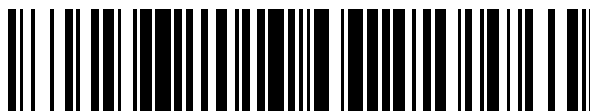


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 590 925**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **18.05.2007** **E 11161778 (3)**

97 Fecha y número de publicación de la concesión europea: **22.06.2016** **EP 2426220**

54 Título: **Microorganismos marcados y métodos de marcado**

30 Prioridad:

19.05.2006 US 747682 P
02.03.2007 US 904721 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
24.11.2016

73 Titular/es:

DUPONT NUTRITION BIOSCIENCES APS
(100.0%)
Langebrogade 1
1411 Copenhagen K , DK

72 Inventor/es:

BARRANGOU, RODOLPHE;
FREMAUX, CHRISTOPHE;
HORVATH, PHILIPPE y
ROMERO, DENNIS

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 590 925 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

Microorganismos marcados y métodos de marcado

5 Campo de la invención

La presente invención proporciona métodos para preparar un ingrediente alimenticio utilizando bacterias marcadas.

10 En algunas realizaciones especialmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. Se exponen en esta memoria los microorganismos marcados usando los métodos que se enseñan en el presente documento.

15 En algunas realizaciones preferidas, los microorganismos marcados son bacterias. En algunas realizaciones especialmente preferidas, las bacterias marcadas son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias marcadas son miembros de otros géneros.

Antecedentes de la invención

20 Las cepas microbianas, especialmente las utilizadas como cultivos iniciadores en la industria de alimentos/bebidas fermentadas están muy seleccionadas y se caracterizan por funcionalidades específicas. Normalmente, estos cultivos iniciadores se adquieren comercialmente como organismos vivos y permanecen a menudo viables en el producto final. De esta manera, es posible aislar, identificar y caracterizar las cepas del cultivo iniciador cultivando los microorganismos presentes en los productos finales. Además, es posible a continuación utilizar las cepas de cultivos iniciadores en otros productos, incluyendo productos de los competidores. Es difícil vigilar el uso de dichas cepas por terceros, incluyendo competidores. A su vez, existe una necesidad en la técnica de métodos para etiquetar fácilmente cultivos para identificar su(s) fuente(s) y vigilar su uso en diversos productos. Sumario de la invención

30 La presente invención proporciona métodos que utilizan bacterias marcadas. En algunas realizaciones especialmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. Se enseñan también en la presente memoria los microorganismos marcados utilizando los métodos que se muestran en el presente documento. Los microorganismos marcados pueden ser bacterias. Las bacterias marcadas pueden ser miembros de los géneros *Streptococcus*, o miembros de otros géneros.

35 La presente invención proporciona un proceso para preparar un ingrediente alimenticio que comprende obtener una bacteria marcada mediante un método que comprende las etapas de:

- 40 (a) exponer una bacteria progenitora que comprende un locus CRISPR a un bacteriófago para introducir un espaciador de repeticiones adicionales en el locus CRISPR para producir una bacteria marcada, en el que la unidad espaciadora de repetición adicional proporciona la marca,
- (b) seleccionar un mutante insensible a bacteriófago;
- (c) comparar un locus CRISPR o una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago;
- 45 (d) seleccionar una bacteria marcada que comprende dicha unidad espaciadora de repetición adicional en el locus CRISPR que no está presente en la bacteria progenitora; y que comprende además la etapa de añadir la bacteria marcada, o un cultivo celular que comprende dicha bacteria marcada, a dicho ingrediente alimenticio.

50 Métodos para marcar o etiquetar una bacterias que comprende las etapas de: (a) exponer una bacteria progenitora a un bacteriófago; (b) seleccionar un mutante insensible a bacteriófago; (c) comparar un locus CRISPR o una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago; y (d) seleccionar una bacteria marcada que comprenda una unidad espaciadora de repetición adicional en el locus CRISPR que no está presente en la bacteria progenitora que se expone en el presente documento. La bacteria marcada puede obtenerse o es obtenible mediante el método de acuerdo con la presente invención. Se enseñan en el presente documento cultivos celulares que comprenden la bacteria marcada.

60 Se enseñan en el presente documento el producto alimenticio y/o los productos alimenticios que comprenden la bacteria marcada y/o los cultivos celulares que comprenden al menos una especie bacteriana. En realizaciones más adicionales, la presente invención proporciona procesos para preparar alimentos y/o productos alimenticios que comprenden la bacteria marcada y/o los cultivos celulares que comprenden al menos una especie bacteriana etiquetada. En algunas realizaciones, la presente invención proporciona métodos para preparar alimentos y/o comida que comprende la etapa de añadir la bacteria marcada o el cultivo de células a dicho producto alimenticio o alimento. Los inventores describen el alimento y/o la comida u obtenible utilizando los métodos de la presente invención.

65

Los métodos para generar variantes CRISPR, comprenden las etapas de: (a) exponer una bacteria progenitora a un bacteriófago; (b) seleccionar una bacteria resistente a bacteriófago; (c) comparar el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago; (d) seleccionar una bacteria marcada que comprende una unidad espaciadora de repetición adicional en el locus CRISPR que no está presente en la bacteria progenitora; y (e) aislar y/o clonar y/o secuenciar la unidad espaciadora de repetición adicional que se expone en el presente documento. Se describen las variantes CRISPR obtenidas u obtenibles utilizando los métodos que se exponen en el presente documento.

Los métodos para identificar una bacteria marcada que comprenden las etapas de: (a) seleccionar la bacteria para una unidad espaciadora de repetición adicional en un locus CRISPR; (b) determinar la secuencia de nucleótidos de la unidad espaciadora de repetición adicional; (c) comparar la secuencia de nucleótidos de la unidad espaciadora de repetición adicional con la base de datos de bacterias marcadas obtenida u obtenible mediante el método de la presente invención; y (d) identificar una secuencia de nucleótidos en la base de datos de bacterias marcadas que se corresponda con la unidad espaciadora de repetición adicional que se expone también en el presente documento.

En algunas realizaciones, se comparan el extremo 5' y/o el extremo 3' del locus CRISPR de la bacteria progenitora. En algunas realizaciones particularmente preferidas, se compara al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) del extremo 5' del locus CRISPR. En otras realizaciones más, se compara al menos el último espaciador CRISPR (por ejemplo, el último núcleo del espaciador CRISPR) en el extremo 3' del locus CRISPR. En realizaciones adicionales preferidas de los métodos de la presente invención, los métodos comprenden la etapa de seleccionar una bacteria marcada que comprende una unidad espaciadora de repetición adicional en el extremo 5' y/o en el extremo 3' del locus CRISPR que no está presente en la bacteria progenitora. En algunas realizaciones, los métodos de la presente invención comprenden exponer la bacteria progenitora a dos o más bacteriófagos tanto simultánea como secuencialmente.

En algunas realizaciones adicionales, el locus CRISPR o al menos una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago se comparan amplificando el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y/o el mutante insensible a bacteriófago. En algunas realizaciones preferidas, la amplificación se lleva a cabo usando la PCR. En algunas realizaciones adicionales, el CRISPR o al menos una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago se comparan secuenciando el locus CRISPR o una parte del mismo procedente de bacteria progenitora y/o el mutante insensible a bacteriófago. En algunas realizaciones particularmente preferidas, el locus CRISPR o al menos una parte del mismo procedente de la bacteria progenitora se comparan amplificando y a continuación secuenciando el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y/o el mutante insensible al bacteriófago. En algunas realizaciones, la unidad espaciadora de repetición adicional tiene al menos 44 nucleótidos de longitud. En algunas realizaciones adicionales, se selecciona una bacteria marcada que comprende dos o tres unidades espaciadoras de repetición adicionales. En algunas realizaciones preferidas, la unidad espaciadora de repetición adicional comprende al menos una secuencia de nucleótidos que tiene al menos un 95 % de identidad, o más preferentemente, aproximadamente un 100 % de identidad con una repetición CRISPR en el locus CRISPR de la bacteria progenitora. En algunas realizaciones preferidas alternativas, la unidad espaciadora de repetición adicional comprende al menos una secuencia de nucleótidos que tiene al menos un 95 % de identidad, preferentemente, un 100 % de identidad con una secuencia de nucleótidos en el genoma del bacteriófago utilizado para la selección de la bacteria marcada.

En algunas realizaciones alternativas preferidas, la unidad espaciadora de repetición adicional comprende una primera secuencia de nucleótidos que tiene al menos aproximadamente un 95 % de identidad, o más preferentemente, aproximadamente un 100 % de identidad con una repetición CRISPR en el locus CRISPR de la bacteria progenitora y una segunda secuencia de nucleótidos que tiene al menos una secuencia de nucleótidos que tiene al menos aproximadamente un 95 % de identidad, o más preferentemente, aproximadamente un 100 % de identidad con una secuencia de nucleótidos en el genoma del bacteriófago utilizado para la selección de la bacteria marcada.

Los métodos para identificar una bacteria marcada pueden comprender la etapa adicional de comparar la unidad espaciadora de repetición adicional con una base de datos de secuencias de bacteriófagos y/o una base de datos de secuencias bacterianas.

En algunas realizaciones preferidas, la bacteria progenitora es adecuada para su uso como cultivo iniciador, un cultivo probiótico y/o un suplemento de la dieta. En algunas realizaciones preferidas, la bacteria progenitora se selecciona entre el grupo de géneros que consiste en: *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*. En algunas realizaciones, la presente invención encuentra uso con cultivos celulares, incluyendo, pero no de manera limitativa, cultivos iniciadores, cultivos probióticos y/o suplementos de la dieta.

En algunas realizaciones, el bacteriófago se selecciona entre el grupo de familias de virus que consisten en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, y Tectiviridae.

En algunas realizaciones, la presente divulgación proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible procedente de un bacteriófago, en el que dicha secuencia comprende la SEQ ID NO:3 y/o 4. En algunas realizaciones, la presente divulgación proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible de un bacteriófago, en el que la secuencia comprende la SEQ ID NO:3 y/o 4 localizado en la dirección 3' (por ejemplo, directamente en la dirección 3') de la primera repetición CRISPR en al menos un locus CRISPR. En otras realizaciones más, la presente divulgación proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible procedente de un bacteriófago, en el que la secuencia comprende la SEQ ID NO:9. En algunas realizaciones preferidas, el *S. thermophilus* que comprende una secuencia obtenida u obtenible de un bacteriófago, en el que la secuencia comprende la SEQ ID NO:9 en la dirección 3' (por ejemplo, directamente después) de la primera repetición CRISPR en al menos un locus CRISPR. En otras realizaciones más, la presente divulgación proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible procedente de un bacteriófago, en el que dicha secuencia comprende la SEQ ID NO:11. En otras realizaciones adicionales, el *S. thermophilus* comprende una secuencia obtenida u obtenible de un bacteriófago, en el que dicha secuencia comprende la SEQ ID NO:11 en la dirección 3' (por ejemplo, directamente en la dirección 3') de la primera repetición CRISPR en al menos un locus CRISPR.

Los métodos para etiquetar una bacteria que comprende las etapas de: exponer al menos una bacteria progenitora que comprende al menos una parte de un locus CRISPR para al menos una secuencia de ácido nucleico exógeno para producir al menos una bacteria marcada que comprende un locus CRISPR modificado, en el que el locus CRISPR modificado comprende al menos una unidad espaciadora de repetición adicional diferente al locus CRISPR de la bacteria progenitora y en el que la unidad espaciadora de repetición adicional comprende una marca; y comparar al menos una parte del locus CRISPR de la bacteria progenitora y el locus CRISPR modificado de dicha bacteria marcada. En algunas realizaciones, la secuencia de ácido nucleico exógeno se selecciona entre bacteriófagos, plásmidos, megaplásmidos, elementos de transposición, transposones, y secuencias de inserción que se exponen también en el presente documento. El ácido nucleico exógeno puede comprender al menos una parte del genoma de al menos un bacteriófago. En algunas realizaciones alternativas particularmente preferidas, la bacteria marcada es un mutante insensible a bacteriófago. En algunas realizaciones adicionales se comparan el extremo 5' y/o el extremo 3' del locus CRISPR de la bacteria progenitora con el locus CRISPR modificado de la bacteria marcada. En realizaciones adicionales, se compara el extremo 5' y/o el extremo 3' de al menos la primera repetición CRISPR y/o al menos el primer espaciador CRISPR del locus CRISPR de la bacteria progenitora con el locus CRISPR de la bacteria marcada. En algunas realizaciones más, los métodos comprenden además la etapa de seleccionar al menos una bacteria marcada. En realizaciones más adicionales, la bacteria progenitora se expone simultánea o secuencialmente a dos o más bacteriófagos. La bacteria marcada comprende al menos una unidad espaciadora de repetición adicional. En algunas realizaciones alternativas, al menos una parte del locus CRISPR de la bacteria progenitora y al menos una parte del locus CRISPR modificado de la bacteria marcada se comparan amplificando al menos una parte del locus CRISPR y al menos una parte del locus CRISPR modificado, para producir una secuencia del locus CRISPR amplificada y una secuencia del locus CRISPR modificada amplificada. En algunas realizaciones preferidas, amplificar comprende el uso de la reacción en cadena de la polimerasa. En algunas realizaciones alternativas preferidas, al menos una parte del locus CRISPR de la bacteria progenitora y al menos una parte del locus CRISPR modificado de la mencionada bacteria se comparan secuenciando al menos una parte del locus CRISPR y al menos una parte del locus CRISPR modificado. En algunas realizaciones adicionales, los métodos comprenden además la etapa de secuenciar la secuencia del locus CRISPR amplificada y la secuencia del locus CRISPR modificada. En algunas realizaciones preferidas, la unidad espaciadora de repetición adicional tiene al menos aproximadamente 44 nucleótidos de longitud. En algunas realizaciones adicionales, la unidad espaciadora de repetición adicional comprende al menos una secuencia de nucleótidos que tiene al menos un 95 % de identidad, con una repetición CRISPR en el locus CRISPR de la bacteria progenitora. En algunas realizaciones adicionales, la unidad espaciadora de repetición adicional comprende al menos una secuencia de nucleótidos que tiene al menos un 95 % de identidad, con una secuencia de nucleótidos en el genoma de al menos un bacteriófago. En otras realizaciones adicionales, la bacteria marcada comprende al menos una secuencia de nucleótidos adicional que tiene al menos un 95 % de identidad con una secuencia de nucleótidos en el genoma de al menos un bacteriófago. En algunas realizaciones preferidas, la bacteria progenitora es un cultivo industrialmente útil. En algunas realizaciones particularmente preferidas, la bacteria progenitora comprende un cultivo seleccionado procedente de cultivos iniciadores, cultivos probióticos, y cultivos del suplemento de la dieta. En algunas realizaciones adicionales particularmente preferidas, la bacteria progenitora se selecciona entre *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*. En algunas realizaciones adicionales, el al menos un bacteriófago se selecciona entre el grupo de familias de virus que consisten en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, y Tectiviridae.

Se exponen también en la presente memoria las bacterias marcadas utilizando los métodos que se muestran en el presente documento. En algunas realizaciones preferidas, la bacteria marcada es un cultivo industrialmente útil. Se enseñan también en la presente memoria los cultivos celulares que comprenden las bacterias marcadas utilizando los métodos que se muestran en el presente documento. El cultivo celular que comprende las bacterias marcadas puede comprender un cultivo industrialmente útil.

En algunas realizaciones particularmente preferidas, la bacteria marcada comprende un cultivo seleccionado entre cultivos iniciadores, cultivos probióticos y cultivos del suplemento de la dieta. En realizaciones más adicionales, la presente invención proporciona un alimento y/o alimentación que comprende una bacteria marcada obtenida utilizando los métodos que se muestran en el presente documento.

La presente invención proporciona también métodos para preparar alimentos y/o alimentación que comprenden el uso de bacterias marcadas en el que las bacterias marcadas se añaden al alimento o a la alimentación. En algunas realizaciones adicionales, el cultivo celular que comprende la bacteria marcada comprende un cultivo industrialmente útil. En algunas realizaciones particularmente preferidas, la bacteria marcada comprende un cultivo seleccionado entre cultivos iniciadores, cultivos probióticos y cultivos del suplemento de la dieta.

Se exponen en la presente memoria el alimento y la comida que comprende la bacteria marcada obtenida utilizando los métodos que se muestran en el presente documento.

Los métodos para generar al menos una variante CRISPR que comprende una marca, que comprende las etapas de: exponer una bacteria progenitora que comprende a al menos una parte de un locus CRISPR a al menos una bacteriófago para producir un cultivo de bacteriófagos resistentes a variantes de bacterias que comprenden un locus CRISPR modificado; seleccionar un bacteriófago resistente a una bacteria variante; comparar el locus CRISPR o una parte del mismo de la bacteria progenitora y el locus CRISPR modificado de la variante insensible al bacteriófago, para identificar variantes insensibles a bacteriófagos que comprenden al menos una marca, en el que al menos una marca comprende al menos un fragmento de ácido nucleico adicional en el locus CRISPR modificado que está ausente del locus CRISPR de la bacteria progenitora; seleccionar las variantes insensibles a bacteriófagos que comprenden una marca; y analizar al menos una marca también se exponen en el presente documento. Se puede llevar a cabo el análisis utilizando cualquier método adecuado conocido en la técnica. El análisis puede seleccionarse mediante aislamiento, clonación y secuenciación. Se expone también en la presente memoria al menos una variante CRISPR obtenida utilizando los métodos que se muestran en el presente documento. Se enseñan también en la presente memoria los cultivos celulares que comprenden al menos una variante CRISPR producida utilizando los métodos que se muestran en el presente documento.

Se expone también en la presente memoria el alimento y/o la comida que comprende al menos una variante CRISPR producida utilizando los métodos que se muestran en el presente documento.

Los métodos para identificar una bacteria marcada que comprende al menos un locus CRISPR, que comprende las etapas de: seleccionar una bacteria marcada en presencia de una etiqueta en el locus CRISPR; determinar la secuencia de nucleótidos de la etiqueta; comparar la secuencia de nucleótidos de la etiqueta con las secuencias de nucleótidos presentes en al menos una base de datos; y se describe en el presente documento identificar una secuencia de nucleótidos en la base de datos que comparte homología con la etiqueta. La base de datos puede comprender las secuencias de nucleótidos de las bacterias marcadas. Las bases de datos pueden seleccionarse a partir de las bases de datos de las secuencias de bacteriófagos y de las bases de datos de las secuencias bacterianas.

Descripción de las figuras

La Figura 1 ilustra una realización de la presente invención en la que una secuencia marcadora y una repetición CRISPR se integran en un extremo del locus CRISPR. El panel A muestra un locus y elementos CRISPR, incluyendo repeticiones (R), espaciadores (S), la secuencia líder en la dirección 5' y la secuencia remolque en la dirección 3', con la repetición terminal (RT) adyacente a la secuencia remolque, y los genes *cas* genes en la proximidad (4 genes *cas* denominados *cas 1* a *cas 4* en este ejemplo, no está dibujado a escala). Los genes *cas* pueden estar en cualquier extremo, o repartidos y presentes en ambos extremos. Los genes *cas* pueden localizarse en cualquiera de las dos hebras de ADN. El panel B muestra una secuencia de fago en negro, utilizándose un fragmento de la secuencia (Sn) como espaciador adicional (es decir, secuencia marcadora). El panel C muestra la inserción de un nuevo espaciador (Sn) (es decir, una secuencia marcadora) en un extremo del locus CRISPR (cercano a la secuencia líder en este ejemplo en el extremo 5' del locus CRISPR), entre dos repeticiones. El panel D proporciona una comparación del contenido del locus CRISPR entre la bacteria progenitora y la bacteria mutante (es decir, la bacteria marcada), con un nuevo espaciador (Sn) (es decir, la secuencia marcadora) integrado en un extremo del locus CRISPR (cercano a la secuencia líder en este ejemplo), entre repeticiones. El nuevo espaciador (Sn) constituye la secuencia marcadora que es específica de la bacteria mutante (es decir, bacteria marcada). En algunas realizaciones, este proceso da como resultado la adición de uno o más espaciadores procedentes de la secuencia del fago.

La Figura 2 ilustra una realización de la presente invención en la que dos secuencias marcadoras y una

repetición CRISPR se integran en un extremo del locus CRISPR. El panel A muestra un locus y elementos CRISPR, incluyendo repeticiones (R), espaciadores (S), la secuencia líder en la dirección 5' y la secuencia remolque, con la repetición terminal (RT) adyacente a la secuencia remolque, y los genes *cas* genes en la proximidad (4 genes *cas* denominados *cas 1* a *cas4* en este ejemplo, no está dibujado a escala). Los genes *cas* pueden estar en cualquier extremo, o repartidos y presentes en ambos extremos. Los genes *cas* pueden localizarse en cualquiera de las dos hebras de ADN. El panel B muestra una secuencia de fago en negro, utilizándose un fragmento de la secuencia (Sn y Sn') como espaciador adicional (es decir, secuencias marcadoras). El panel C ilustra la inserción de los nuevos espaciadores (es decir, secuencias marcadoras) (Sn y Sn') en el mismo extremo del locus CRISPR (cercano a la secuencia líder en este ejemplo en el extremo 5'), cada uno entre dos repeticiones. El panel D proporciona una comparación del contenido del locus CRISPR entre la bacteria progenitora y la bacteria mutante (es decir, la bacteria marcada), con dos espaciadores (Sn y Sn') integrados en el mismo extremo del locus CRISPR (cercano a la secuencia líder en este ejemplo en el extremo 5'), cada uno entre repeticiones. Los nuevos espaciadores Sn y Sn' constituyen la secuencia marcadora que es específica del mutante. En algunas realizaciones, este proceso da como resultado la adición de uno o más espaciadores procedentes de la secuencia del fago.

La Figura 3 proporciona una representación esquemática del locus CRISPR1 de *S. thermophilus* DGCC7710 y las variantes marcadas. El locus CRISPR1 de DGCC7710(WT) está en la parte superior. En esta Figura, las secuencias con flecha representan genes en la proximidad de CRISPR1. La región espaciadora de repetición de WT está en la parte intermedia, con repeticiones (rombos negros), espaciadores (secuencias en gris numeradas), secuencia líder (L, secuencia blanca), y las repeticiones terminales (rombo negro, T) indicadas. En la parte inferior, las variantes marcadas se representan con 1, 2, o 4 unidades espaciadoras de repetición añadidas (rombos negros asociados con la secuencia blanca S+n). La combinación de las unidades espaciadoras de repetición añadidas representa la marca. En esta Figura, se utiliza el nuevo sistema de numeración de los genes *cas*. Anteriormente, los genes *cas* se numeraban *cas1*, *cas2*, *cas3*, y *cas4*. Ahora, los genes *cas* se numeran *cas5*, *cas1*, *cas6*, y *cas7*. De esta manera, para correlacionar las designaciones en la Figuras 1 y 2, con las de las Figuras 3 y 4, *cas1* en las Figuras 1 y 2 se denomina *cas5* en las Figuras 3 y 4; *cas2* en las Figuras 1 y 2 es *cas1* en las Figuras 3 y 4; *cas3* en las Figuras 1 y 2 es *cas6* en las Figuras 3 y 4; y *cas4* en las Figuras 1 y 2 es *cas7* en las Figuras 3 y 4.

La Figura 4 proporciona una representación esquemática del locus CRISPR1 de *S. thermophilus* DCGG7710 y la ubicación de los cebadores de detección de la PCR. Se utilizan los mismos símbolos que en la Figura 3. Se indican las posiciones del cebador directo y el cebador inverso. Como se indica también en la descripción de la Figura 3 anterior, se utiliza el nuevo sistema de numeración de los genes *cas*. Anteriormente, los genes *cas* se numeraban *cas1*, *cas2*, *cas3*, y *cas4*. Ahora, los genes *cas* se numeran *cas5*, *cas1*, *cas6*, y *cas7*. De esta manera, para correlacionar las designaciones en la Figuras 1 y 2, con las de las Figuras 3 y 4, *cas1* en las Figuras 1 y 2 se denomina *cas5* en las Figuras 3 y 4; *cas2* en las Figuras 1 y 2 es *cas1* en las Figuras 3 y 4; *cas3* en las Figuras 1 y 2 es *cas6* en las Figuras 3 y 4; y *cas4* en las Figuras 1 y 2 es *cas7* en las Figuras 3 y 4.

La Figura 5 proporciona una comparación de secuencias de parte de la secuencia del locus CRISPR1 del aislado obtenido a partir del producto de leche fermentada con la secuencia del locus CRISPR1 de DGCC7710 (SEQ ID NO:1), con la secuencia de repetición consenso CRISPR1 (SEQ ID NO:74) y con la secuencia del espaciador adicional en DGCC_{phi2972}^{S41} (SEQ ID NO:75). Se muestra la secuencia líder de CRISPR1 en minúsculas, las secuencias de repetición están recuadradas, y las otras secuencias corresponden a las secuencias del espaciador.

Descripción detallada de la invención

La presente invención proporciona métodos para marcar bacterias.

En algunas realizaciones especialmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros.

Existe una necesidad de métodos y composiciones útiles para la identificación de cepas bacterianas específicas, a fin de determinar su origen. Aunque es factible insertar un oligonucleótido sintético en una cepa para etiquetar o marcar esta utilizando tecnologías de ADN recombinante, la cepa marcada se considerará un organismo genéticamente modificado y sería probable enfrentarse a problemas normativos en aplicaciones comerciales.

Además, la preparación de cultivos requiere mucha mano de obra, ocupa mucho espacio y equipo, y existe un considerable riesgo de contaminación con bacterias y/o fagos deteriorantes durante las etapas de propagación. El fracaso de los cultivos bacterianos debido a la infección por bacteriófagos (fagos) y la multiplicación es un problema principal con el uso industrial de cultivos bacterianos. Existen muchos tipos diferentes de fagos y continúan emergiendo nuevas cepas. De esta manera, existe una necesidad de métodos y composiciones para realizar un seguimiento y vigilar las bacterias utilizadas en dichos cultivos.

Cuando una población bacteriana se infecta con un bacteriófago virulento, muchas de las células son destruidas por el bacteriófago. Sin embargo, se producen a menudo mutantes espontáneos resistentes a fagos. Estas bacterias resistentes a bacteriófagos corresponden a una subpoblación de bacterias que son capaces de soportar y sobrevivir a la infección por bacteriófagos. Estas bacterias resistentes se denominan en el presente documento "mutantes

resistentes a bacteriófagos", "mutantes insensibles a bacteriófagos", "BIM", "bacterias etiquetadas", "bacteria marcada", "bacterias marcadas", o "bacteria marcada"

Como se describe en el presente documento, cuando un bacteriófago infecta una bacteria, una o más secuencias que se originan a partir del genoma bacteriano se integran en (por ejemplo, en el interior) del locus CRISPR de la bacteria, mientras que en otras realizaciones, la integración se produce en otras localizaciones en el genoma de la bacteria. De hecho, en algunas realizaciones, "mutantes resistentes a bacteriófagos"/"mutantes insensibles a bacteriófagos"/"BIM" tienen secuencias de bacteriófagos integradas en CRISPR mientras que en otras realizaciones, las cepas tienen otro tipo de mutación cromosómica. Sin embargo, se pretende que la bacteria marcada se produzca debido a la integración de una secuencia de fago en CRISPR. De esta manera, la secuencia derivable o derivada de bacteriófago es nueva en el locus CRISPR de la bacteria y proporciona una "marca" o "etiqueta" que es identificable por su localización y/o secuencia y/o la secuencia adyacente. Se ha encontrado también que una secuencia duplicada (por ejemplo, una repetición CRISPR duplicada) que se origina a partir de la bacteria progenitora se integra también de forma iterativa, secuencial, simultánea o sustancialmente simultánea junto con la secuencia que se origina a partir del genoma del bacteriófago.

Además, en algunas realizaciones, la infección independiente de un cultivo (por ejemplo, un cultivo puro) de una cepa bacteriana dada que utiliza el mismo bacteriófago virulento conduce a la integración de una o más secuencias de bacteriófagos diferentes en el locus CRISPR de la cepa bacteriana. En algunas realizaciones preferidas, la integración de diferentes secuencias de bacteriófagos en el locus CRISPR de la cepa bacteriana es un acontecimiento aleatorio. Sin embargo, en algunas realizaciones diferentes, la integración no es una incidencia aleatoria. Una vez que se integra, esta se mantiene y por tanto se convierte en un medio sólido para marcar y/o hacer un seguimiento a una bacteria. De esta manera, la una o más secuencias que se originan a partir del genoma del bacteriófago no son solo nuevas para el locus CRISPR de la bacteria progenitora sino que son también una marca que es única para cada bacteria. De esta manera, la presente invención proporciona composiciones y métodos para etiquetar (es decir, marcar) y/o identificar una bacteria. Además, los métodos y composiciones de la presente invención son particularmente ventajosos debido a que el método es un método 'natural' que no da como resultado un organismo genéticamente modificado. De acuerdo con ello, las bacterias marcadas o etiquetadas preparadas de acuerdo con los métodos de la presente invención no se considera que están genéticamente modificadas, ya que las bacterias se han creado mediante un proceso biológico natural de infección por el bacteriófago.

La presente divulgación proporciona métodos para el uso de una secuencia obtenida u obtenible procedente de un bacteriófago (por ejemplo, en la fabricación de una bacteria marcada) para marcar y/o identificar una bacteria, en la que la secuencia se integra en un extremo del locus CRISPR de una bacteria progenitora. En algunas realizaciones preferidas, la presente divulgación proporciona métodos para el uso de una secuencia obtenida u obtenible procedente de un bacteriófago (por ejemplo, en la fabricación de una bacteria marcada) para marcar y/o identificar una bacteria, en la que la secuencia comprende: (i) al menos una secuencia que es homóloga (por ejemplo, idéntica) a una repetición CRISPR en el locus CRISPR de la bacteria; y (ii) una secuencia marcadora. En realizaciones adicionales, la presente divulgación proporciona métodos para el uso de una secuencia para marcar y/o identificar una bacteria (por ejemplo, en la fabricación de una bacteria marcada), en el que la secuencia es obtenida u obtenible: (a) exponiendo una bacteria progenitora a un bacteriófago; (b) seleccionando un mutante insensible a bacteriófago; (c) comparando el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago; y (d) seleccionando una secuencia en el locus CRISPR o una parte del mismo del mutante insensible a bacteriófago que no está presente en la bacteria progenitora.

En algunas realizaciones adicionales, la presente divulgación proporciona secuencias de ácidos nucleicos (por ejemplo, una secuencia de ácido nucleico recombinante o una secuencia de ácido nucleico aislado) que consiste esencialmente en al menos un gen o proteína. En algunas realizaciones, la secuencia de ácido nucleico es ADN, mientras que en otras realizaciones, es ARN. El ácido nucleico procedente de cualquier origen adecuado encuentra uso en la presente invención, incluyendo el ácido nucleico genómico, sintético o recombinante (por ejemplo, ADNc). Sin embargo, en algunas realizaciones particularmente preferidas, la secuencia es una secuencia de ácido nucleico que se produce naturalmente. En algunas realizaciones, la secuencia de nucleótidos es bicatenaria, mientras que en otras realizaciones, es monocatenaria. En algunas realizaciones adicionales, la secuencia de ácido nucleico representa la hebra de sentido directo, mientras que en otras realizaciones, representa la hebra de sentido contrario o combinaciones de las mismas. Las secuencias de ácido nucleico recombinante preparadas utilizando cualquier técnica recombinante adecuada conocida en la técnica encuentran uso en la presente invención. En algunas realizaciones preferidas, la secuencia de ácido nucleico diana es, o se deriva de un gen. En algunas realizaciones particularmente preferidas, la secuencia de ácido nucleico es una "secuencia de ácido nucleico exógena", que se introduce en una bacteria progenitora utilizando cualquier método adecuado conocido en la técnica, incluyendo, pero no de forma limitativa, métodos naturales y recombinantes. En algunas realizaciones muy particularmente preferidas, la secuencia de ácido nucleico exógena comprende al menos una parte de un genoma de bacteriófago. En algunas realizaciones adicionales particularmente preferidas, la secuencia de ácido nucleico exógena se introduce en una bacteria progenitora mediante la exposición de la bacteria a al menos un bacteriófago.

En algunas realizaciones, la secuencia de ácido nucleico y los ácidos nucleicos abarcados por la presente divulgación están sustancialmente aislados o purificados. Como se usa en el presente documento, los términos "aislado" o "sustancialmente purificado" se refieren a moléculas de ácidos nucleicos, fragmentos biológicamente activos, homólogos o derivados de los mismos que están sustancial o esencialmente exentos de componentes que se encuentran normalmente asociados con el ácido nucleico en su estado natural. Dichos componentes incluyen otros materiales celulares, medios de cultivo procedentes de producción recombinante, y diversos productos químicos utilizados para sintetizar químicamente los ácidos nucleicos. En algunas realizaciones, una secuencia de ácido nucleico o un ácido nucleico "aislado" está exento normalmente de las secuencias de ácidos nucleicos que flanquean el ácido nucleico de interés en el ADN genómico del organismo a partir del cual se derivó el ácido nucleico (por ejemplo, las secuencias de codificación presentes en los extremos 5' o 3'). Sin embargo, en algunas realizaciones, la molécula incluye algunas bases o restos adicionales que no afectan negativamente las características básicas de la composición.

La(s) secuencia(s) de ácidos nucleicos son de utilidad en células genomanipuladas (por ejemplo, una célula receptora). En algunas realizaciones, la secuencia de ácido nucleico se inserta en el ADN (por ejemplo, ADN plásmido o ADN genómico) de una célula receptora que utiliza cualquier método adecuado conocido en la técnica (por ejemplo, recombinación homóloga). En otras realizaciones, la(s) secuencia(s) de ácido nucleico encuentran uso como moldes a partir de los que se modifica (por ejemplo, muta) el ADN de una célula (por ejemplo, una célula receptora) tal como un ADN plásmido o ADN genómico, en condiciones tales que se crean secuencia(s) de ácido nucleico en el ADN de la célula. En algunas realizaciones preferidas, la(s) secuencias de ácidos nucleicos se clonan (por ejemplo, en una construcción, plásmido o vector) que se usa a continuación para transformar la célula utilizando cualquier método adecuado conocido en la técnica.

La presente divulgación proporciona métodos y composiciones que utilizan variantes, homólogos, derivados y fragmentos de los mismos. El término "variante" se usa en el presente documento en referencia a secuencias de polipéptidos o nucleótidos que se producen naturalmente que difieren de una secuencia natural. Como se usa en el presente documento, el término "fragmento" se refiere a un polipéptido o secuencia de nucleótido que comprende una fracción de una secuencia natural. En algunas realizaciones, los fragmentos comprenden una o más secciones grandes contiguas de una secuencia o una pluralidad de pequeñas secciones. En algunas realizaciones, la secuencia comprende también otros elementos. Por ejemplo, en algunas realizaciones, es una proteína de fusión que incluye otra secuencia de proteína. En algunas realizaciones preferidas, la secuencia comprende al menos aproximadamente un 50 %, más preferentemente, al menos aproximadamente un 65 % o más, preferentemente al menos aproximadamente un 80 %, más preferentemente al menos aproximadamente un 85 %, más preferentemente al menos un 90 %, más preferentemente al menos un 95 %, más preferentemente al menos un 96 %, más preferentemente al menos aproximadamente un 97 %, más preferentemente al menos aproximadamente un 98 % o lo más preferente al menos aproximadamente un 99 % de la secuencia natural.

En algunas realizaciones particularmente preferidas, el fragmento es un fragmento funcional. Como se usa en el presente documento, un "fragmento funcional" de una molécula se refiere a un fragmento que retiene o que posee sustancialmente la misma actividad biológica que la molécula intacta. En realizaciones particularmente preferidas, los fragmentos funcionales retienen al menos aproximadamente un 10 %. En otras realizaciones, al menos aproximadamente un 25 %, aproximadamente un 50 %, aproximadamente un 75 %, aproximadamente un 80 %, aproximadamente un 85 %, aproximadamente un 90 %, aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 %, o aproximadamente un 99 % de la actividad biológica de la molécula intacta. En realizaciones alternativas, el fragmento retiene aproximadamente un 50 %, más preferentemente aproximadamente un 60 %, más preferentemente aproximadamente un 70 %, más preferentemente aproximadamente un 80 %, más preferentemente aproximadamente un 85 %, más preferentemente aproximadamente un 90 %, más preferentemente aproximadamente un 95 %, más preferentemente aproximadamente un 96 %, más preferentemente aproximadamente un 97 %, más preferentemente aproximadamente un 98 %, o lo más preferente, aproximadamente una actividad del 99 % del polipéptido natural o la secuencia de nucleótidos.

Como se usa en el presente documento, el término "homólogo" se refiere a una entidad que tiene una determinada homología con las secuencias de aminoácidos sujeto y las secuencias de nucleótidos sujeto. Como se usa particularmente en el presente documento, el término "homología" es sinónimo de "identidad". En el presente contexto, se considera que una secuencia homóloga incluye una secuencia de aminoácidos, que es al menos aproximadamente un 75 %, aproximadamente un 85 %, o aproximadamente un 90 % idéntica, preferentemente al menos aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 % o aproximadamente un 99 % idéntica a la secuencia sujeto. Aunque la homología se puede considerar también en términos de similitud (es decir, restos aminoácidos que tienen similares propiedades/funciones químicas), en el contexto de la presente invención, se prefiere expresar la homología en términos de identidad de secuencias. En algunas realizaciones preferidas, una secuencia homóloga comprende una secuencia de nucleótidos, que es al menos aproximadamente un 75 %, aproximadamente un 85 % o aproximadamente un 90 % idéntica, preferentemente al menos aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 % o aproximadamente un 99 % idéntica a la secuencia sujeto (es decir, la secuencia de interés utilizada como referencia). En algunas realizaciones, se llevaron a cabo comparaciones de homología visualmente, aunque

encuentran uso otros métodos conocidos en la técnica (por, con la ayuda de programas de comparación de secuencias fácilmente disponibles). Los programas informáticos comercialmente disponibles son capaces de calcular el porcentaje de homología (% de homología) entre dos o más secuencias, y de esta manera son de utilidad en la presente invención. De hecho, están comercialmente disponibles de forma fácil métodos y sistemas para dichos análisis.

Descripción detallada de la invención

La presente invención proporciona métodos para marcar bacterias.

En algunas realizaciones especialmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. La presente divulgación proporciona también microorganismos marcados usando los métodos que se muestran en el presente documento. En algunas realizaciones preferidas, los microorganismos marcados son bacterias. En algunas realizaciones particularmente preferidas, las bacterias marcadas son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias marcadas son miembros de otros géneros.

A. Loci CRISPR y CRISPR

CRISPR (repeticiones palindrómicas cortas separadas entre sí regularmente agrupadas); conocidas también como SPIDR - Repeticiones directas intercaladas con separación) constituye una familia de loci de ADN descrita recientemente que suele ser específica de una especie bacteriana concreta. El locus CRISPR es una clase distinta de repeticiones de secuencias cortas intercaladas (SSR) que se reconocieron en primer lugar en *E. coli* (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; y Nakata et al., J. Bacteriol. 171:3553-3556 [1989]). Se han identificado SSR intercaladas similares en *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, y *Mycobacterium tuberculosis* (See, Groenen et al., Mol. Microbiol. 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; y Mojica et al., Mol. Microbiol. 17:85-93 [1995]). Los loci CRISPR difieren de otros SSR en la estructura de las repeticiones, que se han denominado repeticiones separadas regularmente cortas (SRSR) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; y Mojica et al., Mol. Microbiol. 36:244-246 [2000]). Las repeticiones son elementos cortos que se producen en las agrupaciones que están siempre separadas por secuencias únicas intervinientes con una longitud constante (Mojica et al., [2000], *supra*). Aunque las secuencias de repetición están muy concentradas entre cepas, el número de repeticiones intercaladas y las secuencias de las regiones espaciadoras difieren de una cepa a otra (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]).

Los loci CRISPR consisten en repeticiones de ADN palindrómico cortas y parcialmente conservadas, normalmente de 24 a 40 pb. Se ha notificado que estas repeticiones se producen en un intervalo de 1 a 249. Aunque se han detectado elementos aislados, se disponen generalmente en agrupaciones (de hasta aproximadamente 20 o más por genoma) de unidades repetidas separadas por secuencias únicas intervinientes de 20-58 pb. Hasta la fecha, se han encontrado hasta 20 loci CRISPR distintos en un único cromosoma. Los CRISPR son generalmente homogéneos en un genoma dado siendo la mayoría de ellos idénticos. Sin embargo, existen ejemplos de heterogeneidad en, por ejemplo, las Archaea (Mojica et al., [2000], *supra*).

Como se usa en el presente documento, el término "locus CRISPR" se refiere al segmento de ADN que incluye todas las repeticiones CRISPR, comenzando con el primer nucleótido de la primera repetición CRISPR y finalizando con el último nucleótido de la última repetición CRISPR (terminal). En algunas realizaciones alternativas, es de utilidad "al menos una parte" de al menos un locus CRISPR. De esta manera, se pretende que la presente invención abarque realizaciones en las que se usa al menos un locus CRISPR completo, así como realizaciones en el que al menos una parte (es decir, parte de al menos un locus CRISPR).

Aunque la función biológica del loci CRISPR es desconocida, se han propuesto algunas hipótesis. Por ejemplo, se ha propuesto que puedan estar implicados en la unión del cromosoma a una estructura celular o en la replicación del cromosoma y el reparto del replicón (Jansen et al., OMICS 6:23-33 [2002]; Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; y Pourcel et al., Microbiol., 151:653-663 [2005]). Mojica et al. (Mojica et al., J. Mol. Evol., 60:174-182 [2005]) teorizan que CRISPR puede estar implicado en conferir inmunidad específica contra el ADN extraño y Pourcel et al. (*supra*) teorizan que los CRISPR son estructuras que son capaces de tomar partes de ADN extraño como parte de un mecanismo de defensa. Bolotin et al. (*supra*) sugieren que los elementos espaciadores de CRISPR son las trazas de invasiones anteriores por elementos extracromosómicos, y teorizan que proporcionan inmunidad a la frente a la infección por un fago, y más generalmente, la expresión del ADN extraño codificando una ARN de sentido contrario. Bolotin et al. (*supra*) sugieren también que los genes *cas* son necesarios para la formación de CRISPR. Sin embargo, no se pretende que la presente invención esté limitada por cualquier mecanismo, función, teoría concreta, ni medios de acción.

B. Identificar los loci CRISPR

Se conocen en la técnica diversos métodos para identificar loci CRISPR. Por ejemplo, Jensen *et al.* (Jensen *et al.*, [2002], *supra*) describen una solución informática en la cual se investigan secuencias de nucleótidos para los motivos CRISPR utilizando el programa PATSCAN en el servidor de la Mathematics and Computer Science Division at the Argonne National Laboratory, Argonne, IL, USA. El algoritmo que se utilizó para identificar los motivos CRISPR era $p1 = a \dots bc\dots d \ p1 \ c\dots d \ p1 \ c \dots d \ p1$, donde a y b eran el límite del tamaño inferior y superior de la repetición y $p1$ y c and d eran el límite del tamaño inferior y superior de las secuencias del espaciador. Los valores de a , b , c y d pueden variarse desde aproximadamente 15 a aproximadamente 70 pb en incrementos de aproximadamente 5 pb. En algunas realizaciones preferidas, se identificaron los loci CRISPR utilizando gráficas de puntos (por ejemplo, utilizando el programa informático Dotter).

Cualquier método conocido adecuado en la técnica es de utilidad en el análisis de la similitud de la secuencia. Por ejemplo, se puede llevar a cabo el análisis utilizando NCBI BLAST con una base de datos de genomas microbianos y el GenBank, como se conoce en la técnica. Además, se incluyen secuencias de nucleótidos, incluyendo las proporcionadas en el presente documento en las bases de datos (por ejemplo, GenBank o el sitio web del genoma JGI).

En realizaciones adicionales, los métodos de la presente invención utilizan procedimientos de amplificación (Véase por ejemplo, Monica *et al.*, [2005], *supra*; y Pourcel *et al.*, [2005], *supra*). Se puede conseguir la amplificación de la región deseada del ADN mediante cualquier método conocido en la técnica, incluyendo la reacción en cadena de la polimerasa (PCR). "Amplificación" se refiere a la producción de copias adicionales de una secuencia de ácido nucleico. Esto se lleva a cabo generalmente utilizando las tecnologías de la PCR. La "reacción en cadena de la polimerasa" ("PCR") es bien conocida por los expertos en la técnica. En la presente invención, los cebadores de oligonucleótidos se diseñan para su uso en las reacciones de la PCR para amplificar todo o parte de un locus CRISPR. El término "cebador" se refiere a un oligonucleótido, que se produce tanto naturalmente como en una digestión de restricción purificada o se produce sintéticamente, que es capaz de actuar como un punto de inicio de la síntesis cuando se coloca en condiciones en las que se induce la síntesis de un producto de extensión del cebador que es complementario con una hebra de ácido nucleico (es decir, en presencia de nucleótidos y un agente inductor, tal como la ADN polimerasa y a una temperatura y pH adecuados). En algunas realizaciones, el cebador es monocatenario para una eficacia máxima en la amplificación, aunque en otras realizaciones, el cebador es bicatenario. En algunas realizaciones, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebear la síntesis de los productos de extensión en presencia del agente inductor. La longitud exacta de los cebadores depende de muchos factores, incluyendo la temperatura, la fuente del cebador, y el uso del método. Los cebadores de la PCR tienen normalmente al menos aproximadamente 10 nucleótidos de longitud, y más normalmente al menos aproximadamente 20 nucleótidos de longitud. Los métodos para diseñar y llevar a cabo la PCR son bien conocidos en la técnica, e incluyen, pero no de forma limitativa, métodos que utilizan cebadores emparejados, cebadores anidados, cebadores monoespecíficos, cebadores degenerados, cebadores específicos de gen, cebadores específicos de vector, cebadores emparejados de forma parcialmente incorrecta, etc.

En algunas realizaciones preferidas de la presente invención, un locus CRISPR o una parte del mismo procedente de una bacteria progenitora y una bacteria marcada se comparan utilizando cualquier método adecuado conocido en la técnica. En algunas realizaciones preferidas de la presente invención, el locus CRISPR o una parte del mismo procedente de la bacteria progenitora se comparan amplificando el locus CRISPR o una parte del mismo. Además de los métodos de amplificación bien conocido (por ejemplo, la PCR, la reacción en cadena de la ligasa, etc.), otros métodos, incluyendo, pero no de forma limitativa, los métodos de amplificación isotérmica, son de utilidad en la presente invención. Los métodos de amplificación isotérmica bien conocidos que son de utilidad en la presente invención incluyen, pero no de forma limitativa, la amplificación por desplazamiento de hebra (SDA), Q-beta-replicasa, y la amplificación de secuencia basada en ácidos nucleico (NASBA), y la replicación de secuencias autosostenida.

En algunas otras realizaciones preferidas diferentes de la presente invención, el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y la bacteria marcada se comparan secuenciando el locus CRISPR o una parte del mismo procedente de la bacteria progenitora y la bacteria marcada. En algunas realizaciones alternativas, se comparan mediante amplificación y posteriores secuenciación de los loci CRISPR o una parte de los mismos. En algunas realizaciones, se compara un extremo de los loci CRISPR procedentes de bacterias progenitoras y bacterias marcadas, mientras que en otras realizaciones, se comparan ambos extremos 5' y 3' de los loci. En algunas realizaciones preferidas, se compara un extremo (por ejemplo, el extremo 5') de los loci CRISPR. En otras realizaciones más, se compara al menos la última repetición CRISPR en el extremo 3' del locus CRISPR y/o al menos el último espaciador CRISPR (por ejemplo, el último núcleo del espaciador CRISPR) en el extremo 3' del locus CRISPR y/o al menos la primera repetición CRISPR en el extremo 5' del locus CRISPR y/o al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' del locus. En algunas realizaciones preferidas, se comparan al menos la primera repetición CRISPR en el extremo 5' del locus CRISPR y/o al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' del locus CRISPR. En algunas realizaciones preferidas adicionales, se comparan al menos el último espaciador CRISPR (por ejemplo, el último núcleo del espaciador CRISPR) en el extremo 3' del locus CRISPR y/o al menos el

primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' del locus CRISPR. En algunas realizaciones preferidas más, se compara al menos el primer separador CRISPR (por ejemplo, el primer núcleo del separador CRISPR) en el extremo 5' de los loci CRISPR.

En algunas realizaciones, los loci CRISPR comprende ADN, mientras que en otras realizaciones, los loci CRISPR comprenden ARN. En algunas realizaciones, el ácido nucleico es de origen genómico, mientras que en otras realizaciones, es de origen sintético o recombinante. En algunas realizaciones, los loci CRISPR son bicatenarios, mientras que en otras realizaciones, son monocatenarios, representando tanto la hebra de sentido directo como la de sentido contrario o las combinaciones de las mismas. En algunas realizaciones, se preparan los loci CRISPR mediante el uso de técnicas de ADN recombinante (por ejemplo ADN recombinante), como se describe en el presente documento.

En el contexto de la presente invención, el locus CRISPR se orienta basándose en la orientación 5' - 3' de los genes cas. Los genes cas (asociados a CRISPR) son usualmente adyacentes a los loci CRISPR. Por ejemplo, en el cromosoma de la cepa *S. thermophilus* CNRZ1066, el locus CRISPR1 se localiza en la dirección 3' de los genes cas *str0657*, *str0658*, *str0659*, y *str0660*. El locus CRISPR1 se orienta colinealmente a los genes cas. De esta manera, los genes cas se localizan en la dirección 5' de CRISPR1. La secuencia no codificante localizada entre el codón de detención del último gen cas y el primer nucleótido de la primera repetición de CRISPR se localiza en la dirección 5' de CRISPR y se denomina en el presente documento "secuencia líder de CRISPR". La secuencia líder CRISPR se localiza en el extremo 5' del locus CRISPR. La secuencia no codificante en el lado opuesto del locus CRISPR se denomina en el presente documento "secuencia remolque de CRISPR". La secuencia remolque de CRISPR se encuentra inmediatamente después del último nucleótido de la última repetición de CRISPR. Esta última repetición de CRISPR se denomina también "repetición terminal". Las secuencias remolques de CRISPR y las repeticiones terminales se localizan en el extremo 3' del locus CRISPR. Por ejemplo, la secuencia líder de CRISPR1 en la cepa CNRZ1066 tiene la secuencia 5'-CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATAAAAACGTCAAATTTTCATTTGA G-3' (SEQ ID NO: 12), y la secuencia remolque de CRISPR tiene la secuencia 5'-TTGATTCAACATAAAAAGCCAGTTCAATTGAACTTGGCTTT-3' (SEQ D NO: 13). La secuencia líder de CRISPR corresponde a las posiciones 625038 a 625100, y la secuencia remolque de CRISPR corresponde a las posiciones 627845 a 627885 en el genoma de *S. thermophilus* CNRZ1066 (CP000024).

Como se usa en el presente documento el término "parte del mismo" en el contexto de un locus CRISPR significa al menos aproximadamente 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos o incluso aproximadamente 100 o más nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) de un locus CRISPR.

En algunas realizaciones adicionales, el término "parte del mismo" en el contexto de un locus CRISPR significa los aproximadamente primeros 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos, o aproximadamente 100 o más nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) en la dirección 3' desde el primer nucleótido de la primera repetición de CRISPR al extremo 5' de un locus CRISPR o en la dirección 5' del último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR. En algunas realizaciones preferidas, el término "parte del mismo" se refiere a al menos aproximadamente los primeros 44 nucleótidos en la dirección 3' del primer nucleótido de la primera repetición de CRISPR en el extremo 5' de un locus CRISPR o al menos aproximadamente 44 nucleótidos en la dirección 5' del último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR.

En algunas realizaciones, el tamaño mínimo de la secuencia duplicada es aproximadamente de 24 nucleótidos y el tamaño mínimo de la secuencia marcadora es aproximadamente de 20 nucleótidos. De esta manera, en algunas realizaciones preferidas, el término "parte del mismo", en el contexto de un locus CRISPR, significa al menos 44 nucleótidos.

En algunas realizaciones, el tamaño máximo de la secuencia duplicada es aproximadamente de 40 nucleótidos y el tamaño máximo de la secuencia es aproximadamente de 58 nucleótidos. De esta manera, en algunas realizaciones, el término "parte del mismo", cuando se usa en el contexto de un locus CRISPR significa al menos aproximadamente 98 nucleótidos. En algunas realizaciones preferidas, el término "parte del mismo", en el contexto de un locus CRISPR, significa al menos 44-98 nucleótidos.

La presente invención proporciona también variantes de CRISPR, así como los métodos para generar variantes de CRISPR. En realizaciones adicionales, las variantes de CRISPR se aíslan, clonan y/o secuencian utilizando métodos conocidos en la materia. En algunas realizaciones, las variantes de CRISPR son de utilidad como dianas con fines

de detección y/o identificación, mientras que en realizaciones alternativas, las variantes de CRISPR son de utilidad en la resistencia frente a moléculas de ácidos nucleicos diseñadas mediante ingeniería genética.

C. Extremo de un locus CRISPR

Cuando se comparan el locus CRISPR o una parte del mismo de la bacteria progenitora y una bacteria marcadora, se comparan al menos aproximadamente 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos o aproximadamente 100 nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) de un locus CRISPR. En algunas realizaciones preferidas, se comparan al menos aproximadamente 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos o aproximadamente 100 o más nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) en uno o ambos extremos de un locus CRISPR.

En algunas realizaciones preferidas, se comparan aproximadamente al menos los primeros 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos o aproximadamente 100 o más nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) en el extremo 5' de un locus CRISPR o en el extremo 3' de un locus CRISPR. En algunas realizaciones preferidas, se comparan al menos aproximadamente los primeros 44 nucleótidos en el extremo 5' de un locus CRISPR o al menos aproximadamente los 44 nucleótidos en el extremo 3' de un locus CRISPR.

En algunas realizaciones aproximadamente al menos los primeros 10 nucleótidos, aproximadamente 20 nucleótidos, aproximadamente 24 nucleótidos, aproximadamente 30 nucleótidos, aproximadamente 40 nucleótidos, aproximadamente 44 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 60 nucleótidos, aproximadamente 70 nucleótidos, aproximadamente 80 nucleótidos, aproximadamente 90 nucleótidos, aproximadamente 98 nucleótidos, o aproximadamente 100 o más nucleótidos (por ejemplo, al menos aproximadamente 44-98 nucleótidos) en la dirección 3' desde el primer nucleótido de la primera repetición de CRISPR al extremo 5' de un locus CRISPR o en la dirección 5' del último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR. En algunas realizaciones preferidas, se comparan al menos aproximadamente los primeros 44 nucleótidos en la dirección 3' desde el primer nucleótido de la primera repetición de CRISPR en el extremo 5' de un locus CRISPR o al menos aproximadamente 44 nucleótidos en la dirección 5' del último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR.

En algunas realizaciones, el tamaño mínimo de la secuencia duplicada es aproximadamente de 24 nucleótidos y el tamaño mínimo de la secuencia marcadora es aproximadamente de 20 nucleótidos. En algunas realizaciones preferidas, se comparan al menos 44 nucleótidos. En algunas realizaciones alternativas, el tamaño máximo de la secuencia duplicada es aproximadamente de 40 nucleótidos y el tamaño máximo de la secuencia marcadora es aproximadamente de 58 nucleótidos. En algunas realizaciones preferidas, se comparan al menos 98 nucleótidos. En algunas realizaciones preferidas alternativas, se comparan al menos aproximadamente 44-98 nucleótidos.

D. Repetición de CRISPR

Como se usa en el presente documento, el término "repetición de CRISPR" tiene el significado convencional que se usa en la técnica (es decir, repeticiones directas cortas múltiples, que muestran ninguna o muy poca variación de la secuencia en un locus CRISPR dado). Como se usa en el presente documento, "repetición de CRISPR" es sinónima del término "CRISPR".

Un locus CRISPR comprende una o más repeticiones de CRISPR que son espaciadores de CRISPR. De esta manera, la repetición de CRISPR corresponde a la secuencia repetida en un locus CRISPR. Por ejemplo, excepto para la repetición terminal, la secuencia de repetición típica de la secuencia CRISPR1 del *S. thermophilus* es:

5'-gtttttgtactctcaagatttaagtaactgtacaac-3' (SEQ ID NO: 14)

Se han observado variaciones puntuales de esta secuencia de repetición para las secuencias de repetición en un locus CRISPR de una cepa dada y para secuencias de repetición en un locus CRISPR de cepas de una especie dada, pero son muy raras. En comparación con esta secuencia de repetición típica, la secuencia de repetición terminal de un locus CRISPR dado muestra siempre la misma variación en su extremo 3'. Se han observado también variaciones puntuales de esta secuencia de repetición terminal, pero son raras. Se pueden producir de forma natural repeticiones de CRISPR en la bacteria progenitora. Los números de registro del GenBank de las secuencias de

CRISPR1 incluyen: CP000023, C000024, DQ072985, DQ072986, DQ072987, DQ072988, DQ072989, DQ072990, DQ072991, DQ072992, DQ072993, DQ072994, DQ072995, DQ072996, DQ072997, DQ072998, DQ072999, DQ073000, DQ073001, DQ073002, DQ073003, DQ073004, DQ073005, DQ073006, DQ073007, DQ073008, y AAGS01000003.

Como se describe con más detalle en el presente documento, una secuencia duplicada derivada, derivable, obtenida u obtenible procedente de una bacteria progenitora. En algunas realizaciones preferidas, la secuencia comprende el ADN genómico de una bacteria progenitora. En algunas realizaciones particularmente preferidas, la repetición de CRISPR duplicada (por ejemplo, en el mismo locus CRISPR) se integra de forma iterativa, secuencial, simultánea o sustancialmente simultánea junto con la secuencia marcadora en la bacteria progenitora para dar lugar a una bacteria marcada.

El número de nucleótidos en una repetición es generalmente de aproximadamente 20 a aproximadamente 40 pares de bases (por ejemplo, aproximadamente 36 pares de bases). Sin embargo, no se pretende que la presente invención esté limitada a cualquier intervalo concreto comprendido en aproximadamente 20 y aproximadamente 40 pares de bases. De hecho, se pretende que cada limitación numérica máxima dada a lo largo de esta memoria descriptiva incluya cada limitación numérica inferior, como si dichas limitaciones numéricas inferiores se escribieran expresamente en el presente documento. Cada limitación numérica mínima dada a lo largo de esta memoria descriptiva incluirá cada limitación numérica mayor, como si dichas limitaciones numéricas mayores se escribieran expresamente en el presente documento. Cada intervalo numérico dado a lo largo de esta memoria descriptiva incluirá cada intervalo numérico más estrecho que se encuentra comprendido en dicho intervalo numérico más amplio, como si dichos intervalos numéricos más estrechos se escribieran expresamente en el presente documento.

En realizaciones adicionales, el número de repeticiones varía de aproximadamente 1 a aproximadamente 250. Sin embargo, no se pretende que la presente invención esté limitada a cualquier intervalo concreto comprendido entre aproximadamente 1 y aproximadamente 250 repeticiones. De hecho, como se ha indicado anteriormente, se pretende que cada limitación numérica máxima dada a lo largo de esta memoria descriptiva incluya cada limitación numérica menor, como si dichas limitaciones numéricas menores se escribieran expresamente en el presente documento. Cada limitación numérica mínima dada a lo largo de esta memoria descriptiva incluirá cada limitación numérica mayor, como si dichas limitaciones numéricas mayores se escribieran expresamente en el presente documento. Cada intervalo numérico dado a lo largo de esta memoria descriptiva incluirá cada intervalo numérico más estrecho que se encuentra comprendido en dicho intervalo numérico más amplio, como si dichos intervalos numéricos más estrechos se escribieran expresamente en el presente documento. De hecho, se pretende que esto se aplique a todos los intervalos numéricos proporcionados en el presente documento.

En algunas realizaciones, las repeticiones de CRISPR comprenden ADN, mientras que en otras realizaciones, las repeticiones de CRISPR comprenden ARN. En algunas realizaciones, el ácido nucleico es de origen genómico, mientras que en otras realizaciones, es de origen sintético o recombinante. En algunas realizaciones, los genes de repetición de CRISPR son bicatenarios o monocatenarios representando tanto la hebra de sentido directo como la hebra de sentido contrario o las combinaciones de las mismas. En algunas realizaciones, los genes de repetición de CRISPR se preparan mediante el uso de técnicas de ADN recombinante (por ejemplo, ADN recombinante) como se describe en el presente documento.

En algunas realizaciones, una o más de las repeticiones de CRISPR se usan para genomanipular una célula (por ejemplo, una célula receptora). Por ejemplo, en algunas realizaciones, la(s) repetición(es) se insertan en el ADN de una célula (por ejemplo, ADN plásmido y/o genómico de una célula receptora), utilizando cualquier método adecuado conocido en la técnica. En realizaciones adicionales, la(s) repetición(es) de CRISPR son de utilidad como a partir del cual modificar (por ejemplo, mutar) el ADN de una célula (por ejemplo, ADN plásmido y/o ADN genómico de una célula receptora), de tal manera que se crean o genomanipulan repetición(es) de CRISPR en el ADN de la célula. En realizaciones adicionales, la(s) repetición(es) de CRISPR están presentes en al menos una construcción, al menos un plásmido, y/o al menos un vector, etc. En realizaciones adicionales, las repeticiones de CRISPR se introducen en la célula utilizando cualquier método adecuado conocido en la técnica.

En algunas realizaciones adicionales, se utilizan uno o más genes o proteínas cas junto con, o en combinación con una o más, preferentemente dos o más repeticiones de CRISPR y opcionalmente uno o más espaciadores de CRISPR. En algunas realizaciones particularmente preferidas, el(los) gen(es) o la(s) proteína(s) cas y la(s) repetición(es) de CRISPR (s) forman una combinación funcional, como se describe a continuación.

E. Espaciador de CRISPR

Como se usa en el presente documento, un "espaciador de CRISPR" abarca secuencias espaciadoras no repetitivas que se localizan entre repeticiones (es decir, repeticiones de CRISPR) de los loci CRISPR. En algunas realizaciones de la presente invención, un "espaciador de CRISPR" se refiere a un segmento de ácido nucleico que está flanqueado por dos repeticiones de CRISPR. Se ha descubierto que las secuencias espaciadoras de CRISPR tienen a menudo similitudes significativas con una variedad de moléculas de ADN móviles (por ejemplo, bacteriófagos y plásmidos). En algunas realizaciones preferidas, los espaciadores de CRISPR se localizan entre dos repeticiones de

CRISPR idénticas. En algunas realizaciones, los espaciadores de CRISPR se identifican por el análisis de la secuencia de los tramos de ADN localizados entre dos repeticiones de CRISPR.

De manera interesante, se ha descubierto que las células que transportan los espaciadores de CRISPR no pueden ser infectadas por moléculas de ADN que contienen secuencias homólogas a los espaciadores (Mojica *et al.* [2005], *supra*). En la mayoría de realizaciones, el espaciador de CRISPR está presente naturalmente entre dos repeticiones directas cortas múltiples idénticas que son palindrómicas.

En algunas realizaciones, el espaciador de CRISPR es homólogo al ácido nucleico diana o a un producto de la transcripción del mismo o una secuencia identificada. Aunque en algunas realizaciones, la homología se toma en consideración en términos de similitud, en el contexto de la presente invención, en algunas realizaciones preferidas, la homología se expresa en términos de identidad de secuencias. En realizaciones preferidas, el análisis de secuencias homólogas incluye un espaciador de CRISPR, que en algunas realizaciones es al menos aproximadamente 70 %, aproximadamente 75 %, aproximadamente 80 %, aproximadamente 85 %, aproximadamente 90 %, aproximadamente 91 %, aproximadamente 92 %, aproximadamente 93 %, aproximadamente 94 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 % o aproximadamente 99 % idéntico a la secuencia de ácido nucleico diana o a un producto de transcripción de la misma o una secuencia identificada (por ejemplo, una secuencia de interés). En algunas realizaciones, el espaciador de CRISPR es un 100 % idéntico a la secuencia de ácido nucleico diana.

El número de espaciadores de CRISPR en unos loci o en un locus CRISPR dado puede variar entre especies. En algunas realizaciones preferidas, el número de repeticiones varía de aproximadamente 1 a aproximadamente 248. Sin embargo, no se pretende que la presente invención esté limitada a cualquier intervalo concreto comprendido entre aproximadamente 1 y aproximadamente 140 repeticiones. De hecho, como se ha indicado anteriormente, se pretende que cada limitación numérica máxima dada a lo largo de esta memoria descriptiva incluya cada limitación numérica menor, como si dichas limitaciones numéricas menores se escribieran expresamente en el presente documento. Cada limitación numérica mínima dada a lo largo de esta memoria descriptiva incluirá cada limitación numérica superior, como si dichas limitaciones numéricas mayores se escribieran expresamente en el presente documento. Cada intervalo numérico dado a lo largo de esta memoria descriptiva incluirá cada intervalo numérico más estrecho que se encuentra comprendido en dicho intervalo numérico más amplio, como si dichos intervalos numéricos más estrechos se escribieran expresamente en el presente documento. De hecho, se pretende que esto se aplique a todos los intervalos numéricos proporcionados en el presente documento.

En algunas realizaciones, los espaciadores de CRISPR se identifican por el análisis de la secuencia de los tramos de ADN localizados entre dos repeticiones.

Como se describe en el presente documento, la presente invención proporciona métodos, y los inventores describen composiciones que facilitan el uso de uno o más genes o proteínas cas en combinación con una o más, preferentemente dos o más repeticiones de CRISPR adecuadas para conferir especificidad de inmunidad a al menos un espaciador de CRISPR en una célula receptora. En algunas realizaciones preferidas, al menos unos genes o proteínas cas y al menos una repetición de CRISPR se usan en combinaciones funcionales para conferir especificidad de inmunidad a al menos un espaciador de CRISPR en una célula.

Como se usa en el presente documento, el término "especificidad de inmunidad" significa que se confiere inmunidad frente a una secuencia de ácido nucleico específica o producto de transcripción de la misma, utilizando un espaciador de CRISPR específico o secuencia espaciadora de pseudo-CRISPR. Como se ha indicado en el presente documento, un espaciador de CRISPR dado no confiere resistencia frente a cualquier secuencia de ácido nucleico o producto de transcripción de la misma pero solo para aquellas secuencias frente a las cuales el espaciador de CRISPR o el espaciador de pseudo-CRISPR es homólogo (por ejemplo, aquellas que son un 100 % idénticas).

En algunas realizaciones, el(los) espaciador(es) de CRISPR se obtienen de un organismo donante que es diferente de la célula receptora. En algunas realizaciones preferidas, las células donantes y receptoras son cepas bacterianas, especies, y/o géneros diferentes. En algunas realizaciones preferidas, los al menos unos genes o proteínas cas y/o al menos unas repeticiones de CRISPR se obtienen de un organismo diferente que el del organismo receptor. En algunas realizaciones preferidas, se transfieren al menos dos repeticiones de CRISPR. En algunas realizaciones preferidas, los espaciadores de CRISPR se obtienen a partir de un organismo que es heterólogo para el receptor o una célula donante adicional a partir de la cual se obtienen los al menos unos genes y/o proteínas cas y/o al menos una repetición CRISPR. En algunas realizaciones preferidas, los espaciadores de CRISPR se obtienen a partir de un organismo que es homólogo para el receptor o una célula donante adicional a partir de la cual se obtienen los al menos unos genes y/o proteínas cas y/o al menos una repetición CRISPR. En algunas realizaciones preferidas, el(los) espaciador(es) de CRISPR se diseña(n) y produce(n) utilizando métodos recombinantes conocidos en la técnica. De hecho, se pretende que los espaciadores de CRISPR se produzcan utilizando cualquier método adecuado conocido en la técnica.

En algunas realizaciones, los espaciadores de CRISPR son heterólogos para la célula receptora a partir de la cual se obtienen al menos unos genes o proteínas cas y/o la al menos una (en algunas realizaciones, preferentemente, dos o más) repetición o repeticiones de CRISPR. En algunas realizaciones alternativas, los espaciadores de CRISPR son homólogos a la célula receptora que tiene al menos unos genes o proteínas cas y/o se obtienen al menos una (en algunas realizaciones, preferentemente, dos o más) repetición o repeticiones de CRISPR. De hecho, se pretende que cualquiera de los elementos utilizados en los métodos sea heterólogo u homólogo. En algunas realizaciones, donde se usan múltiples elemento (por ejemplo, cualquier combinación de espaciador(es) de CRISPR, repetición(es) de CRISPR, gen(es) cas, y proteína(s) Cas, algunos elementos son homólogos entre sí y algunos elementos son heterólogos entre sí (por ejemplo, en algunas realizaciones, el(los) espaciador(es) de CRISPR y los genes cas son homólogos, pero la(s) repetición(es) CRISPR es/son heteróloga(s). De esta manera, en algunas realizaciones, el espaciador de CRISPR no está asociado naturalmente con la repetición de CRISPR y/o los genes cas y/o la combinación funcional de la repetición de CRISPR -gen cas. De hecho, se pretende que cualquier combinación de elementos heterólogos y homólogos sea de utilidad en la presente invención. En otras realizaciones adicionales, las células donantes y receptoras son heterólogas, mientras que en realizaciones adicionales, son homólogas. Se pretende también que los elementos contenidos en las células donantes y receptoras sean homólogos y/o heterólogos. Los elementos (por ejemplo, espaciadores de CRISPR) se introducen en ADN plásmido y/o genómico de la célula receptora utilizando cualquier método adecuado conocido en la técnica.

En algunas realizaciones preferidas, se usa al menos un espaciador de CRISPR para genomanipular una célula (por ejemplo, una célula receptora). En realizaciones adicionales, se usan los espaciadores de CRISPR como un molde con el que se modifica (por ejemplo, se muta) el ADN plásmido y/o genómico de una célula (por ejemplo, una célula receptora), de tal manera que se crean espaciadores de CRISPR en el ADN de la célula. En algunas realizaciones, el(los) espaciador(es) se clona(n) en al menos una construcción, plásmido u otro vector, con lo cual, la célula receptora se transforma a continuación, utilizando cualquier método adecuado conocido en la técnica. Los espaciadores de CRISPR están flanqueados por dos repeticiones de CRISPR (es decir, un espaciador de CRISPR tiene al menos una repetición de CRISPR en cada lado).

Aunque no se pretende que la presente invención esté limitada a ningún mecanismo concreto, teoría ni hipótesis, se contempla que, adicionalmente, un espaciador de CRISPR dado esté desde el extremo 5' del locus CRISPR comprendiendo el(los) gen(es) cas y/o la secuencia líder, cuanto menor sea la resistencia conferida por este espaciador CRISPR. De esta manera, en algunas realizaciones de la presente invención, uno o más de los primeros 100 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en otras realizaciones, uno o más de los primeros 50 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en realizaciones adicionales, uno o más de los primeros 40 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en realizaciones más adicionales, uno o más de los primeros 30 espaciadores de CRISPR desde el extremo 5' del locus CRISPR, en realizaciones más adicionales, uno o más de los primeros 20 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en realizaciones más adicionales, uno o más de los primeros 15 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, y en algunas realizaciones preferidas, uno o más de los primeros 15 espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, y en algunas realizaciones preferidas, uno o más de los 10 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados. Como se ha indicado en el presente documento, diferentes bacterias tienen diferentes números de espaciadores de CRISPR, de esta manera, en algunas realizaciones, están modificados diversos espaciadores.

F. Núcleo del espaciador de CRISPR

En algunas realizaciones, para un tipo de CRISPR específico comprendido en una especie microbiana, el espaciador CRISPR está representado por una longitud definida predominante, aunque puede variar el tamaño. Se ha encontrado que los tipos de CRISPR descritos hasta la fecha contienen un espaciador con una longitud predominante de entre aproximadamente 20 pb y aproximadamente 58 pb.

Como se usa en el presente documento, "núcleo del espaciador CRISPR" se refiere a la longitud del espaciador más corto observado comprendido en un tipo de CRISPR. De esta manera, por ejemplo, dentro del CRISPR1 de *S. thermophilus*, la longitud predominante del espaciador es de 30 pb, con una minoría de espaciadores entre 28 pb y 32 pb. De esta manera, en el CRISPR1 de *S. thermophilus*, el núcleo del espaciador de CRISPR se define como un tramo continuo de 28 pb.

En algunas realizaciones preferidas de la presente invención, el núcleo del espaciador de CRISPR es homólogo al ácido nucleico diana, un producto de la transcripción del mismo, o una secuencia identificada sobre la longitud de la secuencia del núcleo. Aunque la homología se toma en consideración también en términos de similitud, en algunas realizaciones preferidas, la homología se expresa en términos de identidad de secuencias. De esta manera, en algunas realizaciones, una secuencia homóloga abarca un núcleo del espaciador CRISPR, que puede ser al menos aproximadamente un 90 % idéntico; o al menos aproximadamente 91, aproximadamente 92, aproximadamente 93, aproximadamente 94, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98 o aproximadamente 99 % idéntica a la secuencia de ácido nucleico diana, un producto de la transcripción de la misma, una secuencia identificada sobre la longitud de la secuencia del núcleo. En algunas realizaciones particularmente

preferidas, el núcleo del espaciador de CRISPR es idéntico en un 100 % al ácido nucleico diana, un producto de la transcripción del mismo, o una secuencia identificada sobre la longitud de la secuencia del núcleo.

Durante el desarrollo de la presente invención, se analizaron las secuencias de CRISPR de varias cepas de *S. thermophilus*, incluyendo las cepas industriales estrechamente relacionadas y las variantes resistentes a fagos. Se observaron diferencias en el número y tipo de espaciadores principalmente en el locus CRISPR1. De forma notable, la sensibilidad del fago parecía estar correlacionada con el contenido del espaciador de CRISPR1. De forma específica, el contenido del espaciador era casi idéntico entre cepas progenitoras y derivados resistentes a fagos, excepto para los espaciadores adicionales presentes en el último. Estos hallazgos sugieren una relación potencial entre la presencia de espaciadores adicionales y las diferencias observadas en la sensibilidad al fago de una cepa dada. Esta observación impulsó la investigación del origen y función de los espaciadores adicionales presentes en los mutantes resistentes a fagos.

G. Espaciador de pseudo-CRISPR

Como se usa en el presente documento, las expresiones "espaciador de pseudo-CRISPR", "proespaciador", y "protoespaciador" se refieren a una secuencia de ácido nucleico presente en un organismo (por ejemplo, un organismo donante, incluyendo, pero no de forma limitativa, un bacteriófago), que es preferentemente esencial para funcionar y/o para la supervivencia y/o para la replicación y/o para la infectividad, y que forma una secuencia del espaciador CRISPR. En algunas realizaciones, los espaciadores pseudo-CRISPR son de utilidad en la producción de secuencias de espaciadores de CRISPR que son complementarias a u homólogas del espaciador de pseudo-CRISPR.

En algunas realizaciones, al menos varios espaciadores de pseudo-CRISPR y espaciador(es) de CRISPR que es/son complementarios u homólogos de al menos un(os) espaciador(es) de pseudo-CRISPR se utilizan para genomanipular una célula receptora. En algunas realizaciones, los espaciadores pseudo-CRISPR o espaciador(es) de CRISPR que es/son complementarios u homólogos de uno o más espaciador(es) de pseudo-CRISPR se insertan en el ADN plásmido y/o genómico de una célula receptora utilizando cualquier método adecuado conocido en la técnica.

En algunas realizaciones adicionales, se usan los espaciadores de pseudo-CRISPR como un molde mediante el cual se modifica (por ejemplo, se muta) el ADN plásmido y/o genómico de una célula receptora (por ejemplo, una célula receptora), de tal manera que se crean espaciadores de CRISPR en el ADN plásmido y/o genómico de la célula. En algunas realizaciones adicionales, los espaciadores de pseudo-CRISPR o espaciador(es) de CRISPR que es/son complementarios u homólogos de uno o más espaciador(es) de pseudo-CRISPR se clonan en una construcción, plásmido y/o vector, etc., se introduce(n) en la célula hospedadora utilizando cualquier método adecuado conocido en la técnica.

H. Proteínas Cas y genes cas

Como se usa en el presente documento, el término "gen cas" tiene el significado convencional que se usa en la técnica y se refiere a uno o más genes cas que se acoplan, se asocian o se acercan a o están en proximidad a los loci de CRISPR flanqueantes.

Una revisión comprehensiva de la familia de la proteína Cas se presenta en Haft *et al.* (Haft *et al.*, Comput. Biol., 1, 6 e60 [2005]). Como se describe en la anterior, se describen 41 familias de genes (cas) asociados a CRISPR, además de cuatro familias de genes anteriormente conocidos. Como se ha indicado, los sistemas CRISPR pertenecen a diferentes clases, con diferentes modelos de repetición, conjuntos de genes, e intervalos de especies. De hecho, el número de genes cas en un locus CRISPR dado puede variar entre especies.

En algunas realizaciones, uno o más de los genes y/o proteínas cas se producen naturalmente en una célula receptora y/o uno o más espaciadores heterólogos se integra(n) o inserta(n) en los loci CRISPR adyacentes en uno o más de los genes o proteínas cas.

En algunas realizaciones, uno o más de los genes y/o proteínas cas es(son) heterólogos para la célula receptora y uno o más de los espaciadores es/son homólogos o heterólogos. En algunas realizaciones preferidas, los espaciadores se integran o insertan en los loci CRISPR adyacentes a uno o más genes o proteínas cas.

Los loci CRISPR se encuentran normalmente cerca de cuatro genes denominados *cas 1* a *cas 4*. La disposición más común de estos genes es *cas3-cas-cas1-cas2*. La proteína Cas3 parece ser una helicasa, mientras que Cas4 se asemeja a la familia RecB de exonucleasa y contiene un motivo rico en cisteína, que sugiere la unión del ADN. Cas 1 es generalmente muy básica y es la única proteína Cas que se encuentra consistentemente en todas las especies que contienen loci CRISPR. Queda por caracterizar Cas2. *cas1-4* se caracterizan normalmente por su estrecha proximidad a los loci CRISPR y su amplia distribución a través de las especies bacterianas y especies de archaea. Aunque no todos los genes *cas1-4* se asocian con todos los loci CRISPR, se encuentran todos en múltiples subtipos.

Además, existe otra agrupación de tres genes asociados con las estructuras CRISPR en muchas especies bacterianas, denominados aquí *cas1B*, *cas5* and *cas6* (Véase, Bolotin *et al.*, [2005], *supra*). En algunas realizaciones, el gen *cas* se selecciona entre *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* y/o *cas6*. En algunas realizaciones, el gen *cas* es *cas1*. En otras realizaciones más, el gen *cas* se selecciona entre fragmentos, variantes, homólogos y/o derivados de los mismos de *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* y/o *cas6*. En algunas realizaciones adicionales, es de utilidad una combinación de dos o más genes *cas*, cualquier combinación es adecuada. Se señala que la nomenclatura de los genes *cas* está en proceso de cambio. De esta manera, el texto en el presente documento debe tomarse en contexto.

El término "proteína Cas" abarca también una pluralidad de proteínas Cas (por ejemplo, entre aproximadamente 2 y aproximadamente 12 proteínas Cas, más preferentemente, entre aproximadamente 3 y aproximadamente 11 proteínas Cas, más preferentemente, entre aproximadamente 4 y aproximadamente 10 proteínas Cas, más preferentemente, entre aproximadamente 4 y aproximadamente 9 proteínas Cas, entre aproximadamente 4 y aproximadamente 8 proteínas Cas, y más preferentemente, entre aproximadamente 4 y aproximadamente 7 genes de proteínas; tal como 4, 5, 6, o 7 proteínas Cas).

En algunas realizaciones, las proteínas Cas están codificadas por genes *cas* que comprenden el ADN, mientras que en otras realizaciones, el *cas* comprende ARN. En algunas realizaciones, el ácido nucleico es de origen genómico, mientras que en otras realizaciones, es de origen sintético o recombinante. En algunas realizaciones, los genes *cas* que codifican las proteínas Cas son bicatenarios representando tanto la hebra de sentido directo como la hebra de sentido contrario bicatenarias o monocatenarias o las combinaciones de las mismas. En algunas realizaciones, los genes *cas* se preparan mediante el uso de técnicas de ADN recombinante (por ejemplo ADN recombinante), como se describe en el presente documento.

I. Bacteriófago

Como se usa en el presente documento, el término "bacteriófago" (o "fago") tiene su significado convencional como se entiende en la técnica (es decir, un virus que infecta selectivamente una o más especies bacterianas). Muchos bacteriófagos son específicos de un género o especie o cepa concreta de bacterias. En algunas realizaciones preferidas, los fagos son capaces de infectar bacterias progenitoras y/o células hospedadoras. En algunas realizaciones, los bacteriófagos son virulentos para la bacteria progenitora. En algunas realizaciones, los fagos son líticos, mientras que en otras realizaciones, los fagos son lisógenos.

Un fago lítico es aquel que permite la ruta lítica mediante el agotamiento del ciclo lítico, en lugar de la introducción en la ruta lisógena. Un bacteriófago lítico experimenta la replicación vírica que conduce a la lisis de la membrana celular, la destrucción de la célula, y la liberación de partículas de la progenie de bacteriófago capaz de infectar otras células.

Un bacteriófago lisógeno es aquel que es capaz de iniciar la ruta lisógena, en la que el bacteriófago se vuelve una parte pasiva durmiente del genoma de la célula antes de completar su ciclo lítico.

Los bacteriófagos de utilidad en la presente invención incluyen, pero no de forma limitativa a bacteriófagos que pertenecen a cualquiera de las siguientes familias de virus: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, o Tectiviridae. En algunas realizaciones, los bacteriófagos que infectan bacterias son patógenos para las plantas y/o los animales (incluyendo seres humanos).

En algunas realizaciones particularmente preferidas, los bacteriófagos de la presente invención incluyen, pero no de forma limitativa, aquellos bacteriófagos capaces de infectar una bacteria que comprenden naturalmente uno o más loci CRISPR. Se han identificado loci CRISPR en más de 40 procariotas (Véase, por ejemplo, Jansen *et al.*, *Mol. Microbiol.*, 43:1565-1575 [2002]; yd Mojica *et al.*, [2005]) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema*, y *Thermotoga*.

En algunas realizaciones, los bacteriófagos incluyen, pero no de forma limitativa, aquellos bacteriófagos capaces de infectar bacterias que pertenecen a los siguientes géneros: *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella* y *Xanthomonas*.

En realizaciones más adicionales, los bacteriófagos incluyen, pero no de forma limitativa, aquellos bacteriófagos capaces de infectar (o de transducir) bacterias acidolácticas, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (por ejemplo, *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc*, y

Oenacoccus.

En otras realizaciones adicionales, los bacteriófagos incluyen, pero no de forma limitativa, aquellos bacteriófagos capaces de infectar *Lactococcus lactis* (por ejemplo, *L. lactis* subsp. *lactis* y *L. lactis* subsp. *cremoris*, y *L. lactis* subsp. *lactis* biovar *diacetylactis*), *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*, *Bifidobacterium infantis*, *Lactobacillus paracasei*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus gasseri*, *Lactobacillus johnsonii* o *Bifidobacterium longum*.

En algunas realizaciones particularmente preferidas, los bacteriófagos incluyen, pero no de forma limitativa, aquellos bacteriófagos capaces de infectar bacterias que comprenden uno o más loci CRISPR heterólogos. En algunas realizaciones, las bacterias comprenden uno o más loci CRISPR heterólogos, y/o uno o más genes cas heterólogos, y/o una o más repeticiones de CRISPR heterólogos, y/o uno o más espaciadores de CRISPR heterólogos.

La infección de bacterias por fagos es el resultado de la inyección o transferencia de ADN de fagos en células. En algunas realizaciones, la infección da como resultado la expresión (es decir, la transcripción y traducción) del ácido nucleico del bacteriófago en el interior de la célula y la continuación del ciclo vital del bacteriófago. En algunas realizaciones que implican el bacteriófago recombinante, se expresan también las secuencias recombinantes en el genoma del fago (por ejemplo, los ácidos nucleicos indicadores)

Se ha descubierto que las secuencias espaciadoras de CRISPR en procariotas tienen a menudo similitudes significativas con una variedad de moléculas de ADN incluyendo elementos genéticos tales como cromosomas, bacteriófagos y plásmidos conjugativos). Se ha notificado que las células que transportan estos espaciadores de CRISPR no se pueden infectar por moléculas de ADN que contienen secuencias homólogas a los espaciadores (Véase, Mojica *et al.*, [2005]).

En algunas realizaciones de la presente invención, las bacterias progenitoras se exponen (por ejemplo, iterativamente, secuencialmente, simultáneamente o de forma sustancialmente simultánea) a más de un bacteriófago. En algunas realizaciones preferidas, las bacterias se exponen a mezclas de uno o más (por ejemplo, algunos) fagos diferentes. En algunas realizaciones preferidas alternativas, las bacterias progenitoras son sensibles a cada uno de los bacteriófagos a los cuales se exponen.

En algunas realizaciones, cada una de las secuencias marcadoras de cada uno de los bacteriófagos y/o cada una de las secuencias duplicadas (por ejemplo, la repetición de CRISPR duplicada) de la bacteria progenitora se integra en el mismo locus CRISPR. En otras realizaciones, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran en uno o ambos extremos del mismo locus CRISPR. En otras realizaciones adicionales, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran en el extremo 5' y/o el extremo 3' del mismo locus CRISPR. En algunas realizaciones preferidas, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integra en el extremo 5' del mismo locus CRISPR.

En algunas realizaciones, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas procedentes de la bacteria progenitora se integra iterativamente, simultáneamente o de forma sustancialmente simultánea. En realizaciones en las que cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran secuencialmente, la primera secuencia marcadora y/o la primera secuencia duplicada se integra en la bacteria progenitora. Una segunda secuencia marcadora procedente de un segundo bacteriófago y/u otra secuencia duplicada se integran a continuación en la bacteria progenitora. En algunas realizaciones preferidas, la secuencia marcadora y/o la secuencia duplicada se integran en el extremo 5' del mismo locus CRISPR.

En algunas realizaciones preferidas, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integra en un extremo (por ejemplo, el extremo 5') del mismo locus CRISPR adyacente (es decir, próximo a) entre sí. De esta manera, en algunas realizaciones, cada una de las secuencias marcadoras y/o las secuencias duplicadas se integra secuencialmente, por lo cual, las primeras secuencias se integran en la bacteria progenitora en un extremo (por ejemplo, comprendido en o en el extremo 5' y/o el extremo 3') En algunas realizaciones preferidas, una segunda secuencia marcadora y/o una secuencia duplicada se integran a continuación en la bacteria progenitora adyacente (por ejemplo, directamente adyacente) a la primera pareja de secuencias. En algunas realizaciones, las segundas secuencias se integran en la bacteria progenitora adyacente (por ejemplo, directamente adyacentes) al extremo 5' o al extremo 3' de las primeras secuencias. En algunas realizaciones preferidas, las segundas secuencias se integran en la bacteria progenitora adyacente (por ejemplo, directamente adyacentes) al extremo 3' de las primeras secuencias. En realizaciones en las que se proporcionan secuencias adicionales, estas se integran a continuación.

En algunas realizaciones, cada una de las secuencias se integran adyacentes (es decir, próximas) entre sí comprendidas dentro o en el extremo 3' y/o en el extremo 5' del mismo locus CRISPR de la bacteria progenitora. En algunas realizaciones preferidas, cada una de las secuencias se integra adyacentes (es decir, próximas) entre sí en el extremo 5' del mismo locus CRISPR de la bacteria progenitora. En algunas realizaciones particularmente

preferidas, cada una de las secuencias se integran adyacentes (es decir, próximas a) entre sí en la dirección 5' del extremo 5' del locus CRISPR de la bacteria progenitora. En algunas realizaciones alternativamente preferidas, cada una de las secuencias se integran adyacentes (es decir, próximas a) entre sí, en una localización que está en la repetición 5' de CRISPR del locus CRISPR de la bacteria progenitora. En algunas realizaciones particularmente preferidas, cada una de las secuencias se integran adyacentes (es decir, próximas a) entre sí en la dirección 5' de la primera repetición 5' de CRISPR del locus CRISPR de la bacteria progenitora.

J. Bacteria progenitora

Como se usa en el presente documento, las expresiones "bacteria progenitora", "bacterias progenitoras" y "cepa progenitora" se refiere a cualquier bacteria/bacterias/cepas que se expone(n) a uno o más bacteriófagos virulentos. En algunas realizaciones particularmente preferidas, las bacterias progenitoras son sensibles al fago virulento. En algunas realizaciones preferidas, la cepa progenitora se infecta por el bacteriófago. En algunas realizaciones particularmente preferidas, la infección por el fago convierte a la bacteria/bacterias/cepa progenitoras o a una subpoblación de las mismas en insensible a una infección adicional por el bacteriófago. En algunas realizaciones preferidas, la infección de una "bacteria progenitora" por uno o más bacteriófagos da como resultado la creación de una cepa marcada que se puede seleccionar basándose en su insensibilidad al bacteriófago. En algunas realizaciones preferidas, "mutantes resistentes a bacteriófagos" son bacterias que están marcadas o etiquetadas de acuerdo con los métodos de la presente invención. En algunas realizaciones, las bacterias progenitoras son cepas bacterianas naturales. En algunas realizaciones preferidas, las bacterias progenitoras son cepas naturales de bacterias que no se han infectado previamente con ningún bacteriófago. En algunas realizaciones preferidas, las bacterias progenitoras son cepas naturales de bacterias que no se han marcado previamente, mientras que en algunas realizaciones alternativas, las bacterias progenitoras son mutantes resistentes a bacteriófagos que se han marcado previamente.

En algunas realizaciones particularmente preferidas, la bacteria progenitora se selecciona entre cualquier bacteria que comprende naturalmente uno o más loci CRISPR. Se han identificado loci CRISPR en más de 40 procariotas (Jansen *et al.* [2002] *supra*, Mojica *et al.*, [2005], *supra*; y Haft *et al.*, [2005], *supra*) incluyendo pero sin limitación, *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halococcus*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* y *Thermotoga*.

En algunas realizaciones, la bacteria progenitora comprende uno o más espaciadores de CRISPR heterólogos, una o más repeticiones de CRISPR heterólogas, y/u uno o más genes *cas* heterólogos. En algunas realizaciones alternativas, la bacteria progenitora comprende uno o más loci CRISPR heterólogos, preferentemente, uno o más loci CRISPR completos. En algunas realizaciones adicionales la bacteria progenitora comprenden uno o más loci CRISPR y comprende también una o más espaciadores de CRISPR heterólogos, una o más repeticiones de CRISPR heterólogas, y/o uno o más genes *cas* heterólogos. En algunas realizaciones adicionales, la bacteria progenitora comprende naturalmente uno o más loci CRISPR y comprende también uno o más loci CRISPR heterólogos, preferentemente, uno o más loci CRISPR completos.

En algunas realizaciones preferidas, la subpoblación resistente a fago creada mediante la exposición de las bacterias progenitoras a al menos un fago es un cultivo puro. Sin embargo, no se pretende que la presente invención esté limitada a cultivos puros de cepas bacterianas, variantes, o fagos. De hecho, se pretende que la presente invención abarque cultivos mixtos de células y fagos. En algunas realizaciones, el cultivo mixto es una mezcla de diferentes mutantes que corresponden a diferentes episodios de integración en el mismo y/o diferentes loci CRISPR.

Aunque no se pretende que la presente invención esté limitada, los géneros bacterianos progenitores preferidos son *Streptococcus* y *Lactobacillus*. De hecho, se pretende que cualquier especie bacteriana será de utilidad en la presente invención, incluyendo, pero no de forma limitativa *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Corynebacterium*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Enterococcus*, *Pediococcus*, *Leuconostoc*, *Oenococcus*, y/o *Xanthomonas*. En algunas realizaciones, las bacterias progenitoras son o se derivan de bacterias acidolácticas, incluyendo, pero no de forma limitativa *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (por ejemplo, *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc*, y/o *Oenococcus*. En realizaciones adicionales, las bacterias progenitoras son o se derivan de *Lactococcus lactis* (por ejemplo, *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris*, y *L. lactis* subsp. *lactis* biovar *diacetylactis*), *L. delbrueckii* subsp. *bulgaricus*, *L. helveticus*, *L. acidophilus*, *L. casei*, *L. paracasei*, *L. salivarius*, *L. plantarum*, *L. reuteri*, *L. gasseri*, *L. johnsonii*, *Bifidobacterium lactis*, *B. infantis*, *B. longum*, y/o *Streptococcus thermophilus*.

En la presente invención, la bacteria progenitora es una "bacteria de calidad alimentaria" (es decir, una bacteria que se usa y por lo general se considera segura para el uso en la preparación y/o producción de alimentos y/o alimentación). En algunas realizaciones preferidas, la bacteria progenitora es adecuada para su uso como cultivo

iniciador, un cultivo probiótico y/o un suplemento de la dieta. En realizaciones adicionales, la bacteria progenitora es de utilidad en la fermentación de la carne (por ejemplo, (por ejemplo, ternera, cerdo, cordero, y aves de corral) incluyendo, pero no de forma limitativa, bacterias acidolácticas, *Pediococcus cerevisiae*, *Lactobacillus plantarum*, *L. brevis*, *L. sakei*, *L. curvatus*, *Micrococcus species*, *Pediococcus pentosaceus*, *Staphylococcus xylosus*, *S. vitulinus* y mezclas de las mismas (véase por ejemplo, Knorr (ed.), Food Biotechnology, en 538-39 [1987]; y Pederson, Microbiology of Fermented Foods, en 210-34, 2d ed., [1379]; y la patente de Estados Unidos N.º 2.225.783.

En realizaciones más adicionales, la bacteria progenitora encuentra uso en la fermentación de vegetales (por ejemplo, zanahorias, pepinos, tomates, pimientos, y col) incluyendo, pero no de forma limitativa, *L. plantatum*, *L. brevis*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, y las mezclas de las mismas (Véase, por ejemplo, Knorr, *supra*; Pederson, *supra*; y la patente de Estados Unidos números 3.024.116, 3.403.032, 3.932. 674, y 3.897.307). En realizaciones más adicionales, la bacteria progenitora es de utilidad en la fermentación de la masa formada a partir de cereales (por ejemplo, trigo, centeno, arroz, avena, cebada, y maíz). En otras realizaciones más, la bacteria progenitora es de utilidad en la producción de vino mediante la fermentación del zumo de fruta (por ejemplo, zumo de uva). En algunas realizaciones adicionales, la bacteria progenitora es de utilidad en la fermentación de leche (por ejemplo, *L. delbrueckii subsp. bulgaricus*, *L. acidophilus*, *S. thermophilus*, y las mezclas de los mismos (véase, Knorr, *supra*; y Pederson *supra*, en las páginas 105-35). En algunas realizaciones preferidas, la bacteria progenitora es de utilidad en la producción de queso, incluyendo, pero no de forma limitativa *L. delbrueckii subsp. bulgaricus*, *L. helveticus*, *L. lactis subsp. lactis*, *L. lactis subsp. cremoris*, *L. lactis subsp. lactis biovar diacetylactis*, *S. thermophilus*, *Bifidobacterium Enterococcus*, etc., y las mezclas de los mismos (Véanse, por ejemplo, Knorr, *supra*, y Pederson, *supra*, en 135-51). En realizaciones más adicionales, la bacteria progenitora es de utilidad en la fermentación de los huevos, incluyendo, pero no de forma limitativa *Pediococcus pentosaceus*, *Lactobacillus plantarum*, y las mezclas de los mismos (Véase, Knorr, *supra*). En algunas realizaciones, la bacteria progenitora es de utilidad en la fermentación para producir diversos productos, incluyendo, pero no de forma limitativa los quesos cheddar y cottage (por ejemplo, *L. lactis subsp. lactis*, *L. lactis subsp. cremoris*), yogur (*L. delbrueckii subsp. bulgaricus*, y *S. thermophilus*), queso suizo (por ejemplo, *S. thermophilus*, *L. lactis*, y *L. helveticus*), queso azul (*Leuconostoc cremoris*), queso italiano (*L. bulgaricus* and *S. thermophilus*), viili (*L. lactis subsp. cremoris*, *L. lactis subsp. lactis biovar diacetylactis*, *Leuconostoc cremoris*), yakult (*L. casei*), caseína (*L. lactis subsp. cremoris*), natto (*Bacillus subtilis var. natto*), vino (*Leuconostoc oenos*), sake (*Leuconostoc mesenteroides*), polimixina (*Bacillus polymyxa*), colistina (*Bacillus colistrium*), bacitracina (*Bacillus licheniformis*), ácido L-glutámico (*Brevibacterium lactofermentum* y *Microbacterium ammoniaphilum*), y acetona y butanol (*Clostridium acetobutyricum*, y *Clostridium saccharoperbutylacetonicum*). En algunas realizaciones preferidas, las especies bacterianas progenitoras se seleccionan entre *S. thermophilus*, *L. delbrueckii subsp. bulgaricus* y/o *L. acidophilus*.

En otras realizaciones adicionales, las bacterias progenitoras son de utilidad en métodos que incluyen, pero no de forma limitativa la producción de antibióticos, la producción de aminoácidos, la producción de disolventes, y la producción de otros materiales económicamente útiles. En otras realizaciones más, las bacterias progenitoras son de utilidad en composiciones cosméticas, terapéuticas, y/o farmacéuticas. En algunas realizaciones, las composiciones tienen actividades concretas, incluyendo, pero no de forma limitativa, la regeneración de la piel, incluyendo, pero no de forma limitativa, las propiedades antiarrugas, borrar las cicatrices antiguas, reparar los tejidos dañados por quemaduras, promover la cicatrización de la piel, eliminar las manchas pigmentarias, etc. En algunas realizaciones, las composiciones o bien promueven, o bien inhiben el crecimiento de las uñas, el cabello o los cabellos. En algunas realizaciones adicionales, las composiciones comprenden al menos un cultivo microbiano y/o una bacteria marcada y/o un cultivo celular producido usando los métodos y composiciones de la presente invención

En realizaciones adicionales, las bacterias progenitoras son mutantes insensibles a bacteriófagos. De esta manera, en algunas realizaciones, las bacterias progenitoras son insensibles a uno o más bacteriófagos. En algunas realizaciones preferidas, la bacteria progenitora no es un mutante insensible a bacteriófagos para el bacteriófago al que se va a exponer durante el uso de la presente invención.

K. Secuencia marcadora

Como se usa en el siguiente documento, el término "secuencia marcadora" se refiere a una parte de una "unidad espaciadora de repetición adicional" que se deriva, es derivable, se obtiene o es obtenible a partir del genoma de uno o más bacteriófago(s) que la bacteria progenitora se expone a o está de acuerdo con los métodos de la presente invención y se usa como una marca o una etiqueta (por ejemplo, una única marca o una única etiqueta). En algunas realizaciones preferidas, la secuencia marcadora tiene al menos aproximadamente 20 nucleótidos de longitud, mientras que en algunas realizaciones más preferidas, la secuencia marcadora tiene aproximadamente 20 a aproximadamente 58 nucleótidos de longitud. Sin embargo, en algunas realizaciones alternativas, se genera una "marca" utilizando elementos genéticos procedentes de fuentes diferentes a fagos. Por ejemplo, en algunas realizaciones, la marca se proporciona por plásmidos, elementos que se pueden transponer, ácido nucleico aislado, etc. De hecho, en algunas realizaciones, la marca es una única secuencia sintética no funcional. De esta manera, no se pretende que la presente invención este limitada a marcas de ácidos nucleicos que se generan únicamente a partir de ácidos nucleicos de fagos.

La secuencia marcadora es normalmente una secuencia que se produce naturalmente en el bacteriófago. Preferentemente, la secuencia marcada tiene al menos aproximadamente un 90 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, o aproximadamente 99 % de identidad con la secuencia que se produce naturalmente en el bacteriófago (por ejemplo, el genoma del bacteriófago a partir del cual se deriva, es derivable, se obtiene o es obtenible). En algunas realizaciones ligeramente preferidas, la secuencia marcadora tiene un 100 % de identidad con la secuencia que se produce naturalmente en el bacteriófago (por ejemplo, el genoma del bacteriófago a partir del cual se deriva, es derivable, se obtiene o es obtenible).

En algunas realizaciones, la secuencia marcadora tiene menos de aproximadamente un 40 %, aproximadamente 30 %, aproximadamente 20 %, aproximadamente 10 %, aproximadamente 5 %, aproximadamente 4 %, aproximadamente 3 %, aproximadamente 2 %, aproximadamente 1 %, o aproximadamente 0 % de identidad con cualesquiera espaciadores de CRISPR diferente o núcleos de espaciadores de CRISPR en el uno o más loci CRISPR de la bacteria marcada. En algunas realizaciones, la secuencia marcadora tiene menos de aproximadamente un 40 %, aproximadamente 30 %, aproximadamente 20 %, aproximadamente 10 %, aproximadamente 5 %, aproximadamente 4 %, aproximadamente 3 %, aproximadamente 2 %, aproximadamente 1 % o aproximadamente 0 % de identidad con cualquier otra secuencia en el uno o más loci de CRISPR de la bacteria marcada.

En algunas realizaciones alternativas, la secuencia marcadora es una secuencia que es idéntica a una secuencia (por ejemplo, un espaciador de CRISPR) en el locus CRISPR de la bacteria. En algunas realizaciones alternativas, la secuencia marcadora es una secuencia que es casi idéntica a una secuencia (por ejemplo un espaciador de CRISPR) en el locus CRISPR de la bacteria en que este contiene uno o más polimorfismos de nucleótidos únicos (por ejemplo, uno o dos polimorfismos de nucleótidos únicos).

En algunas realizaciones, al menos una secuencia marcadora se integra en la bacteria progenitora. En algunas realizaciones alternativas, se integra también al menos una secuencia duplicada (por ejemplo, una secuencia de repetición de CRISPR duplicada) que se deriva, es derivable, se obtiene o es obtenible a partir del genoma de la bacteria progenitora o uno o más de los plásmidos de la bacteria progenitora (por ejemplo, megaplásmidos). En algunas realizaciones particularmente preferidas, al menos una secuencia duplicada se copia o replica procedente del genoma de la bacteria progenitora. En algunas realizaciones, la secuencia de repetición de CRISPR en un locus CRISPR se duplica y la secuencia marcadora se integra en el genoma de la bacteria inmediatamente después (por ejemplo, en la dirección 3') de la nueva repetición de CRISPR duplicada. Sin embargo, no se pretende que la presente invención esté limitada a cualquier mecanismo específico o teoría de la acción.

En algunas realizaciones ligeramente preferidas, la al menos una secuencia duplicada es una secuencia de repetición de CRISPR que tiene al menos aproximadamente un 90 %, aproximadamente 95 %, aproximadamente 96 %, aproximadamente 97 %, aproximadamente 98 %, o aproximadamente 99 % de identidad con las repeticiones de CRISPR en el uno o más loci CRISPR de la bacteria progenitora y/o la bacteria marcada. En algunas realizaciones particularmente preferidas, la al menos una secuencia duplicada es una secuencia de repetición de CRISPR que tiene al menos aproximadamente un 100 % de identidad con las repeticiones de CRISPR en el uno o más loci CRISPR de la bacteria progenitora y/o la bacteria marcada. En algunas realizaciones, la secuencia marcadora tiene al menos aproximadamente 24 nucleótidos de longitud, mientras que en algunas realizaciones preferidas, la secuencia duplicada tiene entre aproximadamente 24 a aproximadamente 40 nucleótidos de longitud.

En algunas realizaciones preferidas, la al menos una secuencia marcadora y la al menos una secuencia duplicada están integradas en la bacteria progenitora. En algunas realizaciones, cada vez se integra una secuencia marcadora en el genoma de la bacteria progenitora, se acompaña por la integración iterativa, secuencial, simultánea o sustancialmente simultánea de al menos una secuencia duplicada. De acuerdo con ello, una pareja de secuencias que comprenden la secuencia marcadora y la secuencia duplicada se integran en la bacteria progenitora, resultando por tanto una bacteria marcada. Sin embargo, no se pretende que la presente invención esté limitada por cualquier mecanismo, o medios de acción específicos.

En algunas realizaciones preferidas, la al menos una secuencia marcadora y la al menos una secuencia duplicada se integran adyacentes entre sí. En algunas realizaciones particularmente preferidas, la al menos una secuencia marcadora y la al menos una secuencia duplicada se integran directamente adyacentes entre sí de tal manera que no existen nucleótidos intervinientes entre las secuencias.

En algunas realizaciones, la secuencia duplicada se une, enlaza o fusiona con un extremo (por ejemplo, el extremo 5' o el extremo 3') de la secuencia marcadora. En algunas realizaciones preferidas, la secuencia duplicada se une, enlaza o fusiona con el extremo 5' de la secuencia marcadora. En algunas realizaciones particularmente preferidas, la fusión de una secuencia duplicada con una secuencia marcadora forma una unidad espaciadora de repetición de CRISPR. De acuerdo con ello, en algunas realizaciones, después de la integración de una unidad espaciadora de repetición de CRISPR, la secuencia duplicada es la primera secuencia en el extremo 5' del locus CRISPR y la secuencia marcadora es la segunda secuencia (es decir, la siguiente) del locus CRISPR, después de la secuencia duplicada. En otras realizaciones preferidas, la secuencia dentro de una unidad espaciadora de repetición de

CRISPR está directamente unida, directamente enlazada o directamente fusionada de tal forma que no hay nucleótidos intercalados entre la secuencia duplicada y la secuencia marcadora.

En algunas realizaciones especialmente preferidas, una unidad espaciadora de repetición de CRISPR se integra en el genoma de la bacteria progenitora para producir una bacteria marcada. En algunas realizaciones preferidas, la secuencia duplicada se deriva, es derivable, se obtiene o es obtenible a partir del genoma de la bacteria progenitora. En algunas realizaciones adicionales, la secuencia marcadora se deriva, es derivable, se obtiene o es obtenible a partir del genoma del bacteriófago que se utiliza para infectar la bacteria progenitora.

En algunas realizaciones adicionales, las múltiples unidades espaciadoras de repetición CRISPR están integradas en el genoma de la bacteria progenitora. En algunas realizaciones, las múltiples unidades espaciadoras de repetición de CRISPR comprenden una primera unidad espaciadora de repetición de CRISPR que comprende una secuencia duplicada y una secuencia marcadora y una segunda unidad espaciadora de repetición de CRISPR que comprende una segunda secuencia duplicada y una segunda secuencia marcadora. En algunas realizaciones preferidas, la segunda secuencia duplicada tiene típicamente la misma secuencia (por ejemplo, más de aproximadamente un 95 %, aproximadamente un 96 %, aproximadamente un 97 %, aproximadamente un 98 %, aproximadamente un 99 %, o aproximadamente un 100 % de identidad) con la primera secuencia duplicada. En algunas realizaciones, la secuencia marcadora tiene típicamente una secuencia diferente (por ejemplo, menos de aproximadamente un 40 %, aproximadamente un 30 %, aproximadamente un 20 %, aproximadamente un 10 %, aproximadamente un 5 %, aproximadamente un 4 %, aproximadamente un 3 %, aproximadamente un 2 %, aproximadamente un 1 % o aproximadamente un 0 % de identidad) con la primera secuencia marcadora. Este también es el caso de las realizaciones que contienen otras secuencias integradas de unidades espaciadoras de repetición CRISPR.

En algunas realizaciones preferidas, la configuración de las múltiples unidades espaciadoras de repetición CRISPR es típicamente:

[secuencia duplicada - secuencia marcadora]_n

en la que n = 2, 3, 4, 5, o ≥6.

En algunas realizaciones especialmente preferidas, la configuración de las múltiples unidades espaciadoras de repetición CRISPR es típicamente:

[repetición CRISPR - secuencia marcadora]_n

en la que n = 2, 3, 4, 5, o ≥6.

En algunas preferidas, la configuración de las múltiples unidades espaciadoras de repetición CRISPR es:

5'-[secuencia duplicada - secuencia marcadora]_n-3'

en la que n = 2, 3, 4, 5, o ≥6.

En algunas realizaciones especialmente preferidas, la configuración de las múltiples unidades espaciadoras de repetición CRISPR es:

5'-[repetición CRISPR - secuencia marcadora]_n

en la que n = 2,3,4, 5, o ≥6.

En algunas realizaciones preferidas, las múltiples unidades espaciadoras de repetición CRISPR están integradas en la bacteria progenitora.

En algunas realizaciones, la parte de la secuencia marcadora de la unidad espaciadora de repetición CRISPR está integrada adyacente a: (i) una secuencia duplicada que sea homóloga (por ejemplo, idéntica) a una secuencia natural en la bacteria progenitora; (ii) una secuencia duplicada que sea homóloga (por ejemplo, idéntica) a una secuencia natural del locus CRISPR de la bacteria progenitora; o (iii) lo más preferiblemente, una secuencia duplicada que sea homóloga (por ejemplo, idéntica) de una repetición CRISPR natural en el locus CRISPR de la bacteria progenitora.

Después de cada exposición de una bacteria progenitora a un bacteriófago dado en experimentos independientes, la secuencia marcadora de cada una de las bacterias marcadas se presenta en una secuencia de nucleótidos diferente, creando de esta forma una secuencia que es única para cada bacteria. Así, tras la exposición de un bacteriófago dado, la secuencia marcadora que está integrada en una bacteria progenitora se selecciona entre el genoma del bacteriófago. Como se ha indicado anteriormente, no se pretende que la presente invención esté

limitada por eventos de integración aleatoria, ni tampoco ningún mecanismo ni medio de acción completo.

Este hallazgo sorprendente se utilizó para el desarrollo de la presente invención, ya que la secuencia marcadora seleccionada proporciona una marca o etiqueta única en la bacteria marcada. Sorprendentemente, también se ha descubierto que cuando la misma bacteria progenitora se expone al mismo bacteriófago, la secuencia marcadora que está integrada en experimentos independientes/diferentes es de una secuencia diferente, dando como resultado de esta forma una etiqueta única en la bacteria marcada después de cada exposición.

En algunas realizaciones, una secuencia marcadora seleccionada aleatoriamente se identifica en la bacteria marcada debido a una o más de las siguientes propiedades de la secuencia marcadora: (1) la ubicación de la secuencia marcadora en uno o más loci CRISPR del mutante insensible al bacteriófago (como se indica en el presente documento, la secuencia marcadora se suele encontrar en uno o ambos extremos 5' y/o 3' (más preferiblemente en el extremo 5') del locus CRISPR de la bacteria marcada; (2) la secuencia marcadora tiene un elevado grado de homología o identidad (por ejemplo, 100 % de identidad) respecto de una secuencia en el genoma del bacteriófago a la que fue expuesta la bacteria progenitora; y/o (3) la secuencia marcadora está fusionada, vinculada o unida a, (por ejemplo, directamente fusionada, vinculada o unida a) al menos una secuencia (por ejemplo, una repetición CRISPR; es decir, una unidad espaciadora de repetición CRISPR) que está duplicada a partir del genoma de la bacteria progenitora. Típicamente, como se describe en el presente documento, esta unidad espaciadora de repetición CRISPR está situada en uno y/o en ambos extremos (por ejemplo, el extremo 5' y/o 3'; más preferiblemente en el extremo 5') del locus CRISPR de la bacteria marcada. Así, en algunas realizaciones, las unidades espaciadoras de repetición CRISPR se integran en ambos extremos del locus CRISPR de la bacteria progenitora de tal forma que las secuencias se encuentran en el extremo 5' y el extremo 3' del locus CRISPR. En algunas realizaciones, una de las secuencias duplicadas es la primera secuencia del extremo 5' del locus CRISPR y la secuencia marcadora está situada inmediatamente después de la secuencia duplicada. En algunas realizaciones, la otra secuencia duplicada es la última secuencia del extremo 3' del locus CRISPR y la secuencia marcadora está situada inmediatamente antes de la secuencia duplicada.

En algunas realizaciones preferidas, la secuencia o secuencias marcadoras y/o la secuencia o secuencias duplicadas de la unidad espaciadora de repetición CRISPR se integra en un extremo del locus CRISPR de la bacteria progenitora de tal forma que la secuencia o secuencias se encuentran en el extremo 3' del locus CRISPR. En algunas realizaciones adicionales, la secuencia duplicada es la última secuencia del extremo 3' del locus CRISPR y la secuencia marcadora está situada inmediatamente antes de la secuencia duplicada. En algunas realizaciones preferidas, la secuencia o secuencias marcadoras y/o la secuencia o secuencias duplicadas se integran en un extremo del locus CRISPR de la bacteria progenitora de tal forma que las secuencias se encuentran en el extremo 5' del locus CRISPR. En algunas realizaciones, la secuencia duplicada es la primera secuencia del extremo 5' del locus CRISPR y la secuencia marcadora está situada inmediatamente después de la secuencia duplicada.

Tal como se describe en el presente documento, la secuencia o secuencias marcadoras es una marca específica de la cepa en el sentido de que la secuencia marcadora que se integra o inserta desde el bacteriófago en la bacteria progenitora es diferente cada vez que la bacteria progenitora (por ejemplo, la misma bacteria progenitora) se expone al bacteriófago (por ejemplo, al mismo bacteriófago). Así, la secuencia marcadora es de utilidad como marca única para una cepa bacteriana dada.

En algunas realizaciones, la secuencia o secuencias marcadoras y/o la secuencia o secuencias duplicadas se integran en uno o más loci CRISPR. En algunas realizaciones alternativas, la secuencia o secuencias marcadoras y/o la secuencia o secuencias duplicadas se integran en uno o más loci CRISPR diferentes. En realizaciones adicionales, dos o más secuencia(s) marcadoras y/o secuencia(s) duplicadas diferentes se integran en un locus CRISPR. En otras realizaciones adicionales, dos o más secuencia(s) marcadoras y/o secuencia(s) duplicadas diferentes se integran cada una de ellas en dos o más loci CRISPR diferentes.

L. Loci CRISPR marcados

El genoma de *Streptococcus thermophilus* LMG18311 contiene 3 loci CRISPR; las secuencias de 36 pb repetidas son diferentes en CRISPR1 (34 repeticiones), CRISPR2 (5 repeticiones), y CRISPR3 (una sola secuencia). Sin embargo, están perfectamente conservadas en cada locus. Las repeticiones de CRISPR1 y CRISPR2 están separadas entre sí por 33 y 4 secuencias de 30 pb de longitud. Todas estas secuencias intercaladas son diferentes entre sí. También se diferencias de las descubiertas en la cepa CNRZ1066 (41 secuencias intercaladas en CRISPR1) y en la cepa LMD-9 (16 en CRISPR1 y 8 en CRISPR3), que son ambas *S. thermophilus*.

La cepa DGCC7710 de *Streptococcus thermophilus* (depositada en la "Collection Nationale de Cultures de Microorganismes" francesa con número CNCM 1-2423) contiene al menos 3 loci CRISPR: CRISPR1, CRISPR2, y CRISPR3. En las cepas CNRZ1066 y LMG18311 para las que se conoce la secuencia del genoma completo (Bolotin *et al.*, [2004] *supra*), CRISPR1 se sitúa en el mismo locus cromosómico: (entre str0660 (o stu0660) y str0661 (o stu0661)). En la cepa DGCC7710, CRISPR1 se sitúa entre genes muy similares. CRISPR1 de la cepa DGCC7710 contiene 33 repeticiones, incluyendo la repetición terminal) y, por tanto, 32 espaciadores. Cada uno de estos

espaciadores son diferentes entre sí. Aunque la mayoría de estos separadores son nuevos (es decir, no estaban anteriormente identificados dentro del loci CRISPR), cuatro separadores cercanos al remolque de CRISPR1 son idénticos a los separadores de CRISPR1 ya conocidos. Estos cuatro incluyen: el 28° separador de DGCC7710, que es idéntico en un 100 % con el 31° separador de rCRISPR1 de la cepa CNRZ1575 (número de registro de GenBank DQ072991); el 30° separador de DGCC7710, que es idéntico en un 100 % con el 27° separador de rCRISPR1 de la cepa CNRZ703 (número de registro de GenBank DQ072990); el 31° separador de DGCC7710, que es idéntico en un 100 % con el 28° separador de rCRISPR1 de la cepa CNRZ703 (número de registro de GenBank DQ072990); y el 32° separador de DGCC7710, que es idéntico en un 100 % con el 30° separador de rCRISPR1 de la cepa CNRZ703 (número de registro de GenBank DQ072990). La secuencia CRISPR1 (5'-3') de la cepa DGCC7710 se muestra en la SEQ ID NO:1.

La cepa DGCC7778 de *Streptococcus thermophilus* se aisló como un mutante natural resistente a fago que utiliza DGCC7710 como cepa progenitora, y el fago D858 como cepa virulenta. El CRISPR1 de la cepa DGCC7778 contiene 35 repeticiones, incluyendo la repetición terminal) y, por tanto, 34 espaciadores. Cuando se compara con la secuencia CRISPR1 de DGCC7710, la secuencia CRISPR1 de DGCC7778 tiene dos nuevos espaciadores adicionales adyacentes, así como dos repeticiones adicionales que flanquean los nuevos espaciadores, en un extremo del locus CRISPR (es decir, cerca de la secuencia líder). El resto de los espaciadores del locus CRISPR1 están inalterados. La secuencia CRISPR1 (5'-3') de la cepa DGCC7778 se muestra en la SEQ ID NO:2.

Así, en el caso de DGCC7778, el primer espaciador (5'-caacacattcaacagattaatgaagaatac-3' [SEQ ID NO:3] y el segundo espaciador (5'-tccactcacgtacaaatagtgtactc-3' [SEQ ID NO:4]) constituyen la marca específica de la cepa que identifica la cepa marcada. Durante el desarrollo de la presente invención, se demostró que la secuencia de ambos espaciadores nuevos existe dentro del genoma del fago D858. La secuencia del segundo espaciador nuevo se encuentra entre las posiciones 25471 y 25442 pb (es decir, de la hebra menos) del genoma de D858, con el otro emparejamiento incorrecto (96,7 % de nucleótidos idénticos sobre 30 nucleótidos). La secuencia del primer espaciador se encuentra entre las posiciones 31481 y 31410 pb (es decir, de la hebra más) del genoma de D858 (100 % de nucleótidos idénticos sobre 30 nucleótidos). Aunque no se pretende que la presente invención esté limitada a ningún mecanismo ni teoría particular, el hecho es que se necesitan dos nuevos espaciadores presentes en el locus CRISPR1 de DGCC7778 para transmitir resistencia a la cepa DGCC7778 contra el fago D858, se contempla que el separador "2" se inserte en primer lugar en el locus CRISPR1 de DGCC77710 (33 repeticiones y 32 espaciadores), en un espaciadores de este locus CRISPR, junto con una repetición. Esta inserción da lugar a un mutante insensible al bacteriófago (cepa intermedia), marcada con este nuevo espaciador adicional (es decir, que tiene ahora 34 repeticiones y 33 espaciadores). Este espaciador se deriva del genoma de D858 pero un error en la replicación o un error de transcripción inversa probablemente acaecido durante el proceso de inserción, condujo a una mutación puntual. Debido al emparejamiento imperfecto (es decir, 1 emparejamiento incorrecto) entre este espaciador recientemente adquirida y la secuencia del fago diana, la eficacia de la resistencia de esta cepa intermedia al fago D858 fue baja.

Sin embargo, un segundo episodio de introducción del espaciador realizada en esta cepa intermedia (es decir, la cepa más resistente al fago D858 que la cepa progenitora DGCC7710 pero no "completamente" resistente debido al emparejamiento incorrecto), llevó a la inserción de un segundo espaciador nuevo (el espaciador "1" como aparece en DGCC7778) en el mismo extremo del locus CRISPR1, junto con una repetición. Esta segunda inserción da lugar a un nuevo mutante insensible al bacteriófago, que se aisló y nombró como DGCC7778. DGCC7778 es más resistente a D858 que la cepa intermedia y mucho más resistente que la cepa progenitora DGCC7710, debido a la presencia del espaciador "1," que es idéntico en un 100 % a la secuencia del fago mutada.

La cepa DGCC7710-RH1 de *Streptococcus thermophilus* se aisló como un mutante natural resistente a fago que utiliza DGCC7710 como cepa progenitora, y el fago D858 como cepa virulenta. El CRISPR1 de la cepa DGCC7710-RH1 contiene 34 repeticiones, incluyendo la repetición terminal) y 33 espaciadores. Cuando se compara con la secuencia CRISPR1 de la cepa DGCC7710 de *Streptococcus thermophilus*, la secuencia CRISPR1 de la cepa DGCC7710-RH1 de *Streptococcus thermophilus* tiene dos nuevos espaciadores adicionales (es decir, la secuencia mercadora) y una repetición adicional que flanquea el nuevo espaciador, en un extremo del locus CRISPR (es decir, cerca de la secuencia líder), en el extremo 5' del locus CRISPR. El resto de los espaciadores del locus CRISPR1 están inalterados. La secuencia CRISPR1 (5'-3') de la cepa DGCC7710-RH1 es:

caaggacagtattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataaccactt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcaagattcgaattgt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaactaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgattgacaatctgctgaccactgtatc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacactggcaggcttattactcaacagcga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttcctgttcttittgttatctttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcatcttccgttttgttgcgaatcct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaacagattcaa
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcgaaggta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatcttaattacaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgtagcgtcatgtgttaacttaatta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgctccaatcctgtaataacttactcg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagggatgatgctatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattctcatccggtaactgctcaagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaaggcgatagaaggagacaacatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaatcatttcaacttcat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagatcagcaagcaagcgtttagttact
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataattttaaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataattatggtagcttaattatcattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcaggttcgagttaccgrrtc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatcagggtcaactaacaattatgct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaatctgttttaggtacattt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatcagacaagagttaaaatggctt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctacttttacagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagctgtaaagct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatcacaaggatgatccaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggttat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcattgagcgtatattgattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatttggcccttcttggcccttgactag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcattgtgccattgagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTttagtcaacataaaaagccagttcaattgaacttggcctt
 (SEQ ID NO:5)

La secuencia líder de este CRISPR1 es: 5' caaggacagtattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' (SEQ ID NO:6). La secuencia integrada que comprende las repeticiones CRISPR se muestra en mayúsculas, mientras que los espaciadores CRISPR se muestran en minúscula. En esta secuencia, la repetición terminal tiene la secuencia 5' gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO:7), mientras que la secuencia remolque tiene la secuencia: 5' ttgattcaacataaaaagccagttcaattgaactggcttt 3' (SEQ ID NO:8). Así, para la cepa DGCC7710-RH1 de *S. thermophilus*, el espaciador (5'-tcaacaattgcaacatcttataaccactt-3' [SEQ ID NO:9]) constituye la secuencia marcadora específica de cepa que identifica esta cepa mutante, es decir, la bacteria marcada). La secuencia del nuevo espaciador (es decir, la secuencia marcadora) está presente dentro del genoma del fago D858.

La secuencia del espaciador se encuentra entre las posiciones 31921 y 31950 pb (es decir, de la cepa positiva) del genoma de D858 (y tiene un 100 % de identidad con la secuencia genómica de D858 para más de 30 nucleótidos). El nuevo espaciador (es decir, la secuencia marcada) que está integrada en el locus CRISPR1 de la cepa DGCC7710-RH1 de *S. thermophilus* transmite resistencia al fago D858 para esta cepa.

La cepa DGCC7710-RH2 de *S. thermophilus* se aisló como un mutante natural resistente a fago que utiliza la cepa DGCC7710 de *S. thermophilus* como cepa progenitora, y el fago D858 como fago virulento. El CRISPR1 de la cepa DGCC7710-RH2 de *S. thermophilus* contiene 34 repeticiones, incluyendo la repetición terminal) y 33 espaciadores. Cuando se compara con la secuencia CRISPR1 de la cepa DGCC7710 de *S. thermophilus*, la secuencia CRISPR1 de la cepa DGCC7710-RH2 de *S. thermophilus* tiene dos nuevos espaciadores adicionales (es decir, la secuencia marcadora) y una repetición adicional que flanquea el nuevo espaciador, en un extremo del locus CRISPR (es decir, cerca de la secuencia líder), en el extremo 5' del locus CRISPR. El resto de los espaciadores del locus CRISPR1 están inalterados. La secuencia CRISPR1 (5'-3') de la cepa DGCC7710-RH2 es:

caaggacagttattgatttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACttacgtttgaaaagaatatcaaatcaatga
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgtttgacagcaaatcaagattcgaattgt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACcgatttgacaatctgctgaccactgttctc
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACacactggcaggcttattactcaacagoga
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACctgttcctgttctttgtgtatctttc
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACttcattcttccgtttgttgcgaatcct
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACcctagagtggaacttagaacaagattcaa
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACataatgccgttgaattacacggcaaggta
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgagcgagctcgaataatcctaattacaag
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgttgcgtacgtcatgtgttaactgtattta
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACggcgctccaatcctgattataacttactcg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaacacagcaagacaaggagatgatgctatg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACcgacacaagaagctatcgaagagttcaag
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACacaattcttcaatcggttaactgtcaagt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaatgaaggcattagaaggagacacatg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACcgatatttaaatcatttataacttcat
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgcagtatcagcaagcagctgttattact

GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACataaactatgaattttataatttttaaga
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaataatttatggtatagcttaatatcattg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACtgcatcagcagcttcgatttaccgttctc
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACtctatacagggtcaactaacaattatgct
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaatcgttcaattctgttttaggtacatt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaatcaatcagacaagagttaaaatggtctt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACgcttagctgtccaatccacgaactgtggatg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACcaaccaacggtaacagctacttttacagt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACataaactgaaggataggagctgttaagctt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACtaatgctacatctcaaggatgatccaga
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaagtagttgatgacctatacaatggtttat
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACacctaagaagcatttgagcgtatattgattg
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaattttgcccttcttggcccttgactag
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAACaccattagcaatcatttggccattgagt
 GTTTTTGTA CTCTCAAGATTTAAGTA ACTGTACAGTttgattcaacataaaaagccagttcaattgaacttggcttt
 (SEQ ID NO:10)

- 5 La secuencia líder es 5'caaggacagttattgatttataatcactatgtgggtataaaaacgtcaaaatttcatttgag3' (SEQ ID NO:6). Las secuencias integradas que comprenden las repeticiones CRISPR se muestran en mayúscula, mientras que el espaciador CRISPR (es decir, la secuencia marcadora) se muestra en minúscula. Esta repetición terminal tiene la secuencia 5' gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO: 7), y la secuencia remolque tiene la secuencia 5' ttgattcaacataaaaagccagttcaattgaacttggcttt 3' (SEQ ID NO:8). Así, en el caso de la cepa DGCC7710-RH2 de *S. thermophilus*, el espaciador (5'-ttacgtttgaaaagaatatcaaatcaatga-3' [SEQ ID NO:11]) constituye la marca específica de la cepa que identifica esta cepa mutante, es decir, la bacteria marcada). También se demostró que la secuencia del nuevo espaciador existe dentro del genoma del fago D858. La secuencia del espaciador se sitúa entre las posiciones 17215 y 17244 pb (es decir, de la cepa positiva) del genoma de D858 (y tiene un 100 % de identidad con la secuencia genómica de D858 para más de 30 nucleótidos). El nuevo espaciador que está integrado en el locus CRISPR1 de la cepa DGCC7710-RH2 de *S. thermophilus* transmite resistencia al fago D858 para esta cepa.

- Además de los loci CRISPR marcados naturalmente desarrollados anteriormente descritos, en algunas realizaciones, las bacterias marcadas se producen mediante técnicas de ADN recombinante como se conoce en la técnica. Por ejemplo, en algunas realizaciones, los oligonucleótidos sintéticos se producen y se introducen en un cultivo de bacterias progenitoras para producir bacterias marcadas. Tampoco se pretende que la presente invención se limite a loci CRISPR marcados, ya que los loci adicionales son de utilidad en las realizaciones marcadoras.

M. Tipificado

- 25 Los inventores también se describen métodos y composiciones para identificar (por ejemplo, tipificar) una bacteria marcada. En algunas realizaciones, la identificación implica la amplificación (por ejemplo, usando PCR) del locus CRISPR o una parte del mismo. En algunas realizaciones, un primer cebador se diseña un primer cebador para hibridarse con una secuencia que está situada antes de la primera repetición de un locus CRISPR. Por ejemplo, en algunas realizaciones, el primer cebador se hibrida con parte de la secuencia líder común del locus CRISPR. En realizaciones alternativas, el primer cebador se hibrida con un gen adyacente que está situado antes del locus CRISPR. En algunas realizaciones, un segundo cebador se hibrida después del primer espaciador de CRISPR o el núcleo del primer espaciador de CRISPR. En algunas realizaciones, el segundo cebador se hibrida en el remolque o incluso en un gen adyacente posterior. En algunas realizaciones preferidas, el segundo cebador se hibrida dentro del locus CRISPR. En algunas realizaciones preferidas alternativas, el segundo cebador se hibrida al menos

parcialmente con un espaciador de CRISPR posterior o con el núcleo del espaciador de CRISPR.

En algunas realizaciones especialmente preferidas, después de la amplificación, la secuencia marcadora se identifica utilizando cualquier método(s) adecuado conocido en la técnica. Por ejemplo, en algunas realizaciones, la secuencia marcadora se identifica determinando el patrón de restricción del producto de amplificación. De acuerdo con algunas realizaciones, una vez que el ADN que comprende el locus CRISPR o una parte del mismo se ha amplificado, se digiere con una o más enzimas de restricción.

En algunas realizaciones preferidas adicionales, las secuencias marcadoras se identifican usando métodos de secuenciación que son conocidos en la técnica. En otras realizaciones adicionales, los métodos de hibridación bien conocidos en la técnica son de utilidad en la presente invención. En algunas realizaciones, los métodos que abarcan técnicas de hibridación conocidas en la técnica son de utilidad para detectar y/o diferenciar cepas bacterianas, incluyendo pero sin limitación, la transferencia Southern, ensayos de movilidad con desplazamiento, ensayos de secuenciación con matrices de oligonucleótidos, espoligotipación, hibridación con fluorescencia in situ (FISH), ensayos de rastreo de heterodupletes, y ensayos de movilidad de heterodupletes.

En algunas realizaciones preferidas adicionales, la secuencia marcadora identificada se compara con una secuencia de al menos una base de datos de secuencias de fago y/o al menos una base de datos de secuencias bacterianas. En algunas realizaciones, la secuencia marcadora se corresponde con una o más secuencias de la base de datos de secuencias de fago pero no con las secuencias de la base de datos de secuencias bacterianas. Se contempla que, a medida que se preparen nuevas secuencias marcadas utilizando los métodos proporcionados en el presente documento, otras bases de datos de marcas sean de utilidad en la presente divulgación.

N. Bacterias marcadas

Tal como se usa en el presente documento, los términos "bacteria marcada", "bacterias marcadas", "bacteria etiquetada" y "bacterias marcadas" se utilizan indistintamente con referencia a una bacteria progenitora o bacterias progenitoras, en la que uno o más loci CRISPR o una parte del mismo se han modificado (por ejemplo, mutado) de tal forma que no sea sensible a uno o más bacteriófagos a los que se ha expuesto.

Tal como se describe con más detalle en el presente documento, en algunas realizaciones, las bacterias marcadas se exponen a uno o más bacteriófagos (por ejemplo, bien iterativamente, secuencialmente o bien simultáneamente), de tal forma que acumula una o más modificaciones genómicas dentro de uno o más loci de CRISPR de tal forma que se vuelve insensible a cada uno de los bacteriófagos a los que ha estado expuesto.

Para infectar células, un bacteriófago inyecta o transfiere su ácido nucleico al interior de la célula en la que el ácido nucleico existe de forma independiente al genoma de la célula. En algunas realizaciones, la infección da como resultado la expresión (es decir, la transcripción y traducción) del ácido nucleico del bacteriófago en el interior de la célula y la continuación del ciclo de vida del bacteriófago.

En algunas realizaciones de la presente invención, tras la exposición al bacteriófago, la bacteria marcada tiene poca o ninguna susceptibilidad a la infección y/o multiplicación por el bacteriófago cuando se compara con la bacteria progenitora. Tal como se usa en el presente documento, la expresión "susceptibilidad reducida a la infección y/o multiplicación por el bacteriófago" significa que el nivel de infección y/o multiplicación por el bacteriófago en la bacteria marcada no ocasiona un efecto perjudicial sobre la bacteria marcada.

Así, en algunas realizaciones de la presente invención, una bacteria progenitora no muere después de la exposición al bacteriófago, debido a la mutación de la bacteria progenitora de tal forma que se vuelve insensible al bacteriófago.

En algunas realizaciones, la bacteria marcada es insensible o prácticamente insensible a la infección adicional y/o multiplicación por el bacteriófago. En realizaciones adicionales, la bacteria marcada es insensible o prácticamente insensible a uno o más de los mecanismos que utiliza el bacteriófago para infectar y/o multiplicarse en una bacteria. En otras realizaciones adicionales, la bacteria marcada es insensible o prácticamente insensible a todos los mecanismos que utiliza el bacteriófago para infectar y/o multiplicarse en una bacteria. En otras realizaciones adicionales, la bacteria marcada desarrolla uno o más mecanismos que atenúan, inactivan o destruyen el bacteriófago durante el ciclo de infección. En algunas realizaciones adicionales, la presente invención proporciona cepas marcadas seleccionadas mediante procedimientos de cribado convencional que son conocidos en la técnica para aislar mutantes insensibles a bacteriófagos.

Como se ha indicado anteriormente, además de los loci CRISPR marcados naturalmente desarrollados anteriormente descritos, en algunas realizaciones, las bacterias marcadas se producen mediante técnicas de ADN recombinante como se conoce en la técnica. Por ejemplo, en algunas realizaciones, los oligonucleótidos sintéticos se producen y se introducen en un cultivo de bacterias progenitoras para producir bacterias marcadas. Tampoco se pretende que la presente invención se limite a loci CRISPR marcados, ya que los loci adicionales son de utilidad en las realizaciones marcadoras.

O. Cultivos

Los cultivos, especialmente los cultivos iniciadores, se utilizan ampliamente en la industria alimentaria en la fabricación de productos fermentados incluyendo productos lácteos (por ejemplo, yogur y queso) productos cárnicos, productos de panadería, vino, y productos vegetales. En particular, los cultivos iniciadores tienen amplia utilidad en la fabricación de muchos productos de leche fermentada, queso y mantequilla. Estos cultivos iniciadores bacterianos transmiten rasgos específicos a diferentes productos lácteos realizando numerosas funciones. En algunas realizaciones especialmente preferidas, los cultivos utilizados en la presente invención son cultivos "industrialmente útiles". Tal como se usa en el presente documento, este término se refiere a cualquier cultivo bacteriano que es de utilidad en cualquier industria, que incluye aunque sin limitación, la producción de alimentos, comida, cosméticos, sustancias farmacéuticas, nutracéuticos, probióticos, enzimas, metabolitos, etc. Así, no se pretende que la presente invención se limite a ningún cultivo o industria en concreto, ya que la presente invención es de utilidad en numerosas aplicaciones.

Los cultivos de bacterias no concentrados comerciales se denominan en la industria como "cultivos madre", y se propagan en el centro de producción (por ejemplo, de productos lácteos) antes de añadirse a un material de partida comestible (por ejemplo, leche) para su fermentación. El cultivo iniciador que se propaga en el centro de producción para su inoculación a un material de partida comestible se denomina "iniciador a granel".

Los cultivos iniciadores adecuados para su uso en la presente invención incluyen cualquier organismo que es adecuado para su uso en la industria alimentaria, cosmética y/o farmacéutica. En algunas realizaciones preferidas, el cultivo iniciador es de utilidad en la industria láctea. Así, los cultivos de bacterias acidolácticas se suelen utilizar en la fabricación de productos de leche fermentada (por ejemplo, suero lácteo y crema agria) y en la fabricación de mantequilla y queso (por ejemplo, brie, havarti, cheddar, Monterey jack, etc.).

Tal como se usa en el presente documento, la expresión "bacteria acidoláctica" se refiere a bacterias gram positivas, microaerófilas o anaerobias que fermentan azúcar con la producción de ácidos, incluyendo ácido láctico como ácido producido principal, ácido fórmico y ácido propiónico. Las bacterias acidolácticas más útiles desde el punto de vista industrial incluyen especies de *Lactococcus* (por ejemplo, *Lactococcus lactis*), *Lactobacillus*, *Bifidobacterium*, *Streptococcus*, *Leuconostoc*, *Pediococcus*, and *Propionibacterium*. En algunas realizaciones, la presente invención proporciona cultivos iniciadores que comprenden al menos una especie de bacteria acidoláctica que comprende al menos una especie de bacteria acidoláctica tal como *L. lactis*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Streptococcus thermophilus* o combinaciones de las mismas. Los cultivos iniciadores de bacterias acidolácticas se utilizan habitualmente en la industria alimentaria como cepas de cultivos mixtos que comprenden una o más especies. En algunas realizaciones que comprenden cultivos de cepas mixtas (por ejemplo, cultivos iniciadores de yogur) que comprenden cepas de *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Streptococcus thermophilus*, se establece una relación simbiótica entre las especies en la que la producción de ácido láctico es mayor en comparación con los cultivos de una sola cepa de bacterias acidolácticas (Véase por ejemplo, Rajagopal et al., 1. Dairy Sci., 73:894-899 [1990]).

En algunas realizaciones especialmente preferidas, el cultivo iniciador es una especie de bacteria acidoláctica, que incluyendo pero sin limitarse a las cepas de *Bifidobacterium*, *Brevibacterium*, o *Propionibacterium*. Los cultivos iniciadores adecuados del grupo de bacterias acidolácticas incluyen pero no están limitadas a las cepas de uso habitual *Lactococcus*, *Streptococcus*, *Lactobacillus* (por ejemplo, *Lactobacillus acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc*, y *Oenococcus*. Las especies del *Lactococcus* incluyen pero no están limitadas a los muy utilizados *Lactococcus lactis*, incluidos *Lactococcus lactis* subsp. *lactis* y *Lactococcus lactis* subsp. *cremoris*, y *Lactococcus lactis* subsp. *lactis* biovar. Otras especies de bacterias acidolácticas incluyen *Leuconostoc*, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Lactobacillus helveticus*. Además, las cepas probióticas como *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei* son de utilidad en la potenciación del sabor y proporcionan beneficios para la salud. Los cultivos termófilos de bacterias acidolácticas habitualmente utilizados en la fabricación de quesos italianos tales como Pasta filata o parmesano incluyen *S. thermophilus* y *L. delbrueckii* subsp. *bulgaricus*. En algunas realizaciones, se añaden otras especies de *Lactobacillus* (por ejemplo, *L. helveticus*) durante la fabricación para obtener un sabor deseado.

En algunas realizaciones, el cultivo iniciador comprende o consiste en una cepa genéticamente modificada (preparada de acuerdo con los métodos deseados en el presente documento) o una de las cepas de bacterias acidolácticas anteriores o cualquier otra cepa cultivada adecuada como iniciador. Como conocen los expertos en la técnica, la selección de organismos para los cultivos iniciadores utilizados en la presente invención depende del tipo concreto de los productos a preparar y tratar. Así, por ejemplo, para la fabricación de queso y mantequilla, los cultivos mesófilos de especies de *Lactococcus*, especies de *Leuconostoc* y especies de *Lactobacillus* son de amplia utilidad, mientras que las cepas termófilas de las especies de *Streptococcus* y las especies de *Lactobacillus* son de amplia utilidad para el yogur y otros productos lácteos fermentados.

En algunas realizaciones, el cultivo iniciador un iniciador desecado, mientras que en otras realizaciones, se trata de un iniciador concentrado y, en otras realizaciones, se trata de un cultivo congelado. En algunas realizaciones preferidas, el cultivo desecado comprende al menos una bacteria acidoláctica. En algunas realizaciones, el cultivo

iniciador se utiliza en inoculación directa. En algunas realizaciones preferidas, el cultivo es un cultivo iniciador concentrado utilizado en inoculación directa.

En algunas realizaciones, el cultivo iniciador bacteriano comprende una cepa o especie bacteriana (es decir, es un cultivo puro). Así, en estas realizaciones, prácticamente todo, o una parte significativa, del cultivo iniciador bacteriano comprende la misma cepa o especie bacteriana. Sin embargo, en algunas realizaciones alternativas, el cultivo iniciador comprende más de una o varias cepas o especies bacterianas (es decir, se trata de un cultivo mixto definido)

Los cultivos iniciadores preparados cualquier técnica adecuada conocida en la materia son de utilidad en la presente invención (Véase por ejemplo, la patente de Estados Unidos N.º 4.621.058). Por ejemplo, los cultivos iniciadores preparados mediante la introducción de un inóculo (por ejemplo, un cultivo bacteriano) en un medio de cultivo para producir un medio inoculado e incubar el medio inoculado para producir un cultivo iniciador. Sin embargo, no se pretende que la presente invención esté limitada a ningún método concreto para preparar los cultivos iniciadores.

Los cultivos iniciadores desecados preparados cualquier técnica adecuada conocida en la materia son de utilidad en la presente invención (Véase por ejemplo, la patente de Estados Unidos números. 4.423.079 y 4.140.800). En algunas realizaciones, los cultivos iniciadores desecados utilizados en la presente invención son preparaciones sólidas (por ejemplo, comprimidos, aglomerados, cápsulas, polvo, gránulos, harinas, cualquiera de los cuales se pueden humedecer, secar mediante pulverización, criodesecar o liofilizar en algunas realizaciones). En algunas realizaciones alternativas, los cultivos iniciadores desecados de la presente invención son tanto un aglomerado congelado a baja temperatura como un polvo liofilizado. Estos cultivos iniciadores desecados se pueden preparar mediante cualquier método adecuado conocido en la técnica.

En algunas realizaciones, los cultivos iniciadores utilizados en la presente invención comprenden concentrados que tienen concentraciones sustancialmente elevadas de al menos una especie de bacteria. En algunas realizaciones preferidas, los concentrados se diluyen con agua o se resuspenden en agua u otro diluyente adecuado (por ejemplo, un medio de cultivo adecuado, aceite mineral, o aceite vegetal) para su uso en la presente invención. En algunas realizaciones, los cultivos iniciadores desecados de la presente invención se preparan usando métodos bien conocidos en la técnica incluyendo, pero sin limitación, centrifugación, filtración, o una combinación de estas técnicas.

P. Productos

La presente invención es de utilidad en la producción de varios productos, incluyendo pero sin limitación alimentos y/o productos alimenticios. Así, se contempla que cualquier producto elaborado o que comprende un cultivo bacteriano es de utilidad en la presente invención. Estos incluyen, pero no se limitan a, fruta, legumbres, hortalizas y cultivos forrajeros incluyendo productos derivados, grano, y productos derivados de grano, productos lácteos y productos derivados de los lácteos, carne, aves de corral, marisco, enzimas y metabolitos.

Tal como se usa en el presente documento, el término "alimento" se utiliza en un sentido amplio e incluye comida, productos alimenticios, ingredientes alimenticios, suplementos dietéticos, y alimentos funcionales. Aunque el término incluye alimento para seres humanos, se pretende que el término también abarque alimento para animales no humanos (es decir, "pienso"). Sin embargo, en algunas realizaciones preferidas, la presente invención proporciona alimentos para consumo humano. Tal como se usa en el presente documento, el término "ingrediente alimenticio" incluye una formulación adecuada para su adición a alimentos. En algunas realizaciones, las formulaciones se utilizan a bajos niveles en una amplia variedad de productos que requieren por ejemplo, acidificación o emulsión. Tal como se usa en el presente documento, el término "alimento funcional" se refiere a alimentos que pueden no solamente proporcionar un efecto nutritivo y/o una satisfacción por su sabor, sino que también pueden proporcionar un beneficio adicional al consumidor. En algunas realizaciones, la bacteria de la presente invención comprende o se añade a un ingrediente alimenticio, suplemento o alimento funcional. Se contempla que el alimento se proporcione en cualquier forma adecuada, incluyendo pero sin limitación, soluciones, sólidos, geles, emulsiones, etc., dependiendo del uso y/o del modo de aplicación y/o el modo de administración. Así, no se pretende que la presente invención esté limitada a ningún alimento de ninguna forma particular. En algunas realizaciones, la bacteria de la presente invención es de utilidad en las numeras preparaciones de productos alimenticios, incluyendo pero sin limitarse a productos de confitería, productos lácteos, productos cárnicos, productos de aves de corral, productos de pescado, y productos de panadería. En algunas realizaciones, las bacterias se utilizan como ingredientes en refrescos, zumos de frutas, bebidas que comprenden suero lácteo, infusiones saludables, bebidas de cacao, bebidas lácteas, bebidas que incluyen bacterias acidolácticas, queso, yogur, yogur líquido y vino. En algunas realizaciones adicionales, la presente invención proporciona métodos para preparar alimentos, incluyendo métodos que comprenden premezclar las bacterias de acuerdo con la presente invención con un ingrediente alimenticio (por ejemplo, un material de partida de un alimento). En algunas realizaciones preferidas, el alimento proporcionado en el presente documento es un producto lácteo. En algunas realizaciones especialmente preferidas, el producto lácteo se selecciona entre yogur, queso (por ejemplo, un requesón, un queso duro, un queso semiduro, un queso de pasta ácida), un suero de mantequilla, quark, una crema agria, kefir, una bebida a base de suero de leche fermentado, kumis, una bebida láctea y un yogur batido.

Tal como se usa en el presente documento, los términos "comida" y "alimento" incluyen, pero no se limitan a material vegetal bruto y procesado, así como material no vegetal. Se pretende que el alimento/comida sea adecuado para su consumo por cualquier animal, ser humano o no humano. En algunas realizaciones preferidas, el alimento/comida es de utilidad para el ganado (por ejemplo, ganado vacuno, oveja, cerdos, etc.), aves de corral (por ejemplo, pollos y pavos), pescado, reptiles o crustáceos. Así, no se pretende que la presente invención esté limitada a ningún alimento/comida para ningún organismo en particular.

Parte experimental

Se proporcionan los siguientes ejemplos para demostrar e ilustrar adicionalmente determinadas realizaciones y aspectos preferidos de la presente invención y no deben tomarse como limitantes de su alcance.

En la divulgación experimental que sigue se aplican las siguientes abreviaturas: °C (grados centígrados); H₂O (agua); aa (aminoácido); pb (pares de bases); kb (pares de kilobases); kD (kilodaltons); g y gm (gramos); µg y ug (microgramos); mg (miligramos); ng (nanogramos); µl y ul (microlitros); ml (mililitros); mm (milímetros); nm (nanómetros); µm y um (micrómetro); M (molar); mM (milimolar); µM y uM (micromolar); sec y s (segundos); min(s) (minuto/minutos); h(s) (hora/horas); MOI (multiplicidad de la infección); EOP (eficiencia de la siembra en placas); UFP (unidades formadoras de placa); MgCl₂ (cloruro de magnesio); NaCl (cloruro de sodio); PAGE (electroforesis en gel de acrilamida); PBS (suero salino tamponado con fosfato [NaCl 150 mM, tampón fosfato de sodio 10 mM, pH 7,2]); SDS (dodecil sulfato de sodio); Tris (tris(hidroximetil)aminometano); p/v (peso en volumen); v/v (volumen en volumen); Promega (Promega, Inc., Madison, WI); ATCC (American Type Culture Collection, Manassas, VA); Difco (Difco Laboratories, Detroit, MI); GIBCO BRL o Gibco BRL (Life Technologies, Inc., Gaithersburg, MD); y Sigma (Sigma Chemical Co., St. Louis, MO).

La presente invención utiliza, salvo que se indique otra cosa, técnicas convencionales de química, biología molecular, microbiología, ADN recombinante e inmunología, que están dentro de las capacidades de los expertos en la materia. Dichas técnicas son bien conocidas de los expertos en la materia.

Tal como se usa en el presente documento, DGCC7710 también se denomina como WT"; DGCC7710RH1 también se denomina como "DGCC7710-RH1" y "RH1"; DGCC7710RH2 también se denomina como "DGCC7710-RH2" y "RH-2."

Ejemplo 1

Marcado de *Streptococcus thermophilus* DGCC7710 usando el bacteriófago D2972 mediante la inserción de una sola unidad espaciadora de repetición en el interior de CRISPR1

En este ejemplo, *S. thermophilus* cepa DGCC7710 (depositada en la "Collection Nationale de Cultures de Microorganismes" francesa con número CNCM I-2423) se marcó mediante inserción "natural" dentro de su locus CRISPR1 de una unidad espaciadora de repetición adicional donde el espaciador procede del bacteriófago D2972. El locus CRISPR1 de DGCC7710 contiene 33 repeticiones (incluyendo la repetición terminal) y 32 espaciadores (véase, número de registro de GenBank: EF434469). El bacteriófago D2972 se aisló a partir de un producto lácteo fermentado usando la cepa DGCC7710. Su genoma se ha secuenciado completamente (véase, número de registro de GenBank: AY699705). La secuencia del locus CRISPR1 de DGCC7710 es:

caaggacagttatgatttataatcactatgtgggtataaaacgtcaaaatttcatttgag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcaagattcgaattgi
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgaattgacaatctgctgaccactgttate
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccctgttctttgttattctttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcttattcttccgttttgttgcgaatcct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaactagaaacagattcaa
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaaggcca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatcttaattacaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggaactattta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgctcccaatcctgattatacttactc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggaggtatgctatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacagaacgtatgcaagagttcaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttctcggtaactgctcaagtg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatlaaggcatalagaaggaggaacaacatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgaattttaaatactttcataacttcat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttattact
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaatttataattttaaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttattgtatagcttaatactattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcattcagcagcttcgaattaccgtttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcttataatcagggtcaactaacaattatgct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaatctgttttaggtacattt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacagacaaggttaaatggctctt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctacttttacagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggatagagcttgaaggtct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatcacaaggatgaicccaga

 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagtigatgacctctacaatggtttat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccatagaagcatttgagcgtatattgattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatttggcccttcttgcctttagctag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttggccattgagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTtatttcaacataaaagccagticaatg
 aacttggctt (SEQ ID NO:1)

- 5 En la primera etapa, la cepa progenitora (DGCC7710) se expuso al bacteriófago donante (D2972). DGCC7710 se precultivó en medio de tipo leche esterilizada (10 % p/v de leche en polvo en agua, esterilizada durante 20 min a 110 °C) a 42 °C durante 6 horas. El precultivo se utilizó para inocular 10 ml de medio de tipo leche esterilizada a aproximadamente un 0,05 % (p/v). Se añadieron aproximadamente 10⁷ bacteriófagos D2972 al medio de tipo leche inoculado (recuento final de bacteriófagos de aproximadamente 10⁶ ufp/ml). A continuación, la mezcla se cultivó a 42 °C durante 16 horas. Tras la incubación diluciones del cultivo se sembraron en placas con medio M17-glucosa (0,5 % p/v) para obtener colonias aisladas tras su incubación a 42 °C durante 24 horas. Un número de colonias se repicaron y se hicieron crecer por separado en medio de tipo leche esterilizada a 42 °C durante 18 horas para proporcionar una solución madre de aislados.
- 10
- 15 A continuación, cada cepa marcada se identificó mediante análisis de secuencia en su loci CRISPR1. En este Ejemplo, una cepa marcada es una variante de la cepa progenitora, en la que la variante contiene una unidad espaciadora de repetición adicional dentro del locus CRISPR1 (Véase la Figura 3). Típicamente, la parte espaciadora de la unidad adicional tiene aproximadamente 30 nucleótidos de longitud, y su secuencia es idéntica a una consecuencia del fago donante. Los aislados se cultivaron por separado a 42 °C durante 18 horas en medio
- 20 M17-glucosa. Las células se recogieron y se extrajo su ADN. Para cada aislado, la región del cromosoma correspondiente al extremo líder del locus CRISPR1 se amplificó mediante la PCR usando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagctctc-3' [SEQ ID NO:15]; Véase, Bolotin et al., Microbiol., 151:2551-2561 [2005]); y cebador inverso CR1-89R5 (5'-acaaaaacggaagaatgaagctg-3' [SEQ ID NO:16]). La mezcla reacción de la PCR (volumen final de 25 µl contenía): tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada una de las cuatro
- 25 dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa Taq 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron las siguientes: predesnaturalización a 98 °C durante 5 min, después 33 ciclos alternando desnaturalización a 94 °C durante 30 s, hibridación a 56 °C durante 30 s, y elongación a 72 °C durante 1 min; seguido por una etapa de elongación final a 72 °C durante 4 min. El ácido nucleico de cada fragmento de PCR se determinó usando una metodología de "cebador directo" y secuenciación clásica, como es conocido en la técnica.

Cada secuencia se comparó con la obtenida para la cepa progenitora para identificar la presencia de una secuencia adicional.

Mediante experimentos independientes, se crearon múltiples cepas marcadas a partir de DGCC7710 utilizando D2972 como donante. Algunas de estas cepas marcadas se describen en la Tabla 1-1. Todas se diferencian de la cepa progenitora por una única unidad espaciadora de repetición adicional en el extremo líder del locus CRISPR1. En todos los casos, la parte del separador de la unidad adicional es un 100 % idéntica a una subsecuencia del bacteriófago donante D2972 (esta subsecuencia del bacteriófago donante también se denomina como "proespaciador"). Típicamente, las cepas marcadas descritas en este Ejemplo tienen un locus CRISPR1 de 34 repeticiones y 33 separadores; 32 separadores se originan a partir de la DGCC7710, el espaciador suplementario procede de D2972.

Tabla 1-1. Descripción de cepas marcadas en el locus CRISPR1 derivado de DGCC7710 usando D2972 como donante				
Cepa marcada	Cepa progenitora	Fago donante	ADN donante insertado	
			Secuencia	Ubicación*
DGCC7710 _{phi 2972} ^{+S40}	DGCC7 710	D2972	TCTGGAAAGCATATTGAGGGA GCTACTCTT (SEQ ID NO:17)	27974-28003 (+)
DGCC7710 _{phi 2972} ^{+S41}	DGCC7 710	D2972	TCTAATCCCACTAGGAATAGT GGGTAGTAA (SEQ ID NO:18)	25693-25722 (+)
DGCC7710 _{phi 2972} ^{+S43}	DGCC7 710	D2972	TTATAACATAACGGTTAGTTG GCCTCTAT (SEQ ID NO:19)	23410-23382 (-)
DGCC7710 _{phi 2972} ^{+S44}	DGCC7 710	D2972	AAGGAGCTAGCCACATTTCCG CAATTGATA (SEQ ID NO:20)	23334-23363 (+)
DGCC7710 _{phi 2972} ^{+S45}	DGCC7 710	D2972	CAGCTTGAAATGTTTATTGAA GCAGCAGTG (SEQ ID NO:21)	24624-24653 (+)
DGCC7710 _{phi 2972} ^{+S46}	DGCC7 710	D2972	AAATCAGTTTTTTGTTTCAGAA ACTTGTCT (SEQ ID NO:22)	25582-25611 (+)
*(+/-): indica la cepa del cromosoma del fago				

Ejemplo 2

Marcado de *Streptococcus thermophilus* DGCC77710 usando el bacteriófago D858 mediante la inserción de una sola unidad espaciadora de repetición en el interior de CRISPR1

En este Ejemplo, *S. thermophilus* cepa DGCC77710 se marcó mediante inserción "natural" dentro de su locus CRISPR1 de una unidad espaciadora de repetición adicional donde el espaciador procede del bacteriófago D858. El bacteriófago D858 se aisló a partir de un producto lácteo fermentado en la cepa DGCC7710. D858 es un bacteriófago que pertenece a la familia de virus *Siphoviridae*. La secuencia de su genoma se ha determinado completamente (número de registro de GenBank: EF529515). Este fago es virulento para *S. thermophilus* es la cepa DGCC7710.

Su genoma se ha secuenciado completamente.

En primer lugar, la cepa progenitora (DGCC77710) se expuso al bacteriófago donante (D858). DGCC77710 se precultivó en medio de tipo leche esterilizada (10 % p/v de leche en polvo en agua, esterilizada durante 20 min a 110 °C) a 42 °C durante 6 horas. El precultivo se utilizó para inocular 10 ml de medio de tipo leche esterilizada a aproximadamente un 0,05 % (p/v). Se añadieron aproximadamente 10⁷ bacteriófagos D858 al medio de tipo leche inoculado (recuento final de bacteriófagos de aproximadamente 10⁶ ufp/ml). A continuación, la mezcla se cultivó a 42 °C durante 16 horas. Tras la incubación diluciones del cultivo se sembraron en placas con medio M17-glucosa (0,5 % p/v) para obtener colonias aisladas tras su incubación a 42 °C durante 24 horas. Un número de colonias se repicaron y se hicieron crecer por separado en medio de tipo leche esterilizada a 42 °C durante 18 horas para proporcionar una solución madre de aislados.

A continuación, cada cepa marcada se identificó mediante análisis de secuencia en su locus CRISPR1. En este Ejemplo, una cepa marcada es una variante de la cepa progenitora que contiene una unidad espaciadora de repetición adicional dentro de su locus CRISPR1 (Véase la Figura 3). Típicamente, la parte espaciadora de la unidad adicional tiene aproximadamente 30 nucleótidos de longitud, y su secuencia es idéntica a una consecuencia del fago donante. Los aislados se cultivaron por separado a 42 °C durante 18 horas en medio M17-glucosa. Las células se recogieron y se extrajo su ADN. Para cada aislado, la región del cromosoma correspondiente al extremo líder del locus CRISPR1 se amplificó mediante la PCR usando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:15], Bolotin *et al.*, [2005], *supra*); y cebador inverso CR1-89R5 (5'-acaaaaacggasgaatgaagttg-3' [SEQ ID NO:20]). La mezcla reacción de la PCR (volumen final de 25 µl contenía): tampón exento de Mg 1x (Promega). MgCl₂ 2,5 mM, cada una de las cuatro dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa *Taq* 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron las siguientes: predesnaturalización a 98 °C durante 5 min, después 33 ciclos alternando desnaturalización a 94 °C durante 30 s,

hibridación a 56 °C durante 30 s, y elongación a 72 °C durante 1 min; seguido por una etapa de elongación final a 72 °C durante 4 min. El ácido nucleico de cada fragmento de PCR se determinó usando una metodología de "cebador directo" y secuenciación clásica, como es conocido en la técnica. Cada secuencia se comparó con la obtenida para la cepa progenitora para identificar la presencia de una secuencia adicional.

Se describe en la Tabla 2-1 una cepa marcada a partir de DGCC7710 usando D858 como donante. Se diferencia de la cepa progenitora por una única unidad espaciadora de repetición adicional en el extremo líder del locus CRISPR1. La parte del separador de la unidad adicional es un 100 % idéntica a una subsecuencia del bacteriófago donante D858.

Tabla 2-1. Descripción de cepas marcadas en el locus CRISPR1 derivado de DGCC7710 y uso de D858 como donante				
Cepa marcada	Cepa progenitora	Fago donante	ADN donante insertado	
			Secuencia	Ubicación*
DGCC7710 _{ph} ^{+S42} 1858	DGCC7 710	D858	TCGATAAATCAGCCAAAGTATT AAGTGGIT (SEQ ID NO:23)	27560-27589 (+)

*(+/-): indica la cepa del cromosoma del fago

Ejemplo 3

Marcado de *Streptococcus thermophilus* DGCC7710 usando el bacteriófago D2972 mediante la inserción de múltiples unidades espaciadoras de repetición en el interior de CRISPR1

En este Ejemplo, *S. thermophilus* cepa DGCC7710 se marcó mediante inserción "natural" dentro de su locus CRISPR1 mediante la adición de múltiples unidades espaciadoras de repetición donde los espaciadores proceden del bacteriófago D2972.

En primer lugar, la cepa progenitora (DGCC7710) se expuso al bacteriófago donante (D2972). DGCC7710 se precultivó en medio de tipo leche esterilizada (10 % p/v de leche en polvo en agua, esterilizada durante 20 min a 110 °C) a 42 °C durante 6 horas. El precultivo se utilizó para inocular 10 ml de medio de tipo leche esterilizada a aproximadamente un 0,05 % (p/v). Se añadieron aproximadamente 10⁷ bacteriófagos D2972 al medio de tipo leche inoculado (recuento final de bacteriófagos de aproximadamente 10⁶ ufp/ml). A continuación, la mezcla se cultivó a 42 °C durante 16 horas. Tras la incubación diluciones del cultivo se sembraron en placas con medio M17-glucosa (0,5 % p/v) para obtener colonias aisladas tras su incubación a 42 °C durante 24 horas. Un número de colonias se repicaron y se hicieron crecer por separado en medio de tipo leche esterilizada a 42 °C durante 18 horas para proporcionar una solución madre de aislados.

A continuación, cada cepa marcada se identificó mediante análisis de secuencia en su locus CRISPR1. En este Ejemplo, una cepa marcada es una variante de la cepa progenitora que contiene múltiples unidades espaciadoras de repetición adicional dentro de su locus CRISPR1 (Véase la Figura 3). Típicamente, la parte espaciadora de cada unidad adicional tiene aproximadamente 30 nucleótidos de longitud, y su secuencia es idéntica a una consecuencia del fago donante. Los aislados se cultivaron por separado a 42 °C durante 18 horas en medio M17-glucosa. Las células se recogieron y se extrajo su ADN. Para cada aislado, la región del cromosoma correspondiente al extremo líder del locus CRISPR1 se amplificó mediante la PCR usando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:15], Bolotin *et al.*, 2005, *supra*); cebador inverso CR1-89R5 (5'-acaaaaacggaagaatgaagttg-3' [SEQ ID NO: 16]). La mezcla reacción de la PCR (volumen final de 25 µl contenía): tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada una de las cuatro dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa *Taq* 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron las siguientes: predesnaturalización a 98 °C durante 5 min, después 33 ciclos alternando desnaturalización a 94 °C durante 30 s, hibridación a 56 °C durante 30 s, y elongación a 72 °C durante 1 min; seguido por una etapa de elongación final a 72 °C durante 4 min. El ácido nucleico de cada fragmento de PCR se determinó usando una metodología de "cebador directo" y secuenciación clásica, como es conocido en la técnica. Cada secuencia se comparó con la obtenida para la cepa progenitora para identificar la presencia de una secuencia adicional.

Mediante experimentos independientes, se crearon múltiples cepas marcadas a partir de DGCC7710 utilizando D2972 como donante. Algunas de estas cepas marcadas se describen en la Tabla 3-1. Todas se diferencian de la cepa progenitora con múltiples unidades espaciadoras de repetición adicional en el extremo líder del locus CRISPR1. En todos los casos, la parte del separador de cada unidad adicional es un 100 % idéntica a una subsecuencia del bacteriófago donante D2972.

Tabla 3-1. Descripción de cepas marcadas en el locus CRISPR1 derivado de DGCC7710 usando D2972 como donante				
Cepa marcada	Cepa progenitora	Fago donante	ADN donante insertado	
			Secuencia	Ubicación
DGCC7710 ^{+S46+S47} phi2972	DGCC7710	D2972	AAATCAGTTTTTTGTTTCAGAAACTTG TTCT (SEQ ID NO:24)	33045-33073 (+)
			TTGTCTATTACGACAACATGGAAGA TGAT (SEQ ID NO:25)	25582-25611 (+)
DGCC77710 ^{+S48+S49} phi2972	DGCC77710	D2972	TTTTGAGAAAGTCTTTAACGATGCA GTAGC (SEQ ID NO:26)	25967-25938 (-)
			TAATAGTTTACCAAATCATCTTTATT CCAA (SEQ ID NO:27)	6008-6037 (+)
DGCC77710 ^{+S50+S51} phi2972	DGCC7710	D2972	GAAGTTGAAATAATTCGAGAAATAG AACTC (SEQ ID NO:28)	34105-34134 (+)
			TGGAAACCAAGAAATGCAATAGAAT GGAAG (SEQ ID NO:29)	29246-29275 (+)
DGCC7710 ^{+S52+S54} phi2972	DGCC7710	D2972	CTGATTGTTAATGTACGAGGGCTCC AGCCA (SEQ ID NO:30)	31582-31611 (+)
			CTCAGTCGTTACTGGTGAACCAGTTT CAAT (SEQ ID NO:31)	21732-21703 (-)
DGCC7710 ^{+S53+S54} phi2972	DGCC7710	D2972	TGTTTCAAGGTTTCGGGTCCAAGTAT CATT (SEQ ID NO:32)	29647-29618 (-)
			TTTTCCGTCTTCTTTTITAGCAAAGA TACG (SEQ ID NO:33)	16681-16652 (-)
DGCC7710 ^{+S61+S62} phi2972	DGCC7710	D2972	GATTCGTGGCGATATTCGTCTTACGT TTGA (SEQ ID NO:34)	31709-31737 (+)
			ACATATCGACGTATCGTGATTATCCC ATT (SEQ ID NO:35)	17182-17211 (+)
DGCC7710 ^{+S55+S63+S41} phi2972	DGCC7710	D2972	CTGGAAAGCATATTGAGGGAGCTAC TCIT (SEQ ID NO:36)	25693-25722 (+)
			GTATATCGAAGAACGACTGAAAGAG CTTGA (SEQ ID NO:37)	1114-1142 (+)
			TCTAATCCCACTAGGAATAGTGGGT AGTAA (SEQ ID NO:38)	27381-27409 (+)
*(+/-): indica la cepa del cromosoma del fago				

Ejemplo 4

- 5 Marcado de *Streptococcus thermophilus* DGCC77710 usando el bacteriófago D2972 mediante la inserción iterativa de unidades espaciadoras de repetición en el interior de CRISPR1

En este Ejemplo, *S. thermophilus* cepa DGCC7710 se marcó con un medio "natural" mediante inserción iterativa dentro de su locus CRISPR1 mediante la adición de unidades espaciadoras de repetición adicionales donde los espaciadores proceden del bacteriófago D2972 y de bacteriófagos derivados de D2972.

En la primera iteración, la cepa progenitora (DGCC7710) se expuso al bacteriófago donante (D2972) y se aisló una cepa marcada, que se caracterizó usando la misma metodología que se ha descrito en el Ejemplo 1. En comparación con DGCC7710, esta cepa marcada (denominada DGCC7710_{phi2972}^{S6}) tenía una unidad espaciadora de repetición adicional como se describe en la Tabla 4 en su locus CRISPR1

Debido a la inserción de una unidad espaciadora de repetición en el locus CRISPR1 de la cepa DGCC7710_{phi2972}^{S6}, el bacteriófago donante D2972 ya no resultó virulento frente a DGCC7710_{phi2972}^{S6}, y no se puede utilizar como bacteriófago donante para esta cepa. Este problema se superó usando un fago donante mutante derivado de D2972 que incluye una modificación específica dentro de su genoma (es decir, un "fago mutado"). Este fago mutado se seleccionó exponiendo el bacteriófago donante a la cepa marcada, de tal forma que una modificación (es decir) del fago progenitor lo transforma en virulento para la cepa marcada.

DGCC7710_{phi2972}^{S6} se precultivó en medio de tipo lácteo a 42 °C durante 18 horas. A continuación, un medio de tipo lácteo se inoculó con el precultivo de DGCC7710_{phi2972}^{S6} a una concentración de aproximadamente 10⁶ ufc/ml y con una suspensión de D2972 y una MOI (multiplicidad de infección) mayor de 100. El cultivo se incubó a 42 °C durante 18 horas y a continuación se centrifugó durante 10 min a 10.000xg. El sobrenadante se recogió y se filtró usando un filtro de 0,45 µm. Se utilizaron diluciones adecuadas del sobrenadante filtrado para inocular un medio de glucosa agar M17 sembrado con una capa de DGCC7710_{phi2972}^{S6} usando métodos bien conocidos en la técnica. Las placas

de agar sembradas se incubaron durante 24 horas a 42 °C. Una placa aislada se repicó y se cultivó en DGCC7710^{phi2972}^{S6} en medio glucosa M17 durante 6 horas a 42 °C. Una suspensión de este nuevo bacteriófago, denominado D4724, se obtuvo mediante filtración del cultivo a través de un filtro de 0,45 µm. Se verificó la virulencia de D4724 contra DGCC7710^{phi2972}^{S6}.

A continuación, la cepa DGCC7710^{phi2972}^{S6} se expuso al bacteriófago donante D4724 y se aisló una cepa marcada, que se caracterizó usando la misma metodología descrita en el Ejemplo 1. En comparación con DGCC77710 esta cepa marcada (denominada DGCC1710^{phi2972}^{S6}^{phi4724}^{S15}) tenía unidades espaciadoras de repetición en su locus 2 de CRISPR1 como se describe en la Tabla 4-1.

Para los fines de una tercera iteración, se seleccionó un segundo fago mutante, denominado D4733, mediante estímulo de DGCC7710^{phi2972}^{S6}^{phi4724}^{S15} con D4724 usando la misma metodología que para obtener D4724. El bacteriófago D4733 es virulento contra DGCC7710^{phi2972}^{S4}^{phi4724}^{S15}. Tras exposición de virulencia de DGCC7710^{phi2972}^{S6}^{phi4724}^{S15} al bacteriófago donante D4733, se aisló una cepa marcada que se aisló y se caracterizó usando la misma metodología descrita en el Ejemplo 1. En comparación con DGCC77710, esta cepa marcada, denominada DGCC7710^{phi2972}^{S6}^{phi4724}^{S15}^{phi4733}^{S29}, tiene 3 unidades espaciadoras de repetición adicionales en su locus CRISPR1 como se describe en la Tabla 4-1.

Tabla 4-1. Descripción de cepas marcadas iterativamente en el locus de CRISPR1 Locus a partir de DGCC7710 usando D2972 y los fagos mutados D4724 y D4733 como bacteriófagos donantes					
Cepa marcada	Cepa progenitora	Fago donante	ADN donante insertado	Ubicación	
DGCC7710 _{ph} i2972 ^{+S6}	DGCC7 710	D2972	Secuencia GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:39)	34521-34492 (-)	
DGCC7710 _{ph} i2972 ^{+S6} ^{+S20} _{phi4724}	DGCC7 710 _{ph} i2972 ^{+S6}	D4724	GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:40) TTATATCGAAGAACGACTGAA AGAGCTTGA (SEQ ID NO:41)	34521-34492 (-) 1113-1142 (+)	
DGCC7710 _{ph} i2972 ^{+S6} ^{+S20} _{phi4724} ^{+S29} _{phi4733}	DGCC7710 _{ph} i2972 ^{+S6} ^{+S20} _{phi4724}	D4733	GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:42) TTATATCGAAGAACGACTGAA AGAGCTTGA (SEQ ID NO:43) ATTGGCATGATTTCAAATTITAA TTGGGAT (SEQ ID NO:44)	34521-34492 (-) 1113-1142 (+) 32136-32164 (+)	
*(+/-): indica la cepa del cromosoma del fago					

Ejemplo 5

Marcado de *Streptococcus thermophilus* DGCC3198 usando el bacteriófago D4241 mediante la inserción de una sola unidad espaciadora de repetición en el interior de CRISPR1

En este Example, *S. thermophilus* cepa DGCC3198 (también conocida como LMD-9 y depositada en la American Type Culture Collection como ATCC BAA-365) se marcó mediante inserción "natural" dentro de su locus CRISPR1 de una unidad espaciadora de repetición adicional donde el espaciador procede del bacteriófago D4241. El locus CRISPR1 de DGCC3198 contiene 17 repeticiones (incluyendo la repetición terminal) y 16 espaciadores (número de registro de GenBank: CP000419). El bacteriófago D4241 se aisló a partir de un producto lácteo fermentado usando la cepa DGCC3198. La secuencia del locus CRISPR1 de DGCC3198 es:

```
caagaacagttattgatttataatcactatgtgggtatgaaatctcaaaatcatttgag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACatgatgatgaagtatcgtaactactaaac
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctcacctcaaatcttagagctggactaaa
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACatgctgaaaaataaccgaccatcattact
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcgaagctatcagttaaggctaaaacctat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctgctaaatagatttctgcaccattgta
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcattcgtgaaaaaatatcgtgaaataggcaa
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctaggctcatctaaagataaatcagtagc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAActaaaacatggggcggcgtaaatagttaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcacaaccagcaaaagagagcggcagacaacatt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctaacacagggttagaggatgttatactt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctagaagctcaagcggtaaaagttagtgcg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAccttgaggcgaagccctcggcgttcattt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcaactaccaagcaaatcagcaatcaataagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcctataagtgaatcagcgtagggaatagc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcactagtcggtatattacccatagacgcta

GTTTTGTACTCTCAAGATTTAAGTAACTGTACAAcactagttactattatcaacgattccaacgg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTttgaticaacataaaagcgggtcaattg
aactggcctt (SEQ ID NO:45)
```

La cepa progenitora (DGCC3198) se expuso al bacteriófago donante (D4241) como se describe en el Ejemplo 1. Usando los métodos descritos en el Ejemplo 1, se obtuvieron los aislados que se caracterizaron adicionalmente. Las reacciones de la PCR y la determinación de la secuencia de ADN se realizaron como se describe en el Ejemplo 1, salvo que el cebador inverso utilizado en la PCR fue CR122-R3, con la siguiente secuencia: 5'-gctctaagatttgaggtagg-3' (SEQ ID NO:46).

Se crearon múltiples cepas marcadas a partir de DGCC3198 utilizando D4241 como donante. Estas cepas marcadas se describen en la Tabla 5-1. Todas estas cepas marcadas se diferencian de la cepa progenitora por una única unidad espaciadora de repetición adicional en el extremo líder del locus CRISPR1. Para las tres nuevas secuencias espaciadoras descritas en la Tabla 5-1, se realizó una búsqueda de homología con las secuencias disponibles en la base de datos pública que presentaran las mejores puntuaciones de homología con las subsecuencias del bacteriófago DT1 de *S. thermophilus* (número de registro de GenBank AF085222), confirmando que la nueva secuencia separadora procedía del bacteriófago.

Tabla 5-1. Descripción de cepas marcadas en el locus CRISPR1 derivado de DGCC3198 usando D4241 como bacteriófago donante			
Cepa marcada	Cepa progenitora	Fago donante	Secuencia del ADN donante insertado
DGCC3198 _{phi} 4241 ^{+S64}	DGCC3198	D4241	ACCAAGTAGCATTGAGCAAAGATAGATTG (SEQ ID NO:47)
DGCC3198 _{phi} 4241 ^{+S65}	DGCC3198	D4241	TAGATCTCATGAGTGGCGACAGTGAGCTT (SEQ ID NO:48)
DGCC3198 _{phi} 4241 ^{+S66}	DGCC73198	D4241	TACCATCTTGGGATAGGTACTGGTCATGCC (SEQ ID NQ:49)

Ejemplo 6

Marcado de *Streptococcus thermophilus* DGCC3198 usando el bacteriófago D4241 mediante la inserción de una sola unidad espaciadora de repetición en el interior de CRISPR3

En este Ejemplo, *S. thermophilus* cepa DGCC3198 se marcó mediante inserción "natural" dentro de su locus CRISPR3 de una unidad espaciadora de repetición adicional donde el espaciador procede del bacteriófago D4241. El locus CRISPR3 de DGCC3198 contiene 9 repeticiones (incluyendo la repetición terminal) y 8 espaciadores

(número de registro de GenBank: CP000419). La secuencia del locus CRISPR3 de DGCC3198 es:

```

taaatggtaataagfatagatagctctgagttattcaagactatcttttagtatttagtatttctgtatgaagtgaatgggataatcatttttagagagtaga
ttataaggattgafagaggaggaatgaattgcttgacatatgattatagaaataatciaataggtgacagtcacatcttgctaaaacgttgatataaag
gatttttaaggatataataataataaaaaatgaattatttgaagctgaagtcagctgagattaatagtgcgattacgaaatctgtagaaaagatacctacga
gGTTTTAGAGCTGTGTGTTTTCGAATGGTTCCAAAAACggtagaaaagggttcactgtacgagtacttaGTTTTAG
AGCTGTGTTGTTTTCGAATGGTTCCAAAAACtcaatgagtggtatccaagacgaaaacttaGTTTTAGAGCTGTGT
TGTTCGAATGGTTCCAAAAACccttgctgctgctcaccatagcccatataGTTTTAGAGCTGTGTGTTTTCGAA
TGGTTCCAAAAACgttgggaaaccgcagtagccatgattaaGTTTTAGAGCTGTGTGTTTTCGAATGGTTCCA
AAACacagagtacaatatgtcctcattggagacacGTTTTAGAGCTGTGTGTTTTCGAATGGTTCCAAAAACctcat
atcgttagtgctttgtcataaaGTTTTAGAGCTGTGTGTTTTCGAATGGTTCCAAAAACagaactttaicaagataaaac
tactttaaaGTTTTAGAGCTGTGTGTTTTCGAATGGTTCCAAAAACatagtaataatitcaitgaaaataattgtGTTTT
AGAGCTGTGTGTTTTCGAATGGTTCCAAAAACttttgtatcacaaatttcgggtgacatctcttagaactcatcttataaagg
agcttagtattgaatgtgagaaggac (SEQ ID NO:50)

```

La cepa progenitora (DGCC3198) se expuso al bacteriófago donante (D4241) como se describe en el Ejemplo 1. Usando los métodos descritos en el Ejemplo 1, se obtuvieron los aislados que se caracterizaron adicionalmente. Las reacciones de la PCR y la determinación de la secuencia de ADN se realizaron como se describe en el Ejemplo 1, salvo que se utilizaron CR3lead-F1 (5'-ctgagattaatagtgcgattacg-3'; SEQ ID NO:51) y CR3trail-R2 (5'-gctggatattcgataacatgtc-3'; SEQ ID NO:52).

Se creó una cepa marcada a partir de DGCC3198 usando D4241 como donante. Esta cepa marcada se describe en la Tabla 6-1. Esta cepa marcada se diferencia de la cepa progenitora por una única unidad espaciadora de repetición adicional en el extremo líder del locus CRISPR3. Para la nueva secuencias espaciadoras descritas en la Tabla 6, se realizó una búsqueda de homología con las secuencias disponibles en la base de datos pública que presentaran las mejores puntuaciones de homología con las subsecuencias del bacteriófago DT1 de *S. thermophilus* (número de registro de GenBank AF085222 y AF115102), confirmando que la nueva secuencia separadora procedía del bacteriófago.

Tabla 6-1. Descripción de la cepa marcada en el locus CRISPR3 derivado de DGCC3198 usando D4241 como bacteriófago donante			
Cepa marcada	Cepa progenitora	Fago donante	Secuencia del ADN donante insertado
DGCC3198 ^{+S67} 8 _{phi4241}	DGCC3198	D4241	5'-tgcaatttcattagtcttgacgcccttt-3'(SEQ ID NO:53)

Ejemplo 7

Método de la PCR específico para detectar cepas marcadas

En este ejemplo, se describen métodos de la PCR específicos para la detección específica de cepas marcadas. Cuando una cepa se marca de forma natural mediante la adición de una o más secuencia(s) de oligonucleótidos realizada con la adición de espaciadores en uno o más loci de CRISPR1, es necesario poder detectar estas cepas marcadas. Los métodos se basan en la presencia de secuencias específicas de las cepas marcadas que se han introducido dentro de una región identificada con precisión en el cromosoma de la cepa. Por tanto, se diseña un método de amplificación por PCR específico que es específico para las cepas marcadas. Una cepa desprovista de esta secuencia de oligonucleótidos única no da como resultado una amplificación del ADN mediante la PCR, mientras que la PCR que utiliza el ADN de la cepa marcada da como resultado un fragmento de ADN de longitud definida.

Para configurar el método de detección por PCR específica, se diseñaron 2 cebadores. Uno de los cebadores es el cebador directo, que no es específico de la cepa marcada pero lo es del locus CRISPR; se denomina "cebador CRISPR". En *S. thermophilus*, y dependiendo de la cepa marcada, el "cebador CRISPR" tiene una secuencia idéntica para el locus CRISPR1 ("cebador CRISPR1") o para el locus CRISPR3 ("cebador CRISPR3") o para el locus CRISPR2 ("cebador CRISPR2"). Los cebadores CRISPR se seleccionan entre las secuencias que están conservadas entre las cepas de las especies de interés. Para *S. thermophilus*, se recomiendan los siguientes cebadores: cebador CRISPR1, 5'-tgctgagacaacctagtctctc-3' (yc70, Bolotin *et al.*, [2005], *supra* [SEQ ID NO:15]); cebador CRISPR3, 5'-ctgagattaatagtgcgattacg-3'(CR3lead-F1; SEQ ID NO:51). El segundo cebador es el cebador inverso, y es específico de la cepa marcada; se denomina "cebador TAG". Los cebadores TAG son complementarios de una de las unidades espaciadoras de repetición de las cepas marcadas. Preferentemente, el cebador TAG es complementario del separador de la unidad espaciadora de repetición que está más alejada de la secuencia líder del locus CRISPR. La Figura 4 ilustra la ubicación del cebador CRISPR y el cebador TAG.

La Tabla 7-1 proporciona ejemplos de cebadores TAG para la detección de las cepas marcadas relacionadas en los Ejemplos 1 a 6.

Tabla 7-1. Cebadores utilizados para la detección de las cepas marcadas descritas en los Ejemplos 1 a 6.		
Cepa marcada	cebador CRISPR	cebador TAG
DGCC7710 _{phi2972} ^{+S40}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-aagagtagctccctcaatatgc-3' (SEQ ID NO:54)
DGCC7710 _{phi2972} ^{+S41}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-ttactaccactattcctagtg-3' (SEQ ID NO:55)
DGCC7710 _{phi2972} ^{+S43}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-atagaggccaactaacgttat-3' (SEQ ID NO:56)
DGCC7710 _{phi2972} ^{+S44}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-tatcaattgcggaaatgtggct-3' (SEQ ID NO:57)
DGCC7710 _{phi2972} ^{+S45}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-cactgtgtctcaataaacatt-3' (SEQ ID NO:58)
DGCC7710 _{phi2972} ^{+S46}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-agaacaagtttctgaacaaaa-3' (SEQ ID NO:59)
DGCC7710 _{phi858} ^{+S42}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-aaccactaataactttggctga-3' (SEQ ID NO:60)
DGCC7710 _{phi2972} ^{+S46+S47}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-agaacaagtttctgaacaaaa-3' (SEQ ID NO:61)
DGCC7710 _{phi2972} ^{+S48+S49}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-gctactgcctgtaagactt-3' (SEQ ID NO:62)
DGCC7710 _{phi2972} ^{+S50+S51}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-gagttctatttctgaattatt-3' (SEQ ID NO:63)
DGCC7710 _{phi2972} ^{+S52+S4}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-tggctggagccctcgatatta-3' (SEQ ID NO:64)
DGCC7710 _{phi2972} ^{+S53+S54}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-aatgatactggaccggaacc-3' (SEQ ID NO:65)
DGCC7710 _{phi2972} ^{+S61+S62}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-tcaaacgtaagacgaatatcgc-3' (SEQ ID NO:66)
DGCC7710 _{phi2972} ^{+S55+S63+S41}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-aagagtagctccctcaatatgc-3' (SEQ ID NO:54)
DGCC7710 _{phi2972} ^{+S6}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20} _{phi4733} ^{+S29}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC3198 _{phi4241} ^{+S64}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-caatctatcttgcctcaatgc-3' (SEQ ID NO:67)
DGCC3198 _{phi4241} ^{+S65}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-aagctcactgtcgccacttag-3' (SEQ ID NO:68)
DGCC3198 _{phi4241} ^{+S66}	5'-tgctgagacaacctagtcctc-3' (SEQ ID NO:15)	5'-ggcatgccagttacatccaa-3' (SEQ ID NO:69)
DGCC3198 _{phi4241} ^{+S67}	5'-ctgagattaatagtcgattacg-3' (SEQ ID NO:51)	5'-tcaagaactaattggaattgcag-3' (SEQ ID NO:70)

En estos experimentos, la muestra que contiene la cepa a detectar se trató usando cualquier método adecuado conocido en la técnica para clasificar las bacterias del resto de la muestra. Como en el ejemplo en el caso del yogur o muestras de productos lácteos frescos que contienen *S. thermophilus*, la muestra se trató como se describe en Lick *et al.* (Lick *et al.*, *Milchwissenschaft* 50:183-186 [1996]).

Dependiendo de la cantidad de bacteria incluida en el material resultante, las células de *S. thermophilus* se amplificaron mediante cultivo en medio M17-glucosa durante 18 horas a 42 °C. A continuación, el ADN se extrajo de las bacterias y se envió para una PCR específica. Los cebadores de PCR (cebador CRISPR y cebador TAG) se seleccionaron adecuadamente como se describe en la Tabla 7-1. La mezcla de reacción de la PCR (volumen final de 25 µl) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada una de las cuatro dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa *Taq* 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron: predesnaturalización a 98 °C durante 5 min, seguido de 33 ciclos alternando desnaturalización a 94 °C durante 30 s, hibridación a 56 °C durante 30 s, y elongación a 72 °C durante 1 min; seguido por una etapa de alargamiento final a 72 °C durante 4 min. En una PCR de control, el ADN extraído se sometió a una segunda PCR dirigida a los genes del ARN 16S usando los siguientes cebadores universales: BSF8-20, 5'-agagttgatcctggctcag-3' (SEQ ID NO:71) y BSRI 541-20, 5'-aaggaggtgatccagccgca-3' (SEQ ID NO:72; Véase, Wilimotte *et al.*, *FEBS Lett.*, 317:96-100 [1993]).

La mezcla reacción de la PCR (volumen final de 25 µl contenía): tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada una de las cuatro dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa *Taq* 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron las siguientes: predesnaturalización a 95 °C durante 7 min, seguido de 35 ciclos alternando desnaturalización a 95 °C durante 1 min, hibridación a 58 °C durante 1 min 30 s, y elongación a 72 °C durante 2 min 30 s; seguido por una etapa de alargamiento final a 72 °C durante 5 min. Los productos de amplificación de la PCR se analizaron a continuación usando electroforesis en gel de agarosa (1 %, p/v) y se registró el tamaño de los fragmentos del ADN amplificado. Para cada PCR específica, se prepararon controles usando el ADN extraído de la cepa progenitora (DGCC7710 o DGCC3198 dependiendo de la cepa marcada) y de una de las cepas marcadas (DGCC_{phi2972}^{S6}).

Los resultados se presentan en la Tabla 7-2. Cada una de las reacciones de la PCR era específica de la cepa marcada analizada, ya que los productos de la PCR del tamaño adecuado siempre se obtuvieron de las PCR del control y las PCR específicas solamente dieron como resultado fragmentos de ADN amplificados cuando se utilizó ADN de la cepa marcada específica.

Tabla 7-2: Tamaño de los productos de la PCR obtenidos mediante PCR específica y PCR del control para ADN de cepas marcadas, ADN de la cepa progenitora, y DGCC _{phi2972} ^{S6} DNA							
PCR específica ^c	ADN de la cepa marcada específica		ADN de DGCC _{phi2972} ^{S6}		ADN de la cepa progenitora		
	PCR ^a específica	PCR ^a del control	PCR ^a específica	PCR ^a del control	PCR ^a específica del control PCR ^a		
DGCC7710 _{phi2972} ^{+S40}	240	1530	0	1530	0 ^b	1530	
DGCC7710 _{phi2972} ^{+S41}	240	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S43}	240	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S44}	240	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S45}	240	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S46}	240	1530	0	1530	0	1530	
DGCC7710 _{phi858} ^{+S42}	240	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S46+S47}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S48+S49}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S50+S51}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S52+S54}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S53+S54}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S61+S62}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S55+S63+S41}	370	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S6}	240	1530	240	1530	0	1530	
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20}	310	1530	0	1530	0	1530	
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20} _{phi4733} ^{+S29}	370	1530	0	1530	0	1530	
DGCC3198 _{phi4241} ^{+S64}	240	1530	0	1530	0	1530	
DGCC3198 _{phi4241} ^{+S65}	240	1530	0	1530	0	1530	
DGCC3198 _{phi4241} ^{+S66}	240	1530	0	1530	0	1530	
DGCC3198 _{phi4241} ^{+S67}	105	1530	0	1530	0	1530	

^a, tamaño aproximado del fragmento amplificado en pares de bases; ^b, 0 significa que no se ha detectado fragmento de la PCR; ^c, usando el cebador específico de la PCR que se ha indicado en la Tabla 7-1.

Ejemplo 8

Método para identificar cepas marcadas

En este ejemplo se describe un método para detectar la presencia de cepas marcadas en una muestra e identificar su naturaleza. Esto se realiza mediante la amplificación por PCR de uno o más loci de CRISPR y determinación parcial de la secuencia. Este método es de utilizada en diferentes realizaciones, ya que aunque el método descrito en el Ejemplo 7 es útil para detectar una cepa marcada, puede que no sea suficiente para su identificación formal. Además, en algunos casos, la naturaleza de la cepa diana incluida en la muestra es desconocida. Así, el método de la PCR específica no puede usarse para su detección. En su caso, la cepa marcada se puede detectar e identificar mediante el análisis del locus CRISPR modificado.

Las células de *S. thermophilus* incluidas en una muestra se extrajeron de la muestra usando un método adecuado conocido en la técnica y se sembraron en placas en las diluciones adecuadas sobre agar M17-glucosa y se volvieron a incubar durante 24 horas a 42 °C para obtener colonias aisladas. Una o más colonias aisladas se repicaron a continuación y se hicieron crecer en medio M17-glucosa durante 18 horas a 42 °C. De cada cultivo, las células se

- recogieron y se extrajo su ADN. Para cada aislado, la región del cromosoma correspondiente al locus CRISPR1 se amplificó mediante la PCR usando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:15], Bolotin *et al.*, [2005]; más arriba); y el cebador inverso SPIDR-dws (5'-taaacagagcctccctatcc-3' [SEQ ID NO:73]). La mezcla reacción de la PCR (volumen final de 25 µl contenía): tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada una de las cuatro dNTP 2 mM, ADN 10 a 100 ng, cada cebador 0,2 µM, polimerasa Taq 1,25 U (Promega). Las condiciones de ciclación de la PCR fueron las siguientes: predesnaturalización a 98 °C durante 5 min, seguido de 33 ciclos alternando desnaturalización a 94 °C durante 30 s, hibridación a 56 °C durante 30 s, y elongación a 72 °C durante 1 min; seguido por una etapa de alargamiento final a 72 °C durante 4 min. En algunos casos, la región correspondiente al locus CRISPR3 se amplificó usando los siguientes cebadores: CR3lead-F1, 5'-ctgagattaatagtcgattacg-3' (SEQ ID NO:51) y CR3trail-R2, 5'-gctggatattcgataacatgac-3' (SEQ ID NO:52). La secuencia de ácido nucleico de cada fragmento de PCR se determinó usando una metodología de "cebador directo" y secuenciación clásica, como es conocido en la técnica. Cada secuencia se comparó a continuación con las secuencias de los loci de CRISPR disponibles en las bases de datos.
- En un experimento, el locus de CRISPR1 procedente de un aislado obtenido a partir de un producto de leche fermentada se sometió a la PCR y el amplicón resultante se sometió a secuenciación. La secuencia se comparó con la de las secuencias disponibles en las bases de datos. La Figura 5 proporciona los resultados de la comparación. Resulta que la secuencia era idéntica en un 100 % a la de la cepa DGCC7710, con una secuencia adicional de 66 nucleótidos. Esta secuencia de 66 nucleótidos se compone de 36 nucleótidos en su extremo 5' que es idéntica a la de las repeticiones del locus CRISPR1 de *S. thermophilus* y la secuencia restante de 30 nucleótidos es idéntica a una subsecuencia del bacteriófago D2972. Además, esta secuencia adicional de 66 nucleótidos se ubicó inmediatamente después de la secuencia líder de CRISPR1. En consecuencia, el locus CRISPR1 del aislado contiene el locus CRISPR1 de DGCC7710 con una unidad espaciadora de repetición adicional como se describe en la Figura 3. Además, la secuencia restante de 30 nucleótidos es idéntica también a la unidad espaciadora de repetición adicional de la cepa marcada DGCC_{phi2972}^{S41}. Esto indica de forma concluyente que el aislado obtenido del producto de leche fermentada era la cepa marcada DGCC_{phi2972}^{S41}.

REIVINDICACIONES

1. Un proceso para preparar un ingrediente alimenticio que comprende obtener una bacteria marcada mediante un método que comprende las etapas de:

(a) exponer una bacteria progenitora que comprende un locus CRISPR a un bacteriófago para introducir una unidad espaciadora de repeticiones adicionales en el locus CRISPR para producir una bacteria marcada, en el que la unidad espaciadora de repetición adicional proporciona la marca,

(b) seleccionar un mutante insensible a bacteriófago;

(c) comparar un locus CRISPR o una parte del mismo procedente de la bacteria progenitora y el mutante insensible a bacteriófago;

(d) seleccionar una bacteria marcada que comprende dicha unidad espaciadora de repetición adicional en el locus CRISPR que no está presente en la bacteria progenitora;

y que comprende además la etapa de añadir la bacteria marcada, o un cultivo celular que comprende dicha bacteria marcada, a dicho ingrediente alimenticio.

2. El proceso de acuerdo con la reivindicación 1, en el que la bacteria progenitora se selecciona entre el grupo de géneros que consiste en: *Bacillus*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Corynebacterium*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*.

3. El proceso de acuerdo con la reivindicación 1 o reivindicación 2, en el que dicho método comprende exponer la bacteria progenitora a dos o más bacteriófagos tanto simultánea como secuencialmente y en el que se selecciona una bacteria marcada que comprende dos o tres o más unidades espaciadoras de repetición adicionales.

4. El proceso de acuerdo con cualquier reivindicación anterior, en el que la bacteria marcada acumula las unidades espaciadoras de repetición en el interior de más de un loci CRISPR.

5. El proceso de acuerdo con cualquier reivindicación anterior, en el que la bacteria progenitora es *Streptococcus*.

6. El proceso de acuerdo con la reivindicación 5, en el que la bacteria progenitora es *Streptococcus thermophilus*.

7. El proceso de acuerdo con cualquier reivindicación anterior, en el que el bacteriófago se selecciona entre el grupo de familias de virus que consiste en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae y Tectiviridae.

8. El proceso de acuerdo con cualquier reivindicación anterior, en el que el bacteriófago puede infectar dicha bacteria progenitora.

9. El proceso de acuerdo con cualquier reivindicación anterior, en el que un cultivo celular que comprende la bacteria marcada se añade a dicho ingrediente alimenticio.

10. El proceso de acuerdo con la reivindicación 9, en el que dicho cultivo celular es un cultivo iniciador, un cultivo probiótico o un suplemento dietético.

11. El proceso de acuerdo con la reivindicación 10, en el que dicho cultivo celular es un cultivo iniciador.

12. El proceso de acuerdo con la reivindicación 11, en el que el cultivo iniciador es un cultivo congelado.

13. El proceso de acuerdo con la reivindicación 11, en el que el cultivo iniciador es un cultivo iniciador desecado.

14. El proceso de acuerdo con la reivindicación 13, en el que el cultivo iniciador desecado está liofilizado.

15. El proceso de acuerdo con la reivindicación 11, en el que el cultivo iniciador es un cultivo mixto definido.

16. El proceso de acuerdo con la reivindicación 11, en el que el cultivo iniciador es un cultivo iniciador concentrado utilizado en inoculación directa.

17. El proceso de cualquier reivindicación anterior, en el que dicho ingrediente alimenticio es adecuado para su adición a un alimento y en el que el alimento selecciona entre el grupo que consiste en productos de confitería, productos lácteos, productos cárnicos, productos de aves de corral, productos de pescado y productos de panadería.

18. El proceso de acuerdo con la reivindicación 17, en el que dicho alimento se selecciona entre el grupo que consiste en refrescos, zumos de frutas, bebidas que comprenden proteína de suero lácteo, infusiones saludables, bebidas de cacao, bebidas lácteas, bebidas que incluyen bacterias acidolácticas, queso, yogur, yogur líquido y vino.

Figura 1

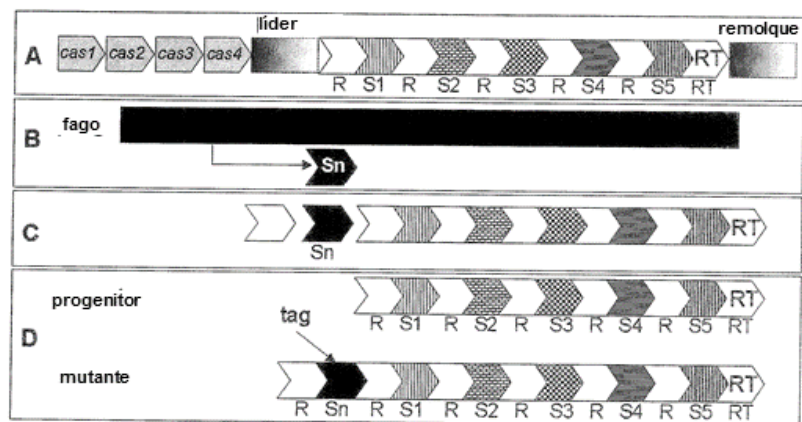


Figura 2

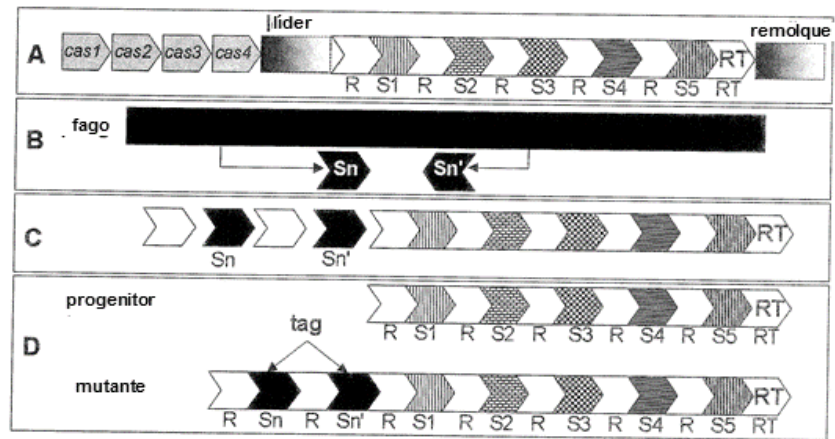


Figura 3

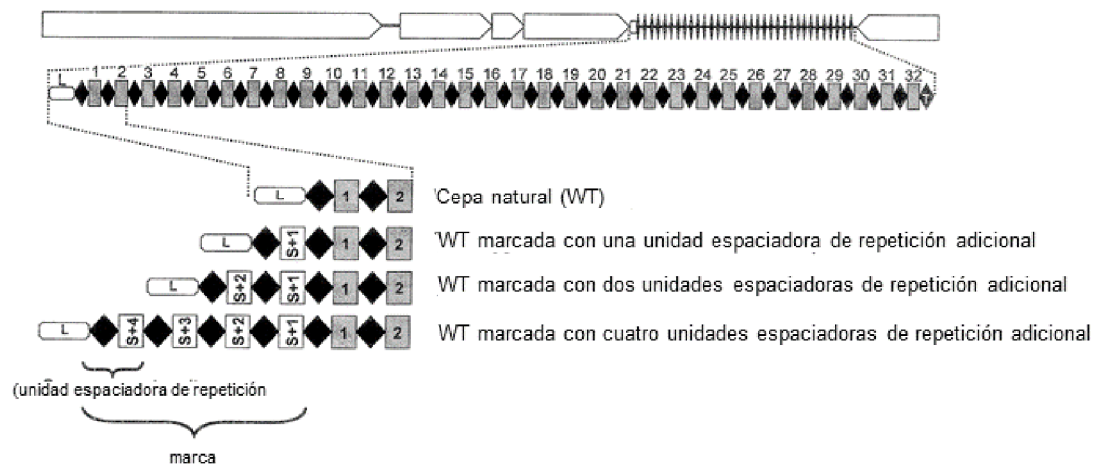


Figura 4

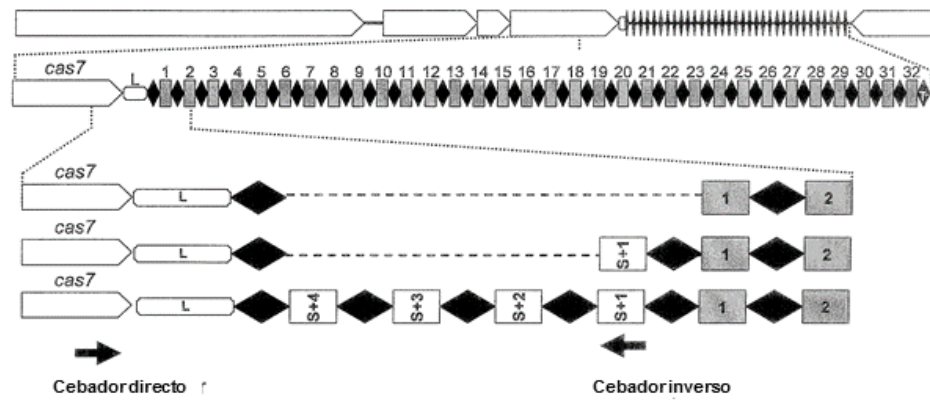


Figura 5

```

DGCC7710 CRISPR1      5'..caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
Isolate CRISPR1 5'      5'..caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
Consensus CRISPR 1 repeat
DGCC77109812972241 additional spacer

DGCC7710 CRISPR1      ..... GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC TGTTT
Isolate CRISPR1      AGTAACTGTACAAC TCTAATCCCACTAGGAATAGTGGGTAGTAA GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC TGTTT
Consensus CRISPR 1 repeat AGTAACTGTACAAC 3' 5' GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC 3'
DGCC77109812972241 additional spacer TCTAATCCCACTAGGAATAGTGGGTAGTAA 3'

DGCC7710 CRISPR1      GACAGCAAATCAAGATTGCAATTGT GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC AATGA.....3'
Isolate CRISPR1      GACAGCAAATCAAGATTGCAATTGT GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC AATGA.....3'
Consensus CRISPR 1 repeat GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC 5' GTTTTGTACTCTCAAGATTAAAGTAACTGTACAAC 3'
DGCC77109812972241 additional spacer

```