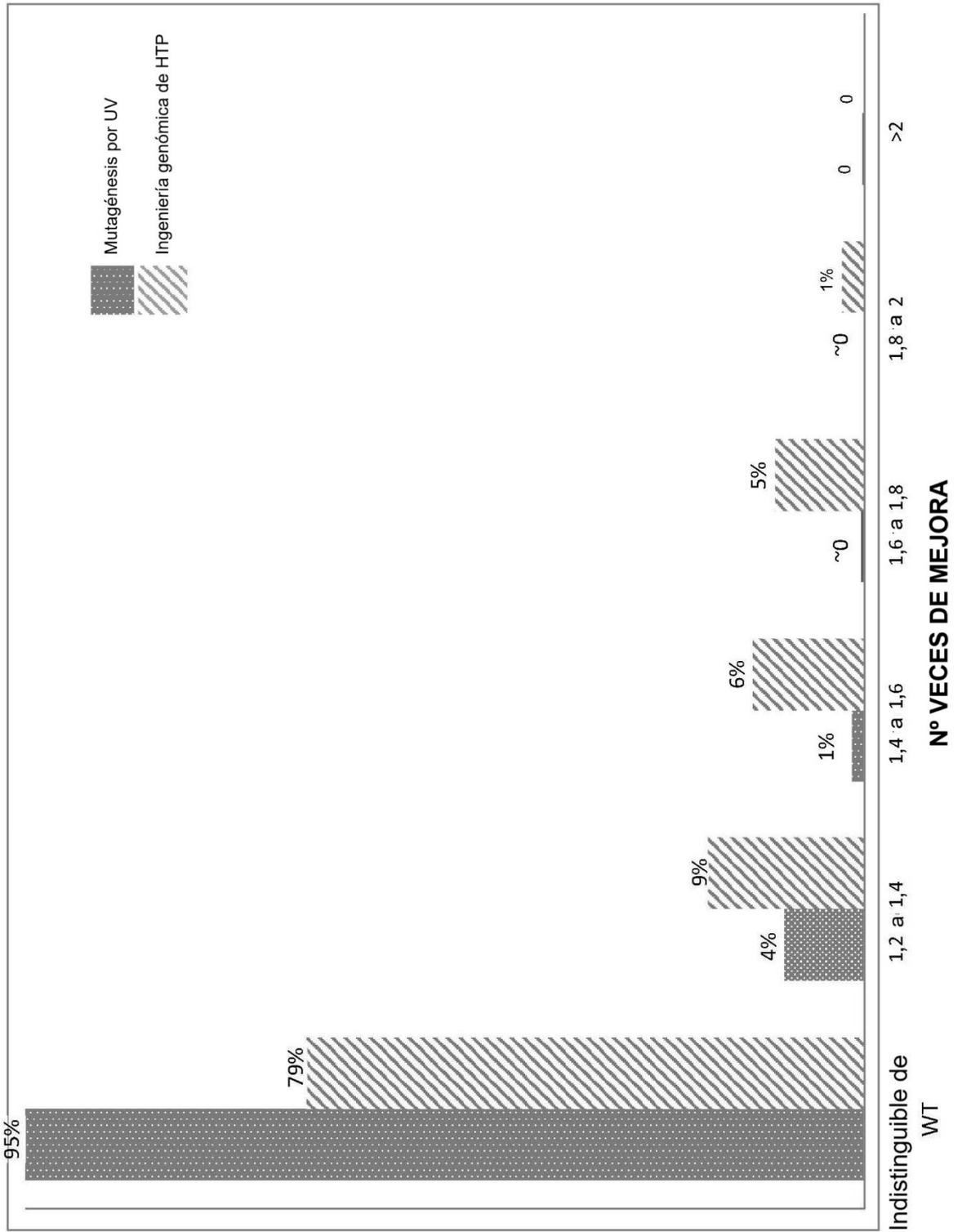


FIGURA 38

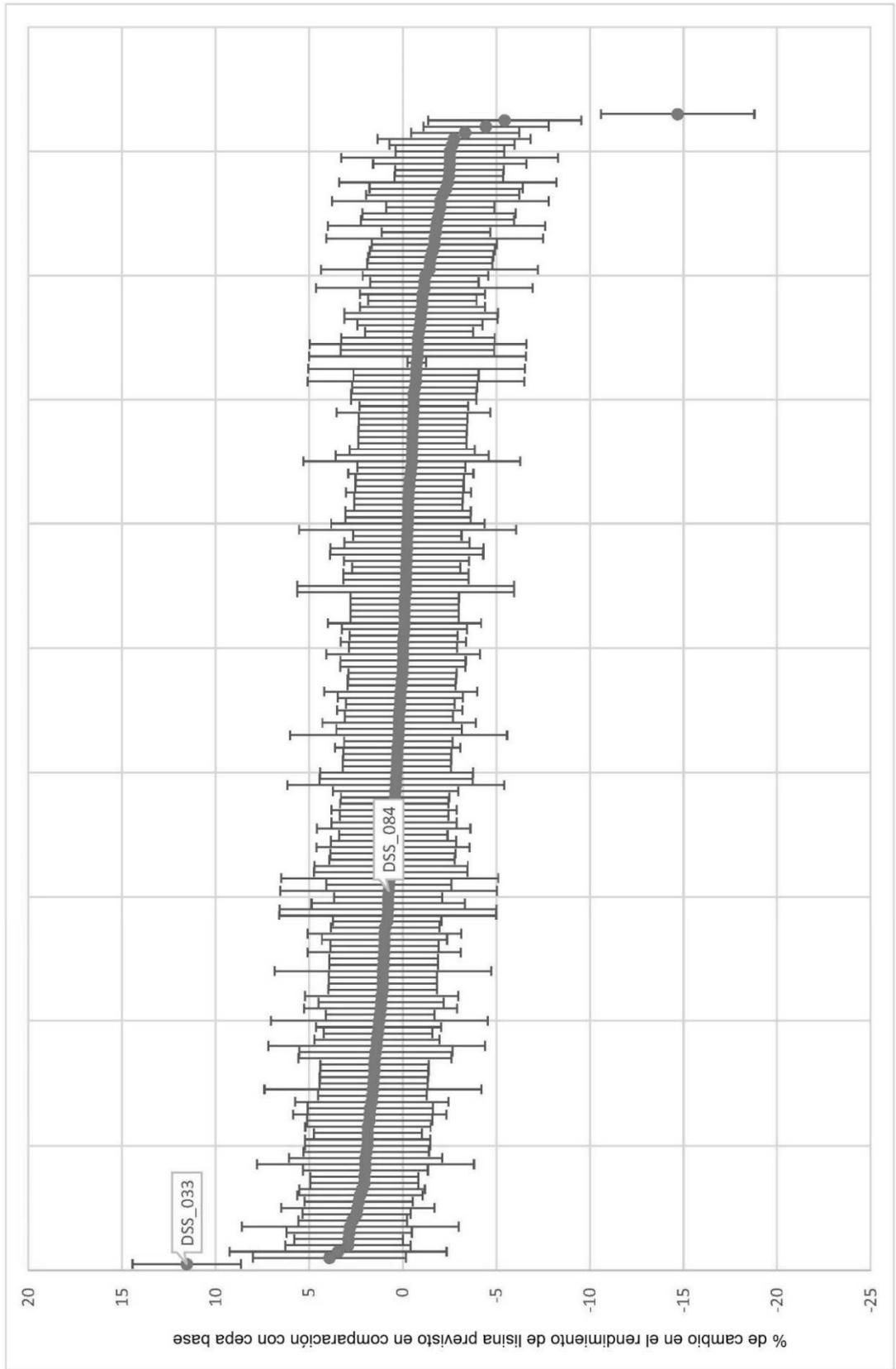


FIGURA 39

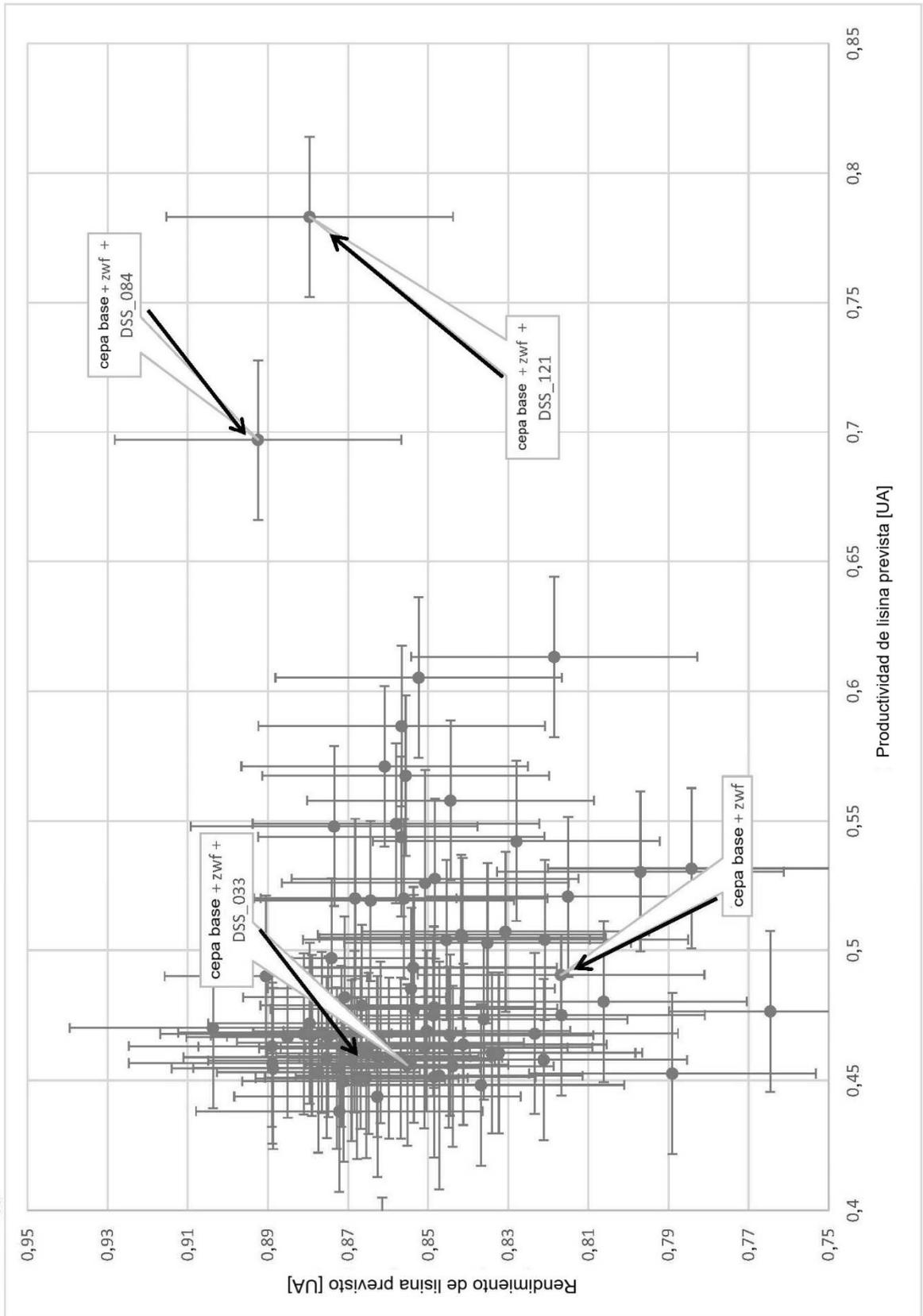


FIGURA 40

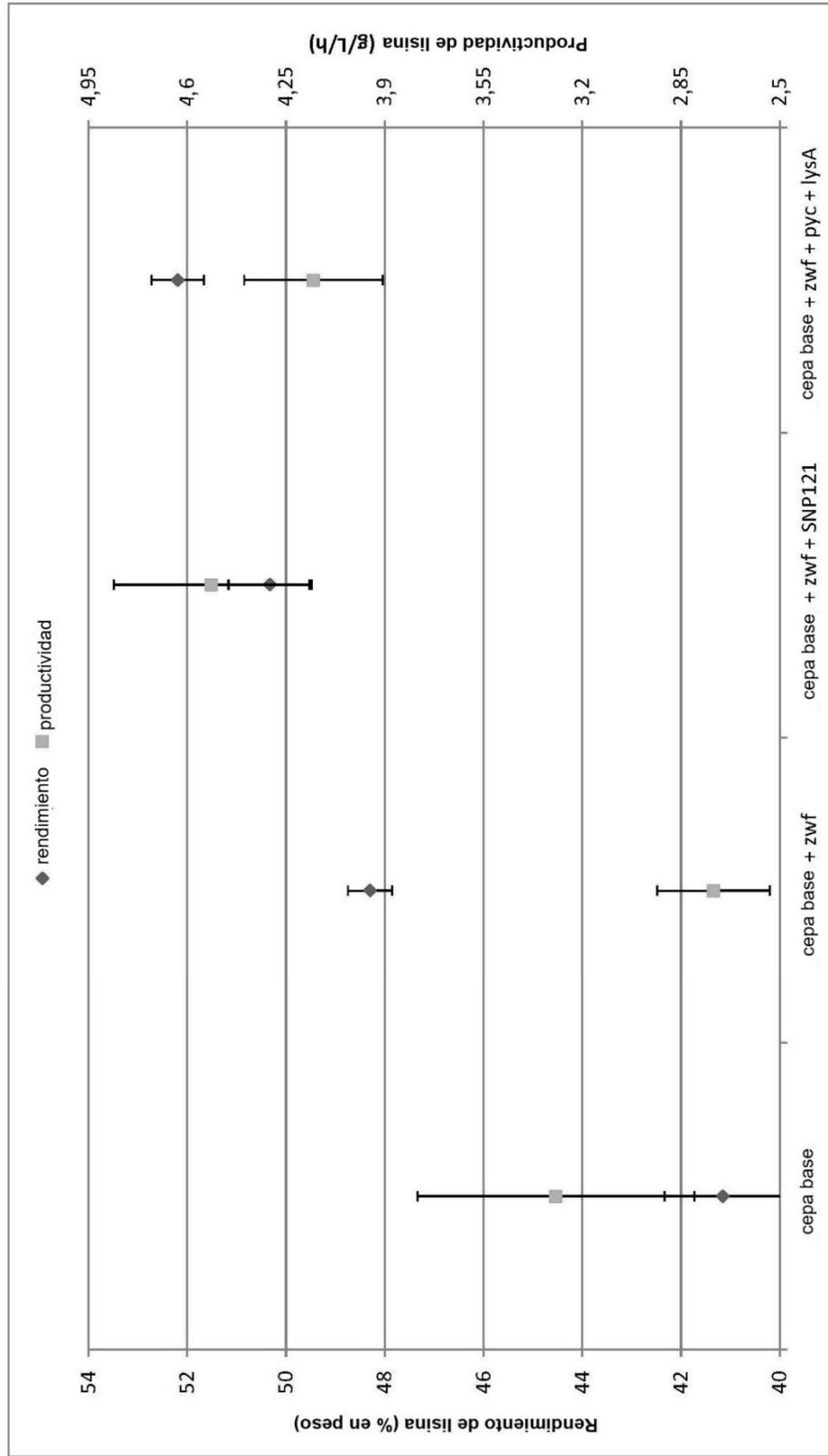


FIGURA 41

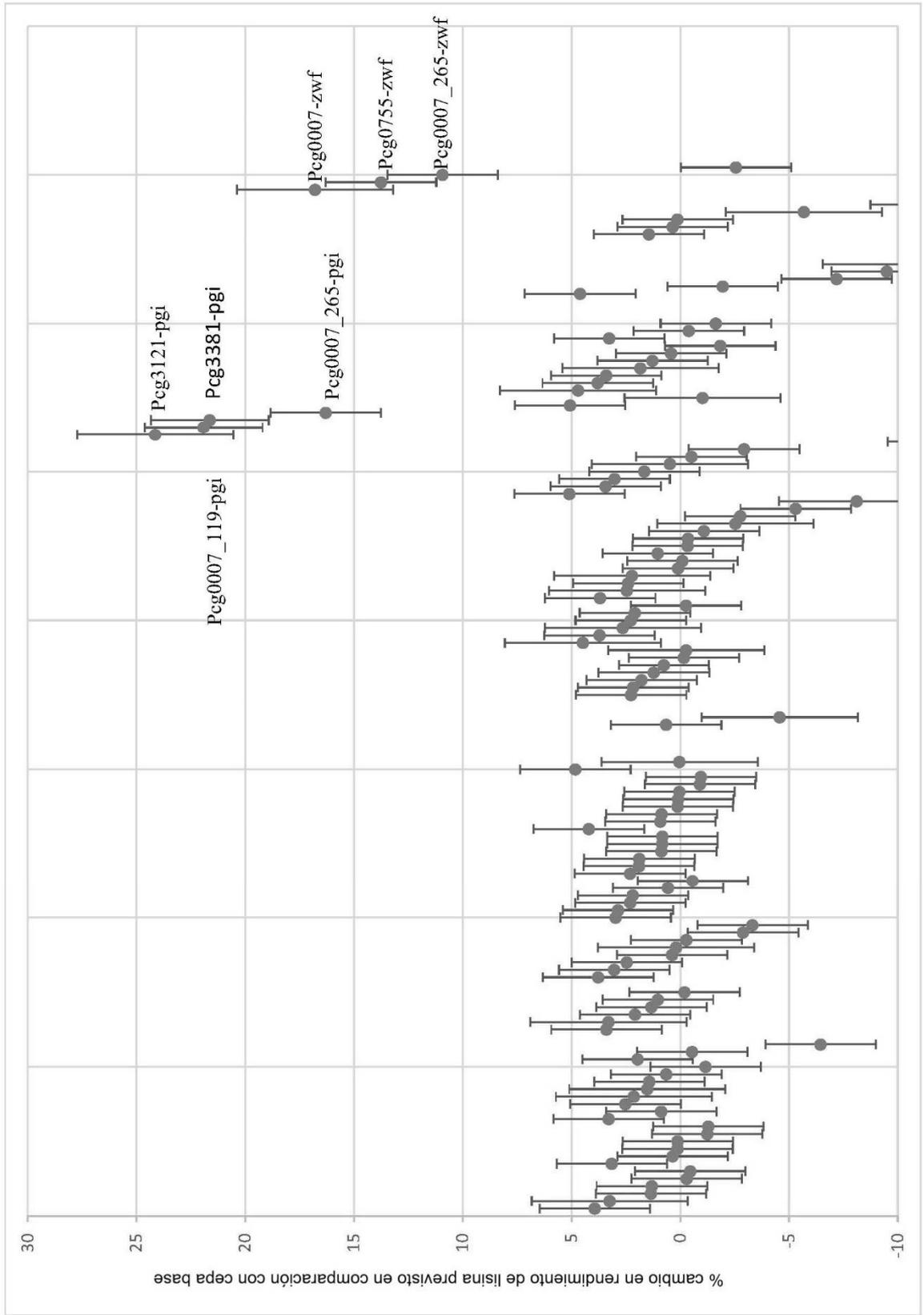


FIGURA 42

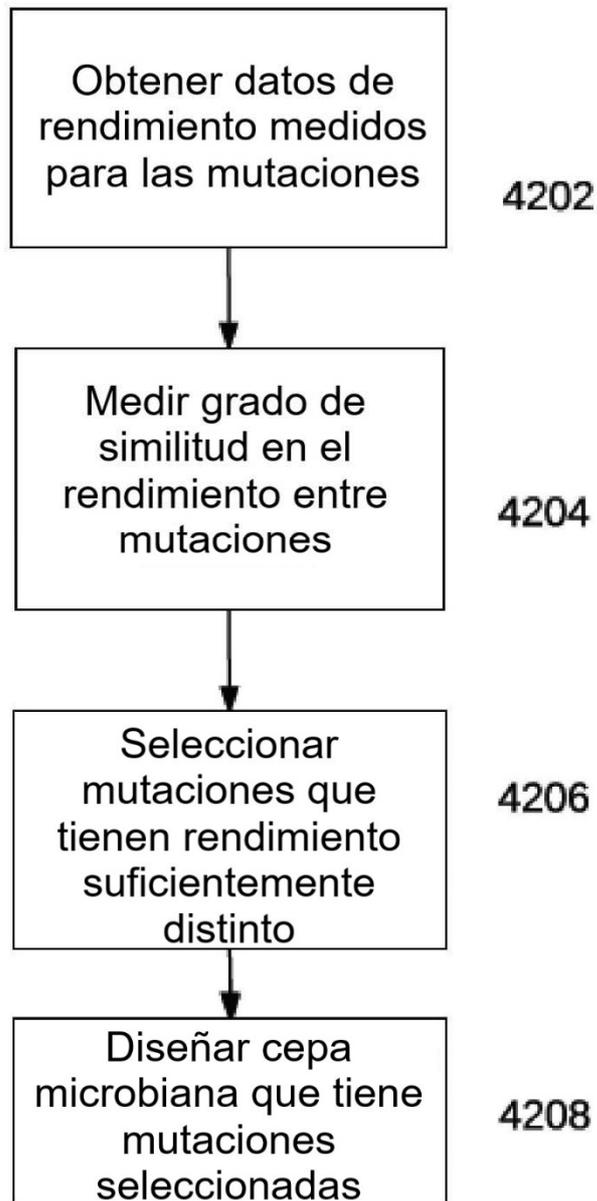


FIGURA 44

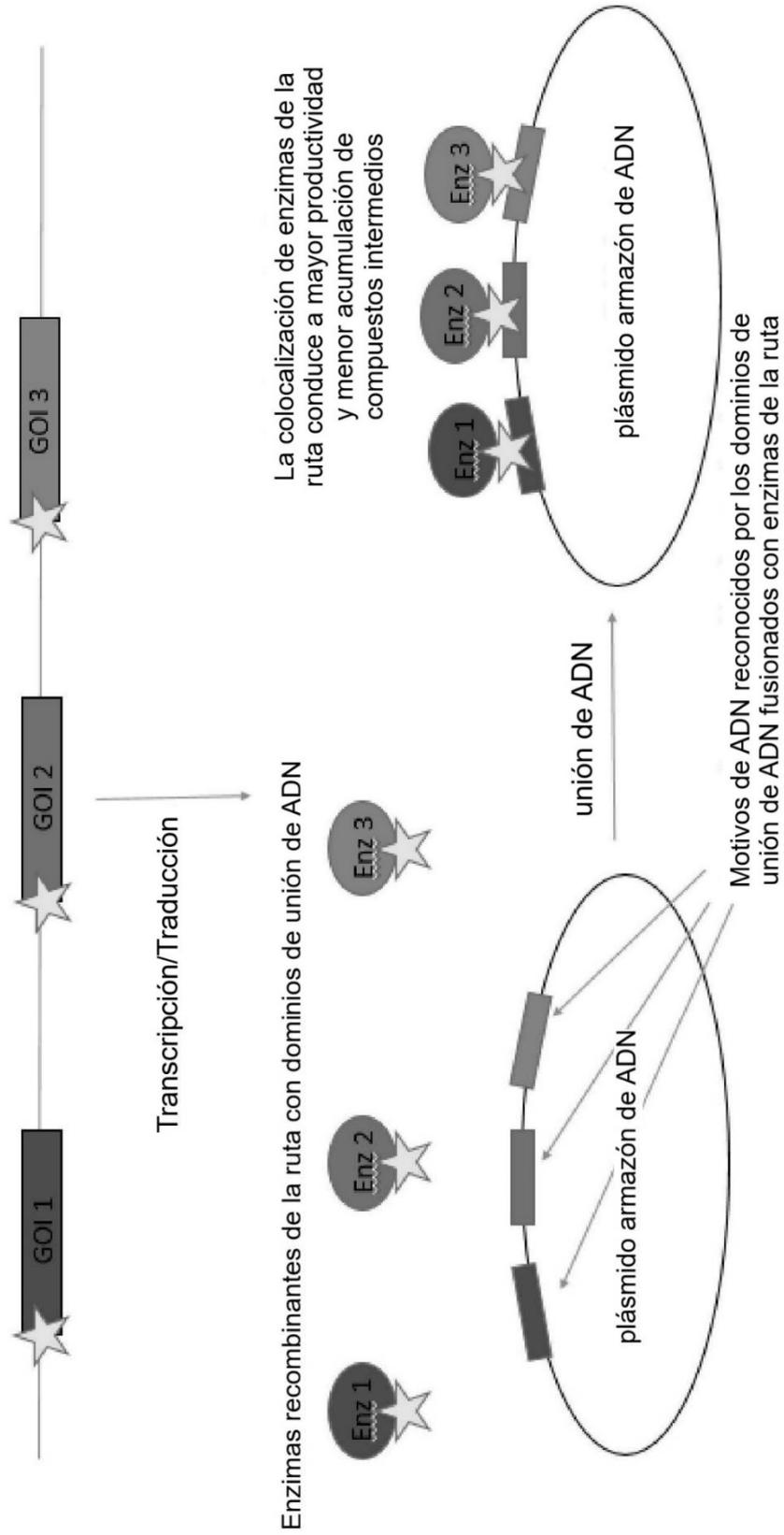


FIGURA 45

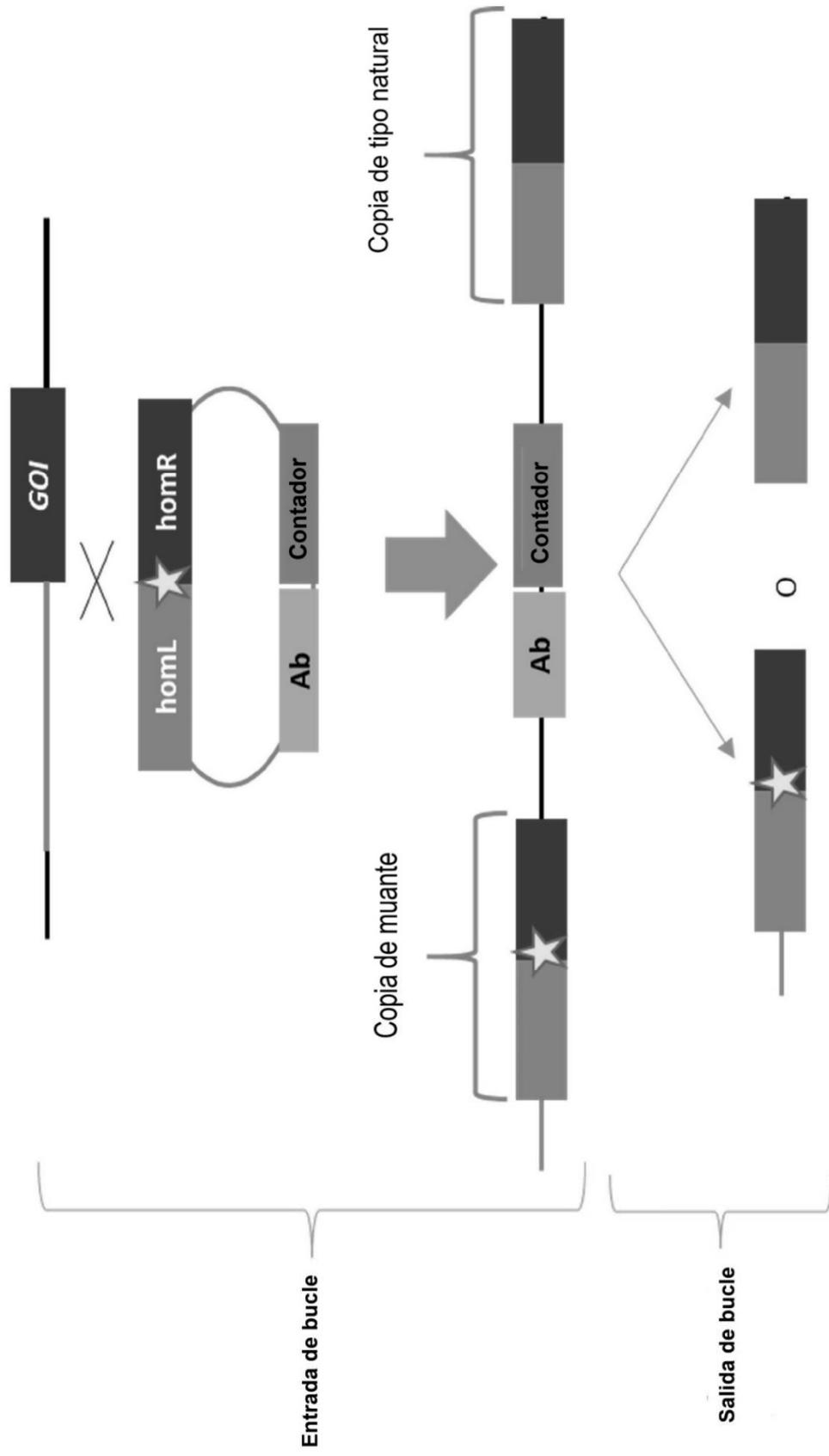


FIGURA 46

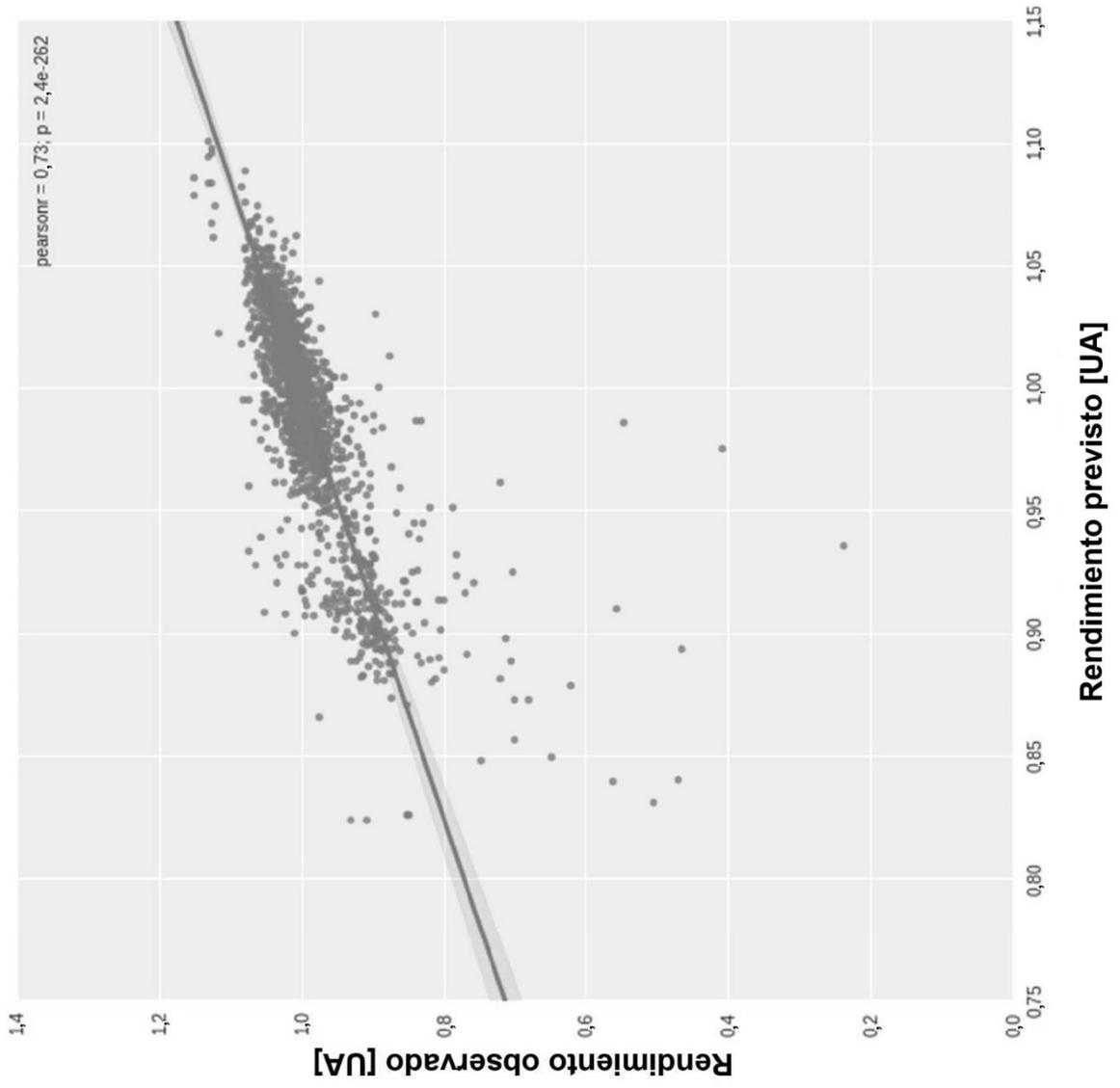


FIGURA 47

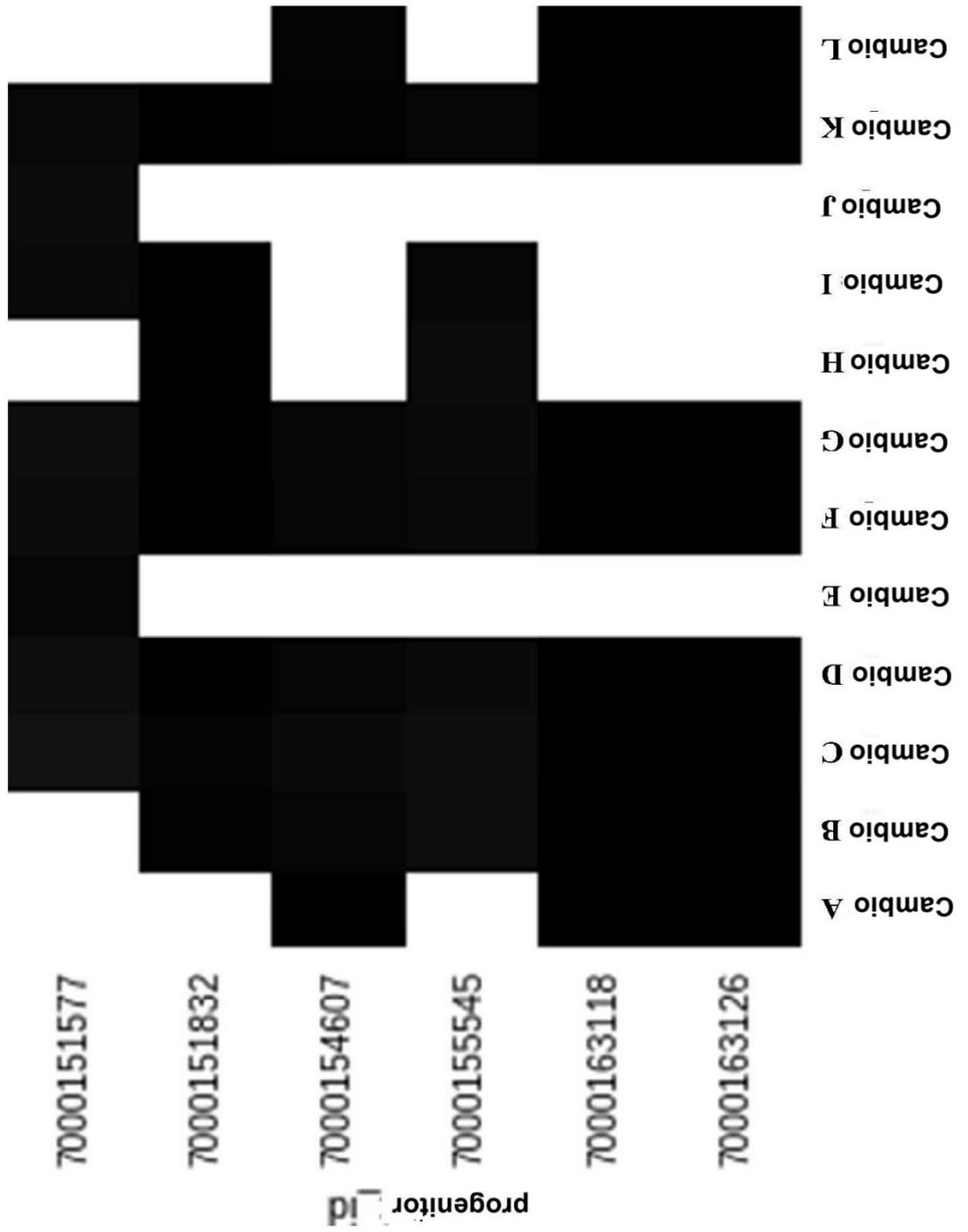


FIGURA 48

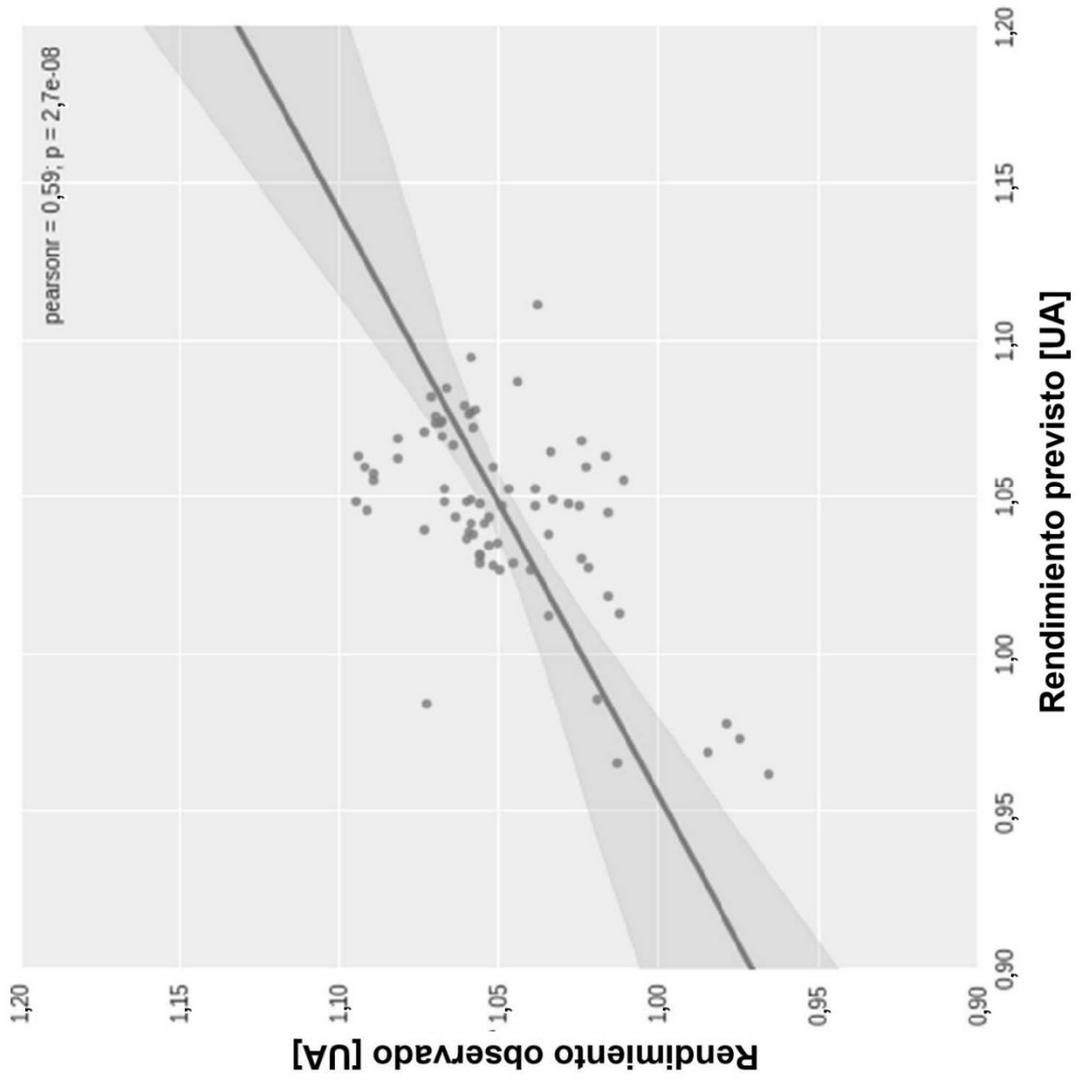
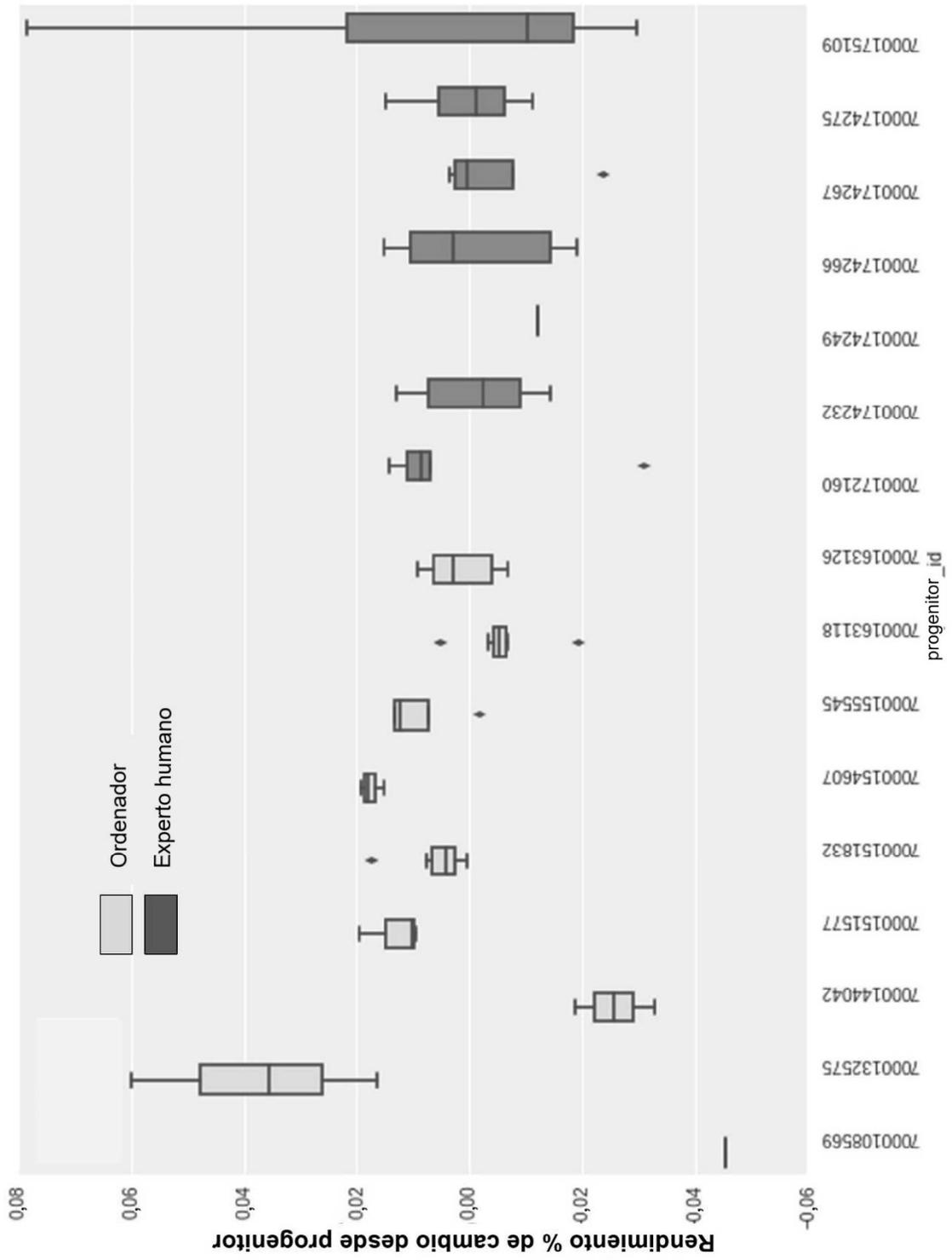


FIGURA 49



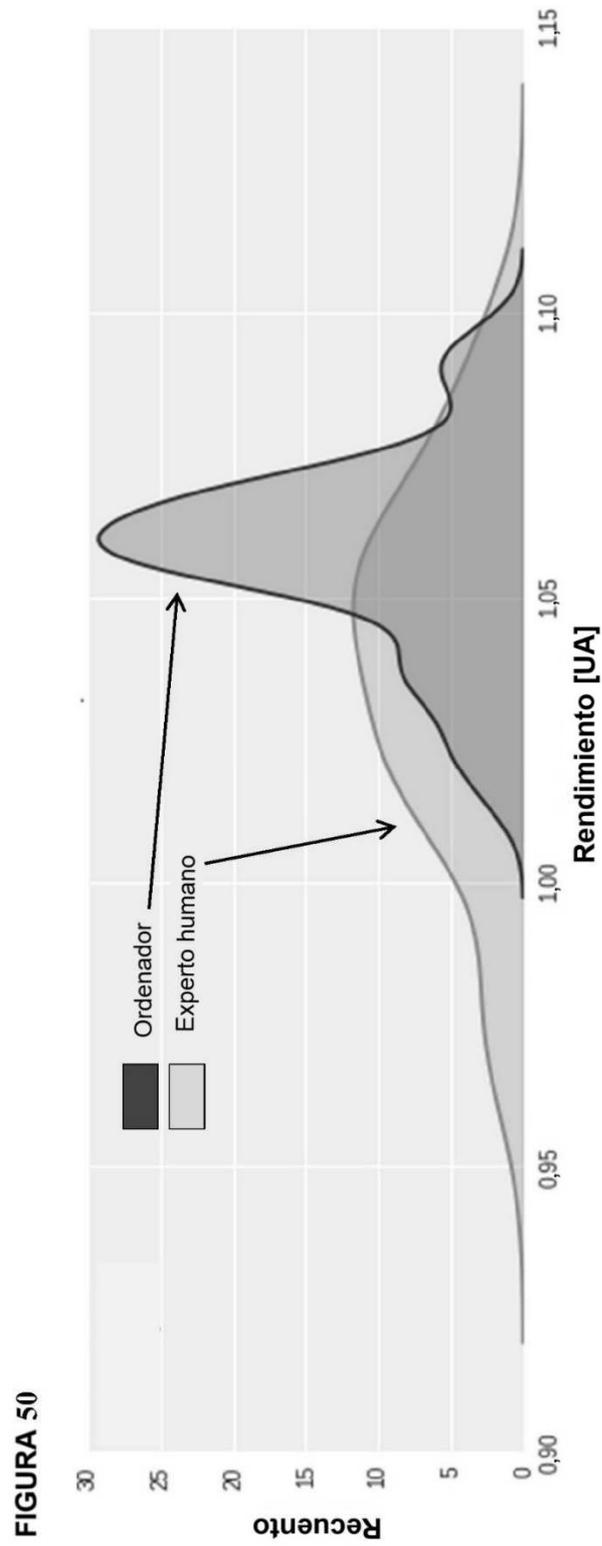


FIGURA 51

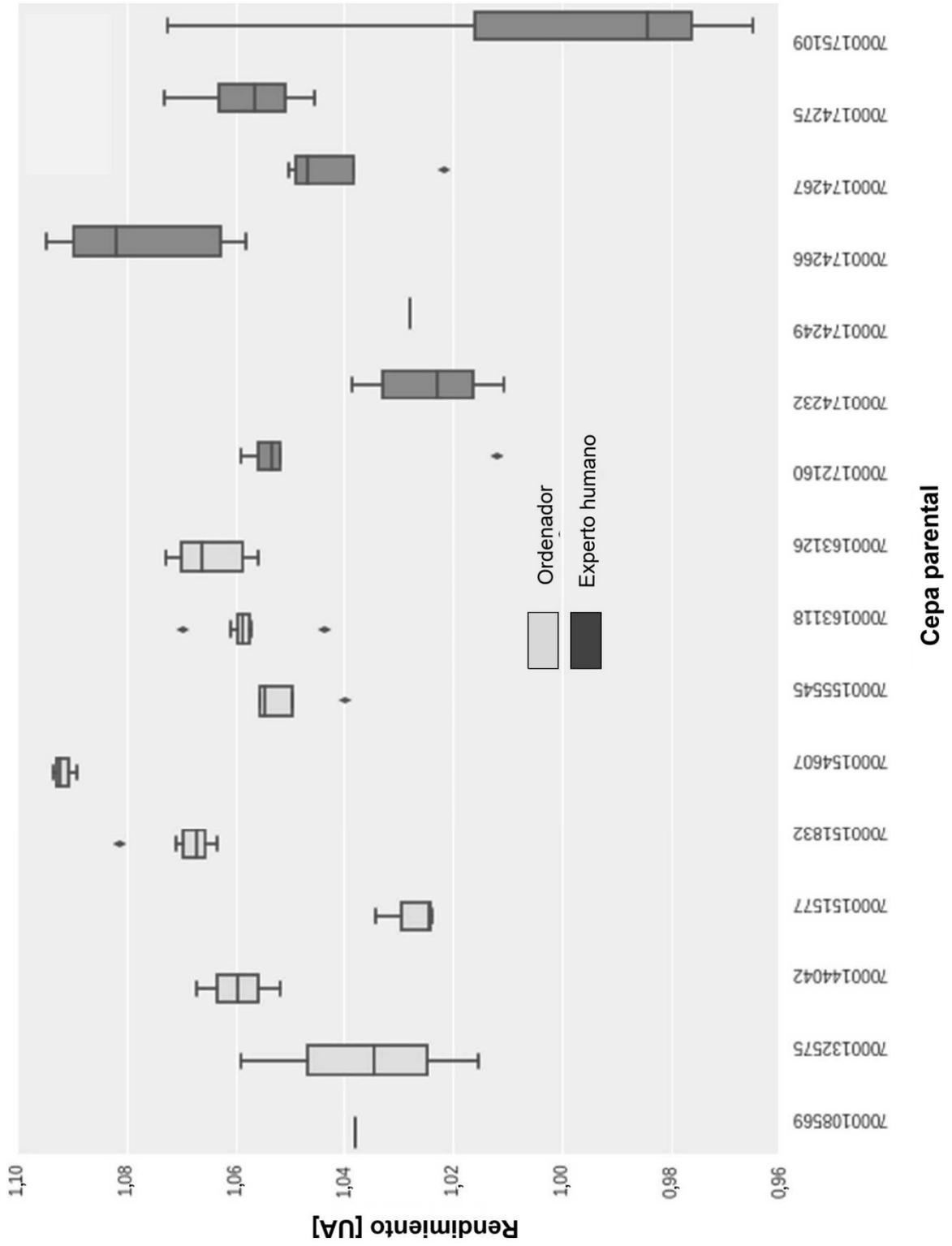


FIGURA 52

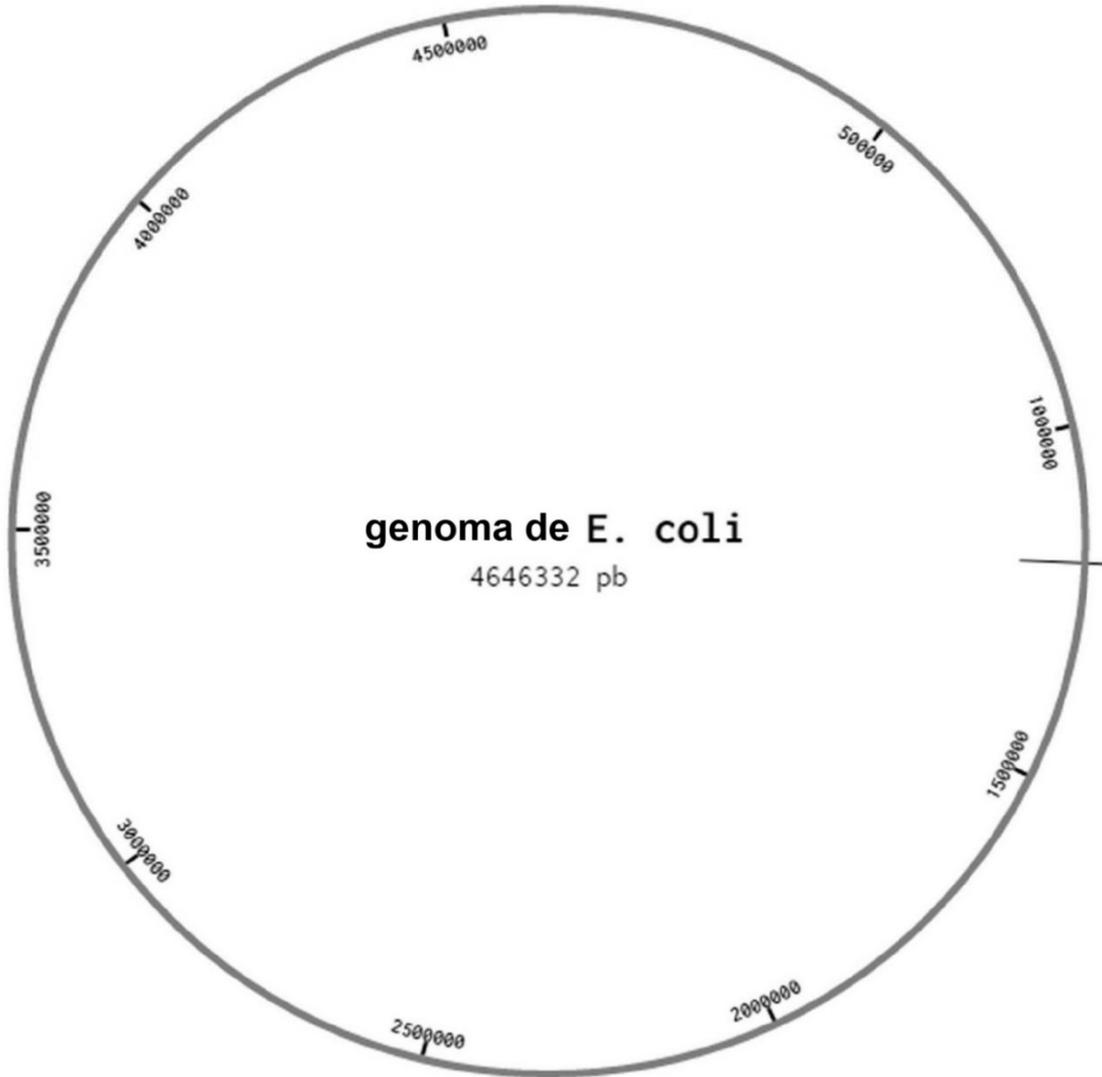
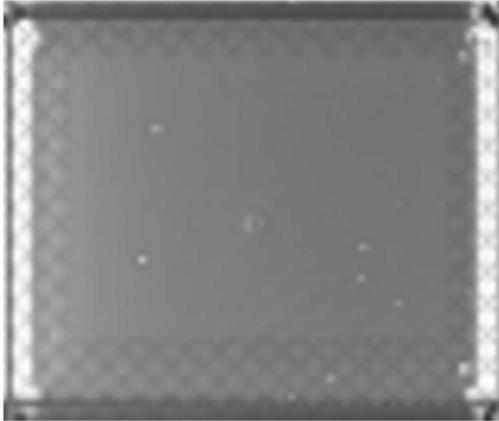


FIGURA 53

aroA-PS en
versión 1



aroA-PS en
versión 2

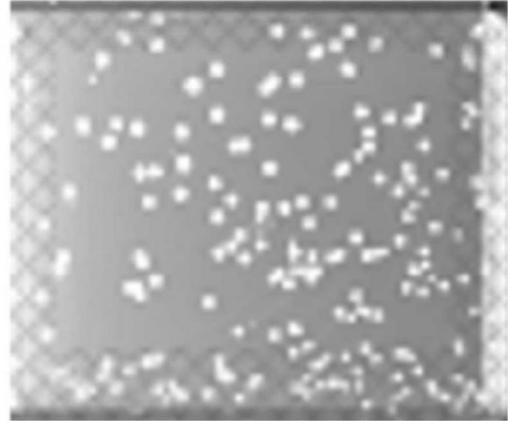


FIGURA 54



FIGURA 55

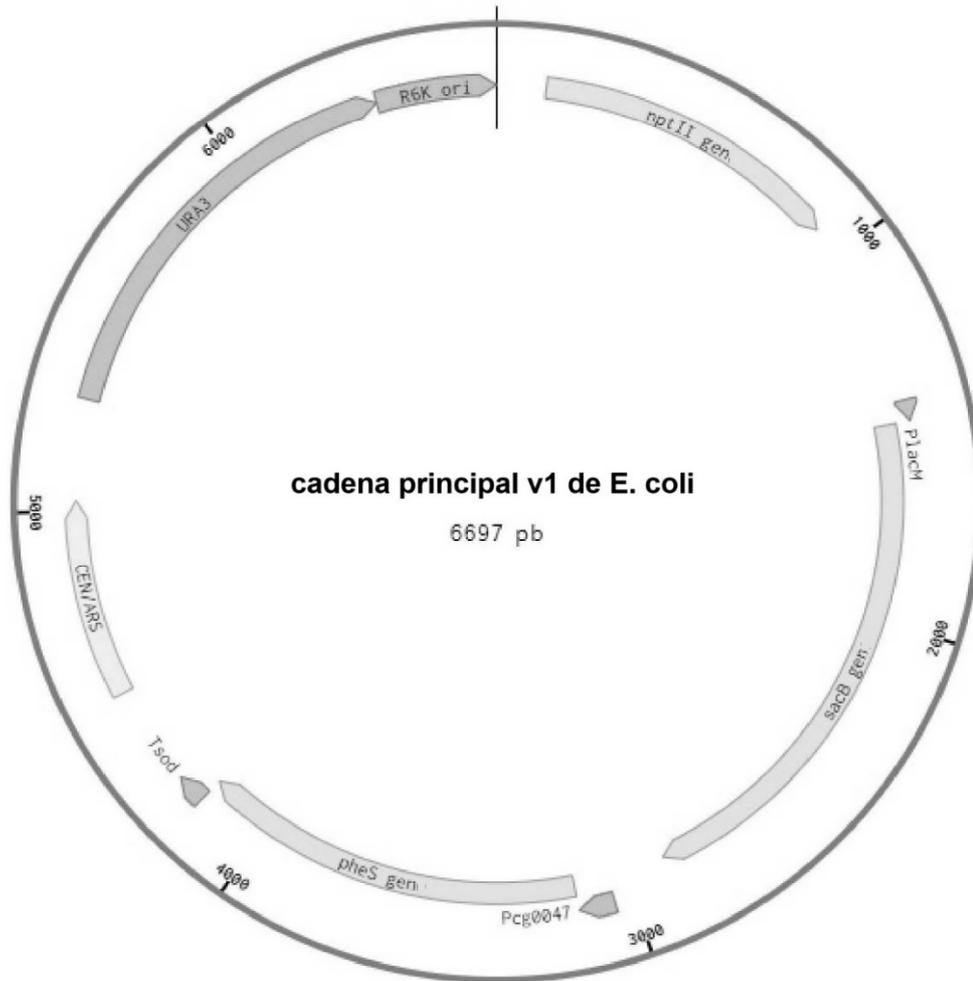


FIGURE 56

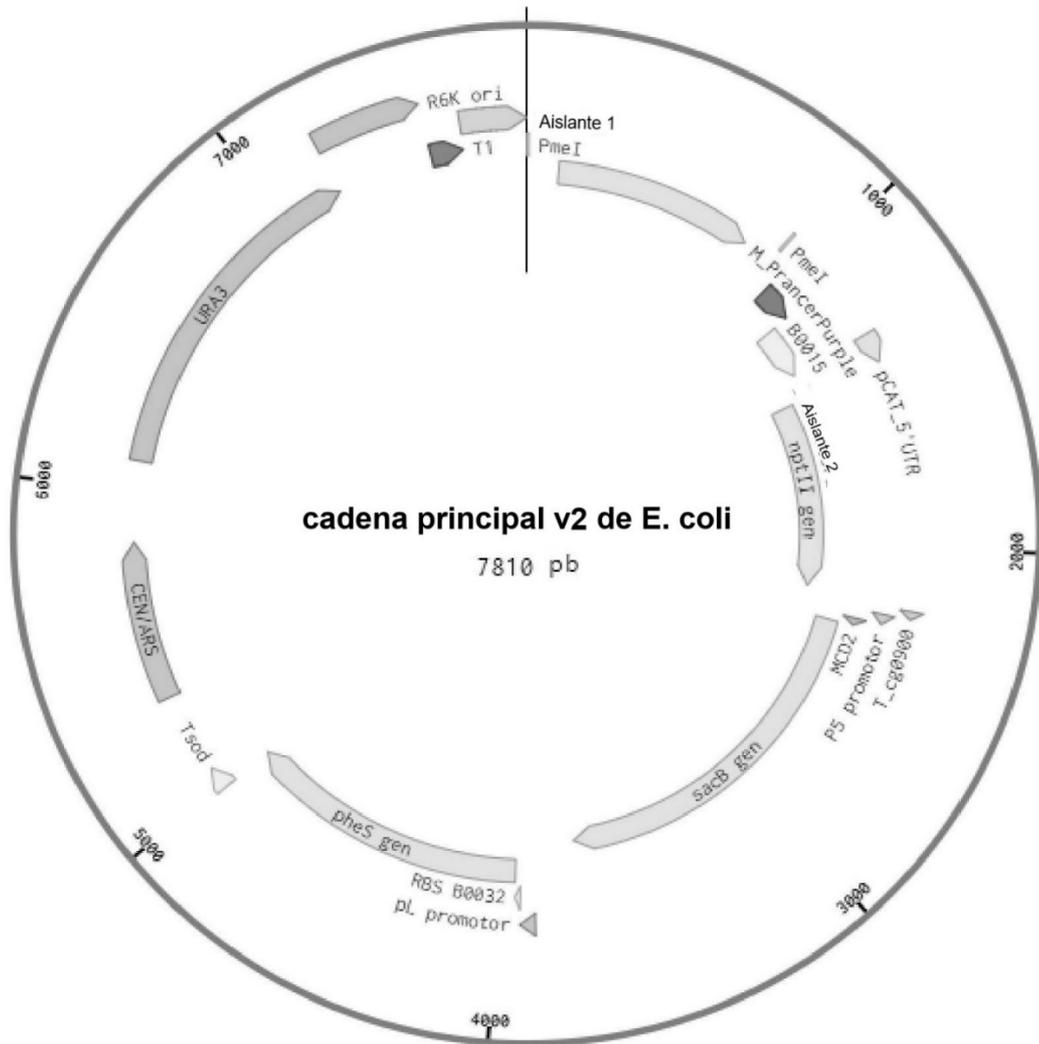


FIGURA 57

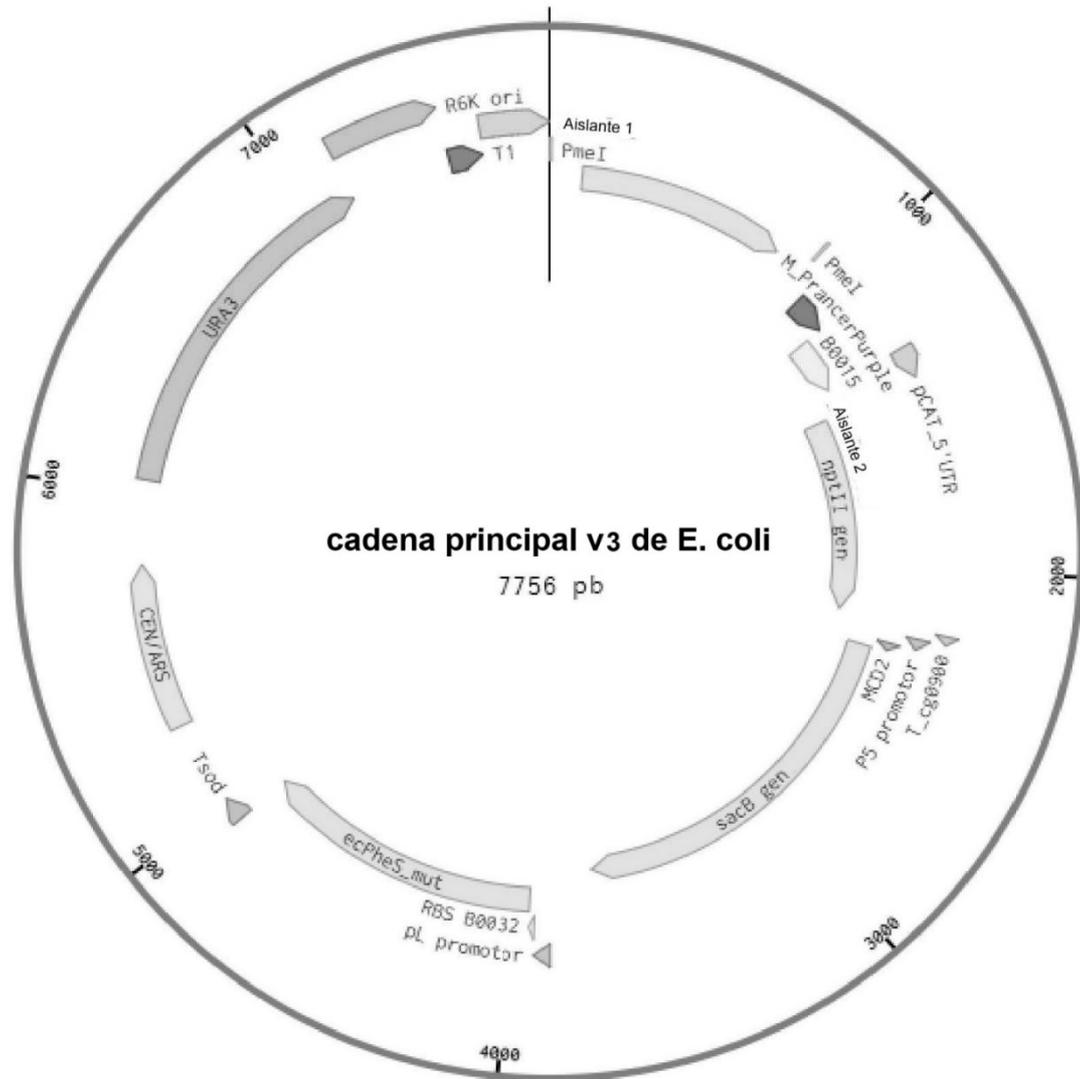


FIGURA 58

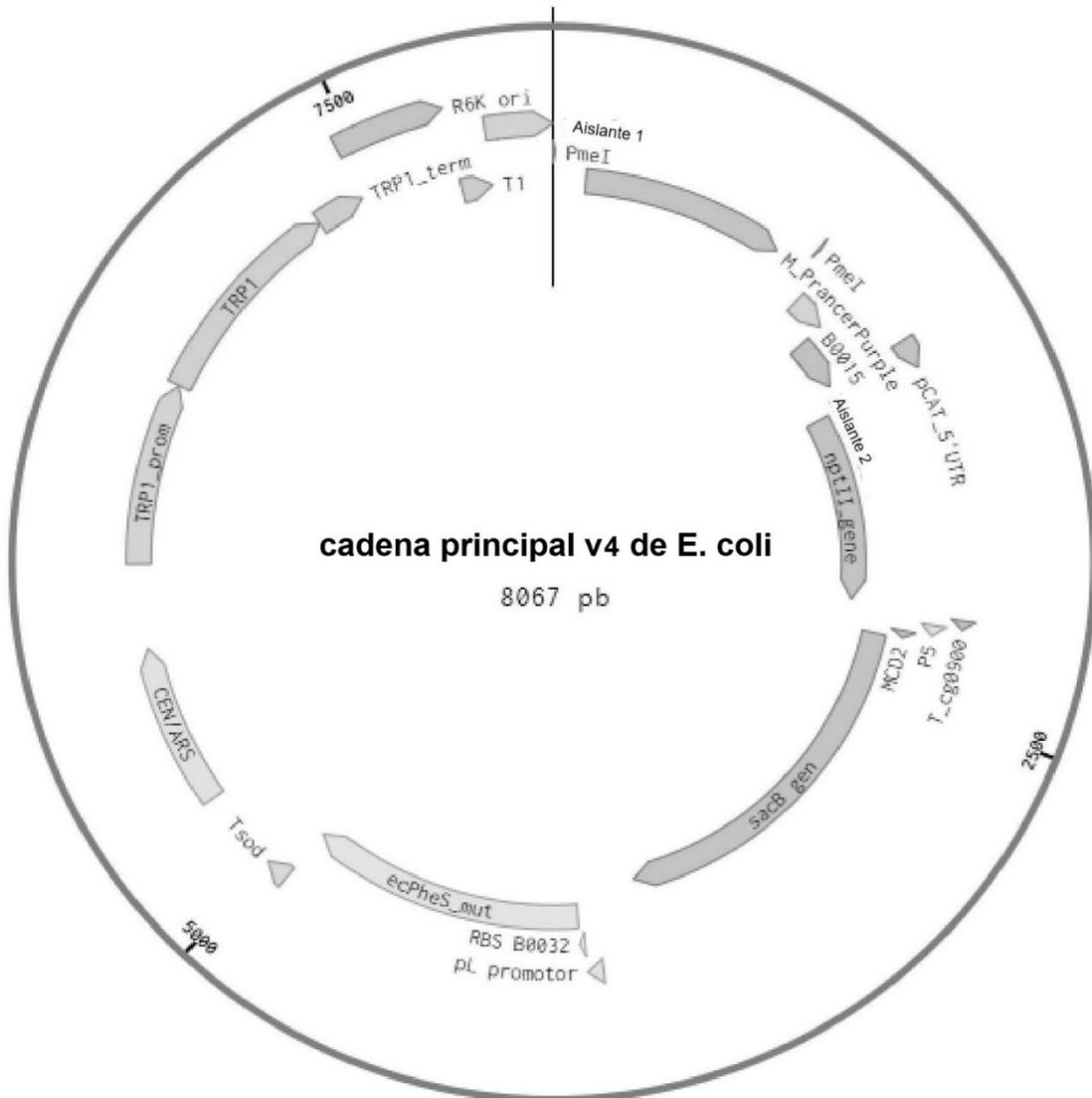


FIGURA 59

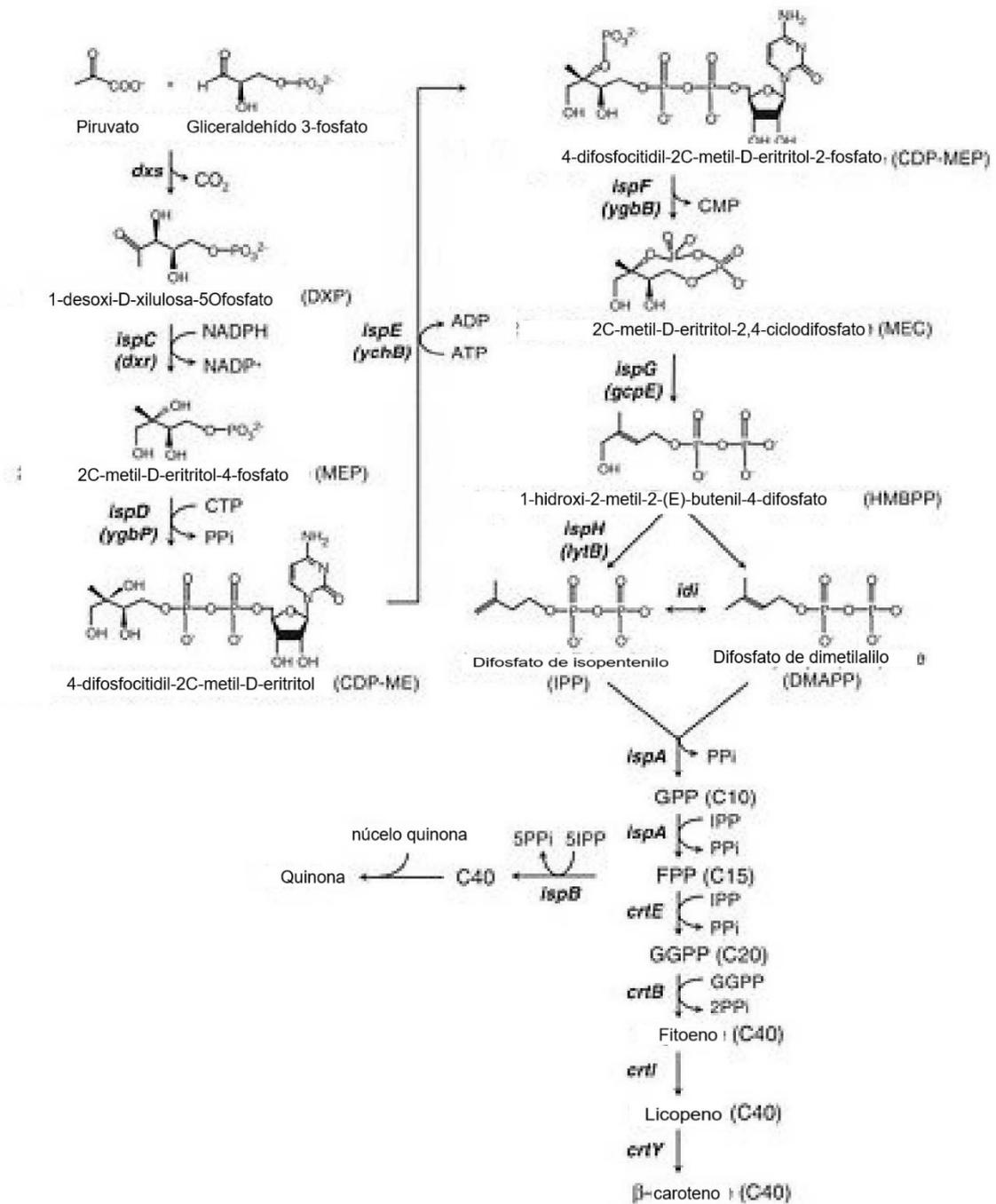


FIGURA 60

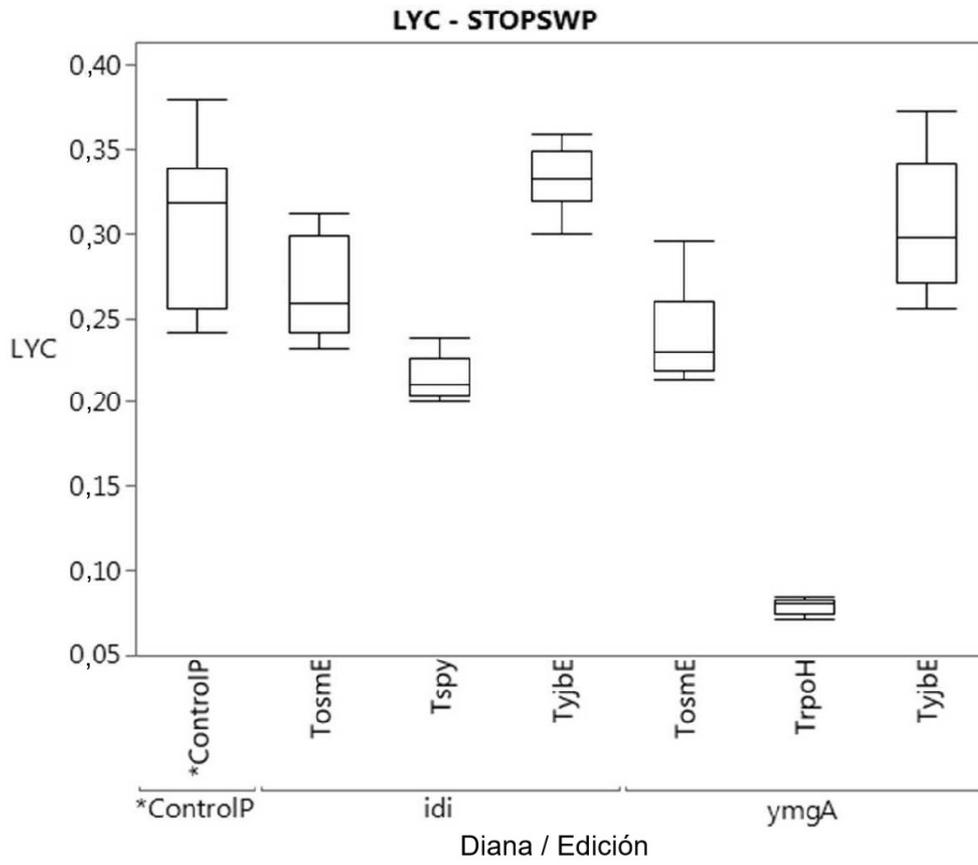


FIGURA 61

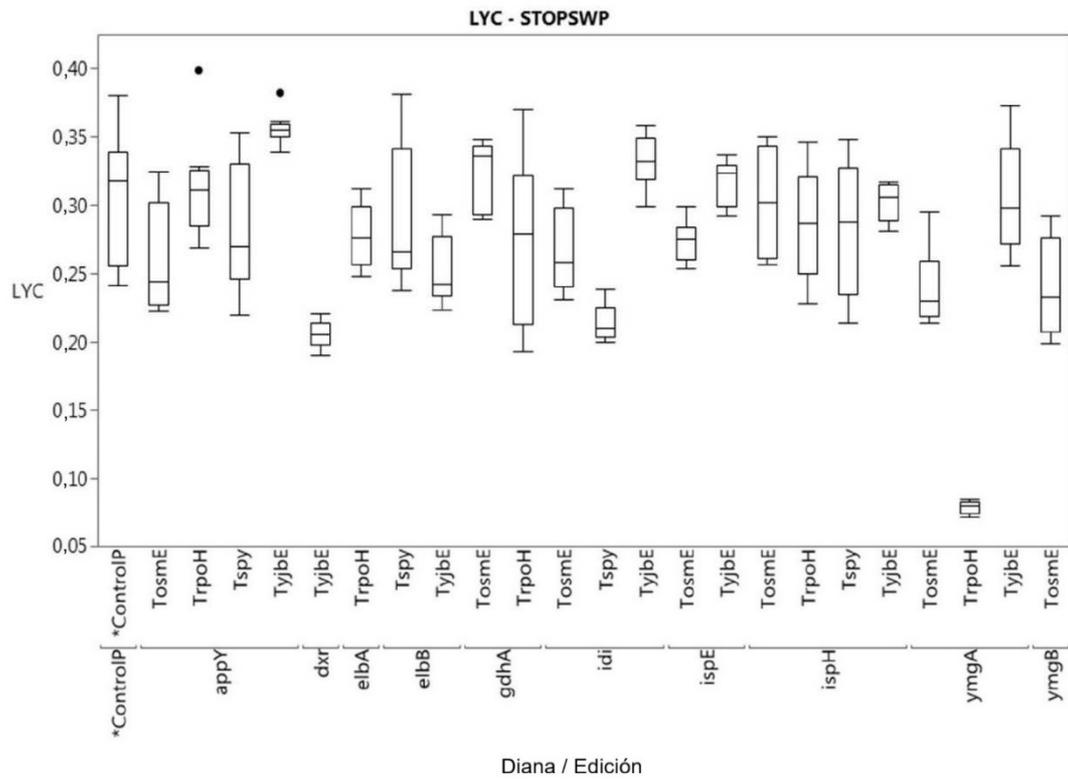


FIGURA 62

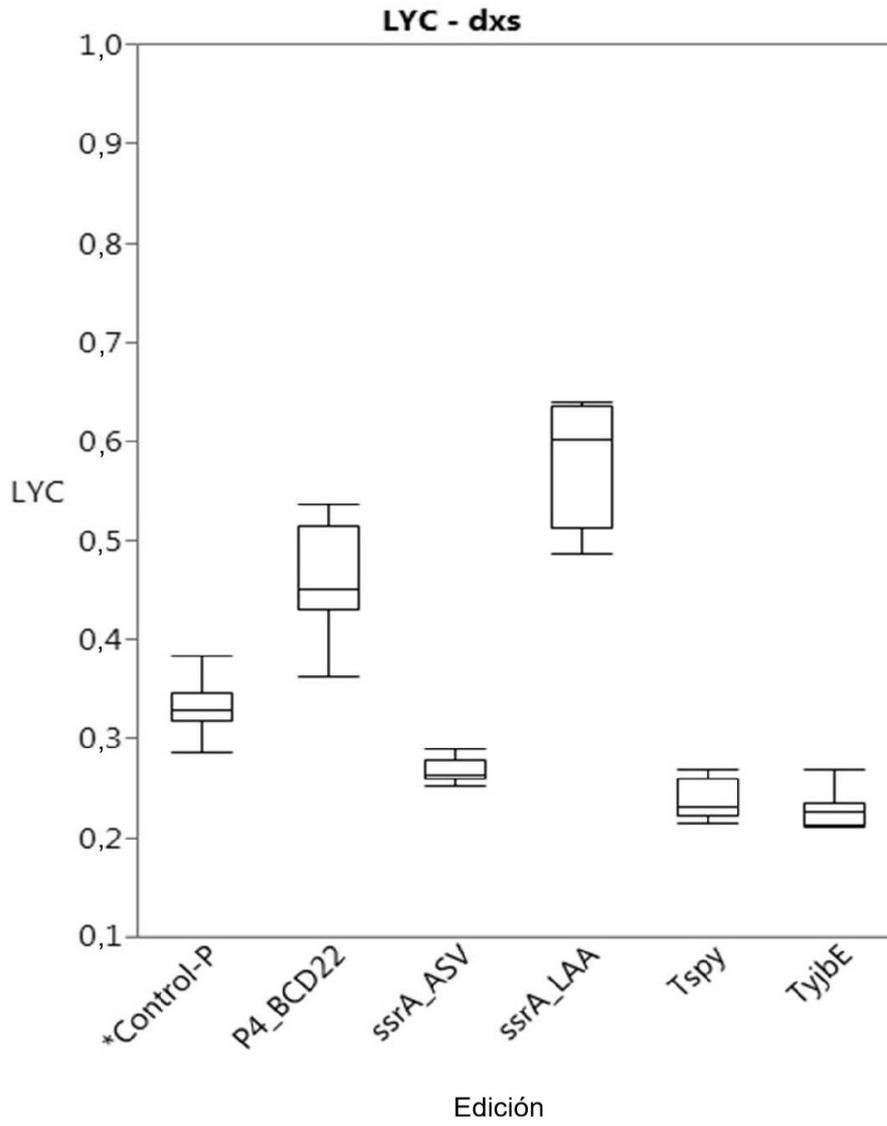


FIGURA 63

