

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 553 079**

51 Int. Cl.:

C12P 19/62 (2006.01)

C12N 15/52 (2006.01)

C12N 15/90 (2006.01)

C12N 1/19 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **03.05.2012 E 12166535 (0)**

97 Fecha y número de publicación de la concesión europea: **07.10.2015 EP 2520658**

54 Título: **Mejora de la producción de espinosinas con proteínas que se unen al oxígeno**

30 Prioridad:

03.05.2011 US 201113100202

03.05.2011 US 201113100220

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

04.12.2015

73 Titular/es:

DOW AGROSCIENCES LLC (100.0%)

9330 Zionsville Road

Indianapolis, IN 46268, US

72 Inventor/es:

HAN, LEI y

MOUNCEY, NIGEL

74 Agente/Representante:

DE ELZABURU MÁRQUEZ, Alberto

ES 2 553 079 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

Mejora de la producción de espinosinas con proteínas que se unen al oxígeno

Campo técnico

5 La invención se aplica al campo técnico de la genética molecular en la que los genes pueden estar integrados en el cromosoma de *Saccharopolyspora spinosa*. Un método de ingeniería metabólica incluye la expresión recombinante de una proteína que se une al oxígeno, tal como un gen de hemoglobina, dentro de una región del DNA cromosómico que da como resultado el aumento de la producción de espinosinas en *Saccharopolyspora spinosa*, preferiblemente no teniendo al mismo tiempo un efecto negativo sobre el crecimiento u otras características metabólicas deseadas de *Saccharopolyspora spinosa*.

10 La invención se refiere al aumento de los títulos de espinosinas de una célula hospedante de *S. spinosa* transformando la célula hospedante de *S. spinosa* con un gen heterólogo que codifica una proteína que se une al oxígeno, como se define en las reivindicaciones que se acompañan.

Antecedentes

15 Como se describe en la Patente de EE.UU. N° 5.362.634, el producto de fermentación A83543 es una familia de compuestos relacionados producida por *Saccharopolyspora spinosa*. Los miembros conocidos de esta familia se han denominado factores o componentes y a cada uno se le ha dado una letra de identificación. Estos compuestos se denominan en adelante espinosina A, B, etc. Los compuestos espinosinas son útiles para el control de arácnidos, nemátodos e insectos, en particular las especies de lepidópteros y dípteros. Los compuestos se consideran compatibles con el medio ambiente con un perfil toxicológico atractivo.

20 Los compuestos espinosinas producidos de forma natural son macrólidos que consisten en una lactona tetracíclica de 21 carbonos que incluye la unión de dos desoxiazúcares: un azúcar neutro (ramnosa) y un aminoazúcar (forosamina). (Véase Kirst et al., (1991)). Si el aminoazúcar no está presente, los compuestos se han denominado pseudoagliconas de A, D, etc., y si el azúcar neutro no está presente, entonces los compuestos se han denominado pseudoagliconas inversas de A, D, etc. Una nomenclatura más preferida es referirse a las pseudoagliconas como espinosina A 17-Psa, espinosina D 17-Psa, etc., y a las pseudoagliconas inversas como espinosina A 9-Psa, espinosina D 9-Psa, etc.

25 Los compuestos espinosinas producidos de forma natural se pueden obtener por fermentación de las cepas de *S. spinosa* NRRL 18395, 18537, 18538, 18539, 18719, 18720, 18743 y 18823 y sus derivadas. Estos cultivos se han depositado y forman parte de la colección de cultivos madre del Midwest Area Northern Regional Research Center, Agricultural Research Service, United States Department of Agriculture, 1815 North University Street, Peoria, 111, 61604.

30 La Patente de EE.UU. N° 5.362.634 y la Patente Europea correspondiente N° 0375316 B1 se refieren a las espinosinas A, B, C, D, E, F, G, H y J. Se dice que estos compuestos son producidos cultivando una cepa del nuevo microorganismo *Saccharopolyspora spinosa* seleccionado de NRRL 18395, NRRL 18537, NRRL 18538 y NRRL 18539.

35 El documento WO 93/09126 se refiere a las espinosinas L, M, N, Q, R, S y T. En él también se describen dos cepas productoras de la espinosina J: NRRL 18719 y NRRL 18720, y una cepa que produce las espinosinas Q, R, S y T: NRRL 18823.

El documento WO 94/20518 y la Patente de EE.UU. N° 5.670.486 se refieren a las espinosinas K, O, P, U, V, W e Y, y sus derivados. En ellos también se describen la cepa NRRL 18743 productora de la espinosina K.

40 El documento WO 99/46387 se refiere a un método de transformación de cepas de *S. spinosa* con un vector de DNA recombinante, o una de sus porciones, que incluye una secuencia de DNA que codifica la expresión de una actividad que es limitante de la velocidad en la vía biosintética de las espinosinas.

45 Feng Liang et al. (*Applied Microbiology and Biotechnology*, 74(2), 7 November 2006, pp. 390-397) describen la expresión de la hemoglobina de *Vitreoscilla* en *Bacillus thuringiensis* y varios hospedantes bacterianos heterólogos para mejorar la densidad celular, el metabolismo oxidante y la producción de proteínas.

Sería ventajoso mejorar la producción de los compuestos espinosinas.

Sumario de la invención

50 La presente invención proporciona un método para aumentar los títulos de espinosinas de una célula hospedante de *S. spinosa*, comprendiendo dicho método transformar la célula hospedante de *S. spinosa* con un gen heterólogo que codifica una proteína que se une al oxígeno.

La presente invención proporciona también un método para mejorar la producción de espinosinas en una célula hospedante de *S. spinosa*, comprendiendo dicho método integrar un gen que codifica una proteína que se une al

oxígeno en el DNA cromosómico de la cepa de especies de *S. spinosa*, en donde la integración del gen que codifica una proteína que se une al oxígeno en el DNA cromosómico comprende:

- a) crear una fagoteca para identificar un locus genómico;
 - 5 b) clonar dos fragmentos de DNA genómico a partir del locus genómico, en donde un fragmento genómico se clona aguas arriba de un polinucleótido que comprende un gen que codifica una proteína que se une al oxígeno y el segundo fragmento genómico se clona aguas abajo de dicho polinucleótido;
 - c) introducir los fragmentos genómicos en una cepa de *S. spinosa* por transformación;
 - d) obtener transconjugantes; y
 - e) cribar dichos transconjugantes para detectar la presencia de dicho polinucleótido.
- 10 La presente invención proporciona también una célula de *S. spinosa* modificada genéticamente, que comprende un gen heterólogo que codifica una proteína que se une al oxígeno, en la que cuando se expresa el gen heterólogo que codifica una proteína que se une al oxígeno, la célula de *S. spinosa* modificada genéticamente es capaz de aumentar la producción de espinosinas. La presente invención proporciona también una cepa que comprende una población de las células de *S. spinosa* modificadas genéticamente como se ha definido anteriormente.
- 15 En realizaciones, la invención proporciona la integración de dos o más genes mejoradores de las espinosinas y/o genes que codifican enzimas biosintéticas de espinosinas. En una realización, se proporciona al menos uno en un sitio neutro en el genoma de la célula hospedante. Se pueden proporcionar cualesquiera polinucleótidos heterólogos adicionales ya sea en el mismo sitio neutro, en una localización genómica diferente o como una construcción de expresión no integrada.
- 20 La presente invención incluye dentro de su alcance la transformación de células hospedantes por métodos distintos de la integración genómica de un polinucleótido y también las células hospedantes transformadas y sus usos. Por ejemplo, la invención incluye dentro de su alcance un método para transformar una célula hospedante de *S. spinosa* con un gen heterólogo de una proteína que se une al oxígeno. En una realización, el gen de la proteína que se une al oxígeno puede ser proporcionado como un casete de expresión. En una realización, el casete de expresión no
- 25 llega a integrarse en el genoma de la célula hospedante. Por lo tanto, en dichas realizaciones, la secuencia de polinucleótido no se expresa en el genoma de la célula hospedante.

Breve descripción de los dibujos

Las realizaciones preferidas de la invención se describen con referencia a los dibujos, en los cuales:

La FIG. 1 representa el plásmido pDAB109000 que contiene el fragmento de DNA sintético de 977 pb con pJ201.

- 30 La FIG. 2 representa el plásmido pDAB 109001 que resultó de la clonación del casete de expresión del gene de la hemoglobina en pIJ773.

La FIG. 3 representa el mapa de la secuencia final de los clones de cósmidos en la agrupación de genes de obscurina. Se muestran los clones de cósmidos solapantes. Las barras rellenas indican el tamaño real de los insertos en el clon del cósmido 1E3 y el clon del cósmido 2N14 con relación a la agrupación de genes de obscurina. Las líneas de

35 puntos indican que sólo un extremo del cósmido se encuentra dentro de la agrupación de genes de obscurina.

La FIG. 4 representa el plásmido pIJ773, que contiene el casete de expresión de resistencia a apramicina (marcado como aac(3) IV).

Descripción detallada

- 40 Existen muchos usos para los genes integrados dentro del cromosoma de *Saccharopolyspora spinosa*. Los genes clonados, bien naturales o heterólogos, se pueden usar para mejorar los rendimientos de las espinosinas naturales y para producir nuevas espinosinas. Se pueden obtener mejores rendimientos, por ejemplo, proporcionando (por ejemplo, integrando en el genoma de una cepa) una copia duplicada del gen para que cualquier enzima sea limitante de la velocidad en dicha cepa. En casos en los que está bloqueada la vía biosintética en una cepa mutante particular, debido a la falta de una enzima requerida, se puede restaurar la producción de las espinosinas deseadas
- 45 proporcionando el gen requerido (por ejemplo, integrando una copia del gen). Cuando se interrumpe una vía biosintética, se puede crear una cepa precursora diferente.

La presente invención ilustra que la expresión de una proteína que se une al oxígeno en *S. spinosa* mejora la producción de una espinosina por la célula de *S. spinosa*. Más específicamente, se ha demostrado que la expresión de la secuencia del gen de la hemoglobina de *Vitreoscilla*, en lo sucesivo "VHb", en *S. spinosa* da como resultado el

50 aumento de la producción de espinosinas. Preferiblemente, para la mejora de cepas de *S. spinosa*, la transformación estable de un polinucleótido se realizó integrando un casete de expresión génica en el genoma de *S. spinosa*. Un método de integración es por recombinación homóloga, preferiblemente usando una parte del DNA cromosómico y

un elemento de inserción. El DNA cromosómico se puede insertar en el genoma de *S. spinosa*. Basándose en esta recombinación y como resultado de su aplicación, en una realización de la invención, los casetes de expresión génica *aac(3) IV* y *VHb* se integraron por separado en el cromosoma de *S. spinosa* en el locus de las policétido-sintasas (PKS) de *obscurina* dando como resultado la inactivación de un gen natural, *obsA*.

- 5 Las siguientes definiciones se utilizan en la presente memoria y deben ser remitidas a la interpretación de las reivindicaciones y la parte descriptiva de la memoria.

Como se usa en la presente memoria, se entiende que los artículos indefinidos "un" y "una" precediendo a un elemento o componente de la invención no son restrictivos respecto al número de casos (es decir, incidencias) del elemento o componente. Por tanto, "un" o "una" se debe leer como que incluye uno o al menos uno, y la forma en singular del elemento o componente incluye también el plural salvo que el número se entienda claramente que es el singular.

Como se usan en la presente memoria, los términos "que comprende" y "que incluye" significan la presencia de las características, números enteros, etapas o componentes establecidos como se citan en las reivindicaciones, pero que no excluyen la presencia o adición de una o más de otras de sus características, números enteros, etapas, componentes o grupos. Esto significa que una composición, una mezcla, un proceso, un método, un artículo o un aparato que "comprende" o "incluye" una lista de elementos no está limitado sólo a aquellos elementos sino que puede incluir otros no expresamente incluidos en la lista o inherentes a ella. Como se usa en la presente memoria, "o" se refiere a un "o" inclusivo y exclusivo. Por ejemplo, una condición A o B es satisfecha por una cualquiera de las siguientes: A es verdadera (o está presente) y B es falsa (o no está presente), A es falsa (o no está presente) y B es verdadera (o está presente) y tanto A como B son verdaderas (o están presentes).

Como se usa en la presente memoria, el término "aproximadamente" se refiere a la modificación de la cantidad de un ingrediente o reaccionante de la invención o cuando se emplea se refiere a la variación en la cantidad numérica que puede ocurrir, por ejemplo, en los procedimientos típicos de medición y manipulación de líquidos utilizados para la preparación de concentrados o el uso de soluciones en la práctica; en un error involuntario en estos procedimientos; en diferencias en la fabricación, fuente o pureza de los ingredientes empleados para preparar las composiciones o llevar a cabo los métodos; y similares. El término "aproximadamente" abarca también cantidades que difieren debido a las diferentes condiciones de equilibrio para una composición que resulta de una mezcla inicial particular. Sean o no modificadas por el término "aproximadamente", las reivindicaciones incluyen equivalentes a las cantidades.

Como se utiliza en la presente memoria, los términos "polipéptido" y "péptido" y "proteína" se usarán indistintamente para referirse a un polímero de dos o más aminoácidos unidos entre sí por un enlace peptídico. En un aspecto, este término también incluye modificaciones del polipéptido posteriores a la expresión, por ejemplo, glicosilaciones, acetilaciones, fosforilaciones y similares. Incluidos en la definición están, por ejemplo, péptidos que contienen uno o más análogos de un aminoácido o aminoácidos marcados y peptidomiméticos. Los péptidos pueden comprender L-aminoácidos. La referencia en la presente memoria a "polipéptido", "péptido" y "proteína" incluye sus fragmentos, derivados y variantes.

Como se usa en la presente memoria, los términos "péptido de interés", "POI", "producto génico", "producto génico diana" y "producto génico con la región codificadora diana" se refieren al producto peptídico/proteínico deseado codificado por el gen expresado. El péptido de interés puede incluir cualquier producto peptídico/proteínico, incluyendo, aunque sin limitación, proteínas, proteínas de fusión, enzimas, péptidos, polipéptidos y oligopéptidos. El péptido de interés tiene una longitud que varía desde 2 a 398 aminoácidos. El producto peptídico/proteínico deseado puede ser un péptido/proteína heterólogo, expresado por un gen extraño (es decir, no natural).

Como se utiliza en la presente memoria, el término "construcción genética" se refiere a una serie de ácidos nucleicos contiguos útiles para modular el genotipo o fenotipo de un organismo. Ejemplos no limitantes de construcciones genéticas incluyen, aunque sin limitación, una molécula de ácido nucleico y un marco de lectura abierto, un gen, un casete de expresión, un vector, un plásmido y similares.

Como se utiliza en la presente memoria, el término "gen endógeno" se refiere a un gen natural en su localización natural en el genoma de un organismo.

Como se utiliza en la presente memoria, un "gen extraño" se refiere a un gen que no se encuentra normalmente en el organismo hospedante, pero que es introducido en el organismo hospedante por transferencia de genes. Los genes extraños pueden comprender genes naturales insertados en un organismo no natural o genes quiméricos. Un gen extraño también puede ser denominado un gen no natural. Un gen extraño está preferiblemente optimizado en sus codones para su expresión en una célula hospedante, preferiblemente cuando la célula hospedante es *S. spinosa*. La optimización adecuada para la expresión en *S. spinosa* se incluye en la Tabla 1.

La referencia en la presente memoria a un "gen" incluye, para los fines de la invención, cualquier secuencia codificadora o secuencia de control, o sus fragmentos. Un gen puede incluir cualquier combinación de secuencia codificadora y secuencia de control, o sus fragmentos. Por lo tanto, un "gen" como se cita en la presente memoria puede ser todo o parte de un gen natural. Una secuencia de polinucleótido como se cita en la presente memoria se puede usar

indistintamente con el término "gen", o puede incluir cualquier secuencia codificadora, secuencia no codificadora o secuencia de control, sus fragmentos y sus combinaciones.

Como se usa en la presente memoria, el término "heterólogo" con respecto a una secuencia dentro de un organismo/genoma particular indica que la secuencia se origina a partir de una especie extraña, o, si procede de la misma especie, está sustancialmente modificada a partir de su forma natural en composición y/o locus genómico y/o número de copias por intervención humana deliberada. Por tanto, cuando el gen es un gen natural, puede ser heterólogo si ha sido mutado, replicado (es decir, aumentado el número de copias) o movido (es decir, su localización dentro de la célula). Así, por ejemplo, la expresión génica heteróloga se refiere al proceso de expresar un gen a partir de un organismo/genoma colocándolo en el genoma de un organismo/genoma diferente.

Como se usa en la presente memoria, el término "recombinante" se refiere a una combinación artificial de dos segmentos de secuencia que estarían de otro modo separados, por ejemplo, por síntesis química o por la manipulación de segmentos aislados de ácidos nucleicos por técnicas de ingeniería genética. "Recombinante" incluye también la referencia a una célula o vector, que ha sido modificado por la introducción de un ácido nucleico heterólogo o una célula procedente de una célula así modificada, pero no abarca la alteración de la célula o vector por episodios que ocurren de forma natural (por ejemplo, mutación espontánea, transformación natural, transducción natural, transposición natural), tales como los que se producen sin intervención humana deliberada. Una célula hospedante recombinante de la presente invención será distinguible de una célula de origen natural de la misma especie en virtud de la presencia en ella de DNA extraño (por ejemplo, marcadores de selección, vectores, genes no naturales), o en virtud de una estructura genética alterada (por ejemplo, eliminación o interrupción de secuencias naturales, cambios en la orientación de las secuencias naturales, promotores alterados u otras secuencias de control, números de copia alterados, localización alterada de secuencias naturales). Además de lo anterior, una célula hospedante recombinante de la invención mostrará mejor producción de espinosinas en comparación con una célula no modificada del mismo tipo. Con referencia a la célula, recombinante puede ser utilizado indistintamente con el término "modificada genéticamente".

Una "cepa" como se denomina en la presente memoria es una población de células que desciende de un solo organismo o que es un material aislado de un cultivo sustancialmente puro. En una realización, una cepa puede comprender al menos 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95% de células hospedantes modificadas genéticamente, tal como se define en la presente memoria.

El término "modificado genéticamente" o "alterado genéticamente" significa la alteración científica de la estructura del material genético de un organismo vivo. Implica la producción y uso del DNA recombinante. En particular, se usa para definir el organismo genéticamente diseñado o modificado a partir del organismo de origen natural. La modificación por ingeniería genética se puede realizar por varios métodos conocidos en la técnica, tales como por ejemplo, sustitución de genes, amplificación de genes, interrupción de genes, transfección, transformación utilizando plásmidos, virus u otros vectores. Un organismo modificado genéticamente, por ejemplo, un microorganismo modificado genéticamente, se denomina también a menudo un organismo recombinante, por ejemplo, un microorganismo recombinante.

Como se utiliza en la presente memoria, el término "interrumpido" o "interrupción" se refiere a un gen que ha sido manipulado o modificado por ingeniería genética o por causas naturales que cambian la actividad de un gen. Dicha actividad génica se puede aumentar o disminuir, en comparación con la actividad génica no interrumpida. Además, dicha interrupción puede alterar o suprimir la expresión y/o función de las proteínas. La interrupción que causa una disminución en la actividad de los genes puede ser una reducción del número de copias de un gen, provocando así la subexpresión del gen. Se dice que un gen está "subexpresado" si está reducido el nivel de transcripción de dicho gen en comparación con un gen no interrumpido (por ejemplo, el gen de tipo natural). Esto se puede medir, por ejemplo, por análisis de transferencia de Northern cuantificando la cantidad de mRNA como indicación de la expresión génica. Como se usa en la presente memoria, un gen se subexpresa si disminuye la cantidad de mRNA generado en al menos 1%, 2%, 5%, 10%, 25%, 50%, 75%, 100%, 200% o incluso más de 500%, en comparación con la cantidad de mRNA generado a partir de un gen no interrumpido (por ejemplo, gen de tipo natural). Alternativamente, la interrupción puede incluir la expresión del gen a partir de una secuencia de control débil, tal como un promotor, en comparación con el promotor natural. En otra realización, se pueden alterar el promotor, la región reguladora y/o el sitio de unión al ribosoma aguas arriba del gen para lograr la expresión reducida. En otra realización, también se puede reducir la expresión disminuyendo la semivida relativa del RNA mensajero. En otra realización, se puede disminuir la actividad de un polipéptido propiamente dicho, por ejemplo empleando una o más mutaciones en la secuencia de aminoácidos del polipéptido, que disminuyen la actividad. Por ejemplo, la alteración de la afinidad del polipéptido para su sustrato correspondiente puede dar lugar a una actividad reducida. Igualmente, se puede disminuir la semivida relativa del polipéptido. En cualquier caso, sea la expresión del gen reducida o la actividad reducida, la reducción se puede conseguir alterando la composición de los medios de cultivos celulares y/o los métodos utilizados para el cultivo. "Expresión reducida" como se utiliza en la presente memoria significa una disminución de al menos 5%, 10%, 25%, 50%, 75%, 100%, 200% o incluso más de 500%, en comparación con la cantidad de mRNA o proteína generados en comparación con el gen no interrumpido (por ejemplo, un gen de tipo natural). "Actividad reducida" como se usa en la presente memoria significa una disminución de al menos 5%, 10%, 25%, 50%, 75%, 100%, 200% o incluso más de 500%, en comparación con la actividad o concentración de la proteína, polinucleótido o gen antes de la interrupción. La actividad de una proteína también se puede reducir poniendo en contacto la pro-

teína con un inhibidor específico o general de su actividad. Los términos "actividad reducida", "actividad disminuida o suprimida" se usan indistintamente en la presente memoria.

En otra realización, la interrupción puede aumentar la actividad o expresión de un gen. Una secuencia de control, tal como el promotor, la región reguladora y/o el sitio de unión al ribosoma aguas arriba del gen se puede alterar para lograr el aumento de la expresión. En otra realización, el aumento de la expresión se puede lograr aumentando el número de copias del gen. En una realización, la sobreexpresión también se puede lograr aumentando la semivida relativa del RNA mensajero. En otra realización, la actividad del polipéptido propiamente dicho se puede aumentar realizando una o más mutaciones en la secuencia de aminoácidos del polipéptido, lo que aumentaría la actividad. Por ejemplo, la alteración de la afinidad del polipéptido para su sustrato correspondiente puede dar como resultado mayor actividad. Igualmente, se puede aumentar la semivida relativa del polipéptido. En cualquiera caso, sea la sobreexpresión del gen o la actividad aumentada, el aumento puede conseguirse alterando la composición de los medios de cultivos celulares y/o los métodos utilizados para el cultivo. "Sobreexpresión" o "actividad aumentada" como se usa en la presente memoria significa un aumento de al menos 5%, 10%, 25%, 50%, 75%, 100%, 200% o incluso más de 500%, en comparación con una proteína, polinucleótido o gen de tipo natural; o la actividad y/o la concentración de la proteína presente antes de la interrupción para aumentar la expresión y/o la actividad. La actividad de una proteína también se puede aumentar poniendo en contacto la proteína con un inhibidor específico o general de su actividad. Los términos "sobreexpresión" y "actividad aumentada" se pueden utilizar indistintamente.

"Alterar" con referencia a los cambios en la expresión o actividad puede incluir mutación, sustitución, delección o adición de una secuencia génica de control o secuencia de polinucleótido o sus fragmentos.

En la presente memoria, mejora o aumento de la producción de espinosinas significa un aumento en la cantidad de una o más espinosinas naturales o la producción de una o más nuevas espinosinas, en comparación con la producción de espinosinas por una célula hospedante del mismo tipo cuando se desarrolla en las mismas condiciones (por ejemplo, fermentación en un matraz con agitación) durante el mismo período de tiempo. Preferiblemente, los métodos y las células hospedantes de la presente invención proporcionan un aumento en el título de las espinosinas naturales de al menos 4%, 4,6%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20 %, 30%, 40%, 50%, 60%, 70%, 80%, 90% o al menos 100%.

La expresión "secuencias de control" se refiere colectivamente a secuencias promotoras, sitios de unión al ribosoma, secuencias de terminación de la transcripción, dominios reguladores aguas arriba, potenciadores y similares, que proporcionan colectivamente la transcripción y traducción de una secuencia codificadora en una célula hospedante. Una cualquiera o más de las "secuencias de control" antes mencionadas seleccionadas del grupo antes mencionado se puede proporcionar en un casete de expresión. No todas estas secuencias de control necesitan siempre estar presentes en un vector recombinante, siempre que el gen deseado sea capaz de ser transcrito y traducido.

En la presente memoria, la referencia a DNA cromosómico o al genoma de la célula hospedante se puede usar indistintamente.

"Recombinación" se refiere a la redistribución de secciones de secuencias de DNA o RNA entre dos moléculas de DNA o RNA. La "recombinación homóloga" se produce entre dos moléculas de DNA que se hibridan en virtud de las secuencias de nucleótidos homólogas o complementarias presentes en cada molécula de DNA.

La referencia en la presente memoria a un "gen mejorador de espinosinas" incluye cualquier secuencia de polinucleótido que, cuando se proporciona en una célula hospedante de *S. spinosa*, tiene el efecto de aumentar la expresión o actividad de una espinosina natural, o permitir la expresión de una nueva espinosina o un derivado de una espinosina natural. Un gen mejorador de espinosinas puede ser una secuencia de control o una secuencia codificadora, o ambas, o sus fragmentos. También están incluidos sus derivados y variantes. En una realización, un gen mejorador de espinosinas puede codificar una proteína que se une al oxígeno. Además, la proteína que se une al oxígeno puede ser cualquier proteína que se una al oxígeno, en particular las que se unen al oxígeno de forma reversible, tales como las globinas. Un gen mejorador de espinosinas puede ser un gen que codifica una enzima biosintética de espinosinas.

Las "proteínas que se unen al oxígeno" que se pueden utilizar en la invención incluyen, aunque sin limitación, hemoglobina de *Vitreoscilla* (VHb), flavohemoproteína de *Alcaligenes eutrophus*, mioglobina de corazón de caballo, hemoproteína de *E. coli*, hemoproteína de *B. subtilis*, flavohemoglobina de levadura, leghemoglobina de soja, leghemoglobina de altramuz y mioglobina de esperma de ballena. Como se ha indicado antes, la proteína que se une al oxígeno también puede ser una proteína que sea endógena al organismo productor de espinosinas. Alternativamente, la proteína que se une al oxígeno puede ser heteróloga del organismo productor de espinosinas.

Un "sitio neutro" para la integración de un casete de expresión génica es aquel que codifica un producto génico que no está implicado en una vía metabólica primaria y no se requiere preferiblemente para la producción de espinosinas. Un sitio neutro puede ser uno que codifica un gen que no es necesario para el desarrollo y/o división de la célula. Por tanto, si se interrumpe el gen en el sitio neutro, una célula hospedante modificada genéticamente presentará un metabolismo primario y un desarrollo comparable al de una célula hospedante del mismo tipo que no ha sido modificada genéticamente, cuando se mantiene en las mismas condiciones. Un sitio neutro puede ser un sitio de un

gen implicado en una vía metabólica secundaria, por ejemplo el locus de la poliketido-sintasa (PKS) de obscurina, dando como resultado preferiblemente la inactivación de un gen natural, *obsA*. La secuencia de un fragmento del gen *obsA* se denomina en la presente memoria SEQ ID NO 8. Cuando el sitio neutro es un gen, la interrupción puede ser en un elemento de control del gen o en la secuencia codificadora del gen, o en ambos. Alternativamente, un sitio neutro puede ser un sitio que no es un gen.

Una vía metabólica primaria incluye las vías esenciales para la supervivencia celular, por ejemplo la glicolisis, la vía de la pentosa-fosfato, el ciclo de TCA y la biosíntesis de aminoácidos. Una vía metabólica secundaria es una que está implicada en la producción de un metabolito secundario, es decir, que no es necesaria para la supervivencia de la célula individual.

Los términos "condiciones restrictivas" o "hibridación en condiciones restrictivas" se refieren a condiciones bajo las cuales una sonda se hibridará preferiblemente con su subsecuencia diana, y en menor grado, o nada en absoluto, a otras secuencias. La "hibridación restrictiva" y las "condiciones de lavado en la hibridación restrictiva" en el contexto de los experimentos de hibridación de ácidos nucleicos, tales como hibridaciones de Southern y de Northern, son dependientes de las secuencias y son diferentes bajo diferentes parámetros ambientales. Una amplia guía para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes*, part I, chapter 2, *Overview of principles of hybridization and the strategy of nucleic acid probe assays*, Elsevier, New York. En general, se seleccionan condiciones de hibridación y de lavado altamente restrictivas para que sean aproximadamente 5°C inferiores al punto de fusión térmico (T_m) para la secuencia específica a una fuerza iónica y un pH definidos. El T_m es la temperatura (bajo una fuerza iónica y un pH definidos) a la cual el 50% de la secuencia diana se hibrida con una sonda perfectamente coincidente. Se seleccionan condiciones muy restrictivas para que sean iguales al T_m para una sonda particular.

Un ejemplo de condiciones de hibridación restrictivas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia de Southern o Northern es 50% de formamida con 1 mg de heparina a 42°C, realizándose la hibridación durante una noche. Un ejemplo de condiciones de lavado altamente restrictivas es con NaCl 0,15 M a 72°C durante aproximadamente 15 minutos. Un ejemplo de condiciones de lavado restrictivas es un lavado con 0,2xSSC a 65°C durante 15 minutos (véase, Sambrook et al., (1989) *Molecular Cloning - A Laboratory Manual* (2nd ed.) Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, para la descripción del tampón SSC). Frecuentemente, un lavado muy restrictivo está precedido de un lavado poco restrictivo para eliminar la señal de la sonda de fondo. Un ejemplo de lavado restrictivo medio para un dúplex de, por ejemplo, más de 100 nucleótidos, es 1xSSC a 45°C durante 15 minutos. Un ejemplo de lavado poco restrictivo para un dúplex de, por ejemplo, más de 100 nucleótidos, es 4-6xSSC a 40°C durante 15 minutos. En general, una relación señal a ruido de 2x (o superior) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica la detección de una hibridación específica. Los ácidos nucleicos que no se hibridan entre sí en condiciones restrictivas son todavía sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto ocurre, por ejemplo, cuando se crea una copia de un ácido nucleico usando la máxima degeneración de codones permitida por el código genético.

La descripción se refiere también a un polinucleótido aislado hibridable en condiciones restrictivas, preferiblemente en condiciones muy restrictivas, con un polinucleótido como el de la presente descripción.

Como se utiliza en la presente memoria, se pretende que el término "hibridación" describa condiciones para hibridación y lavado bajo las cuales secuencias de nucleótidos homólogas entre sí, en al menos aproximadamente 50%, al menos aproximadamente 60%, al menos aproximadamente 70%, más preferiblemente al menos aproximadamente 80%, incluso más preferiblemente al menos aproximadamente 85% a 90%, lo más preferiblemente al menos 95%, se mantienen típicamente hibridadas entre sí.

En una realización, un ácido nucleico de la descripción es al menos 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92 %, 93%, 94%, 95%, 96%, 97%, 98%, 99%, o más homólogo de una secuencia de ácido nucleico mostrada en esta solicitud o de sus complementos.

Otro ejemplo no limitativo de condiciones de hibridación restrictivas es la hibridación en 6x cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45°C, seguido por uno o más lavados en 1xSSC, 0,1% de SDS a 50°C, preferiblemente a 55°C, más preferiblemente a 60°C e incluso más preferiblemente a 65°C.

Las condiciones altamente restrictivas pueden incluir incubaciones a 42°C durante un período de varios días, tal como 2-4 días, utilizando una sonda de DNA marcada, tal como una sonda de DNA marcada con digoxigenina (DIG), seguidas de uno o más lavados con 2xSSC, 0,1% de SDS a temperatura ambiente y uno o más lavados con 0,5xSSC, 0,1% de SDS o con 0,1xSSC, 0,1% de SDS a 65-68°C. En particular, las condiciones altamente restrictivas incluyen, por ejemplo, una incubación de 2 horas a 4 días a 42°C utilizando una sonda de DNA marcada con DIG (preparada, por ejemplo, usando un sistema de marcaje con DIG; Roche Diagnostics GmbH, 68298 Mannheim, Alemania) en una solución, tal como solución DigEasyHyb (Roche Diagnostics GmbH) con o sin 100 µg/mL de DNA de esperma de salmón, o una solución que comprende 50% de formamida, 5xSSC (NaCl 150 mM, citrato trisódico 15 mM), 0,02% de dodecilsulfato de sodio, 0,1% de N-lauroilsarcosina y 2% de reactivo de bloqueo (Roche Diagnostics GmbH), seguido de lavado de los filtros dos veces durante 5 a 15 minutos en 2xSSC y 0,1% de SDS a tempera-

tura ambiente y después lavado dos veces durante 15-30 minutos con 0,5xSSC y 0,1% de SDS o 0,1xSSC y 0,1% de SDS a 65-68°C.

5 En algunas realizaciones, una molécula de ácido nucleico aislada de la descripción que se hibrida en condiciones altamente restrictivas con una secuencia de nucleótidos de la descripción puede corresponder a una molécula de ácido nucleico de origen natural. Como se usa en la presente memoria, una molécula de ácido nucleico "de origen natural" se refiere a una molécula de RNA o DNA que tiene una secuencia de nucleótidos que existe en la naturaleza (por ejemplo, codifica una proteína natural).

10 Un experto en la técnica sabrá qué condiciones aplicar para condiciones de hibridación restrictivas y altamente restrictivas. Las directrices adicionales con respecto a estas condiciones están fácilmente disponible en la técnica, por ejemplo, en Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; y Ausubel et al., (eds.), 1995, *Current Protocols in Molecular Biology* (John Wiley & Sons, N.Y.).

15 En la presente memoria se usa la notación convencional para describir secuencias de polinucleótidos: el extremo izquierdo de una secuencia de polinucleótido monocatenaria es el extremo 5'; la dirección hacia la izquierda de una secuencia de polinucleótido bicatenaria se refiere a la dirección 5'. La dirección de 5' a 3' de la adición de nucleótidos a los transcritos de RNA nuevos se denomina dirección de transcripción. La cadena de DNA que tiene la misma secuencia que un mRNA se denomina "cadena codificadora"; las secuencias en la cadena de DNA que tienen la misma secuencia que un mRNA transcrito a partir del DNA y que se encuentran en 5' hasta el extremo 5' del transcrito de RNA se denominan "secuencias aguas arriba"; las secuencias en la cadena de DNA que tienen la misma secuencia que el RNA y que están en 3' hasta el extremo 3' del transcrito de RNA codificador se denominan "secuencias aguas abajo".

20 La presente descripción incluye dentro de su alcance un método para mejorar la producción de espinosinas en una célula hospedante de *S. spinosa*, que comprende integrar un polinucleótido en el DNA cromosómico de una célula de *S. spinosa*, en el que el polinucleótido comprende un gen que codifica una enzima biosintética de espinosinas. La presente descripción también proporciona una célula hospedante de *S. spinosa* modificada genéticamente que comprende un polinucleótido integrado en ella, en el que el polinucleótido comprende un gen que codifica una enzima biosintética de espinosinas. La presente descripción se refiere también al uso de una célula hospedante de *S. spinosa* que comprende un polinucleótido integrado en ella en la producción de una espinosina, en donde el polinucleótido comprende un gen que codifica una enzima biosintética de espinosinas. En una realización, una célula hospedante de la descripción muestra mejor producción de espinosinas, preferiblemente producción de las espinosinas A/D o producción de las espinosinas J/L en la fermentación en un matraz con agitación.

25 Un polinucleótido que contiene un gen para una enzima biosintética de espinosinas permitiría el suministro o la duplicación de genes que codifican enzimas limitantes de la velocidad en la producción de espinosinas. Este podría ser utilizado para aumentar el rendimiento en cualquier circunstancia cuando una de las actividades codificadas limita la síntesis de la espinosina deseada. Se observó un aumento en el rendimiento de las espinosinas A/D cuando se duplicaron genes unidos a la policétido-sintasa de espinosinas integrando un cósmido que los contienen en *S. spinosa* (Madduri et al., 2001). En otro ejemplo, se logró un aumento del rendimiento de este tipo en fermentaciones de *Streptomyces fradiae* duplicando el gen que codifica una metiltransferasa limitante de la velocidad que convierte macrocina en tilosina (Baltz et al., 1997).

30 En una realización, la presente descripción proporciona el suministro en una célula hospedante de un gen que codifica una enzima biosintética de espinosina, en combinación con el suministro de un gen mejorador de espinosinas, que en una realización puede codificar una proteína que se une al oxígeno. Por tanto, la descripción proporciona mayor producción de espinosinas por la expresión heteróloga de un gen mejorador de espinosinas y la co-expresión de una enzima biosintética de espinosina heteróloga. Se proporcionan métodos, células hospedantes y usos relacionados con esta realización.

35 La presente descripción incluye también dentro de su alcance un método para producir una nueva espinosina o un compuesto intermedio o derivado de espinosina, en una célula hospedante de *S. spinosa*, que comprende integrar un polinucleótido en el DNA cromosómico de una célula hospedante de *S. spinosa*, en el que el polinucleótido es capaz de interrumpir la expresión de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, el polinucleótido comprende un fragmento interno de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, el método puede comprender la introducción de un gen heterólogo mejorador de espinosinas en la célula hospedante. La presente descripción proporciona también una célula hospedante de *S. spinosa* modificada genéticamente que comprende un polinucleótido integrado en ella, en el que el polinucleótido es capaz de interrumpir la expresión de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, el polinucleótido comprende un fragmento interno de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, se puede proporcionar un gen heterólogo mejorador de espinosinas en la célula hospedante. La presente descripción se refiere también al uso de una célula hospedante de *S. spinosa* que comprende un polinucleótido integrado en ella en la producción de una espinosina, en el que el polinucleótido es capaz de interrumpir la expresión de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, el polinucleótido comprende un fragmento interno de un gen natural que codifica una enzima biosintética de espinosinas. En una realización, se puede proporcionar en la célula hospedante un gen heterólogo mejorador de

espinosinas. En una realización, una célula hospedante de *S. spinosa* de la descripción muestra mejor producción de espinosinas, preferiblemente producción de las espinosinas A/D o producción de las espinosinas J/L en la fermentación en un matraz con agitación.

5 En una realización, se puede proporcionar en la célula hospedante modificada genéticamente un gen heterólogo que codifica una enzima biosintética de espinosinas. Por tanto, el método puede comprender adicionalmente transformar la célula con un gen heterólogo que codifica una enzima biosintética de espinosinas. La célula hospedante genéticamente modificada como se ha definido antes puede comprender adicionalmente un gen heterólogo que codifica una enzima biosintética de espinosinas. El gen heterólogo puede codificar la misma enzima biosintética de espinosinas o una enzima biosintética de espinosinas diferente. El gen heterólogo puede codificar una enzima biosintética de
10 espinosinas natural o uno de sus mutantes, derivados, fragmentos o variantes.

Por tanto, la descripción en una realización proporciona un método para producir una nueva espinosina o un compuesto intermedio o derivado de espinosina, en una célula hospedante de *S. spinosa*, que comprende: a) integrar un polinucleótido en el DNA cromosómico de una célula hospedante de *S. spinosa*, en el que el polinucleótido es capaz de interrumpir la expresión de un gen natural que codifica una enzima biosintética de espinosinas; y b) proporcionar
15 en la célula hospedante un gen heterólogo mejorador de espinosinas; c) opcionalmente, proporcionar en la célula hospedante un gen heterólogo que codifica una enzima biosintética de espinosinas. Se describe también una célula hospedante genéticamente modificada que comprende un polinucleótido integrado en ella, en el que el polinucleótido es capaz de interrumpir la expresión de un gen natural que codifica una enzima biosintética de espinosinas y que comprende además un gen heterólogo mejorador de espinosinas y opcionalmente un gen heterólogo que codifica
20 una enzima biosintética de espinosinas. Se describe también el uso de dicha célula hospedante en la producción mejorada de una espinosina.

Compuesto intermedios específicos (o sus derivados naturales) podrían ser sintetizados por cepas mutantes de *S. spinosa* en las que hubieran sido interrumpidos ciertos genes que codifican enzimas para la biosíntesis de espinosinas. Dichas cepas se pueden generar integrando, por ejemplo por recombinación homóloga, una secuencia mutagénica que contiene un fragmento interno del gen diana. Por la integración, se forman dos copias incompletas del gen biosintético, eliminando con ello la función enzimática que codifica. El sustrato para esta enzima, o alguno de sus derivados naturales, se debe acumular por fermentación de la cepa mutante. En una realización, se utilizó dicha estrategia para generar eficazmente una cepa de *Saccharopolyspora erythraea* que producía nuevos derivados de
25 6-desoxieritromicina (Weber & McAlpine, 1992).

Dichas cepas se podrían generar intercambiando la región diana, por recombinación homóloga con doble cruzamiento, con un plásmido mutagénico que contenga el nuevo fragmento entre las secuencias no mutadas que flanquean la región diana. El gen híbrido produciría proteínas con funciones alteradas, bien que carecieran de una actividad o que realizaran una nueva transformación enzimática. Se acumularía un nuevo derivado por fermentación de la cepa mutante. Esta estrategia se utilizó para generar una cepa de *Saccharopolyspora erythraea* que produce un
30 nuevo derivado de anhidroeritromicina (Donadio et al., 1993).

Se clonaron los genes que codifican enzimas biosintéticas de espinosinas y marcos de lectura abiertos "ORF" relacionados para uso en la presente descripción y se determinó la secuencia de DNA de cada uno. Los genes clonados y los ORF se denominan de aquí en adelante *spnA*, *spnB*, *spnC*, *spnD*, *spnE*, *spnF*, *spnG*, *spnH*, *spnI*, *spnJ*, *spnK*, *spnL*, *spnM*, *spnN*, *spnO*, *spnP*, *spnQ*, *spnR*, *spnS*, ORFL15, ORFL16, ORFR1, ORFR2, *S. spinosa gtt*, *S. spinosa gdh*, *S. spinosa epi* y *S. spinosa kre*. Las secuencias de los genes biosintéticos de espinosinas son conocidas en la técnica y están definidas en US2004/002343 (solicitud N° 10/329148) por las bases 21111-28898, 28916-35374, 35419-44931, 44966-59752, 59803-76569, 20168-20995, 18541-19713, 17749-18501, 16556-17743, 14799-16418, 13592-14785, 12696-13547, 11530-12492, 10436-11434, 8967-10427, 7083-8450, 5363-6751, 4168-5325, 3416-4165, 2024-2791, 1135-1971, 76932-77528 y 77729-79984 de la SEQ ID NO: 1, las bases 334-1119 de la SEQ ID NO: 27, las bases 88-1077 de la SEQ ID NO 24, las bases 226-834 de la SEQ ID NO 31 y las bases 1165-1992 de la SEQ ID NO: 24, respectivamente. La referencia en la presente memoria a un gen que codifica una enzima biosintética de espinosinas incluye cualquiera de los genes y de los ORF antes mencionados, así como cualesquiera genes y ORF adicionales conocidos por un experto en la técnica. Por tanto, la presente descripción comprende el uso de uno o más genes seleccionados del grupo antes mencionado, o sus fragmentos, derivados o variantes.
40
45

50 En realizaciones en las que está interrumpido un gen natural que codifica una enzima biosintética de espinosinas, la enzima se puede seleccionar de las anteriormente definidas. Un fragmento interno de un gen puede incluir cualquier parte de una secuencia codificadora o no codificadora de un gen, por ejemplo una secuencia de control, regiones no traducidas, secuencias codificadoras, intrones o exones. Se puede determinar un fragmento dependiendo de la localización cromosómica deseada de la secuencia de polinucleótido que se ha de integrar.

55 La *Saccharopolyspora spinosa* produce una mezcla de nueve compuestos muy relacionados denominados colectivamente "espinosinas". Dentro de la mezcla, las espinosinas A y D, conocidas como espinosad, son los componentes principales y tienen la mayor actividad contra insectos diana claves. Las espinosinas J y L, dos de los componentes minoritarios dentro de la mezcla de espinosinas, son los precursores de espinotoram, el insecticida de espinosina de la segunda generación.

Espinosad es un insecticida producido por Dow AgroSciences (Indianapolis, Ind.) que está constituido principalmente por aproximadamente 85% de espinosina A y aproximadamente 15% de espinosina D. Las espinosinas A y D son productos naturales producidos por fermentación de *Saccharopolyspora spinosa*, como se ha descrito en la Patente de EE.UU. N° 5.362.634. El espinosad es un ingrediente activo de varias formulaciones insecticidas disponible comercialmente de Dow AgroSciences, que incluye los productos para el control de insectos TRACER™, SUCCES™, SPINTOR™ y CONSERVE™. Por ejemplo, el producto TRACER está constituido por aproximadamente 44% a aproximadamente 48% de espinosad (p/v) o aproximadamente 479 gramos de espinosad por litro (4 libras por galón). Se ha establecido la utilidad de los compuestos de espinosina en formulaciones granulares y líquidas para el control de arácnidos, nemátodos e insectos, en particular, las especies de lepidópteros, tisánopteros y dípteros. Las espinosinas A y D también se denominan en la presente memoria espinosinas A/D.

Espinotoram es una mezcla de 5,6-dihidro-3'-etoxi-espinosina J (componente principal) y 3'-etoxi-espinosina L producidas por Dow AgroSciences. La mezcla se puede preparar por etoxilación de una mezcla de espinosina J y espinosina L, seguido por hidrogenación. El doble enlace en 5,6 de la espinosina J y su derivado 3'-etoxi se hidrogenan mucho más fácilmente que el de la espinosina L y su derivado 3'-etoxi, debido al impedimento estérico por el grupo metilo en C-5 en la espinosina L y su derivado 3'-etoxi. Véase, la Patente de EE.UU. N° 6.001.981. Las espinosinas J y L también se denominan en la presente memoria espinosinas J/L.

En esta solicitud se ha demostrado que la expresión de *VHb* en *S. spinosa* da como resultado un aumento en la producción de espinosinas. La expresión de la proteína portadora de oxígeno, *VHb*, produce niveles elevados de la hemoglobina homodimérica bajo condiciones de crecimiento hipóxicas (Zhang et al., 2007). Entre las hemoglobinas, la *VHb* tiene un valor medio de la constante de la velocidad de asociación al oxígeno y una constante bastante alta de la velocidad de disociación del oxígeno (cientos de veces mayor que las de otras hemoglobinas). Esto sugiere que *VHb* es capaz de liberar el oxígeno unido más fácilmente que las demás hemoglobinas (Zhang et al., 2007).

La expresión de *VHb* en varias cepas de *Actinomyces* condujo a un aumento en la producción de antibióticos incluyendo clortetraciclina, monensina, eritromicina y actinorrodina (Zhang et al., 2007; Magnolo et al., 1991). La cepa de *Streptomyces coelicolor* recombinante que expresa *VHb* produjo diez veces más actinorrodina que la cepa de tipo natural a bajos niveles de oxígeno disuelto (OD) (OD inferior al 5% de saturación en aire) (Magnolo et al., 1991). Además, la producción de actinorrodina por la cepa recombinante que expresa *VHb* era menos sensible a las condiciones de aireación. Aunque la producción de eritromicina no se considera en general sensible a los niveles de OD, siempre que los niveles de OD estén por encima de los niveles mínimos requeridos para el crecimiento, la expresión de *VHb* en una cepa productora de eritromicina industrial aumentó significativamente los títulos de eritromicina hasta 7,25 g/L desde 4,25 g/L, mientras que redujo la acumulación de biomasa en condiciones de fermentación en un biorreactor alimentado por lotes a escalas entre 10-15 litros (Brunker et al., 1997; 1998; Minas et al., 1998).

Las realizaciones de la presente descripción proporcionan una célula hospedante genéticamente modificada que alberga un gen mejorador de espinosinas integrado en el genoma de *S. spinosa*.

Se han clonado genes que codifican un gran número de proteínas que se unen al oxígeno y se han determinado sus secuencias. Por ejemplo, las secuencias de polinucleótido conocidas de proteínas de globina útiles en la presente invención incluyen, aunque sin limitación, las que codifican una mioglobina de cianobacterias (Potts et al., 1992, *Science* 256:1690-1692), hemoglobina de *Scapharca inaequalvis* (Gambacurta et al., 1993, *FEBS Lett.* 330:90-94), mioglobina de *Aplysia limacina* (Cutruzzola et al., 1996, *Biochem. J.* 314:83-90), hemoglobina de *Ascaris* (Sherman et al., 1992, *Proc. Natl. Acad. Sci. USA* 89:11696-11700), hemoglobina del nemátodo *Pseudoterranova decipiens* (Dixon et al., 1991, *Natl. Acad. Sci. USA* 88:5655-5659 y Dixon et al., 1992, *J. Mol. Evol.* 35:131-136), hemoglobina de *Paramecium caudatum* (Yamauchi et al., 1992, *Biochem. Biophys. Res. Commun.* 182:195-200), hemoproteína de *Rhizobium meliloti* (David et al., 1988, *Cell* 54:671-683) y *Saccharomyces cerevisiae* (Shimada et al., 1989, *J. Biochem.* 105:417-422). Particularmente adecuadas para su uso en la presente invención son las proteínas que se unen al oxígeno que tienen velocidades k_{off} relativamente altas tal como *VHb* (k_{off} 5600 s⁻¹; Orii and Webster, 1986, *J. Biol. Chem.* 261:3544-3547) o afinidad al oxígeno relativamente baja, tal como la mioglobina de corazón de caballo (K_D 0,79 μM; Wittenberg et al., 1985, en *Nitrogen Fixation Research Progress*. H. J. Evand et al., Eds. Martinus Nijhoff Publishers, Dordrecht, p. 354). Por tanto, las proteínas que se unen al oxígeno preferidas pueden ser las proteínas con una velocidad k_{off} para el oxígeno mayor que 10 s⁻¹, más preferiblemente mayor que 100 s⁻¹, o una K_D para el oxígeno mayor que 0,5 μM, aunque se entenderá que también serán útiles las proteínas que se unen al oxígeno con constantes de velocidad fuera de estos parámetros. Como se indicó anteriormente, las proteínas que se unen al oxígeno preferidas incluyen globinas, tales como hemoglobina, mioglobina y legemoglobinas. Las propiedades de muchas proteínas que se unen al oxígeno, incluyendo globinas, están descritas en la bibliografía. Adicionalmente, las técnicas para determinar las propiedades de unión al oxígeno de una proteína, tal como una globina, son muy conocidas por los expertos en la técnica y se pueden realizar sin excesiva experimentación.

Como se indicó anteriormente dicha proteína que se une al oxígeno para uso en la presente invención, como se ha descrito en la presente memoria a modo de ejemplo de trabajo, es *VHb*. La secuencia completa del gen de *VHb* está descrita en la Patente de EE.UU. N° 5.049.493. Los fragmentos, mutantes, variantes y derivados de *VHb* que se unen al oxígeno también están dentro del alcance de la presente descripción.

También están incluidos fragmentos, mutantes, variantes y derivados de la secuencia de polinucleótido que codifica

VHb, que codifican una proteína que se une al oxígeno. Preferiblemente, la presente descripción proporciona el uso de la secuencia codificadora de VHb del vector de expresión descrito en la SEQ ID NO 1, o sus fragmentos, variantes o derivados.

5 Los mutantes incluyen, aunque sin limitación, "polimorfismo(s) funcional(es)", que como se utiliza en la presente memoria se refiere a un cambio en la secuencia de pares de bases de un gen que produce un cambio cualitativo o cuantitativo en la actividad de la proteína codificada por dicho gen (por ejemplo, un cambio en la especificidad de la actividad; un cambio en el nivel de actividad). El término "polimorfismo funcional" incluye mutaciones, deleciones e inserciones.

10 En general, la etapa de detección del polimorfismo de interés se puede realizar recogiendo una muestra biológica que contenga DNA de la fuente y determinando a continuación en la muestra biológica la presencia o ausencia de DNA que contiene el polimorfismo de interés.

15 La determinación de la presencia o ausencia de DNA que codifica una mutación particular se puede llevar a cabo con una sonda de oligonucleótidos marcada con un grupo detectable adecuado y/o por una reacción de amplificación, tal como una reacción en cadena de polimerasa o una reacción en cadena de ligasa (el producto de la cual se puede detectar entonces con una sonda de oligonucleótidos marcada u otras técnicas). Se sabe que para realizar la presente invención se pueden emplear numerosos formatos de ensayo con diferentes sondas de oligonucleótidos. Véase, por ejemplo, la Patente de EE.UU. N° 4.302.204 de Wahl et al.; la Patente de EE.UU. N° 4.358.535 de Falkow et al.; la Patente de EE.UU. N° 4.563.419 de Ranki et al.; y la Patente de EE.UU. N° 4.994.373 de Stavrianopoulos et al.

20 La amplificación de una secuencia de ácidos nucleicos seleccionada, o diana, se puede realizar por cualquier medio adecuado. Véase, en general, Kwoh et al., *Am. Biotechnol. Lab.* 8, 14-25 (1990). Ejemplos de técnicas de amplificación adecuadas incluyen, aunque sin limitación, reacción en cadena de polimerasa, reacción en cadena de ligasa, amplificación por desplazamiento de cadena (véase en general G. Walker et al., *Proc. Natl. Acad. Sci. USA* 89, 392-396 (1992); G. Walker et al., *Nucleic Acids Res.* 20, 1691-1696 (1992)), amplificación basada en la transcripción (véase D. Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173-1177 (1989)), replicación de secuencias autosostenida (o "3SR") (véase J. Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87, 1874-1878 (1990)), el sistema de Q β replicasa (véase P. Lizardi et al., *BioTechnology* 6, 1197-1202 (1988)), amplificación basada en secuencias de ácidos nucleicos (o "NASBA") (véase R. Lewis, *Genetic Engineering News* 12 (9), 1 (1992)), la reacción en cadena de reparación (o "RCR") (véase R. Lewis, supra) y amplificación de DNA bumerán (o "BDA") (véase R. Lewis, supra). Generalmente se prefiere la reacción en cadena de polimerasa.

35 La reacción en cadena de polimerasa (PCR) se puede llevar a cabo de acuerdo con técnicas conocidas. Véase, por ejemplo, las Patentes de EE.UU. Nos. 4.683.195; 4.683.202; 4.800.159; y 4.965.188. En general, la PCR implica, en primer lugar, tratar una muestra de ácido nucleico (por ejemplo, en presencia de una DNA-polimerasa estable al calor) con un cebador oligonucleotídico para cada cadena de la secuencia específica que se ha de detectar en condiciones de hibridación, de modo que se sintetice un producto de extensión de cada cebador que sea complementario con cada cadena de ácidos nucleicos, siendo los cebadores suficientemente complementarios con cada cadena de la secuencia específica para hibridarse con ella de modo que el producto de extensión sintetizado a partir de cada cebador, cuando se separa de su complemento, pueda servir como un molde para la síntesis del producto de extensión del otro cebador, y luego tratar la muestra en condiciones de desnaturalización para separar los productos de extensión del cebador de sus moldes, si están presentes la secuencia o secuencias que se han de detectar. Estas etapas se repiten cíclicamente hasta que se obtiene el grado deseado de amplificación. La detección de la secuencia amplificada se puede realizar añadiendo al producto de reacción una sonda de oligonucleótidos capaz de hibridarse con el producto de reacción (por ejemplo, una sonda de oligonucleótidos de la presente descripción), llevando la sonda un marcador detectable y detectando luego el marcador de acuerdo con técnicas conocidas, o por visualización directa sobre un gel. Dichas sondas pueden tener una longitud de 5 a 500 nucleótidos, preferiblemente de 5 a 250, más preferiblemente de 5 a 100 o de 5 a 50 ácidos nucleicos. Cuando las condiciones de la PCR permiten la amplificación de todos los tipos alélicos, dichos tipos se pueden distinguir por hibridación con una sonda específica alélica, por digestión con endonucleasas de restricción, por electroforesis en geles con gradientes de desnaturalización u otras técnicas.

50 La reacción en cadena de ligasa (LCR) también se realiza de acuerdo con técnicas conocidas. Véase, por ejemplo, R. Weiss, *Science* 254, 1292 (1991). En general, la reacción se lleva a cabo con dos pares de sondas de oligonucleótidos: un par se une a una cadena de la secuencia que se ha de detectar; el otro par se une a la otra cadena de la secuencia que se ha de detectar. Cada par juntos solapan completamente la cadena correspondiente. La reacción se realiza, en primer lugar, desnaturalizando (por ejemplo, separando) las cadenas de la secuencia que se ha de detectar, haciendo reaccionar a continuación las cadenas con los dos pares de sondas de oligonucleótidos en presencia de una ligasa estable al calor, de modo que se ligan entre sí cada par de sondas de oligonucleótidos, separando luego el producto de reacción y repitiendo luego cíclicamente el proceso hasta que la secuencia ha sido amplificada en el grado deseado. A continuación se puede realizar la detección de la misma manera que se ha descrito anteriormente respecto a la PCR.

60 Las técnicas de amplificación del DNA, tales como las anteriores, pueden implicar el uso de una sonda, un par de

sondas o dos pares de sondas que se unen específicamente al DNA que contiene el polimorfismo funcional, pero no se unen al DNA que no contiene el polimorfismo funcional. Alternativamente, la sonda o el par de sondas podría unirse al DNA que tanto contiene como no contiene el polimorfismo funcional, pero produce o amplifica un producto (por ejemplo, un producto de alargamiento) en el que se puede determinar una diferencia detectable (por ejemplo, un producto más corto, donde el polimorfismo funcional es una mutación por delección). Dichas sondas se pueden generar de acuerdo con técnicas estándares a partir de las secuencias de DNA conocidas en o asociadas a un gen unido a *VHb* o a partir de secuencias que se pueden generar a partir de dichos genes de acuerdo con técnicas estándares.

Se apreciará que las etapas de detección descritas en la presente memoria se pueden realizar directa o indirectamente. Otros medios de determinar indirectamente el tipo alélico incluyen la medición de marcadores polimórficos que están unidos al polimorfismo funcional particular, como se ha demostrado para las VNTR (repeticiones en tándem de número variable).

La biología molecular comprende una amplia variedad de técnicas para el análisis de secuencias de ácidos nucleicos y proteínas. Muchas de estas técnicas y procedimientos constituyen la base de pruebas y ensayos de diagnóstico clínico. Estas técnicas incluyen análisis por hibridación de ácidos nucleicos, análisis con enzimas de restricción, análisis de secuencias genéticas y la separación y purificación de ácidos nucleicos y proteínas (véase, por ejemplo, J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, 2 Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989).

La mayoría de estas técnicas implican realizar numerosas operaciones (por ejemplo, pipeteado, centrifugación y electroforesis) en un gran número de muestras. Dichas operaciones son a menudo complejas y consumen tiempo, y requieren generalmente un alto grado de precisión. Muchas técnicas están limitadas en su aplicación por una falta de sensibilidad, especificidad o reproducibilidad.

El análisis por hibridación de ácidos nucleicos implica generalmente la detección de un número muy pequeño de ácidos nucleicos diana específicos (DNA o RNA) con un exceso de DNA sonda, entre una cantidad relativamente grande de ácidos nucleicos no diana complejos. Una reducción en la complejidad del ácido nucleico en una muestra es útil para la detección de bajo número de copias (es decir, 10.000 a 100.000) de ácidos nucleicos diana. La reducción de la complejidad del DNA se logra en cierto grado por la amplificación de secuencias de ácidos nucleicos diana. (Véase, M.A. Innis et al., *PCR Protocols: A guide to Methods and Applications*, Academic Press, 1990, Spargo et al., 1996, *Molecular & Cellular Probes*, respecto a la amplificación por desplazamiento de cadena (SDA)). Esto es debido a que la amplificación de ácidos nucleicos diana da como resultado un enorme número de secuencias de ácidos nucleicos diana con relación a las secuencias no diana, mejorando así la etapa de hibridación de dianas subsiguiente.

La etapa de hibridación implica colocar la muestra de DNA preparada en contacto con una sonda indicadora específica en las condiciones óptimas ajustadas para que tenga lugar la hibridación entre la secuencia de DNA diana y la sonda. La hibridación se puede realizar en uno cualquiera entre varios formatos. Por ejemplo, el análisis por hibridación de ácidos nucleicos en múltiples muestras se ha realizado en una variedad de formatos de filtros y de soportes sólidos, (véase Beltz et al., *Methods in Enzymology*, Vol. 100, Part et al., Eds., Academic Press, New York, Chapter 19, pp. 266-308, 1985). Un formato, la llamada hibridación por "transferencia de mancha", implica la unión no covalente de los DNA diana a un filtro seguido por la posterior hibridación con una o varias sondas marcadas con radioisótopos. La hibridación por "transferencia de mancha" alcanzó un uso generalizado en las dos últimas décadas durante las cuales se desarrollaron muchas versiones (véase Anderson and Young, en *Nucleic Acid Hybridization – A Practical Approach*, Hames and Higgins, Eds, IRL Press, Washington, D.C., Chapter 4, pp.73-111, 1985). Por ejemplo, el método de transferencia de mancha se ha desarrollado para múltiples análisis de mutaciones genómicas (EP-A 0228075 de Nanibhushan et al.) y para la detección de clones solapantes y la construcción de mapas genómicos (Patente de EE.UU. N° 5.219.726 de Evans).

Técnicas adicionales para realizar análisis de hibridación de ácidos nucleicos de múltiples muestras incluyen dispositivos múltiplex micro-formateados o de matriz (por ejemplo, chips de DNA) (véase M. Barinaga, 253 *Science*, pp.1489, 1991; W. Bains, 10 *Bio/Technology*, pp. 757-758, 1992). Estos métodos unen generalmente secuencias de DNA específicas a áreas específicas muy pequeñas de un soporte sólido, tales como micro-pocillos de un chip de DNA. Estos formatos de hibridación son versiones a microescala de los sistemas de hibridación convencionales por "transferencia de mancha" y en "sándwich".

La hibridación micro-formateada se puede utilizar para realizar "secuenciación por hibridación" (SBH) (véase M. Barinaga, 253 *Science*, pp. 1489, 1991; W. Bains, 10 *Bio/Technology*, pp. 757-758, 1992). La SBH hace uso de todos los oligómeros de n-nucleótidos (n-meros) posibles para identificar n-meros en una muestra de DNA desconocida, que se alinean subsiguientemente por análisis de algoritmos produciendo la secuencia de DNA (véase, Drmanac, Patente de EE.UU. N° 5.202.231).

Existen dos formatos para la realización de la SBH. El primer formato implica crear una matriz de todos los n-meros posibles sobre un soporte, que se hibrida luego con la secuencia diana. El segundo formato implica unir la secuencia diana a un soporte, que se sonda secuencialmente con todos los n-meros posibles. Southern (Solicitud de Patente

del Reino Unido GB 8810400, 1988; E. M. Southern et al., *13 Genomics* 1008, 1992), propuso utilizar el primer formato para analizar o secuenciar DNA. Southern identificó una única mutación puntual conocida utilizando DNA genómico amplificado por PCR. Southern describió también un método para sintetizar una matriz de oligonucleótidos sobre un soporte sólido para la SBH. Drmanac et al., (260 *Science* 1649-1652, 1993) utilizaron un segundo formato para secuenciar varias secuencias cortas de DNA (116 pb). Los DNA diana se unieron a soportes de membrana (formato de "transferencia de mancha"). Cada filtro se hibridó secuencialmente con 272 oligonucleótidos de 10-meros y 1-mero marcados. Se utilizaron amplios intervalos de condiciones restrictivas para lograr una hibridación específica para cada sonda de n-meros. Los tiempos de lavado variaron desde 5 minutos hasta una noche utilizando temperaturas de 0°C a 16°C. La mayoría de las sondas requirieron 3 horas de lavado a 16°C. Los filtros tuvieron que ser expuestos de 2 a 18 horas con el fin de detectar las señales de hibridación.

En general, están disponibles una variedad de métodos para la detección y el análisis de los episodios de hibridación. Dependiendo del grupo indicador (fluoróforo, enzima, radioisótopo, etc.) utilizado para marcar la sonda de DNA, la detección y el análisis se realizan fluorimétricamente, calorimétricamente o por autorradiografía. Observando y midiendo la radiación emitida, tal como la radiación fluorescente o la emisión de partículas, se puede obtener información acerca de los episodios de hibridación. Incluso cuando los métodos de detección tienen muy alta sensibilidad intrínseca, la detección de los episodios de hibridación es difícil debido a la presencia de fondo de materiales unidos no específicamente. Por tanto, la detección de episodios de hibridación depende de cómo se pueda realizar la hibridación específica y sensible. En cuanto a los análisis genéticos, se han desarrollado varios métodos que han intentado aumentar la especificidad y sensibilidad.

Una forma de análisis genético es el análisis centrado en la elucidación de polimorfismos en un solo nucleótido ("SNP"). Factores que favorecen el uso de los SNP son su gran abundancia en el genoma humano (especialmente en comparación con repeticiones cortas en tándem (STR)), su localización frecuente dentro de regiones de genes codificadoras o reguladoras (que pueden afectar a la estructura proteica o a los niveles de expresión) y su estabilidad cuando pasan de una generación a la siguiente (Landegren et al., *Genome Research*, Vol. 8, pp.769-776, 1998).

Un SNP se define como cualquier posición en el genoma que existe en dos variantes y la variante más común se produce menos del 99% de las veces. Con el fin de utilizar los SNP como marcadores genéticos generalizados, es crucial que puedan ser genotipificados fácil, rápida, precisa y rentablemente. Numerosas técnicas están actualmente disponibles para tipificar los SNP (para revisión, véase Landegren et al., *Genome Research*, Vol. 8, pp. 769-776, (1998)), las cuales requieren amplificación de la diana. Dichas técnicas incluyen secuenciación directa (Carothers et al., *BioTechniques*, Vol. 7, pp. 494-499, 1989), polimorfismo de conformación monocatenaria (Orita et al., *Proc. Natl. Acad. Sci. USA*, Vol. 86, pp. 2766-2770, 1989), amplificación específica de alelos (Newton et al., *Nucleic Acids Research*, Vol. 17, pp. 2503-2516, (1989)), digestión por restricción (Day and Humphries, *Analytical Biochemistry*, Vol. 222, pp. 389-395, 1994) y ensayos de hibridación. En su forma más básica, los ensayos de hibridación funcionan discriminando indicadores de oligonucleótidos cortos contra dianas apareadas y desapareadas. Se han desarrollado muchas adaptaciones al protocolo básico. Estas incluyen reacción en cadena con ligación (Wu and Wallace, *Gene*, Vol. 76, pp. 245-254, 1989) y minisequenciación (Syvanen et al., *Genomics*, Vol. 8, pp. 684-692, 1990). Otras mejoras incluyen el uso de la actividad 5'-nucleasa de la DNA-polimerasa Taq (Holland et al., *Proc. Natl. Acad. Sci. USA*, Vol. 88, pp. 7276-7280, 1991), balizas moleculares (Tyagi and Kramer, *Nature Biotechnology*, vol. 14, pp. 303-308, 1996), curvas de desnaturalización por calor (Howell et al., *Nature Biotechnology*, Vol. 17, pp. 87-88, 1999) y "chips" de DNA (Wang et al., *Science*, Vol. 280, pp.1077-1082, 1998).

Un fenómeno adicional que se puede utilizar para distinguir los SNP es las energías de interacción de los ácidos nucleicos o energías de apilamiento de bases resultantes de la hibridación de múltiples sondas específicas diana con una sola diana. (Véase, R. Ornstein et al., "An Optimized Potential Function for the Calculation of Nucleic Acid Interaction Energies", *Biopolymers*, Vol. 17, 2341-2360 (1978); J. Norberg and L. Nilsson, *Biophysical Journal*, Vol. 74, pp. 394-402, (1998); y J. Pieters et al., *Nucleic Acids Research*, Vol. 17, nº 12, pp. 4551-4565 (1989)). Este fenómeno de apilamiento de bases se usa en un formato único en la presente invención para proporcionar diferenciales de Tm altamente sensibles que permiten la detección directa de los SNP en una muestra de ácidos nucleicos.

Se han utilizado métodos adicionales para distinguir secuencias de ácidos nucleicos en organismos relacionados o para secuenciar DNA. Por ejemplo, la Patente de EE.UU. Nº 5.030.557 de Hogan et al., describió que la estructura secundaria y terciaria de un ácido nucleico diana monocatenario puede verse afectada por la unión de oligonucleótidos "auxiliares" además de los oligonucleótidos "sonda" haciendo que se presenta un mayor Tm entre la sonda y el ácido nucleico diana. Sin embargo, dicha patente se limitaba en su procedimiento a la utilización de energías de hibridación sólo para alterar las estructuras secundaria y terciaria de cadenas de RNA de auto-asociación, que si se dejaban inalteradas tenderían a evitar que la sonda se hibridara con la diana.

Respecto a la secuenciación del DNA, K. Khrapko et al., *Federation of European Biochemical Societies Letters*, Vol. 256, nº 1,2, pp. 118-122 (1989), por ejemplo, describieron que la hibridación por apilamiento continuo dio como resultado una estabilización del dúplex. Adicionalmente, J. Kieleczawa et al., *Science*, Vol. 258, pp. 1787-1791 (1992), describieron el uso de tramos contiguos de hexámeros para cebar la síntesis de DNA en donde parecía que los tramos contiguos estabilizaban el cebado. Igualmente, L. Kotler et al., *Proc. Natl. Acad. Sci. USA*, Vol. 90, pp. 4241-4245, (1993) describieron la especificidad de secuencias en el cebado de las reacciones de secuenciación de DNA por el uso de módulos de oligonucleótidos hexámeros y pentámeros. Además, S. Parinov et al., *Nucleic Acids Re-*

5 *search*, Vol. 24, nº 15, pp. 2998-3004, (1996), describieron el uso de oligómeros de apilamiento de bases para la secuenciación del DNA en asociación con microchips de secuenciación de DNA pasiva. Por otra parte, G. Yershov et al., *Proc. Natl. Acad. Sci. USA*, Vol. 93, pp. 4913-4918 (1996), describieron la aplicación de las energías del apilamiento de bases en SBH sobre un microchip pasivo. En el ejemplo de Yershov, se anclaron sondas de DNA de 10-meros a la superficie del microchip y se hibridaron con secuencias diana en conjunción con sondas cortas adicionales, pareciendo que dicha combinación estabilizaba la unión de las sondas. En dicho formato, podrían ser elucidados segmentos cortos de una secuencia de ácidos nucleicos para la secuenciación de DNA. Yershov señaló además que en su sistema se incrementó el efecto desestabilizador de los desapareamientos utilizando sondas más cortas (por ejemplo, 5-meros). El uso de dichas sondas cortas en la secuenciación de DNA proporcionó la capacidad de discernir la presencia de desapareamientos a lo largo de la secuencia que se estaba sondando en lugar de un único desapareamiento en una localización específica del complejo de hibridación sonda/diana. El uso de sondas más largas (por ejemplo, oligos de 8-meros, 10-meros y 13-meros) fue menos funcional para dichos fines.

10 Un ejemplo adicional de metodologías que han utilizado el apilamiento de bases en el análisis de ácidos nucleicos incluye la Patente de EE.UU. Nº 5.770.365 de Lane et al., en donde se describe un método para capturar dianas de ácidos nucleicos utilizando una sonda de captura unimolecular que tiene un bucle monocatenario y una región bicatenaria que actúa junto con una diana de unión para estabilizar la formación de dúplex por las energías de apilamiento.

15 La secuencia de nucleótidos puede ser modificada convenientemente por mutagénesis dirigida al sitio de acuerdo con métodos convencionales. Alternativamente, la secuencia de nucleótidos se puede preparar por síntesis química, incluyendo, aunque sin limitación, el uso de un sintetizador de oligonucleótidos, en donde los oligonucleótidos se diseñan basándose en la secuencia de aminoácidos del polipéptido deseado, y preferiblemente seleccionando los codones que son favorecidos en la célula hospedante en la que se producirá el polipéptido recombinante.

20 Las nuevas espinosinas también se pueden producir por mutagénesis de los genes clonados y sustitución en un organismo productor de espinosinas de los homólogos no mutados por los genes mutados. Preferiblemente, la sustitución comprende un método de introducción de polinucleótidos como se describe en la presente memoria. Se puede usar cualquier método de mutagénesis adecuado, incluyendo, por ejemplo mutagénesis de cualquier enzima biosintética de espinosinas bien directamente o por mutagénesis de un gen que codifica una enzima biosintética de espinosinas. La descripción proporciona mutagénesis de cualquier espinosina de *S. spinosa*. Preferiblemente, la mutagénesis puede implicar, por ejemplo: 1) delección o inactivación de un dominio de cetorreductasa, deshidratasa o enoil-reductasa (KR, DH o ER) de modo que se bloquee una o más de estas funciones y la cepa produzca una espinosina que tenga un núcleo de lactona con un doble enlace, un grupo hidroxilo o un grupo ceto que no está presente en el núcleo de espinosina A (véase Donadio et al., 1993); 2) sustitución de un dominio de AT de modo que se incorpore en el núcleo de lactona un ácido carboxílico diferente (véase Ruan et al, 1997); 3) adición de un dominio de KR, DH o ER a un módulo de PKS existente, de modo que la cepa produzca una espinosina que tenga un núcleo de lactona con un enlace saturado, un grupo hidroxilo o un doble enlace que no está presente en el núcleo de espinosina A; o 4) adición o sustracción de un módulo de PKS completo de modo que el núcleo de lactona cíclica tenga un número mayor o menor de átomos de carbono. Se puede crear una PKS híbrida sustituyendo el dominio de carga de la PKS de espinosina con carga de PKS heteróloga. Véase, por ejemplo, la Patente de EE.UU. Nº 7.626.010. Se ha observado además que las espinosinas por la modificación de los azúcares que están unidos a la cadena principal de la lactona de la espinosina pueden incluir modificaciones del resto de ramnosa y/o forosamina o unión de diferentes desoxiazúcares. El grupo de Salas en España demostró que se pueden producir nuevos compuestos de policétidos sustituyendo la molécula de azúcar existente por diferentes moléculas de azúcar. Rodríguez et al., *J. Mol. Microbiol Biotechnol.* 2000 Jul; 2(3): 271-6. Los ejemplos que siguen en la solicitud ayudan a ilustrar el uso de mutagénesis para producir una espinosina con funcionalidad modificada.

25 El DNA de la región de agrupación de genes de espinosinas se puede utilizar como sonda de hibridación para identificar secuencias homólogas. Por tanto, el DNA clonado de la presente memoria se podría utilizar para localizar otros plásmidos de las genotecas de *Saccharopolyspora spinosa* que solapan la región descrita en la presente memoria, pero que también contuvieran anteriormente el DNA no clonado de las regiones adyacentes en el genoma de *Saccharopolyspora spinosa*. Además, el DNA de la región clonada de la presente memoria se puede utilizar para identificar en otros organismos secuencias no idénticas pero similares. Las sondas de hibridación tienen normalmente una longitud de al menos aproximadamente 20 bases y están marcadas para permitir su detección.

30 En la invención se pueden utilizar varios tipos de mutagénesis con una variedad de fines. Incluyen, aunque sin limitación, mutagénesis puntual aleatoria dirigida al sitio, recombinación homóloga, barajadura del DNA u otros métodos de mutagénesis recursivos, construcción química, mutagénesis que usa moldes que contienen uracilo, mutagénesis dirigida por oligonucleótidos, mutagénesis del DNA modificado con fosforotioato, mutagénesis que usa DNA dúplex con huecos o similares, o cualquiera de sus combinaciones. Métodos adecuados adicionales incluyen reparación de desapareamiento puntual, mutagénesis que usa cepas hospedantes deficientes en reparación, selección por restricción y purificación por restricción, mutagénesis por delección, mutagénesis por síntesis génica total, reparación por rotura de la doble cadena y similares. Las mutagénesis, que incluyen, aunque sin limitación, la participación de construcciones químicas, también están incluidas en la presente descripción. En una realización, la mutagénesis puede ser guiada por información conocida de la molécula de origen natural o de la molécula de origen natural alterada o mutada, incluyendo, aunque sin limitación, de la secuencia, comparaciones de secuencias, propiedades

físicas, estructura cristalina o similares.

Los genes responsables de las vías metabólicas secundarias en *S. spinosa*, no esenciales para las actividades metabólicas primarias del organismo, representan una fuente atractiva para los sitios neutros. De estos, son de particular interés las agrupaciones de genes de policétido-sintasa que no son de espinosinas puesto que la interrupción de las agrupaciones de genes pueden dar como resultado la eliminación de las vías competidoras potenciales para los precursores comunes de acil-CoA y los cofactores esenciales para la biosíntesis y producción de espinosinas. Esto puede producir una sinergia entre los beneficios de los genes diana y la mayor disponibilidad de los precursores. En una realización, la presente invención proporciona la integración de un polinucleótido como se define en la presente memoria en un sitio neutro en el genoma de la célula hospedante, por ejemplo en el locus de la policétido-sintasa de la obscurina, por ejemplo el gen *obsA*.

Los textos y ejemplos encontrados en la presente memoria describen estos procedimientos. Se encuentra información adicional en las siguientes publicaciones y referencias citadas en: Ling et al., *Approaches to DNA mutagenesis: an overview*, *Anal. Biochem.* 254(2): 157-178 (1997); Dale et al., *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, *Methods Mol. Biol.* 57:369-374 (1996); Smith, *In vitro mutagenesis*, *Ann. Rev. Genet.* 19:423-462 (1985); Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, *Science* 229:1193-1201 (1985); Carter, *Site-directed mutagenesis*, *Biochem. J.* 237:1-7 (1986); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, en *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, *Proc. Natl. Acad. Sci. USA* 82:488-492 (1985); Kunkel et al., *Rapid and efficient site-specific mutagenesis without phenotypic selection*, *Methods in Enzymol.* 154, 367-382 (1987); Bass et al., *Mutant Trp repressors with new DNA-binding specificities*, *Science* 242:240-245 (1988); *Methods in Enzymol.* 100:468-500 (1983); *Methods in Enzymol.* 154:329-350 (1987); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment*, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, *Methods in Enzymol.* 100:468-500 (1983); Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, *Methods in Enzymol.* 154:329-350 (1987); Taylor et al., *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, *Nucl. Acids Res.* 13:8749-8764 (1985); Taylor et al., *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, *Nucl. Acids Res.* 13:8765-8787 (1985); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, *Nucl. Acids Res.* 14:9679-9698 (1986); Sayers et al., *Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, *Nucl. Acids Res.* 16:791-802 (1988); Sayers et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) *Nucl. Acids Res.* 16:803-814; Kramer et al., *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, *Nucl. Acids Res.* 12:9441-9456 (1984); Kramer & Fritz, *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, *Methods in Enzymol.* 154:350-367 (1987); Kramer et al., *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, *Nucl. Acids Res.* 16:7207 (1988); Fritz et al., *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, *Nucl. Acids Res.* 16:6987-6999 (1988); Kramer et al., *Point Mismatch Repair*, *Cell* 38:879-887 (1984); Carter et al., *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, *Nucl. Acids Res.* 13:4431-4443 (1985); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, *Methods in Enzymol.* 154:382-403 (1987); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, *Nucl. Acids Res.* 14:5115 (1986); Wells et al., *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, *Phil. Trans. R. Soc. Lond. A* 317:415-423 (1986); Nambiar et al., *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, *Science* 223:1299-1301 (1984); Sakamar and Khorana, *Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, *Nucl. Acids Res.* 14:6361-6372 (1988); Wells et al., *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites*, *Gene* 34:315-323 (1985); Grundstrom et al., *Oligonucleotide-directed mutagenesis by microscale 'shot gun' gene synthesis*, *Nucl. Acids Res.* 13:3305-3316 (1985); Mandecki, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181 (1986); Arnold, *Protein engineering for unusual environments*, *Current Opinion in Biotechnology* 4:450-455 (1993); Sieber, et al., *Nature Biotechnology*, 19:456-460 (2001). W. P. C. Stemmer, *Nature* 370, 389-91 (1994); y, 1. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995). Detalles adicionales sobre muchos de los métodos anteriores se pueden encontrar en *Methods in Enzymology*, Volumen 154, que describe también controles útiles para la localización y resolución de problemas con varios métodos de mutagénesis.

También se pueden utilizar fragmentos de péptidos, polipéptidos, proteínas, polinucleótidos y genes descritos en la presente memoria. Los fragmentos pueden comprender al menos 20%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o al menos 99% de la longitud de la secuencia de referencia (por ejemplo, secuencia de tipo natural o de origen natural). En una realización, los fragmentos pueden ser los que son funcionalmente equivalentes en al menos un aspecto a la secuencia de referencia.

Los términos "homología" o "porcentaje de identidad" se usan indistintamente en la presente memoria. Para los fines de esta invención, se define en la presente memoria que con el fin de determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos secuencias de ácidos nucleicos, las secuencias se alinean para fines de com-

paración óptimos (por ejemplo, se pueden introducir huecos en la secuencia de una primera secuencia de aminoácidos o ácidos nucleicos para alineación óptima con una segunda secuencia de aminoácidos o ácidos nucleicos). Se comparan a continuación los residuos de aminoácidos o nucleótidos en las posiciones de aminoácidos o posiciones de nucleótidos correspondientes. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácidos o nucleótidos que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en dicha posición. El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias (es decir, % de identidad = número de posiciones idénticas/número total de posiciones (es decir, posiciones solapantesx100). Preferiblemente, las dos secuencias tienen la misma longitud. En la presente invención, una variante o derivado de péptido, polipéptido, proteína o polinucleótido o secuencia de genes es uno que comprende una o más deleciones, adiciones o sustituciones, y puede tener al menos 20%, 30%, 40%, 50%, 60 %, 70%, 75%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o al menos 99% de identidad de secuencias con la secuencia de referencia (por ejemplo, de tipo natural o de origen natural). En una variante o derivado, las sustituciones pueden ser sustituciones conservadoras, en las que los aminoácidos o ácidos nucleicos están sustituidos por aminoácidos o ácidos nucleicos que tienen propiedades similares, de tal manera que no se cambie la naturaleza y actividad de la secuencia. Alternativamente, las sustituciones pueden no ser conservadoras, de tal manera que estén sustituidas por unas que tengan propiedades diferentes que a su vez afecten a la naturaleza y propiedades de la secuencia.

Los expertos en la técnica serán conscientes del hecho de que están disponibles varios programas informáticos diferentes para determinar la homología entre dos secuencias. Por ejemplo, se pueden realizar una comparación de secuencias y la determinación del porcentaje de identidad entre dos secuencias utilizando un algoritmo matemático. En una realización preferida, el porcentaje de identidad entre dos secuencias de aminoácidos se determina utilizando el algoritmo de Needleman y Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) que se ha incorporado al programa GAP en el paquete de programas informáticos GCG (disponible en Internet en la página web de Accelrys, más específicamente en <http://www.accelrys.com>), usando bien una matriz Blossom 62 o una matriz PAM250, y un peso de huecos de 16, 14, 12, 10, 8, 6 o 4 y un peso de longitudes de 1, 2, 3, 4, 5 o 6. Los expertos en la técnica apreciarán que todos estos parámetros diferentes proporcionarán resultados ligeramente diferentes, pero que el porcentaje de identidad global de dos secuencias no se altera significativamente cuando se utilizan diferentes algoritmos.

Incluso en otra realización, el porcentaje de identidad entre dos secuencias de nucleótidos se determina usando el programa GAP en el paquete de programas informáticos GCG (disponible en Internet en la página web de Accelrys, más específicamente en <http://www.accelrys.com>), utilizando una matriz NWSgapdna.CMP y un peso de huecos de 40, 50, 60, 70 u 80 y un peso de longitudes de 1, 2, 3, 4, 5 o 6. En otra realización, el porcentaje de identidad entre dos secuencias de aminoácidos o de nucleótidos se determina usando el algoritmo de E. Meyers y W. Miller (*CABIOS*, 4:11-17 (1989) que se ha incorporado en el programa ALIGN (versión 2.0) (disponible en Internet en la página web de Vega, más específicamente ALIGN - IGH Montpellier, o más específicamente en <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) usando una tabla de residuos de peso PAM120, una penalización por longitud de hueco de 12 y una penalización por hueco de 4.

Las secuencias de ácidos nucleicos y proteínas de la presente descripción se pueden usar además como una "secuencia de consulta" para realizar una búsqueda en las bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Dichas búsquedas se pueden realizar usando los programas BLASTN y BLASTX (versión 2.0) de Altschul, et al., (1990) *J. Mol. Biol.* 215:403-10. Las búsquedas de nucleótidos por BLAST se pueden realizar con el programa BLASTN, puntuación = 100, longitud de palabra = 12, para obtener secuencias de nucleótidos homólogas a las moléculas de ácidos nucleicos de la presente descripción.

Las búsquedas de proteínas por BLAST se pueden realizar con el programa BLASTX, puntuación = 50, longitud de palabra = 3, para obtener secuencias de aminoácidos homólogas a las moléculas de proteínas de la presente descripción.

Para obtener alineamientos con huecos con fines comparativos, se puede utilizar Gapped BLAST como ha sido descrito por Altschul et al., (1997) *Nucleic Acids Res.* 25 (17): 3389-3402. Cuando se utilizan los programas BLAST y Gapped BLAST, se pueden usar los parámetros por defecto de los programas respectivos (por ejemplo, BLASTX y BLASTN). (Disponible en Internet en la página web de NCBI, más específicamente en www.ncbi.nlm.nih.gov).

Un casete de expresión para uso en la presente descripción puede comprender una secuencia de polinucleótido, preferiblemente un gen que codifica un gen mejorador de espinosinas y/o un gen que codifica una enzima biosintética de espinosinas. En una realización, el gen mejorador de espinosinas y un gen que codifica una enzima biosintética de espinosinas son como se han definido en la presente memoria. Un casete de expresión puede comprender opcionalmente un marcador seleccionable. Un casete de expresión puede comprender opcionalmente una o más secuencias de control, incluyendo un origen de replicación. Un casete de expresión puede comprender opcionalmente una o más secuencias para permitir la integración (por ejemplo, por recombinación) en un genoma de una célula hospedante. El término "casete de expresión" se puede usar indistintamente con términos tales como vector de expresión y construcción genética.

Vectores de expresión adecuados para su uso en la presente invención incluyen vectores procariotas y eucariotas (por ejemplo, plásmido, fagémido o bacteriófago), vectores de mamíferos y vectores de plantas. Los vectores proca-

riotas adecuados incluyen plásmidos tales como, aunque sin limitación, los comúnmente utilizados para la manipulación del DNA en *Actinomyces* (por ejemplo pSET152, pOJ260, pLJ101, pJV1, pSG5, pHJL302, pSAM2, pKC1250). Dichos plásmidos han sido descritos por Kieser et al., ("*Practical Streptomyces Genetics*", 2000). Otros vectores adecuados pueden incluir plásmidos tales como los capaces de replicación en *E. coli* (por ejemplo, pBR322, ColE1, pSC101, PACYC 184, itVX, pRSET, pBAD (Invitrogen, Carlsbad, Calif.) y similares). Dichos plásmidos han sido descritos por Sambrook (véase "*Molecular Cloning: A Laboratory Manual*", segunda edición, editado por Sambrook, Fritsch & Maniatis, Cold Spring Harbor Laboratory, (1989)) y muchos de dichos vectores están disponibles comercialmente. Los plásmidos de *Bacillus* incluyen pC194, pC221, pT127 y similares, y han sido descritos por Gryczan (en: *The Molecular Biology of the Bacilli*, Academic Press, NY (1982), pp. 307-329). Los plásmidos de *Streptomyces* adecuados incluyen pli101 (Kendall et al., *J. Bacteriol.* 169:4177-4183, 1987) y los bacteriófagos de *Streptomyces* incluyen, aunque sin limitación, tales como ψ C31 (Chater et al., en: *Sixth International Symposium on Actinomycetales Biology*, Akademiai Kiado, Budapest, Hungría (1986), pp. 45-54). Los plásmidos de *Pseudomonas* han sido revisados por John et al., (*Rev. Infect. Dis.* 8:693-704, 1986) e Izaki (*Jpn. J. Bacteriol.* 33:729-742, 1978). En una realización, un casete de expresión para uso en la presente descripción puede comprender una secuencia de polinucleótido que codifica un gen mejorador de espinosinas, como se define en la presente memoria, un terminador de la transcripción y un marcador seleccionable. Preferiblemente, el gen mejorador de espinosinas es una secuencia de polinucleótidos que codifica Vhb o uno de sus fragmentos, derivados o variantes. Independientemente, el terminador de la transcripción puede ser la secuencia terminadora de la transcripción aph. Independientemente, el marcador seleccionable puede ser un gen de resistencia a la apramicina. Preferiblemente, el casete de expresión es como se muestra en la SEQ ID NO. 1.

La supresión de la expresión de genes particulares es una herramienta importante tanto para la investigación como para el desarrollo de organismos modificados genéticamente más ajustados para un fin particular. El silenciamiento génico se puede conseguir por la introducción de un transgén que corresponde al gen de interés en la orientación antisentido con relación a su promotor (véase, por ejemplo, Sheehy et al., *Proc. Natl. Acad. Sci. USA* 85:8805-8808 (1988); Smith et al., *Nature* 334:724-726 (1988)), o en la orientación con sentido con relación a su promotor (Napoli et al., *Plant Cell* 2:279-289 (1990); van der Krol et al., *Plant Cell* 2:291-299 (1990); Patente de EE.UU. Nº 5.034.323; Patente de EE.UU. Nº 5.231.020; y Patente de EE.UU. No. 5.283.184), ambas de las cuales conducen a la expresión reducida del transgén así como al gen endógeno.

Se ha documentado que el silenciamiento génico postranscripcional va acompañado por la acumulación de pequeños fragmentos (20 a 25 nucleótidos) de RNA antisentido, que pueden ser sintetizados a partir de un molde de RNA y representan los determinantes de especificidad y movilidad del proceso (Hamilton & Baulcombe, *Science* 286:950-952 (1999)). Ha quedado claro que en una gama de organismos la introducción de dsRNA (RNA bicatenario) es un componente importante que conduce al silenciamiento génico (Fire et al., *Nature* 391:806-811 (1998); Timmons & Fire, *Nature* 395:854 (1998); el documento WO99/32619; Kennerdell & Carthew, *Cell* 95:1017-1026 (1998); Ngo et al., *Proc. Natl. Acad. Sci. USA* 95:14687-14692 (1998); Waterhouse et al., *Proc. Natl. Acad. Sci. USA* 95:13959-13964 (1998); el documento WO99/53050; Cogoni & Macino, *Nature* 399:166-169 (1999); Lohmann et al., *Dev. Biol.* 214:211-214 (1999); Sánchez-Alvarado & Newmark, *Proc. Natl. Acad. Sci. USA* 96:5049-5054 (1999)). En las bacterias, el gen suprimido no necesita ser un gen bacteriano endógeno, ya que tanto los transgenes informadores como los genes víricos son sometidos a silenciamiento génico postranscripcional por transgenes introducidos (English et al., *Plant Cell* 8:179-188 (1996); Waterhouse et al., supra). Sin embargo, en todos los casos anteriores, se puede preferir alguna similitud de secuencias entre el transgén introducido y el gen que es suprimido. En la presente memoria los términos "célula hospedante" y "célula hospedante recombinante" se usan indistintamente. Dichos términos se refieren no sólo a la célula objeto particular, sino también a la progenie o progenie potencial de dicha célula. Debido a que pueden producirse ciertas modificaciones en generaciones sucesivas debido a mutación o influencias ambientales, dicha progenie puede en efecto no ser idéntica a la célula parental, pero aún así está incluida dentro del alcance del término tal como se usa en la presente memoria.

Los métodos de la invención pueden incluir las etapas adicionales de:

- a) cultivar una célula modificada genéticamente de la invención, en condiciones apropiadas para la producción de espinosinas; y
- b) recoger las espinosinas de un cultivo de la célula.

Se incluyen los ejemplos siguientes para ilustrar la presente invención y no se deben interpretar como limitativos de la misma.

Ejemplos

Construcción de un casete de expresión del gen de hemoglobina

- Se diseñó un casete de expresión del gen de hemoglobina para que contuviera las siguientes características claves:
 - i) el promotor (*promotor aac3 (IV)*) del gen de resistencia a apramicina (originalmente de *Klebsiella pneumoniae* bajo el número de acceso en GenBank X99313; Kieser et al., 2001) que es funcional en *S. spinosa*; ii) el gen *Vhb* de la hemoglobina de *Vitreoscilla* con codones optimizados (Tabla 1) diseñado utilizando una tabla de preferencia de

codones de *S. spinosa* y el programa informático GeneDesigner (DNA 2.0, Menlo Park, California); y iii) la secuencia terminadora de la transcripción *aph* (Pulido y Jiménez, 1987). La selección y disposición de estos elementos génicos garantiza la transcripción estable de *VHb* y la expresión óptima de proteínas, proporcionando de esta manera una expresión fuerte de la proteína hemoglobina. El casete de expresión con el gen de la hemoglobina (SEQ ID NO: 1) se sintetizó por el programa DNA 2.0 (Menlo Park, CA). El fragmento sintético resultante de 977 pb flanqueado por *HindIII* se clonó en el vector pJ201 (Figura 1) y se secuenció por el programa DNA 2.0 para confirmar la exactitud de secuencia antes del envío. El plásmido resultante se denominó pDAB109000.
SEQ ID NO. 1

5 aagcttgaat tcgatatcag cccggtacct cgccgccaac ccgtaccacg gcctctgacg
gctccggcca agcccgcgaa agcgggggcc gggcgccggf ctccaggcgg cggggtacct

ccgaccgcac ggccgctcc ggagtggtec ggaagcggcc tgcggagcgc cctgcggggc

ggctatgat tcactccacc gcctgcgcgt acaggtccgc ctccacctgg atgaacacgt

cccgatcac gccgtaccg tggccccag cgtccaggat gtgtccgtc gccgcgtgc

ccagcacctc cttgatcgcg cccagcagct cctggcccac gatcgggtag tgcgccgccc

10 ccacccccg ctggcagtgc ttaccgcga tcttcttcac cgccggcagg atcgcggca
ggttctcga tftctgcgc gccgccagca ccgtcatcgc cagcgccttc ggctgctcca

ggactcctg ccggcccatg tcgaacagcg gccgcacctc cgggtgcttc ggaacaggt

tctttagaa cgtggctcgt atcgtcacgc cgtgtcctt cagcaccggc accgtgcct

tgatgatgtt gatcgtctgc tggccagca tatgattgca ctccaccgct gatgacatca

gtcgaatata gcacgatcaa cggcactgtt gcaaatagtc ggtggtgata aacttatcat

cccccttgc tgatggagct gcacatgaac ccattcaaag gccggcattt tcagcgtgac

atcattctgt gggccgtacg ctggtactgc aaatacggca tcagttaccg tgagctgcat

ttccgctgc ataaccctgc ttcggggta ttatagcga ttttccgta tatccatct

tttcgcacg atatacagga tttgccaaa gggttcgtgt agacttctt tgggtatcc

aacggcgtca gaagctt

Tabla 1. Comparación de los codones del gen de la hemoglobina de *Vitreoscilla* (*VHb*) y los codones preferidos del gen sintético de la hemoglobina para la expresión óptima en *S. spinosa*. Se prefieren los codones de *VHb* sintéticos en *S. spinosa* y se seleccionaron para sustituir los codones preferidos del gen de hemoglobina de *Vitreoscilla* usando el programa informático *GeneDesigner*.

Aminoácido	Leu	Arg	Thr	Ile	Lys	Ala	Val	Pro	His
Codón de <i>VHb</i> de <i>Vitreoscilla</i>	TTA	CGT	ACC	ATT	AAA	GCC	GTT	CCT	CAT
	TTG	CGC	ACT			GCT	GTA	CCA	
						GCA	GTC		
Codón sintético de <i>VHb</i>	CTG	CGG	ACG	ATC	AAG	GCG	GTG	CCG	CAC
Aminoácido	Asp	Gly	Ser	Asn	Cys	Phe	Gln	Glu	Tyr
Codón de <i>VHb</i> de <i>Vitreoscilla</i>	GAT	GGT	TCT	AAT	TGT	TTT	CAA	GAA	TAT
Codón sintético de <i>VHb</i>	GAC	GGC	TCG	AAC	TGC	TTC	CAG	GAG	TAC

5

Introducción del casete de expresión génica de la hemoglobina en el clon del cósmido 1E3

Se eligió la agrupación de genes de la *obscurina* como sitio para la integración del casete de expresión del gen de la hemoglobina debido a que la interrupción de la agrupación de genes de la *obscurina* no reduce la producción de espinosinas. El casete de expresión del gen de la hemoglobina fue clonado en el cósmido 1E3 (que contiene una agrupación parcial de genes de la *obscurina* con DNA genómico). Este cósmido se manipuló genéticamente para facilitar la integración estable del casete de expresión del gen de la hemoglobina dentro de la agrupación de genes de la *obscurina* del genoma de *S. spinosa*.

10

El casete de expresión del gen de la hemoglobina se clonó primeramente en el vector pIJ773 (John Innes Centre, Norwich, Reino Unido) por un método de clonación convencional. Este casete, además del gen de resistencia a apramicina y *oriT* flanqueado por secuencias de recombinación de FRT fue clonado luego en el cósmido 1E3 dirigido por PCR (Gust et al., 2002). Brevemente, el casete de expresión del gen de la hemoglobina se amplificó primeramente usando pDAB109000 como molde y los siguientes cebadores, aac3-VHb directo (SEQ ID NO: 2 5'-CGCTGAAAAGCTTCTGACGCCG-3') y aac3-VHb inverso (SEQ ID NO: 3 5'-CAACCCGTACCACGGCCTCT-3'). El fragmento de PCR amplificado se digirió con *HindIII/KpnI* para generar dos fragmentos; un fragmento *HindIII/KpnI* de 864 pb y un fragmento *HindIII/KpnI* de 83 pb. El fragmento *HindIII/KpnI* de 864 pb, que lleva el casete de expresión de gen de la hemoglobina incluyendo el terminador de la transcripción, se purificó en gel y se clonó en el vector pIJ773 que había sido digerido con *HindIII/KpnI*. Se seleccionaron un total de 38 transformantes cultivados en medios de Luria-Bertani (LB) que contenían 50 µg/mL de apramicina para su posterior análisis y confirmación. Después de crecimiento durante la noche en LB líquido que contenía apramicina a 50 µg/mL, se recogieron las células por centrifugación a 13000 rpm en una microcentrífuga durante 2 minutos.

15

20

25

Los sedimentos celulares que llevan el plásmido de control pDAB109000 produjeron un color rojizo que indicaba la presencia de la proteína hemoglobina. Sin embargo, sólo siete de los clones recombinantes produjeron sedimentos celulares rojizos visibles que iban desde el rojo claro a un color rojo más intenso. Se seleccionaron estos siete clones y se pasó a la siguiente etapa que incluía el aislamiento de DNA plasmídico y digestiones con enzimas de restricción. Dos de los siete clones, el clon 22 y el clon 30, produjeron los patrones de restricción deseados basados en el análisis de restricción usando las siguientes enzimas *NdeI*, *HindIII/KpnI* y *NdeI/KpnI*. Se secuenciaron la región del gen de la hemoglobina de ambos clones y el fragmento de PCR que llevaba el casete de expresión de gen de la hemoglobina. Se secuenciaron completamente la región del gen de la hemoglobina del clon 22 y el fragmento de PCR y los datos de las secuencias confirmaron que el gen sintético de la hemoglobina deseado se amplificó sin ningún error y fue clonado en pIJ773 (Figura 2). El plásmido resultante fue etiquetado como pDAB109001.

30

35

La integración del casete de expresión génica FRT :: *aac3(IV)* y el casete de expresión del gen sintético de la hemoglobina :: *oriT*:: FRT :: en el clon del cósmido 1E3 se llevó a cabo como se describe por un protocolo modificado dirigido por PCR (Gust et al., 2002). Se incorporaron al protocolo las siguientes modificaciones: i) el *E. coli* hospedante para introducir el clon del cósmido recombinante que llevaba el casete de expresión del gen de la hemoglobina era *E. coli* BW25141/pKD78 adquirido a *The Coli Genetic Stock Center* (CGSC) de la Universidad de Yale; ii) el plásmido de expresión de recombinasa rojo lambda pKD78, derivado de pKD46 (Datsenko and Wanner, 2000); y iii) se diseñaron y utilizaron los siguientes cebadores de PCR, tallo-bucle Strep interno ObsA (SEQ ID NO: 4 5'-

40

GAAGAAGGCGGGCGTCTCGAACTGGTTCGACCTCGGTGAGGAAGGTACCTCCGACCGCACGGC-3') y *ObsA* 5' FRT *aac3* (SEQ ID NO: 5 5'-GGCAATGCGCAGAGTTCGTAGTGCGGGAGCCATTTGATGTGTAGGCTGGAGCTGCTTC-3'). La amplificación resultante produjo un fragmento de 2216 pb que consistía en (FRT ::*aac3(IV)* :: *oriT*:: FRT :: *VHb*) dentro del clon del cósmido 1E3.

- 5 La integración dentro del clon del cósmido 1E3 se diseñó para que ocurriera en el extremo 5' de *ObsA*, lo que da como resultado la eliminación de 600 pb de la secuencia codificadora de *ObsA*. Esto asegura la interrupción completa de la primera enzima policétido-sintasa dentro de la agrupación de genes de la *obscurina*. Una vez integrado, el casete de expresión del gen de la hemoglobina fue orientado de tal manera que su secuencia codificadora estaba en la orientación opuesta a la secuencia codificadora de *ObsA*. Esta direccionalidad impide la lectura en la transcripción a partir del supuesto promotor de *ObsA* que potencialmente podría dar como resultado una transcripción no deseada. Los supuestos clones recombinantes que llevaban el fragmento (FRT :: *aac3(IV)*:: *oriT*:: FRT :: *VHb*) se aislaron y se confirmaron por análisis de restricción. La digestión por restricción usando *BamHI* indicó que los clones de cósmidos recombinantes que llevaban el casete del gen de la hemoglobina tenían diferentes patrones de restricción, en comparación con el clon del cósmido parental, 1E3. Por otra parte, la digestión por restricción usando *NdeI* y *ClaI*, destinada a regenerar la secuencia codificadora del gen sintético de hemoglobina, reveló que sólo los tres clones recombinantes producían un fragmento de 444 pb que llevaba la región codificadora completa del gen sintético de hemoglobina.

Conjugación del casete de expresión génica (FRT :: *aac3(IV)* :: *oriT* :: FRT :: *VHb*) en cepas de *S. spinosa*

- 20 Se completó la conjugación micelial entre la cepa donante, *Escherichia coli* S17, que contenía el casete de expresión génica FRT :: *aac3(IV)* y el casete de expresión del gen sintético de hemoglobina :: *oriT* :: FRT :: dentro de cósmido 1E3 y la cepa productora de espinosinas, DAS-2. La conjugación micelial entre la cepa donante y la cepa receptora de *S. spinosa* se llevó a cabo de acuerdo con el método descrito por Matsushima et al., (1994). Las colonias identificadas inicialmente como resistentes a apramicina y ácido nalidixico en los medios de conjugación se parchearon sobre en medios de agar-agar con infusión de cerebro-corazón (BHI) que contenían 50 µg/mL de apramicina y 25 µg/mL de ácido nalidixico para confirmar el fenotipo de resistencia a antibióticos de las colonias. Se supuso que las colonias que tenían el fenotipo de resistencia a antibióticos deseada contenían el casete FRT :: *aac3(IV)*:: *oriT* :: FRT :: *VHb* integrado dentro del *locus ObsA* de la agrupación de genes de la *obscurina*. Para confirmar la presencia del gen de resistencia a apramicina, *aac3(IV)*, se aisló el DNA genómico de los supuestos transconjugantes utilizando el kit genómico bacteriano *PURELUTE™* (Edge BioSystems, Gaithersburg, MD) de acuerdo con las instrucciones del fabricante. Este DNA genómico (gDNA) se utilizó como molde para la confirmación del tamaño por amplificación con PCR de un fragmento de DNA de *aac3(IV)* de 785 pb utilizando el sistema de PCR *FAILSAFE™* (Epicentre Biotechnologies, Madison, WI).

Confirmación de la presencia de un gen de hemoglobina en transconjugantes

- 35 Para confirmar que el gen sintético de hemoglobina estaba presente en los transconjugantes, se aisló el DNA genómico de un número seleccionado de transconjugantes (por el kit genómico bacteriano *PURELUTE™*) y se utilizó como molde para la amplificación por PCR del gen de la hemoglobina (por el sistema de PCR *FAILSAFE™*). Las cepas recombinantes produjeron un producto de PCR de 395 pb, correspondiente al tamaño de la región codificadora del gen de la hemoglobina destinada a la amplificación. Se purificaron los fragmentos de PCR y se completó la secuenciación del DNA de ambas cadenas para confirmar aún más la presencia del gen *VHb*. El alineamiento de la secuencia derivada de los fragmentos de PCR con la secuencia del gen sintético de hemoglobina sintetizado por el programa DNA 2.0 indicó que estas cepas recombinantes llevaban el gen de la hemoglobina con una secuencia idéntica a la secuencia de DNA diseñada.

Expresión del gen de hemoglobina en condiciones de fermentación

- 45 Se determinó en condiciones de fermentación la expresión del gen sintético de hemoglobina en la cepa recombinante que llevaba el casete de expresión del gen de la hemoglobina. Las cepas recombinantes que contenían el casete de expresión de *VHb* se cultivaron en condiciones de fermentación en matraces con agitación, y se aislaron las correspondientes muestras de RNA total por el kit bacteriano *RIBOPURE™* (Ambion, Austin, TX) y se prepararon para su uso como molde en el ensayo de RT-PCR. La fermentación del mutante de doble cruzamiento se realizó bajo las condiciones descritas por Bums et al., (WO 2003/070908). El análisis del caldo de fermentación para detectar la presencia de factores de espinosinas se puede llevar a cabo bajo las condiciones descritas por Baltz et al., (Patente de EE.UU. N° 6.143.526). Para confirmar la presencia de los factores de espinosinas en el líquido sobrenadante, extractos del caldo de fermentación se secaron en un aparato *SpeedVac* durante una noche seguido por reparto del residuo entre agua y éter. La capa etérea se secó por evaporación bajo corriente de N₂. La muestra se disolvió luego en acetona-d₆ y se transfirió a un tubo de RMN para la adquisición de la RMN protónica 1D. Los perfiles de RMN se compararon con los de los patrones de espinosinas.

La preparación de RNA total se trató con DNasa I (Ambion, Austin, TX) inmediatamente después de la purificación del RNA de acuerdo con las instrucciones del fabricante. La RT-PCR del gen sintético de hemoglobina utilizando el RNA total aislado se llevó a cabo usando el kit de RT-PCR *OneStep* (Qiagen, Valencia, CA) de acuerdo con las instrucciones del fabricante. Los cebadores para la amplificación de la región codificadora del gen sintético de he-

5 moglobina fueron SynHemo directo (SEQ ID NO: 6 5'-CAGACGATCAACATCATCAAGGCG-3') y SynHemo inverso (SEQ ID NO: 7 5'-ACCTGGATGAACACGTCGCC-3'). Las cepas recombinantes que contenían el casete de expresión de *VHb* produjeron un fragmento específico de 395 pb correspondiente a la región codificadora del gen de hemoglobina destinada a la amplificación. En el mismo ensayo, el RNA total aislado de cepas de control no recombinantes no produjo un fragmento de PCR correspondiente en tamaño al gen sintético de hemoglobina. Estos resultados confirmaron que un gen sintético de hemoglobina, activado por el promotor del gen de resistencia a apramicina, se transcribió y expresó en *S. spinosa* bajo condiciones de fermentación.

Efecto de la expresión del gen de la hemoglobina sobre la producción de espinosad

10 Se evaluó el impacto de la expresión del gen de hemoglobina sobre la producción de espinosad (A/D) cultivando las cepas recombinantes bajo condiciones de fermentación en matraces con agitación, tal como se ha descrito anteriormente. Los resultados de la fermentación se compararon con los niveles de la espinosina A/D producidos en la respectiva cepa parental, que no llevaba el gen de la hemoglobina. El número de cepas recombinantes seleccionadas para la evaluación en matraces con agitación se basó en el número de transconjugantes disponibles para el análisis y el deseo de obtener datos que determinarían si la expresión del gen sintético de hemoglobina produciría mayores niveles de espinosad A/D.

15 Se analizaron siete cepas recombinantes derivadas de DAS-2. Todas ellas superaron a la cepa parental DAS-2, en la producción de espinosad. La cepa parental acumuló espinosad dentro del intervalo esperado. El aumento estadísticamente significativo de la producción de espinosad A/D varió de 4,6% a 12,9% para las cepas que contenían el gen *VHb* (Tabla 2). La consistencia entre las cepas recombinantes en superar a la cepa de control en la producción de espinosad en condiciones de fermentación en matraces con agitación dio como resultado un aumento significativo porcentual de títulos de espinosad A/D. Este aumento en los títulos de espinosad indicó que las cepas recombinantes como un todo habían adquirido una mayor capacidad de producción de espinosad debido a la presencia en el genoma del gen sintético de hemoglobina. La expresión del gen sintético de hemoglobina en condiciones de fermentación en matraces con agitación indica que la expresión del gen de la hemoglobina aumenta la producción de espinosad.

Tabla 2: Comparación de las cepas DAS-2 y las cepas recombinantes DAS-2 VHb que contienen un gen de *VHb* integrado cromosómicamente en la producción de espinosad.

Identificación de la cepa	Factores principales de espinosinas	Títulos de espinosinas en relación con el control en porcentajes el día 10 de la fermentación
DAS-2	A/D	100,0
DAS-2 VHb1	A/D	107,4
DAS-2 VHb4	A/D	104,6
DAS-2 VHb5	A/D	108,0
DAS-2 VHb6	A/D	110,5
DAS-2 VHb7	A/D	112,9
DAS-2 VHb8	A/D	110,4
DAS-2 VHB9	A/D	105,6

30 Basándose en los resultados de fermentación descritos anteriormente, las cepas recombinantes siguientes, DAS-2 VHb7 y DAS-2 VHb8, se conservaron criogénicamente en glicerol al 20% usando cultivos vegetativos que crecen activamente. Se llevó a cabo un estudio de la fermentación en matraces con agitación dedicado a: i) evaluar las cepas criopreservadas; ii) asegurar la reproducibilidad de los resultados de fermentación; y iii) proporcionar pruebas del avance de las cepas seleccionadas para un trabajo futuro. En este estudio las cepas de control produjeron espinosad en los niveles esperados, y todo el estudio se llevó a cabo sin ningún tipo de desviaciones inesperadas del protocolo de operación. Cada cepa recombinante superó a las cepas de control tanto durante como al final del ciclo de fermentación lo que confirmó las observaciones iniciales de los estudios de fermentación en matraces con agitación. El aumento en los niveles de espinosad en las cepas recombinantes en relación con las cepas de control fue estadísticamente significativo.

Se utilizó el grupo de genes de policétido-sintasa de obscurina para la integración y expresión de un gen de hemoglobina con codones optimizados en la cepa de *S. spinosa*, DAS-2. La integración y expresión del gen de hemoglobina dieron como resultado una mejora de la producción de A/D en la fermentación en matraces con agitación de *S. spinosa*, DAS-2.

5 Identificación de clones de cósmidos que llevan el gen *ObsA* con policétido-sintasa (PKS) de obscurina

Se completó el cribado de una genoteca de cósmidos usando una sonda de 806 pb procedente del *locus* genómico *ObsA* de *S. spinosa* (SEQ ID NO: 8). Se construyó una genoteca de cósmidos a partir de una cepa de *S. spinosa* por BIOS&T (Montreal, Canadá) usando el vector cósmido *Supercos I* (Stratagene, La Jolla, CA) y el DNA genómico de la cepa de *S. spinosa* que se preparó de acuerdo con el protocolo de aislamiento de DNA genómico descrito por Kieser et al., (2000).
SEQ ID NO 8.

```

caagatcggt gggacctggc cgcacgtttc gacgaggagc ttccggggcg ggggagcctg
gggtcgggcc gtggtggttt cctcaccgag gtcgaccagt tcgacgccgc cttcttcggg
atctcccctc gtgaagccgc cgaactcgac ccccggcagc ggctcacgct ggagctgggc
tggaagcgt tggaagacgc cgggatcgtg ccgggtgctg tgcgcggtga acgggtcgcc
gtgttcgtgg gcgcgatggg ggacgactac gccacgctga gcttcgccga cggcggctcc
cagatcggtc actacacggc gaccggtgtg cagcggagca tgatcgcaa ccggttctg
tacgtgctcg gttgacacgg cccgagcttc gtcgtcgact ccggtcagtc atctccctg
gttgcgggtc acctgccctg ggagagcctg cggcggggcg agtcttcggc cgccttggtc
ggcgggggtca cctgaacct ggctgcggag agcatggcgg ccacggccag gttcggcgcg
cttcgcccg acgggctctg cttcactttc gatgcacgcg cgaacggcta cgcacgcggt
gaaggcggcg gtttcgccgt cctcaagccg ctgcacctg cgctcggcga tggcgacgac
atctactgcg tgatccgcgg gaccgcatg aacaacgacg gtggcggcga tgggctcacc
gtcccgaacg agcgcgccca gcaggcgggtg ctgcgggagg cgtaccggcg agcgggagtg
gaccggccg aggtccagta cgtcga
    
```

15 El cribado de la genoteca con la sonda *ObsA* permitió la identificación de clones de cósmidos que llevaban el extremo 5' de *ObsA* en condiciones estrictas. El marcaje de la sonda, la hibridación y la detección se llevaron a cabo utilizando el kit DIG de marcaje de DNA y el kit DIG de detección de ácidos nucleicos (Roche, Basilea, Suiza) de acuerdo las instrucciones del fabricante. Específicamente, se marcó el fragmento de PCR de 806 pb derivado del extremo 5' de *ObsA* usando el kit de marcaje con cebador aleatorio (Roche). La hibridación se realizó durante una noche a 58°C en el horno de hibridación seguida por lavado dos veces de ambos filtros a 65°C durante un total de 30 minutos utilizando el tampón de lavado que contenía 0,1xSSC y 0,1% de SDS. La sonda hibridada se detectó
20 usando el kit DIG de detección de ácidos nucleicos de Roche basado en el inmunoensayo ligado a enzimas (ELISA) con un conjugado de anticuerpos anti-DIG-AP altamente específico y los sustratos de color NBT (azul de nitrotetrazolio) y BCIP (fosfato de 5-bromo-4-cloro-3-indolilo). Se identificaron varios conjuntos de cósmidos con señales de hibridación fuertes. Se seleccionaron cuatro de los clones, 1E3, 2N14, 3M23 y 4E16, para el aislamiento de DNA de cósmidos y confirmación por secuenciación.

25 La amplificación por PCR de los clones de cósmidos utilizando el cebador directo de *ObsA* (SEQ ID NO: 9 5'-CAAGATCGTTGGGACCTGGCC-3') y el cebador inverso de *ObsA* (SEQ ID NO: 10 5'-TCGACGTACTGG ACCTCGGC-3') dio como resultado un solo fragmento de PCR equivalente a la sonda de tamaño 806 pb. Las reacciones de PCR se realizaron de acuerdo con las instrucciones del fabricante utilizando el sistema de PCR *FAILSAFE™* (Epicentre Biotechnologies, Madison, WI).

30 Los insertos de DNA llevados por los cuatro clones de cósmidos se secuenciaron también en ambos extremos con el fin de cartografiar los clones de cósmidos en la agrupación de genes de obscurina y estimar el tamaño de los inser-

tos de los clones de cósmidos (Figura 3). Se determinó que el clon del cósmido 1E3 tenía un tamaño de 42900 pb y contenía el extremo 5' de *ObsA* en el centro del inserto de DNA. Se determinó que el clon del cósmido 2N14 tenía un tamaño de 40.153 pb y no llevaba la región codificadora completa de *ObsA*. Los insertos en los clones de cósmidos 3M23 y 4E16 contenían el extremo 5' de *ObsA* en un extremo y en el otro extremo las secuencias de DNA más allá de la agrupación de genes de obscurina. No se estimaron los tamaños reales de los clones de cósmidos 3M23 y 4E16. Se eligió el clon del cósmido 1E3 para completar la interrupción de *ObsA* para determinar el impacto de la abolición de la función de los genes de policétido-sintasa sobre las características de *S. spinosa* y la producción de espinosinas.

Modificación por ingeniería genética del clon de cósmidos para la interrupción de *ObsA* dirigida por PCR

- 10 Para interrumpir *ObsA* dentro de la agrupación de genes de obscurina, el casete de interrupción de resistencia a apramicina (FRT-*aac3(IV)-oriT-FRT*) del plásmido pIJ773 (proporcionado por The John Innes Center Plant Biosciences Limited, Norwich, Inglaterra; Figura 4) se integró en el clon del cósmido 1E3 dirigido por PCR.

- 15 La integración del casete de interrupción de *ObsA* (FRT-*aac3(IV)-oriT-FRT*) en el clon del cósmido 1E3 se llevó a cabo de acuerdo con Gust et al., (2002) con las siguientes modificaciones. Se usó *E. coli* BW25141/pKD78 adquirida a The Coli Genetic Stock Center (CGSC) de la Universidad de Yale, que contenía el plásmido de expresión de recombinasa rojo lambda, pKD78, derivado de pKD46 (Datsenko and Wanner, 2000). Se utilizaron los siguientes cebadores largos de PCR, *ObsA* 5' FRT *aac3* (SEQ ID NO: 11 5'-GGCAATGCGCAGAGTTCGTAGTGCGG GAGC-CATTTGATGTGTAGGCTGGAGCTGCTTC-3') y *ObsA* interno *oriT* FRT (SEQ ID NO: 12 5'-GAAGAA GGCGGCGTCAACTGGTTCGACCTCGGTGAGGAAATCCGGGGATCCGTCGACC-3'), para amplificación del fragmento de 1322 pb que llevaba FRT-*aac3(IV)-oriT-FRT* usando pIJ773 como molde (Figura 4). El fragmento amplificado de 1322 pb se purificó y se integró en el cósmido 1E3 de acuerdo con Guts et al., (2002).

- 25 Diez de los clones de *E. coli* recombinantes que se obtuvieron dirigidos por PCR eran resistentes a apramicina y tenían el mismo patrón de digestión con la enzima de restricción *BamHI* (este patrón de digestión es diferente del que tiene el clon del cósmido 1E3 que no contiene el casete de interrupción). La confirmación adicional de los clones recombinantes se llevó a cabo por amplificación usando un par de cebadores que se asocian a las regiones de DNA de 95 pb aguas arriba del codón de iniciación de *ObsA* y 334 pb aguas abajo de la codón de inicio de *ObsA*, respectivamente. Los cebadores son *ObsA* 5' aguas arriba directo (SEQ ID NO: 13 5'-CGACCGGTGTGTCGA TGT-TAGGGT-3') y *ObsA* interno inverso (SEQ ID NO: 14 5'-CTTCCAACGCTTCCCAGCCC-3'). Las reacciones de PCR se completaron de acuerdo con las instrucciones del fabricante utilizando el sistema de PCR FAILSAFE™ (Epicentre Biotechnologies, Madison, WI).

- 35 La amplificación utilizando el DNA genómico de la cepa de espinosad utilizada para crear los cósmidos o el DNA del clon del cósmido 1E3 produjo el fragmento esperado de 429 pb. La amplificación utilizando el DNA de cósmido aislado de nueve (no se incluyó el clon 10 debido a redundancia) de los diez clones de cósmidos recombinantes produjo un único fragmento correspondiente al tamaño esperado de 1538 pb. El fragmento de PCR esperado de 1538 pb se debe a la inserción de 1322 pb del casete de interrupción (FRT-*aac3(IV)-oriT-FRT*) en el extremo 5' de *ObsA* con la delección simultánea de 213 pb del gen *ObsA*. Todos los clones de cósmidos recombinantes produjeron un único fragmento de PCR de 1538 pb y no produjeron el fragmento de 429 pb que observado en el clon del cósmido de control 1E3 lo que indica que cada uno de los clones de cósmidos recombinantes contenía el casete de interrupción (FRT-*aac3(IV)-oriT-FRT*) en el extremo 5' de *ObsA*.

- 40 Interrupción de *ObsA* dentro de la agrupación de genes de la obscurina por integración del casete de interrupción (FRT-*aac3(IV)-oriT-FRT*) en *S. spinosa*

- 45 Se usaron métodos de conjugación para introducir el clon del cósmido recombinante que llevaba el casete de interrupción de *ObsA* (FRT-*aac3(IV)-oriT-FRT*) en cepas de *S. spinosa* con el fin de lograr transconjugantes de todas las cepas dianas para el análisis concluyente del impacto de la interrupción de *ObsA* sobre el crecimiento y la producción de espinosinas.

La conjugación micelial entre la cepa donante que lleva el cósmido 1E3 recombinante y una cepa de *S. spinosa* receptora, NRRL 18538, se llevó a cabo de acuerdo con el método descrito por Matsushima et al., (1994).

- 50 Se produjeron transconjugantes. Casi todos los transconjugantes primarios confirmaron el fenotipo resistente a apramicina deseado lo que indica que los transconjugantes llevaban el gen de resistencia a apramicina, *aac3(IV)*, integrado en la agrupación de genes de policétido-sintasa de obscurina por recombinación homóloga. Se analizaron varios transconjugantes seleccionados para determinar la presencia del fragmento de DNA correspondiente al tamaño de un fragmento de *aac3(IV)* de 785 pb por amplificación por PCR. Los controles negativos que no contenían el casete de interrupción (FRT-*aac3(IV)-oriT-FRT*) no produjeron un amplión por PCR.

El impacto de la interrupción de *ObsA* sobre el crecimiento y producción de espinosinas en *S. spinosa*

- 55 El impacto de la interrupción de *ObsA* sobre la producción de espinosina se evaluó usando un protocolo de matraces con agitación. La fermentación de los transconjugantes se realizó bajo las condiciones descritas por Bums et al., (WO 2003/070908). El análisis del caldo de fermentación para detectar la presencia de factores de espinosinas se

llevó a cabo bajo las condiciones descritas por Baltz et al., (Patente de EE.UU. N° 6.143.526).

5 Para los aislados de *S. spinosa* se realizó una comparación directa del rendimiento de los transconjugantes en relación con sus respectivas cepas parentales. Se evaluaron en matraces con agitación cuatro de los mutantes desactivados de *ObsA* derivados de la cepa NRRL 18538. El título medio de cada mutante desactivado fue mayor que el de la cepa parental. Sin embargo, basándose en el análisis estadístico no parecía que las diferencias fueran significativas (Tabla 2).

Tabla 2. Producción de espinosinas el día 10 por los mutantes desactivados de *ObsA* derivados de NRRL 18538.

Cepa	Factores de espinosinas principales	Títulos de espinosinas en relación con el control en porcentajes el día 10 de la fermentación
NRRL 18538 parental	A/D	100
NRRL 18538 $\Delta ObsA-1$	A/D	108
NRRL 18538 $\Delta ObsA-2$	A/D	114
NRRL 18538 $\Delta ObsA-3$	A/D	108
NRRL 18538 $\Delta ObsA-6$	A/D	107

10 La interrupción de *Osba* en el gen PKS de la obscurina en la cepa A/D, NRRL 18538, no tuvo impacto negativo sobre la producción de espinosinas en comparación con las cepas de control parentales respectivas en las condiciones usuales de fermentación en matraces con agitación. La falta de impacto negativo sobre la producción de espinosinas por la interrupción de *ObsA* califica a la agrupación de genes de policétido-sintasa de obscurina como un sitio neutro para la integración y expresión de genes diana de interés para la mejora de los procesos de producción y fermentación de espinosinas. Este locus genómico sirve como ejemplo de un sitio de integración para la integración de genes dentro del genoma de *S. spinosa*.

15 Lo anterior es ilustrativo de la presente invención, y no debe interpretarse como limitativo de la misma. La invención se define por las siguientes reivindicaciones.

Lista de secuencias

- <110> Dow AgroSciences LLC
- <120> Mejora de la producción de espinosinas con proteínas que se unen al oxígeno
- <130> P202858EP
- 5 <150> US 13/100,220
<151> 2011-05-03

<150> US 13/100,202
<151> 2011-05-03

<160> 14
- 10 <170> PatentIn version 3.5

<210> 1
<211> 977
<212> DNA
<213> Secuencia Artificial
- 15 <220>
<223> Casete de expresión génica de hemoglobina

<400> 1

aagcttgaat tcgatatcag cccggtacct cgccgccaac ccgtaccacg gcctctgacg	60
gctccggcca agcccggaag agcgggggcc gggcgccggg ctccaggcgg ccgggtacct	120
ccgaccgcac ggcgcgttcc ggagtgtcc ggaagcggcc tgcggagcgc cctgcggggc	180
ggctatcgat tcaactccacc gcctgcgcgt acaggtccgc ctccacctgg atgaacacgt	240
ccgcgatcac gccgtacgcc ttgccccacg cgtccaggat gtctgtccgtc gccgcgtcgc	300
ccagcacctc cttgatcgcg ccagcagcct cctggcccac gatcgggtag tgcgccgccg	360
ccacgcccgc ctggcagtgc ttcaccgcga tcttcttcac cgccggcagc atcgccggca	420
ggttctcgat gttctgcgcc gccgccagca ccgtcatcgc cagcgccttc ggctgctcca	480
gcgactcctg ccggcccattg tgaacagcgc gccgcacctc cgggtgcttc gcgaacaggt	540
tctttagaaa cgtggtcgtg atcgtcacgc cgtgctcctt cagcaccggc accgtcgcct	600
tgatgatggt gatcgtctgc tggccagca tatgattgca ctccaccgct gatgacatca	660
gtcgtacata gcacgatcaa cggcactggt gcaaatagtc ggtggtgata aacttatcat	720
ccccttttgc tgatggagct gcacatgaac ccattcaaag gccggcattt tcagcgtgac	780
atcattctgt gggccgtacg ctggtactgc aaatacggca tcagttaccg tgagctgcat	840
tttccgctgc ataaccctgc ttgggggtca ttatagogat tttttcggta tatccatcct	900
ttttcgcacg atatacagga ttttgccaaa gggttcgtgt agactttcct tgggtgatcc	960
aacggcgtca gaagctt	977
- 20 <210> 2
<211> 22
<212> DNA
<213> Secuencia Artificial

<220>
<223> Cebador directo aac3-vhb
- 25 <400> 2
cgctgaaaag cttctgacgc cg 22

<210> 3
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

5 <220>
 <223> Cebador inverso aac3-vhb

<400> 3
 caaccggtac cacggcctct 20

10 <210> 4
 <211> 59
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador tallo-bucle Strep interno obsA

15 <400> 4
 gaagaaggcg gcgtcgaact ggtcgacctc ggtgaggaag gtacctccga cgcacggc 59

<210> 5
 <211> 58
 <212> DNA
 <213> Secuencia Artificial

20 <220>
 <223> Cebador aac3 5' FRT obsA

<400> 5
 ggcaatgcg agagttcgta gtgcgggagc cattgatgt gtaggctgga gctgcttc 58

25 <210> 6
 <211> 24
 <212> DNA
 <213> Secuencia Artificial

<220>
 <223> Cebador directo SynHemo

30 <400> 6
 cagacgatca acatcatcaa ggcg 24

<210> 7
 <211> 20
 <212> DNA
 <213> Secuencia Artificial

35 <220>
 <223> Cebador inverso SynHemo

<400> 7
 acctggatga acacgtccgc 20

<210> 8
 <211> 806
 <212> DNA
 <213> Saccharopolyspora spinosa

45 <400> 8

ES 2 553 079 T3

	caagatcgtt gggacctggc cgcacgtttc gacgaggagc tttccgggcg ggggagcctg	60
	gggtcgggcc gtggtggttt cctcaccgag gtcgaccagt tcgacgccgc cttcttcggg	120
	atctcccctc gtgaagccgc cgaactcgac ccccggcagc ggctcacgct ggagctgggc	180
	tgggaagcgt tggaaagacgc cgggatcgtg ccgggtgcbc tgcgcggtga acgggtcgcc	240
	gtgttcgtgg gcgcgatggg ggacgactac gccacgctga gcttcgccga cggcggctcc	300
	cagatcggtc actacacggc gaccggtgtg cagcggagca tgatcgccaa ccggttgtcg	360
	tacgtgctcg gtgtgcacgg cccgagcttc gtcgtcgact ccggtcagtc atcctccctg	420
	gttgcggttc acctcgccgt ggagagcctg cgcccgggcg agtcttcggc cggcctggtc	480
	ggcgggggtca ccctgaacct ggctgcggag agcatggcgg ccacggccag gttcggcgcg	540
	ctttcgcccg acgggctctg cttcactttc gatgcacgcg cgaacggcta cgcacgcggt	600
	gaaggcggcg gtttcgccgt cctcaagccg ctgcacctcg cgctcggcga tggcgacgac	660
	atctactgcg tgatccgcbg gaccgccatg aacaacgacg gtggcggcga tgggctcacc	720
	gtcccgaacg agcgcgccca gcaggcggtg ctccgaggag cgtaccggcg agcgggagtg	780
	gacccggccg aggtccagta cgtcga	806
	<210> 9	
	<211> 21	
	<212> DNA	
5	<213> Secuencia Artificial	
	<220>	
	<223> Cebador directo obsA	
	<400> 9	
	caagatcgtt gggacctggc c	21
10	<210> 10	
	<211> 20	
	<212> DNA	
	<213> Secuencia Artificial	
	<220>	
15	<223> Cebador inverso obsA	
	<400> 10	
	tcgacgtact ggacctcggc	20
	<210> 11	
	<211> 58	
20	<212> DNA	
	<213> Secuencia Artificial	
	<220>	
	<223> Cebador aac3 5' FRT ObsA	
	<400> 11	
25	ggcaatgcbg agagttcgta gtgcgggagc catttgatgt gtaggctgga gctgctc	58
	<210> 12	
	<211> 59	
	<212> DNA	
	<213> Secuencia Artificial	
	<220>	
30	<223> Cebador FRT oriT interno ObsA	
	<400> 12	

gaagaaggcg gcgtcgaact ggtcgacctc ggtgaggaaa ttccggggat ccgtcgacc 59

<210> 13
<211> 24
<212> DNA
5 <213> Secuencia Artificial

<220>
<223> Cebador directo 5' curso arriba obsA

<400> 13
cgaccggtgt gtcgatgta ggg 24

10 <210> 14
<211> 20
<212> DNA
<213> Secuencia Artificial

<220>
15 <223> Cebador inverso interno obsA

<400> 14
cttccaacgc ttcccagccc 20

REIVINDICACIONES

1. Un método para aumentar los títulos de espinosinas de una célula hospedante de *S. spinosa*, comprendiendo dicho método transformar la célula hospedante de *S. spinosa* con un gen heterólogo que codifica una proteína que se une al oxígeno.
- 5 2. El método de acuerdo con la reivindicación 1, en el que la proteína que se une al oxígeno es hemoglobina de *Vitreoscilla*.
3. El método de acuerdo con la reivindicación 1, en el que la proteína que se une al oxígeno es una proteína globina, preferiblemente seleccionada del grupo que consiste en hemoglobina de *Vitreoscilla*, flavohemoproteína de *Alcaligenes eutrophus*, mioglobina de corazón de caballo, hemoproteína de *E. coli*, hemoproteína de *B. subtilis*, flavohemoglobina de levadura, leghemoglobina de soja, leghemoglobina de altramuz y mioglobina de esperma de ballena.
- 10 4. Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en el que el título de espinosinas es aumentado en al menos 4,6%.
5. Un método de aumentar la producción de espinosinas de acuerdo con una cualquiera de las reivindicaciones 1 a 4, comprendiendo además el método:
- 15 a) cultivar la célula de *S. spinosa* que expresa un gen heterólogo que codifica una proteína que se une al oxígeno, en condiciones apropiadas para la producción de espinosinas; y
- b) recoger espinosinas del cultivo.
6. Un método de acuerdo con una cualquiera de las reivindicaciones precedentes, en donde el gen heterólogo que codifica una proteína que se une al oxígeno está integrado en un cromosoma de la célula de *S. spinosa*.
- 20 7. Un método para mejorar la producción de espinosinas en una célula hospedante de *S. spinosa*, comprendiendo dicho método integrar un gen que codifica una proteína que se une al oxígeno en el DNA cromosómico de la cepa de la especie *S. spinosa*, en donde la integración del gen que codifica una proteína que se une al oxígeno en el DNA cromosómico comprende:
- a) crear una fagoteca para identificar un *locus* genómico;
- 25 b) clonar dos fragmentos de DNA genómico del *locus* genómico, en donde un fragmento genómico se clona aguas arriba de un polinucleótido que comprende un gen que codifica una proteína que se une al oxígeno, y el segundo fragmento genómico se clona aguas abajo de dicho polinucleótido;
- c) introducir los fragmentos genómicos en una cepa de *S. spinosa* por transformación;
- d) obtener transconjugantes; y,
- 30 e) cribar dichos transconjugantes para detectar la presencia de dicho polinucleótido.
8. El método de acuerdo con la reivindicación 7, en el que el polinucleótido contiene un casete de expresión génica que comprende la proteína que se une al oxígeno.
9. El método de acuerdo con la reivindicación 8, en el que el casete de expresión génica contiene un marcador seleccionable de antibióticos.
- 35 10. Un método de acuerdo con una cualquiera de las reivindicaciones 7 a 9, en el que el gen que codifica la proteína que se une al oxígeno está integrado en el DNA cromosómico en el *locus* de la policétido-sintasa de obscurina.
11. Una célula de *S. spinosa* modificada genéticamente, que comprende un gen heterólogo que codifica una proteína que se une al oxígeno, en la que cuando se expresa el gen heterólogo que codifica una proteína que se une al oxígeno, la célula de *S. spinosa* modificada genéticamente es capaz de aumentar la producción de espinosinas.
- 40 12. La célula de *S. spinosa* modificada genéticamente de acuerdo con la reivindicación 11, en la que la célula de *S. spinosa* modificada genéticamente es capaz de producir un título de espinosinas que está aumentado en al menos 4,6% en comparación con el título de espinosinas de una célula de *S. spinosa* no modificada.
13. Una célula de *S. spinosa* modificada genéticamente de acuerdo con la reivindicación 11 o la reivindicación 12, en la que la proteína que se une al oxígeno es una proteína globina, preferiblemente seleccionada del grupo que consiste en hemoglobina de *Vitreoscilla*, flavohemoproteína de *Alcaligenes eutrophus*, mioglobina de corazón de caballo, hemoproteína de *E. coli*, hemoproteína de *B. subtilis*, flavohemoglobina de levadura, leghemoglobina de soja, leghemoglobina de altramuz y mioglobina de esperma de ballena.
- 45 14. Una célula de *S. spinosa* modificada genéticamente de acuerdo con una cualquiera de las reivindicaciones 11 a 13, en la que el gen heterólogo que codifica una proteína que se une al oxígeno está integrado en el genoma de la

célula de *S. spinosa* modificada genéticamente, preferentemente en el DNA cromosómico en el *locus* de la policétido-sintasa de obscurina.

15. Una cepa que comprende una población de células de *S. spinosa* modificadas genéticamente como se define en una cualquiera de las reivindicaciones 11 a 14.

- 5 16. El método o la célula de *S. spinosa* de una cualquiera de las reivindicaciones precedentes, en donde las espinosinas se seleccionan del grupo que consiste en espinosinas A + D, espinosinas J + L, espinosina A y espinosina J.

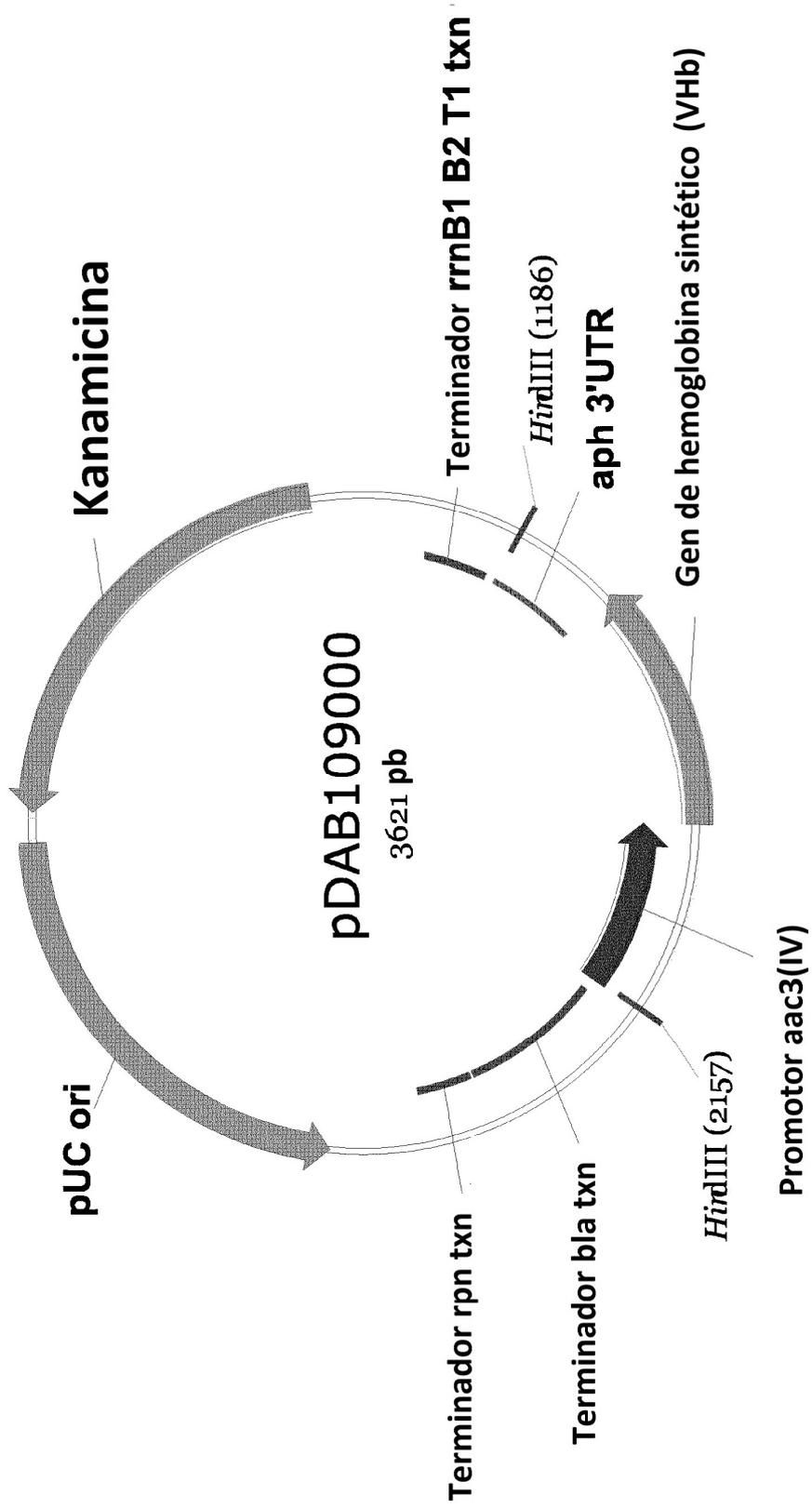


FIGURA 1

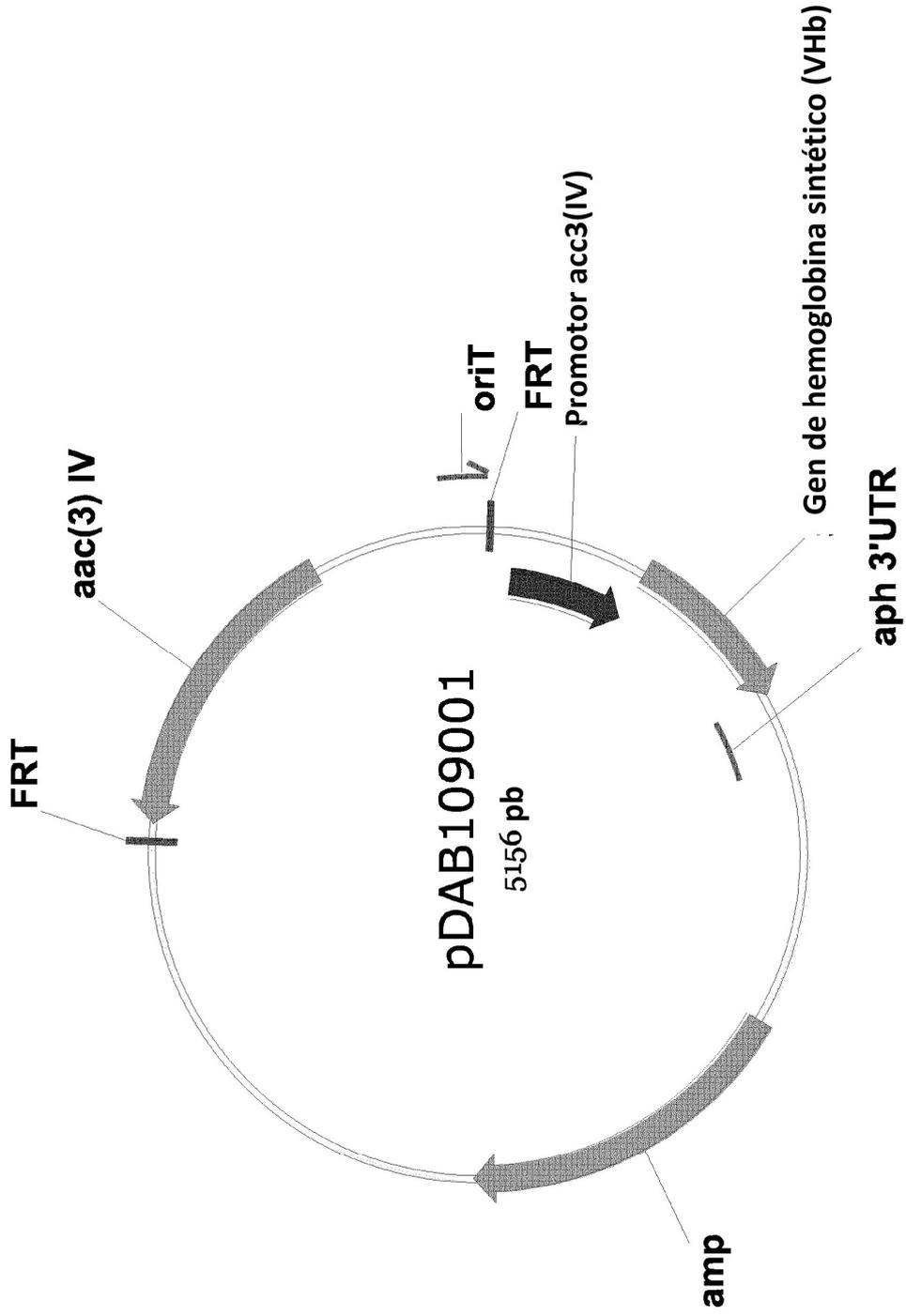


FIGURA 2

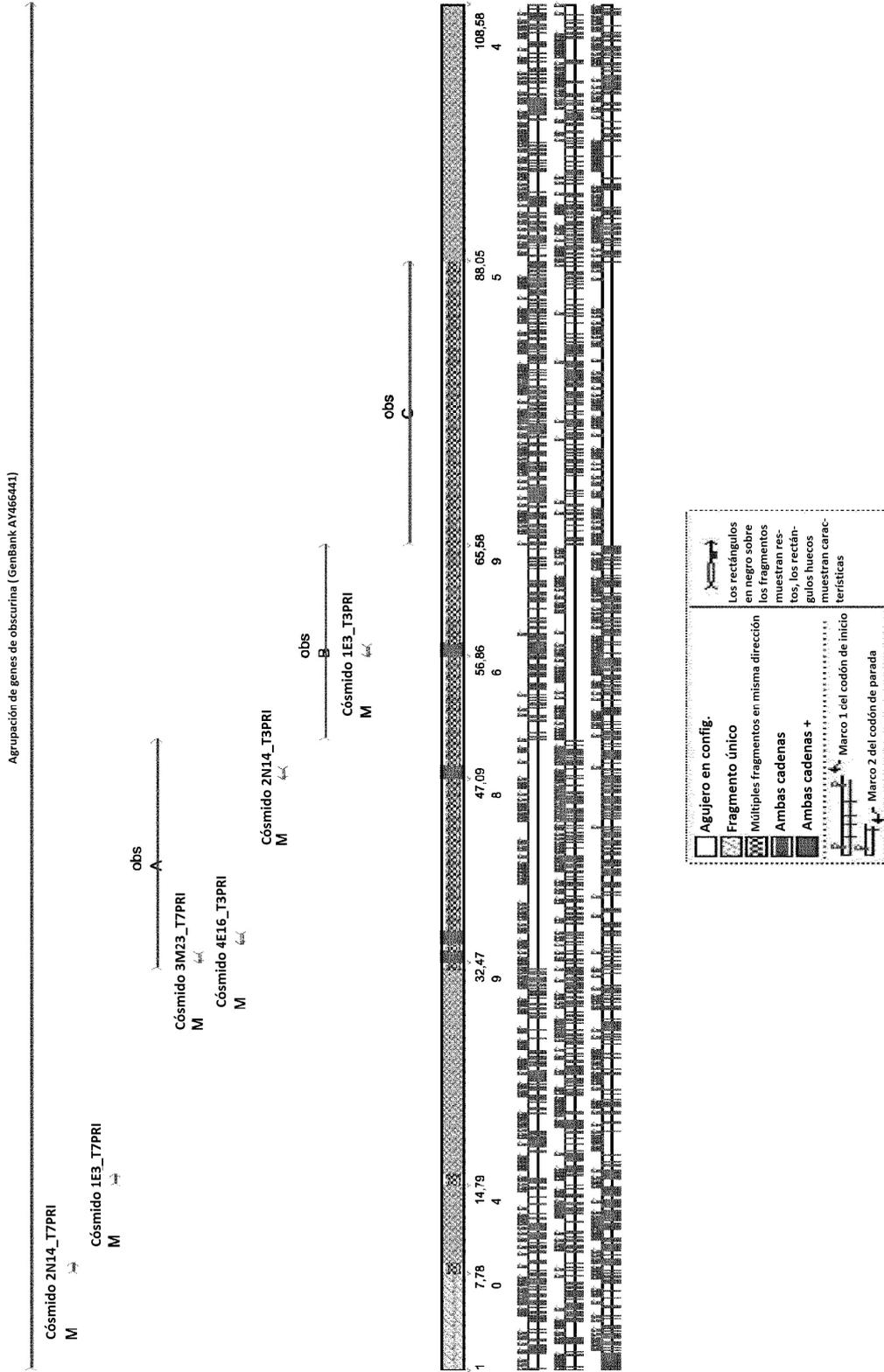


FIGURA 3

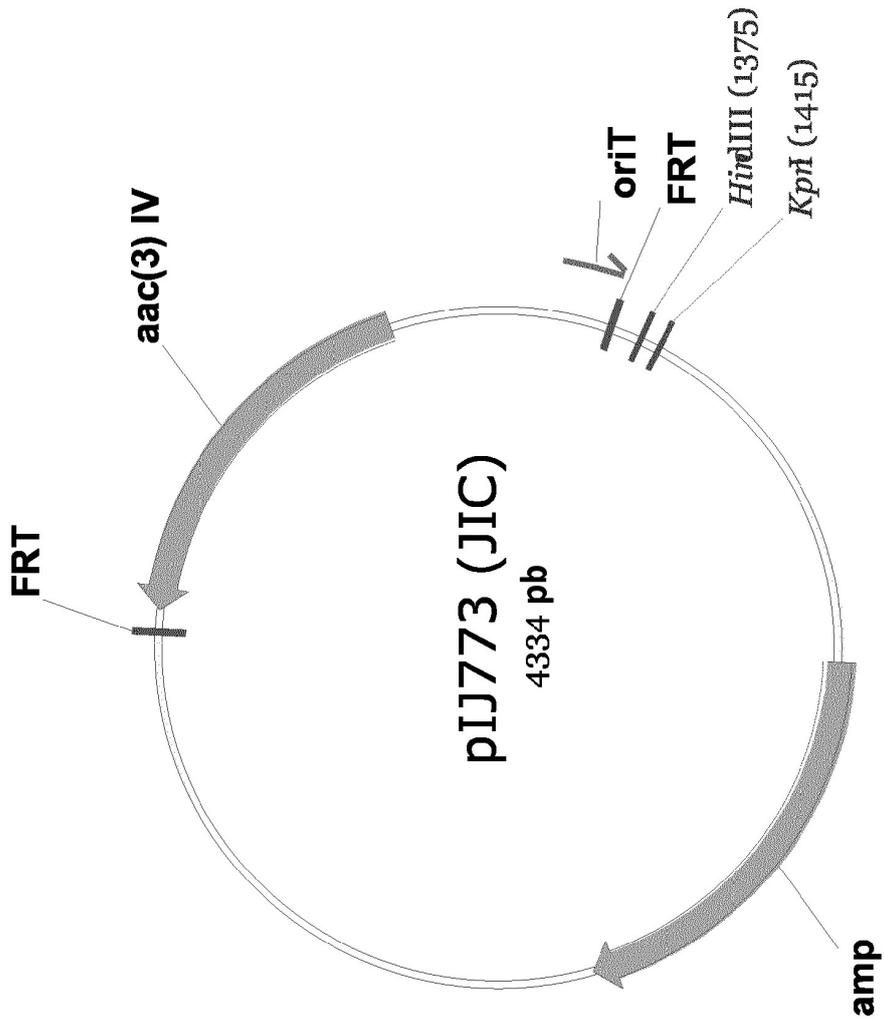


FIGURA 4