

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 373 586**

51 Int. Cl.:
C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07777180 .6**
96 Fecha de presentación: **18.05.2007**
97 Número de publicación de la solicitud: **2018441**
97 Fecha de publicación de la solicitud: **28.01.2009**

54 Título: **MICROORGANISMOS MARCADOS Y MÉTODOS PARA MARCAR.**

30 Prioridad:
19.05.2006 US 747682 P
02.03.2007 US 904721 P

45 Fecha de publicación de la mención BOPI:
06.02.2012

45 Fecha de la publicación del folleto de la patente:
06.02.2012

73 Titular/es:
DANISCO A/S
LANGEBROGADE 1
1001 COPENHAGEN, DK

72 Inventor/es:
BARRANGOU, Rodolphe;
FREMAUX, Christophe;
HORVATH, Philippe y
ROMERO, Dennis

74 Agente: **de Elzaburu Márquez, Alberto**

ES 2 373 586 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 para marcar.

Campo de la invención

5 La presente invención proporciona métodos para marcar microorganismos. En algunas realizaciones preferidas, los microorganismos son bacterias. En algunas realizaciones particularmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. Los microorganismos marcados empleando los métodos descritos en esta memoria, se exponen en esta memoria. 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.

Antecedentes de la invención

15 Las cepas microbianas, especialmente las empleadas como cultivos iniciadores para la fermentación en la industria alimentaria/de bebidas, están muy seleccionadas y caracterizadas para funciones específicas. Típicamente, estos cultivos iniciadores se venden en el comercio como organismos vivos y a menudo siguen siendo viables en el producto final. Por lo tanto, es posible aislar, identificar y caracterizar las cepas de los cultivos iniciadores, mediante el cultivo de los microorganismos presentes en los productos finales. Además, también es posible utilizar a continuación esas cepas de cultivo iniciador en otros productos, incluyendo productos de la competencia. Es difícil controlar el uso de tales cepas por parte de terceros, incluidos los competidores. De hecho, en la técnica existe una falta de métodos para marcar fácilmente los cultivos, con el fin de identificar su(s) fuente(s) y vigilar su uso en diferentes productos.

Sumario de la invención

25 La presente invención proporciona métodos para marcar microorganismos. En algunas realizaciones preferidas, los microorganismos son bacterias. En algunas realizaciones particularmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. Los microorganismos marcados empleando los métodos descritos en esta memoria, también se exponen. Los microorganismos marcados pueden ser bacterias. Las bacterias marcadas pueden ser miembros del género *Streptococcus* o ser miembros de otros géneros.

La presente invención proporciona un método tal y como se describe en la reivindicación 1

30 Los productos alimenticios y/o los piensos que comprenden la bacteria marcada y/o cultivos celulares que comprenden al menos una especie bacteriana marcada, también se describen en esta memoria. Los procedimientos para preparar productos alimenticios y/o piensos que comprenden bacterias marcadas y/o cultivos celulares que comprenden al menos una especie bacteriana marcada, también se describen en esta memoria. Los métodos para preparar alimentos y/o piensos que comprenden la etapa de añadir la bacteria marcada o el cultivo celular a dicho producto alimenticio o pienso, también se describen en esta memoria. Los alimentos y/o los piensos obtenidos u obtenibles empleando los métodos de la presente invención, también se describen.

En realizaciones adicionales, la presente invención proporciona métodos para generar variantes de CRISPR, tal y como se describe en la reivindicación 19. Las variantes de CRISPR obtenidas u obtenibles empleando los métodos de la presente invención, se describen en esta memoria

40 Los métodos para identificar una bacteria marcada que comprenden las etapas de: (a) escrutar la bacteria en busca de una unidad adicional de repetición-espaciador en un locus CRISPR; (b) determinar la secuencia de nucleótidos de la unidad adicional de repetición-espaciador; (c) comparar la secuencia de nucleótidos de la unidad de repetición-espaciador adicional con una base de datos de bacterias marcadas, obtenidas u obtenibles por el método de la presente invención; e (d) identificar una secuencia de nucleótidos en la base de datos de las bacterias marcadas que se empareja con la unidad adicional de repetición-espaciador, también se describen en esta memoria.

45 En algunas realizaciones, se compara el extremo 5' y/o el extremo 3' del locus CRISPR de la bacteria original. En algunas realizaciones particularmente preferidas, se comparan al menos la primera repetición de CRISPR y/o el primer espaciador de CRISPR (*p. ej.*, la primera agrupación de espaciadores de CRISPR) en el extremo 5' del locus CRISPR. En aún otras realizaciones, se compara al menos la última repetición de CRISPR y/o el último espaciador de CRISPR (*p. ej.*, la última agrupación de espaciadores de CRISPR) en el extremo 3' del locus CRISPR. En otras realizaciones preferidas de los métodos de la presente invención, los métodos comprenden la etapa de seleccionar una bacteria marcada que comprende una unidad adicional de repetición-espaciador en el extremo 5' y/o en el extremo 3' del locus CRISPR que no está presente en la bacteria original. En algunas realizaciones, los métodos de la presente invención comprenden la exposición de la bacteria original a dos o más bacteriófagos de forma simultánea o secuencial.

En algunas realizaciones adicionales, se compara el locus CRISPR o al menos una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos, amplificando el locus CRISPR o una porción del mismo procedente de la bacteria original y/o del mutante insensible a los bacteriófagos. En algunas realizaciones preferidas, la amplificación se realiza empleando la PCR. En algunas realizaciones adicionales, se compara el locus CRISPR o al menos una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos, secuenciando el locus CRISPR o una porción del mismo procedente de la bacteria original y/o del mutante insensible a los bacteriófagos. En algunas realizaciones particularmente preferidas, se compara el locus CRISPR o al menos una porción del mismo procedente de la bacteria original y el mutante insensible a los bacteriófagos, amplificando y secuenciando a continuación el locus CRISPR o una porción del mismo procedente de la bacteria original y/o del mutante insensible a los bacteriófagos. En algunas realizaciones, la unidad adicional de repetición-espaciador tiene al menos 44 nucleótidos de longitud. En algunas realizaciones adicionales, se selecciona una bacteria marcada que comprende dos, tres o más unidades adicionales de repetición-espaciador. En algunas realizaciones preferidas, la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene al menos aproximadamente 95% de identidad, o más preferentemente, aproximadamente 100% de identidad con una repetición de CRISPR en el locus CRISPR de la bacteria original. En algunas realizaciones preferidas alternativas, la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene al menos 95% de identidad, preferentemente, 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 adicional de repetición-espaciador comprende una primera secuencia de nucleótidos que tiene al menos aproximadamente 95% de identidad o, más preferentemente, aproximadamente 100% de identidad con una repetición de CRISPR en el locus CRISPR de la bacteria original, y una segunda secuencia de nucleótidos que tiene al menos una secuencia de nucleótidos que tiene al menos aproximadamente 95% de identidad o, con preferencia, aproximadamente 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 adicional de repetición-espaciador con una base de datos de secuencias de bacteriófagos y/o una base de datos de secuencias bacterianas.

En algunas realizaciones preferidas, la bacteria original es adecuada para el uso como cultivo iniciador, cultivo probiótico y/o suplemento dietético. En algunas realizaciones preferidas, la bacteria original se selecciona entre el grupo de géneros consistente 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 en cultivos celulares, incluyendo pero no limitados a cultivos iniciadores, cultivos probióticos y/o suplementos dietéticos.

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

En algunas realizaciones, la presente descripción proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde dicha secuencia comprende SEQ ID NO:3 y/o 4. En algunas realizaciones, la presente descripción proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde la secuencia comprende SEQ ID NO:3 y/o 4, situadas aguas abajo (p. ej., directamente aguas abajo) de la primera repetición de CRISPR en al menos un locus CRISPR. En todavía otras realizaciones, la presente descripción proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde la secuencia comprende SEQ ID NO:9. En algunas realizaciones preferidas, *S. thermophilus* comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde la secuencia comprende SEQ ID NO:9 aguas abajo (p. ej., directamente aguas abajo) de la primera repetición de CRISPR en al menos un locus CRISPR. En todavía otras realizaciones, la presente descripción proporciona *S. thermophilus* que comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde la secuencia comprende SEQ ID NO:11. En todavía otras realizaciones, *S. thermophilus* comprende una secuencia obtenida u obtenible a partir de un bacteriófago, en donde dicha secuencia comprende SEQ ID NO:11 aguas abajo (p. ej., directamente aguas abajo) de la primera repetición de CRISPR en al menos un locus CRISPR.

Los métodos para marcar una bacteria comprenden las etapas de: exponer al menos una bacteria original que comprende al menos una porción de un locus CRISPR, al menos a una secuencia de ácido nucleico exógeno para producir al menos una bacteria marcada que comprende un locus CRISPR modificado, en donde el locus CRISPR modificado comprende al menos una unidad adicional de repetición-espaciador más que el locus CRISPR de la bacteria original y en donde la unidad adicional de repetición-espaciador comprende una marca; y comparar al menos una porción del locus CRISPR de la bacteria original y el locus CRISPR modificado de dicha bacteria marcada. En algunas realizaciones, la secuencia exógena de ácido nucleico se selecciona entre bacteriófagos, plásmidos, megaplásmidos, elementos transponibles, transposones, y también se describen en la memoria secuencias de inserción. El ácido nucleico exógeno puede comprender al menos una porción del genoma de al

menos un bacteriófago. En algunas realizaciones alternativas, particularmente preferidas, la bacteria marcada es un mutante insensible a los bacteriófagos. En algunas realizaciones adicionales, el extremo 5' y/o el extremo 3' del locus CRISPR de la bacteria original, se compara con el locus CRISPR modificado de la bacteria marcada. En realizaciones adicionales, el extremo 5' y/o 3' de al menos la primera repetición de CRISPR y/o al menos el primer espaciador de CRISPR del locus CRISPR de la bacteria original, se comparan con el locus CRISPR modificado de la bacteria marcada. En algunas realizaciones adicionales, los métodos comprenden adicionalmente la etapa de seleccionar al menos una bacteria marcada. En realizaciones aún más adicionales, la bacteria original se expone de forma simultánea o secuencial a dos o varios bacteriófagos. La bacteria marcada comprende al menos una unidad adicional de repetición-espaciador. En algunas realizaciones alternativas, al menos una porción del locus CRISPR de la bacteria original, y al menos una porción del locus CRISPR modificado de la bacteria marcada, se comparan amplificando al menos una porción del locus CRISPR y al menos una porción del locus CRISPR modificado, para producir una secuencia amplificada del locus CRISPR y una secuencia amplificada del locus CRISPR modificado. En algunas realizaciones preferidas, la amplificación comprende el uso de la reacción en cadena de la polimerasa. En algunas realizaciones preferidas alternativas, al menos una porción del locus CRISPR de la bacteria original y al menos una porción del locus CRISPR modificado de dicha bacteria, se comparan mediante la secuenciación de al menos una porción del locus CRISPR y al menos una porción del locus CRISPR modificado. En algunas realizaciones adicionales, los métodos comprenden adicionalmente la etapa de secuenciar la secuencia amplificada del locus CRISPR y la secuencia amplificada del locus CRISPR modificado. En algunas realizaciones preferidas, la unidad adicional de repetición-espaciador comprende al menos aproximadamente 44 nucleótidos. En algunas realizaciones adicionales, la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene al menos 95% de identidad con una repetición de CRISPR en el locus CRISPR de la bacteria original. En algunas realizaciones adicionales, la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene al menos 95% de identidad con una secuencia de nucleótidos en el genoma de al menos un bacteriófago. En todavía realizaciones adicionales, la bacteria marcada comprende adicionalmente al menos una secuencia adicional de nucleótidos que tiene al menos 95% de identidad con una secuencia de nucleótidos en el genoma de al menos un bacteriófago. En algunas realizaciones preferidas, la bacteria original es un cultivo con utilidad industrial. En algunas realizaciones particularmente preferidas, la bacteria original comprende un cultivo seleccionado a partir de cultivos iniciadores, cultivos probióticos y cultivos de suplementos dietéticos. En algunas realizaciones adicionales particularmente preferidas, la bacteria original 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, al menos un bacteriófago se selecciona entre el grupo de familias víricas consistente en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae y Tectiviridae. Las bacterias marcadas obtenidas empleando los métodos descritos en esta memoria, se exponen también en esta memoria. En algunas realizaciones preferidas, la bacteria marcada es un cultivo con utilidad industrial. Los cultivos celulares que comprenden las bacterias marcadas, producidas empleando los métodos descritos en esta memoria, también se exponen en esta memoria. El cultivo celular que comprende las bacterias marcadas, puede comprender un cultivo con utilidad industrial. En algunas realizaciones particularmente preferidas, la bacteria marcada comprende un cultivo seleccionado entre cultivos iniciadores, cultivos probióticos y cultivos de suplementos dietéticos. Los alimentos y/o piensos que comprenden una bacteria marcada obtenida empleando los métodos descritos en esta memoria, se describen en esta memoria.

Los métodos para preparar alimentos y/o piensos que comprenden el uso de bacterias marcadas, en donde las bacterias marcadas se añaden al alimento o al pienso, también se exponen en esta memoria. En algunas realizaciones adicionales, el cultivo celular que comprende bacterias marcadas, comprende un cultivo con utilidad industrial. En algunas realizaciones particularmente preferidas, la bacteria marcada comprende un cultivo seleccionado entre cultivos iniciadores, cultivos probióticos y cultivos de suplementos dietéticos. Los alimentos y/o piensos que comprenden una bacteria marcada obtenida empleando los métodos descritos en esta memoria, también se exponen en esta memoria. Al menos una variante de CRISPR, obtenida empleando los métodos descritos en esta memoria, también se expone en esta memoria. Los cultivos celulares que comprenden al menos una variante de CRISPR producida empleando los métodos proporcionados en esta memoria, también se exponen en esta memoria. Los alimentos y/o piensos que comprenden al menos una variante de CRISPR producida empleando los métodos descritos en esta memoria, también se exponen en esta memoria.

Los métodos para identificar una bacteria marcada que comprende al menos un locus CRISPR, que comprenden las etapas de: escrutar una bacteria marcada en busca de la presencia de un marcador en el locus CRISPR; determinar la secuencia de nucleótidos del marcador; comparar la secuencia de nucleótidos del marcador con la secuencia de nucleótidos presente en al menos una base de datos; e identificar una secuencia de nucleótidos en la base de datos que comparte homología con el marcador, también se exponen en esta memoria. La base de datos puede comprender secuencias de nucleótidos de las bacterias marcadas. La base de datos se puede seleccionar entre las bases de datos de secuencias de bacteriófagos y las bases de datos de secuencias bacterianas.

Descripción de las Figuras

La Figura 1 ilustra una realización de la presente invención en donde una secuencia marcadora y una repetición de CRISPR se integran en un extremo del locus CRISPR. El panel A muestra un locus CRISPR y elementos, que incluyen las repeticiones (R), los espaciadores (S), el líder aguas arriba y el remolque aguas abajo, con la repetición terminal (RT) adyacente al remolque, y los genes *cas* en la proximidad (4 genes *cas* denominados *cas 1* a *cas 4*, en este ejemplo, no están dibujados a escala). Los genes *cas* pueden estar en algún extremo o separados y presentes en ambos extremos. Los genes *cas* se pueden localizar sobre cualquiera de las dos cadenas de ADN. El panel B muestra una secuencia del 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 (próximo al 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 original 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 (próximo al líder, en este ejemplo), entre las repeticiones. El nuevo espaciador (*Sn*) constituye la secuencia marcadora que es específica de la bacteria mutante (es decir, la bacteria marcada). En algunas realizaciones, este procedimiento tiene como resultado la adición de uno o varios espaciadores procedentes de la secuencia del fago.

La Figura 2 ilustra una realización de la presente invención en la que dos secuencias marcadoras y dos repeticiones de CRISPR se integran en un extremo del locus CRISPR. El panel A muestra un locus CRISPR y elementos, que incluyen repeticiones (R), espaciadores (S), el líder aguas arriba y el remolque aguas abajo, con la repetición terminal (RT) adyacente al remolque, y los genes *cas* en la proximidad (4 genes *cas* denominados *cas 1* a *cas 4* en este ejemplo, no están dibujados a escala). Los genes *cas* pueden estar en algún extremo o separados y presentes en ambos extremos. Los genes *cas* se pueden localizar sobre cualquiera de las dos cadenas de ADN. El panel B muestra una secuencia de fago en negro, utilizándose dos fragmentos de la secuencia (*Sn* y *Sn'*) como espaciadores adicionales (es decir, secuencias marcadoras). El panel C ilustra la inserción de los nuevos espaciadores (es decir, las secuencias marcadoras) (*Sn* y *Sn'*), en el mismo extremo del locus CRISPR (próximo al 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 original y la bacteria mutante (es decir, la bacteria marcada), con dos nuevos espaciadores (*Sn* y *Sn'*) integrados en el mismo extremo del locus CRISPR (próximo al 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 procedimiento tiene como resultado la adición de uno o varios 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, los recuadros con forma de flecha representan genes en la proximidad de CRISPR1. La región espaciadora-de repetición de WT está en el centro, indicándose las repeticiones (rombos negros), los espaciadores (recuadros grises numerados), el líder (L, recuadro blanco) y la repetición terminal (T, rombo negro). En la parte inferior, las variantes marcadas se representan con 1, 2 o 4 unidades adicionales de repetición-espaciador (rombos negros asociados con el recuadro blanco S+n). La combinación de las unidades de repetición-espaciador adicionales representa el marcador. En esta figura, se utiliza el nuevo sistema de numeración de los genes *cas*. Previamente, los genes *cas* se habían numerado *cas 1*, *cas 2*, *cas 3* y *cas 4*. Ahora los genes *cas* se numeran *cas 5*, *cas 1*, *cas 6*, y *cas 7*. Por lo tanto, para correlacionar las designaciones en las Figuras 1 y 2 con las de las Figuras 3 y 4, *cas 1* en las Figuras 1 y 2 se refiere a *cas 5* en las Figuras 3 y 4; *cas 2* en las Figuras 1 y 2, es *cas 1* en las Figuras 3 y 4; *cas 3* en las Figuras 1 y 2, es *cas 6* en las Figuras 3 y 4; y *cas 4* en las Figuras 1 y 2, es *cas 7* en las Figuras 3 y 4.

La Figura 4 proporciona una representación esquemática del locus CRISPR1 de *S. thermophilus* DCGG7710 y la posición de los cebadores para la detección con PCR. Se utilizan los mismos símbolos que en la Figura 3. Las posiciones del cebador directo y del cebador inverso se indican. También, tal y como se ha indicado en la descripción de la Figura 3 anteriormente, se utiliza el nuevo sistema de numeración de los genes *cas*. Previamente, los genes *cas* se numeraron como *cas 1*, *cas 2*, *cas 3* y *cas 4*. Ahora, los genes *cas* se numeran *cas 5*, *cas 1*, *cas 6* y *cas 7*. Por tanto, para correlacionar las designaciones en las Figuras 1 y 2 con las de las Figuras 3 y 4, *cas 1* en las Figuras 1 y 2 se refiere a *cas 5* en las Figuras 3 y 4; *cas 2* en las Figuras 1 y 2, es *cas 1* en las Figuras 3 y 4; *cas 3* en las Figuras 1 y 2, es *cas 6* en las Figuras 3 y 4; y *cas 4* en las Figuras 1 y 2, es *cas 7* en las Figuras 3 y 4.

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

Descripción de la Invención

- La presente invención proporciona métodos para marcar microorganismos. En algunas realizaciones preferidas, los microorganismos son bacterias. En algunas realizaciones particularmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. La presente invención también proporciona microorganismos marcados empleando los métodos descritos en esta memoria. 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.
- Existe una falta de métodos y composiciones útiles en la identificación de cepas bacterianas específicas, para determinar su origen. Aunque se puede insertar un oligonucleótido sintético en una cepa para marcarla o etiquetarla, empleando tecnologías de ADN recombinante, la cepa marcada se consideraría que es un organismo genéticamente modificado y es probable que se enfrente a problemas de reglamentación en aplicaciones comerciales.
- Además, la preparación de cultivos es una labor intensiva que requiere mucho espacio y equipamiento, y existe un riesgo considerable de contaminación con bacterias y/o fagos deteriorados, durante las etapas de propagación. La pérdida de cultivos bacterianos debido a una infección con bacteriófagos (fagos) y la multiplicación de los mismos, es un problema principal del uso industrial de cultivos bacterianos. Existen muchos tipos diferentes de fagos y siguen surgiendo nuevas cepas. Por tanto, hay una necesidad de métodos y composiciones para el seguimiento y la vigilancia de las bacterias utilizadas en tales cultivos.
- Cuando una población bacteriana se infecta con un bacteriófago virulento, muchas de las células son destruidas por el bacteriófago. Sin embargo, frecuentemente se producen mutantes espontáneos, resistentes al fago. Estas bacterias resistentes a bacteriófagos se corresponden con una subpoblación de bacterias que son capaces de soportar y sobrevivir una infección con bacteriófagos.
- Tal y como se describe en esta memoria, cuando un bacteriófago infecta una bacteria, una o varias secuencias que proceden del genoma del bacteriófago se integran en (p. ej., dentro) del locus CRISPR de la bacteria, mientras que en otras realizaciones, la integración tiene lugar en otros lugares dentro del genoma de la bacteria. De hecho, en algunas realizaciones, "mutantes resistentes a bacteriófago"/"mutantes insensibles a bacteriófagos"/"BIMs" tienen una secuencia del bacteriófago integrada en CRISPR, mientras que en otras realizaciones, las cepas tienen otro tipo de mutación cromosómica. Sin embargo, se pretende que las bacterias marcadas se produzcan debido a la integración de la secuencia del fago en CRISPR. Por tanto, la secuencia obtenible u obtenida a partir del bacteriófago es nueva para el locus CRISPR de la bacteria y proporciona una "marca" o "etiqueta" que es identificable por su posición y/o por su secuencia y/o por la secuencia adyacente. Se ha encontrado que una secuencia duplicada (p. ej., una repetición de CRISPR duplicada) que procede de la bacteria original, también se integra de forma reiterativa, secuencial, de forma simultánea o sustancialmente simultánea, junto con la secuencia procedente del genoma del bacteriófago.
- Además, en algunas realizaciones, la infección independiente de un cultivo (p. ej., un cultivo puro) de una cepa bacteriana dada, empleando el mismo bacteriófago virulento, conduce a la integración de una o varias secuencias diferentes del bacteriófago en el locus CRISPR de la cepa bacteriana. En algunas realizaciones preferidas, la integración de secuencias diferentes del bacteriófago en el locus CRISPR de la cepa bacteriana, es un hecho aleatorio. Sin embargo, en algunas otras realizaciones, la integración no es un hecho al azar. Una vez que se integra, se mantiene y se convierte de este modo en un medio sólido para marcar y/o rastrear una bacteria. Por tanto, una o varias secuencias procedentes del genoma del bacteriófago no son únicamente nuevas para el locus CRISPR de la bacteria original, sino que también son un marcador que es único para cada bacteria. De este modo, la presente invención proporciona composiciones y métodos para marcar (es decir, señalar) y/o identificar una bacteria. Además, los métodos y las composiciones de la presente invención son particularmente ventajosos porque el método es un método 'natural', que no da lugar a un organismo genéticamente modificado. Por lo tanto, las bacterias marcadas o señaladas, preparadas de acuerdo con los métodos de la presente invención, no se consideran que estén genéticamente modificadas, puesto que las bacterias se han creado mediante un procedimiento biológico natural, de infección con bacteriófago.
- La presente descripción proporciona métodos para el uso de una secuencia obtenida u obtenible a partir de un bacteriófago (p. ej., en la preparación de una bacteria marcada) para marcar y/o identificar una bacteria, en donde la secuencia se integra en un extremo del locus CRISPR de una bacteria original. En algunas realizaciones preferidas, la presente descripción proporciona métodos para el uso de una secuencia obtenida u obtenible a partir de un bacteriófago (p. ej., en la preparación de una bacteria marcada) para marcar y/o identificar una bacteria, en donde la secuencia comprende: (i) al menos una secuencia que es homóloga (p. ej., idéntica) a una repetición de CRISPR en el locus CRISPR de la bacteria; y (ii) una secuencia marcadora. En realizaciones adicionales, la presente descripción proporciona métodos para el uso de una secuencia para marcar y/o para identificar una bacteria (p. ej., en la preparación de una bacteria marcada), en donde la secuencia se obtiene o es obtenible mediante: (a) la exposición de una bacteria original a un bacteriófago; (b) la selección de un mutante insensible a los bacteriófagos; (c) la comparación del locus CRISPR o de una porción del mismo procedente de la bacteria original y el mutante

insensible a los bacteriófagos; y (d) la selección de una secuencia en el locus CRISPR o de una porción del mismo del mutante insensible a los bacteriófagos, que no está presente en la bacteria original.

5 En algunas realizaciones adicionales, las secuencias de la presente descripción (p. ej., una secuencia recombinante o una secuencia de ácido nucleico aislada) consisten esencialmente en al menos un gen o una 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 (p. ej., ADNc). Sin embargo, en algunas realizaciones particularmente preferidas, la secuencia es una secuencia de ácido nucleico presente en la naturaleza. En algunas realizaciones, la secuencia de nucleótidos es bicatenaria, mientras que en otra realización, es monocatenaria. En algunas realizaciones adicionales, la secuencia de ácido nucleico representa la cadena codificante, mientras que en otras realizaciones representa la cadena no codificante o combinaciones de las mismas. Las secuencias de ácido nucleico recombinante preparadas empleando cualquier técnica recombinante adecuada, conocida en la técnica, encuentran uso en la presente invención. En algunas realizaciones preferidas, la secuencia de ácido nucleico es un gen o se obtiene a partir de un gen. En algunas realizaciones particularmente preferidas, la secuencia de ácido nucleico es una "secuencia de ácido nucleico exógeno" que se introduce en una bacteria original empleando cualquier método adecuado conocido en la técnica, incluyendo pero sin estar limitados a los mismos, los métodos naturales y recombinantes. En algunas realizaciones preferidas más particularmente, la secuencia de ácido nucleico exógeno comprende al menos una porción de un genoma de bacteriófago. En algunas realizaciones adicionales particularmente preferidas, la secuencia de ácido nucleico exógeno se introduce en una bacteria original mediante la exposición de la bacteria, al menos a un bacteriófago.

10 En algunas realizaciones, la secuencia de ácido nucleico y los ácidos nucleicos incluidos en la presente descripción, se aíslan o se purifican sustancialmente. Tal y como se emplea en esta memoria, las expresiones "aislado" o "sustancialmente purificado", se refieren a moléculas de ácido nucleico, a fragmentos biológicamente activos, a variantes, a homólogos o a derivados de los mismos que están exentos de forma sustancial o esencial, de componentes que se encuentran normalmente asociados con el ácido nucleico en su estado natural. Tales componentes incluyen otro material celular, medios de cultivo de producción recombinante y diversos compuestos químicos empleados en la síntesis química de los ácidos nucleicos. En algunas realizaciones, una secuencia de ácido nucleico o un ácido nucleico aislado", está típicamente exento de secuencias de ácido nucleico que flanquean el ácido nucleico de interés en el ADN genómico del organismo a partir del cual se va a obtener el ácido nucleico (p. ej., secuencias codificadoras presentes en los extremos 5' o 3'). Sin embargo, en algunas realizaciones, la molécula incluye algunas bases o restos adicionales que no afecten de forma perjudicial a las características básicas de la composición.

15 La(s) secuencia(s) de ácido nucleico encuentra(n) uso en células modificadas genéticamente (p. ej., una célula receptora). En algunas realizaciones, la secuencia de ácido nucleico se inserta en el ADN (p. ej., ADN plasmídico o ADN genómico) de una célula receptora, empleando cualquier método adecuado, conocido en la técnica (p. ej., recombinación homóloga). En otras realizaciones, la(s) secuencia(s) de ácido nucleico encuentra(n) uso como molde a partir del cual se modifica (p. ej., muta) el ADN de una célula (p. ej., una célula receptora), tal como ADN plasmídico o ADN genómico, bajo condiciones tales que la(s) secuencia(s) de ácido nucleico se crea en el ADN de la célula. En algunas realizaciones preferidas, la(s) secuencia(s) de ácido nucleico se clona(n) (p. ej., en una estructura artificial, un plásmido o un vector) que se usa a continuación para transformar la célula, empleando cualquier método adecuado conocido en la técnica.

20 La presente descripción proporciona métodos y composiciones que emplean variantes, homólogos, derivados y fragmentos de los mismos. El término "variante" se emplea en esta memoria haciendo referencia a secuencias de polipéptido o de nucleótido presentes en la naturaleza, que difieren de una secuencia de tipo silvestre. Tal y como se emplea en esta memoria, el término "fragmento" se refiere a una secuencia de polipéptido o de nucleótido que comprende una fracción de una secuencia de tipo silvestre. En algunas realizaciones, los fragmentos comprenden una o varias secciones contiguas de la secuencia o una pluralidad de secciones pequeñas. En algunas realizaciones, la secuencia también comprende otros elementos. Por ejemplo, en algunas realizaciones, es una proteína de fusión que incluye otra secuencia proteica. En algunas realizaciones preferidas, la secuencia comprende al menos aproximadamente 50%, más preferentemente al menos aproximadamente 65%, más preferentemente al menos aproximadamente 80%, más preferentemente al menos aproximadamente 85%, más preferentemente al menos 90%, más preferentemente al menos 95%, más preferentemente al menos 96%, más preferentemente al menos aproximadamente 97%, más preferentemente al menos aproximadamente 98% o, lo más preferible, al menos aproximadamente 99% de la secuencia de tipo silvestre.

25 En algunas realizaciones particularmente preferidas, el fragmento es un fragmento funcional. Tal y como se emplea en esta memoria, un "fragmento funcional" de una molécula, se refiere a un fragmento que conserva o que posee sustancialmente la misma actividad biológica que la molécula intacta. En realizaciones particularmente preferidas, los fragmentos funcionales conservan al menos aproximadamente 10%. En otras realizaciones, al menos aproximadamente 25%, aproximadamente 50%, aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de la actividad biológica de la molécula intacta. En realizaciones alternativas, el fragmento conserva aproximadamente 50%, más preferentemente aproximadamente 60%, más

preferentemente aproximadamente 70%, más preferentemente aproximadamente 80%, más preferentemente aproximadamente 85%, más preferentemente aproximadamente 90%, más preferentemente aproximadamente 95%, más preferentemente aproximadamente 96%, más preferentemente aproximadamente 97%, más preferentemente aproximadamente 98% o, lo más preferible, aproximadamente 99% de la actividad de la secuencia del polipéptido o del nucleótido de tipo silvestre.

Tal y como se emplea en esta memoria, el término "homólogo" se refiere a una entidad que tiene cierta homología con las secuencias de aminoácidos objeto y las secuencias de nucleótidos objeto. Tal y como se emplea particularmente en esta memoria, el término "homología" es sinónimo de "identidad." En el presente contexto, una secuencia homóloga se entiende que comprende una secuencia de aminoácidos, que tiene al menos aproximadamente 75, aproximadamente 85 o aproximadamente 90% de identidad, preferentemente al menos aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad con la secuencia objeto. Aunque la homología también se puede considerar en términos de similitud (es decir, residuos de aminoácidos que tienen propiedades/funciones químicas similares), en el contexto de la presente invención, se prefiere expresar la homología en términos de identidad de la secuencia. En algunas realizaciones preferidas, una secuencia homóloga comprende una secuencia de nucleótidos que tiene una identidad de al menos aproximadamente 75, aproximadamente 85 o aproximadamente 90%, preferentemente una identidad de al menos aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99%, con la secuencia objeto (es decir, la secuencia de interés empleada como referencia). En algunas realizaciones, las comparaciones de la homología se realizan visualmente, aunque otros métodos conocidos en la técnica encuentran uso (p. ej., con la ayuda de programas disponibles para la comparación de secuencias). Los programas informáticos disponibles en el mercado son capaces de calcular el porcentaje de homología (% de homología) entre dos o varias secuencias, y por ello se pueden utilizar en la presente invención. De hecho, los métodos y sistemas para tales análisis se adquieren fácilmente en el mercado.

Descripción detallada de la invención

La presente invención proporciona métodos para marcar microorganismos. En algunas realizaciones preferidas, los microorganismos son bacterias. En algunas realizaciones particularmente preferidas, las bacterias son miembros del género *Streptococcus*, mientras que en otras realizaciones, las bacterias son miembros de otros géneros. La presente descripción también proporciona microorganismos marcados empleando los métodos descritos en esta memoria. 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. CRISPRs y loci CRISPR

Las CRISPRs (Repeticiones palindrómicas cortas agrupadas separadas por espacios de longitud similar, del inglés "Clustered Regularly Interspaced Short Palindromic Repeats"); también conocidas como Repeticiones directas dispersas entre espaciadores (SPIDRs - "Spacer Interspersed Direct Repeats") constituyen una familia de loci de ADN descritos recientemente que son generalmente específicos de una especie bacteriana particular. El locus CRISPR es una clase distinta de repeticiones de secuencias cortas (SSRs) que se reconocieron por primera vez en *E. coli* (Ishino y col., J. Bacteriol., 169:5429-5433 [1987]; y Nakata y col., J. Bacteriol. 171:3553-3556 [1989]). SSRs similares dispersas se han identificado en *Haloflex mediterranei*, *Streptococcus pyogenes*, *Anabaena* y *Mycobacterium tuberculosis* (véase, Groenen y col., Mol. Microbiol. 10:1057-1065 [1993]; Hoe y col., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl y col., Biochim. Biophys. Acta 1307:26-30 [1996]; y Mojica y col., Mol. Microbiol. 17:85-93 [1995]). Los loci CRISPR difieren de otras SSRs por la estructura de las repeticiones, que se han denominado repeticiones cortas, separadas regularmente (SRSRs) (Janssen y col., OMICS J. Integ. Biol., 6:23-33 [2002]; y Mojica y col., Mol. Microbiol. 36:244-246 [2000]). Las repeticiones son elementos cortos que se presentan en agrupaciones que siempre están separados regularmente por secuencias intermedias únicas, con una longitud constante (Mojica y col., [2000], *supra*). Aunque las secuencias repetidas están altamente conservadas entre las cepas, el número de repeticiones intercaladas y las secuencias de las regiones espaciadoras difieren de cepa a cepa (van Embden y col., J. Bacteriol., 182:2393-2401 [2000]). Los loci CRISPR constan de repeticiones de ADN parcialmente palindrómico muy conservado, típicamente de 24 a 40 pb. Estas repeticiones se han descrito en un rango de 1 a 249. Aunque se han detectado elementos aislados, por lo general se organizan en agrupaciones (hasta aproximadamente 20 o más por genoma) de unidades repetidas, separadas por secuencias intercaladas únicas de 20-58 pb. Hasta la fecha, se han encontrado hasta 20 loci distintos de CRISPR dentro de un cromosoma aislado. Las CRISPRs son generalmente homogéneas dentro de un genoma dado, siendo idénticas la mayoría de las mismas. Sin embargo, hay ejemplos de heterogeneidad, por ejemplo, en las Archaea (Mojica y col., [2000], *supra*). Tal y como se emplea en esta memoria, la expresión "locus CRISPR" se refiere al segmento de ADN que incluye todas las repeticiones de CRISPR, empezando por el primer nucleótido de la primera repetición de CRISPR y terminando con el último nucleótido de la última (terminal) repetición de CRISPR. En algunas realizaciones alternativas, se emplea "al menos una porción" de al menos un locus CRISPR. Por tanto, se pretende que la presente invención abarque realizaciones en las que se use al menos un locus CRISPR completo, así como realizaciones en las que se use al menos una porción (es decir, parte de al menos un locus CRISPR). Aunque la función biológica de los loci CRISPR es desconocida, se han propuesto algunas hipótesis. Por ejemplo, se ha propuesto que pueden estar implicadas en la fijación del cromosoma a una estructura celular, o en la replicación

5 cromosómica y la partición del replicón (Jansen y col., OMICS 6:23-33 [2002]; Jansen y col., Mol. Microbiol., 43:1565-1575 [2002]; y Pourcel y col., Microbiol., 151:653-663 [2005]). Mojica y col. (Mojica y col., J. Mol. Evol., 60:174-182 [2005]) plantean la hipótesis de que CRISPR puede estar implicada en conferir inmunidad específica contra ADN ajeno y Pourcel y col. (*supra*) plantean la hipótesis de que las CRISPRs son estructuras que son capaces de recoger pedazos de ADN extraño como parte de un mecanismo de defensa. Bolotin y col. (*supra*) sugieren que los elementos espaciadores de CRISPR son las huellas de invasiones pasadas de elementos extracromosómicos, y plantean la hipótesis de que proporcionan una célula con inmunidad frente a una infección con fagos, y más generalmente la expresión de ADN extraño, codificando un ARN no codificante. Bolotin y col. (*supra*) también sugieren que los genes *cas* son necesarios para la formación de CRISPR. Sin embargo, no se pretende que la presente invención se limite a ningún mecanismo, función, teoría particulares, ni a los medios de actuación.

B. Identificación de los loci CRISPR

15 Se conocen en la técnica diversos métodos para identificar loci CRISPR. Por ejemplo, Jensen y col. (Jensen y col., [2002], *supra*) describen un enfoque basado en informática, en el que en secuencias de nucleótidos, se buscan motivos de CRISPR, utilizando el programa PATSCAN en el servidor de "Mathematics and Computer Science Division" en el Laboratorio Nacional de Argonne, Argonne, IL, EE.UU. El algoritmo que se utilizó para identificar los motivos de CRISPR era $p1 = a...b \ c... \ d \ p1 \ c...d \ p1 \ c...d \ p1$, en donde *a* y *b* eran el límite superior e inferior del tamaño de la repetición y *p1*, *c* y *d* eran el límite superior e inferior del tamaño de las secuencias espaciadoras. Los valores de *a*, *b*, *c* y *d* pueden variar desde aproximadamente 15 hasta aproximadamente 70 pb, en incrementos de aproximadamente 5 pb. En algunas realizaciones preferidas, los loci CRISPR se identifican empleando gráficas de puntos (p. ej., empleando el programa informático de Dotter).

20 Cualquier método conocido en la técnica se emplea en el análisis de la similitud de las secuencias. Por ejemplo, el análisis se puede realizar empleando NCBI BLAST con una base de datos de genomas microbianos y GenBank, como se conoce en la técnica. Además, las secuencias de nucleótidos, incluyendo las proporcionadas en esta memoria, se incluyen en las bases de datos (p. ej., GenBank o la página web de JGI).

25 En realizaciones adicionales, los métodos de la presente invención emplean procedimientos de amplificación (véase, p. ej., Mojica y col., [2005], *supra*; y Pourcel y col., [2005], *supra*). La amplificación de la región deseada del ADN se puede conseguir mediante cualquier método conocido e 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 realiza generalmente empleando tecnologías de PCR. La "reacción en cadena de la polimerasa" ("PCR") es bien conocida por los expertos en la técnica. En la presente invención, se diseñan cebadores de oligonucleótidos para el uso en reacciones PCR para amplificar todo o parte de un locus CRISPR. El término "cebador" se refiere a un oligonucleótido, tanto presente en la naturaleza, como material purificado de la digestión con enzimas de restricción, o producido sintéticamente, que es capaz de actuar como punto de iniciación de la síntesis, cuando se coloca bajo condiciones en las que se induce la síntesis de un producto de extensión del cebador, que es complementario a una hebra de ácido nucleico (es decir, en presencia de nucleótidos y un agente inductor, tal como la polimerasa de ADN, y a una temperatura y pH adecuados). En algunas realizaciones, el cebador es monocatenario para obtener una eficacia máxima en la amplificación, aunque en otras realizaciones, el cebador es bicatenario. En algunas realizaciones, el cebador es un oligodesoxirribonucleótido. El cebador tiene que ser suficientemente largo para cebar la síntesis de los productos de extensión en presencia del agente inductor. La longitud exacta de los cebadores depende de muchos factores, que incluyen la temperatura, la fuente de cebador y el uso del método. Los cebadores de la PCR tienen típicamente al menos aproximadamente 10 nucleótidos de longitud y más típicamente al menos aproximadamente 20 nucleótidos de longitud. Los métodos para diseñar y realizar la PCR son bien conocidos en la técnica e incluyen, pero no están limitados a métodos en los que se emplean cebadores emparejados, cebadores anidados, cebadores específicos aislados, cebadores degenerados, cebadores específicos de un gen, cebadores específicos de un vector, cebadores parcialmente emparejados de forma errónea, etc.

30 En algunas realizaciones preferidas de la presente invención, un locus CRISPR o una porción del mismo procedente de una bacteria original y una bacteria marcada, se comparan empleando cualquier método adecuado conocido en la técnica. En algunas realizaciones preferidas de la presente invención, el locus CRISPR o una porción del mismo procedente de la bacteria original y de la bacteria marcada, se comparan mediante la amplificación del locus CRISPR o una porción del mismo. Además de los métodos bien conocidos de amplificación por ciclación (p. ej., PCR, reacción en cadena de la ligasa, etc.), otros métodos que incluyen pero no están limitados a los métodos de la amplificación isotérmica, se emplean en la presente invención. Los métodos de amplificación isotérmica bien conocidos que se emplean en la presente invención, incluyen pero no están limitados a la amplificación por desplazamiento de la cadena (SDA), Q-beta-replicasa, amplificación de secuencias basada en ácidos nucleicos (NASBA) y replicación de secuencias automantenida.

35 En algunas otras realizaciones preferidas de la presente invención, se compara el locus CRISPR o una porción del mismo procedente de la bacteria original y de la bacteria marcada, mediante secuenciación del locus CRISPR o de una porción del mismo procedente de la bacteria original y de la bacteria marcada. En algunas realizaciones alternativas, se comparan amplificando y secuenciando a continuación los loci CRISPR o una porción de los mismos. En algunas realizaciones, se compara un extremo de los loci CRISPR procedentes de las bacterias originales y

marcadas, mientras que en otras realizaciones, se comparan ambos extremos 5' y 3' de los loci. En algunas realizaciones preferidas, se compara un extremo (p. ej., el extremo 5') de los loci CRISPR. En todavía otras realizaciones, se compara al menos la última repetición de CRISPR en el extremo 3' del locus CRISPR y/o al menos el último espaciador de CRISPR (p. ej., la última agrupación de espaciadores de CRISPR) en el extremo 3' del locus CRISPR y/o al menos la primera repetición de CRISPR en el extremo 5' del locus CRISPR y/o al menos el primer espaciador de CRISPR (p. ej., la primera agrupación de espaciadores de CRISPR) en el extremo 5' del locus CRISPR. En algunas realizaciones preferidas, se compara al menos la primera repetición de CRISPR en el extremo 5' del locus CRISPR y/o al menos el primer espaciador de CRISPR (p. ej., la primera agrupación de espaciadores de CRISPR) en el extremo 5' del locus CRISPR. En algunas realizaciones adicionales preferidas, se compara al menos el último espaciador de CRISPR (p. ej., la última agrupación de espaciadores de CRISPR) en el extremo 3' del locus CRISPR y/o al menos el primer espaciador de CRISPR (p. ej., la primera agrupación de espaciadores de CRISPR) en el extremo 5' del locus CRISPR. En algunas realizaciones adicionales preferidas, se compara al menos el primer espaciador de CRISPR (p. ej., la primera agrupación de espaciadores de CRISPR) en los extremos 5' de los loci CRISPR.

En algunas realizaciones, los loci CRISPR comprenden ADN, mientras que en otras realizaciones, los loci CRISPR comprenden ARN. En algunas realizaciones, el ácido nucleico tiene origen genómico, mientras que en otras realizaciones, tiene origen sintético o recombinante. En algunas realizaciones, los loci CRISPR son bicatenarios, mientras que en otras realizaciones, son monocatenarios, independientemente de que represente la cadena codificante o la no codificante o combinaciones de las mismas. En algunas realizaciones, los loci CRISPR se preparan empleando técnicas de ADN recombinante (p. ej., ADN recombinante), tal y como se describen en esta memoria.

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* (CRISPR-associated) están generalmente próximos a los loci CRISPR. Por ejemplo, en el cromosoma de la cepa CNRZ1066 de *S. thermophilus*, el locus CRISPR1 está localizado aguas abajo de los genes *cas str0657*, *str0658*, *str0659* y *str0660*. El locus CRISPR1 se orienta de forma colineal con los genes *cas*. Por tanto, los genes *cas* están localizados aguas arriba de CRISPR1. La secuencia no codificadora localizada entre el codón de detención del último gen *cas* y el primer nucleótido de la primera repetición de CRISPR, está localizada aguas arriba de CRISPR y se denomina en esta memoria "líder de CRISPR". El líder de CRISPR está localizado en el extremo 5' del locus CRISPR. La secuencia no codificadora en el lado opuesto del locus CRISPR, se denomina en esta memoria "remolque de CRISPR". El remolque de CRISPR comienza justo después del último nucleótido de la última repetición de CRISPR. Esta última repetición de CRISPR también se denomina "repetición terminal". El remolque de CRISPR y las repeticiones terminales se localizan en el extremo 3' del locus CRISPR. Por ejemplo, el líder de CRISPR en la cepa CNRZ1066 tiene la secuencia 5'-CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATAAAAACGTCAAATTTTCATTTGAG-3' (SEQ ID NO:12), y el remolque de CRISPR tiene la secuencia 5'-TTGATTCAACATAAAAAGCCAGTTCAATTGAACCTGGCTTT-3' (SEQ ID NO:13). El líder de CRISPR se corresponde a las posiciones 625038 a 625100, y el remolque de CRISPR se corresponde a las posiciones 627845 a 627885 en el genoma de CNRZ1066 de *S. thermophilus* (CP000024).

Tal y como se emplea en esta memoria, la expresión "una porción 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 (p. ej., al menos aproximadamente 44-98 nucleótidos) de un locus CRISPR.

En algunas realizaciones adicionales, la expresión "una porción del mismo" en el contexto de un locus CRISPR, significa al menos aproximadamente los 10 primeros 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 (p. ej., al menos aproximadamente 44-98 nucleótidos) aguas abajo del primer nucleótido de la primera repetición de CRISPR, en el extremo 5' de un locus CRISPR o aguas arriba desde el último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR. En algunas realizaciones preferidas, la expresión "una porción del mismo" se refiere a al menos a aproximadamente los primeros 44 nucleótidos aguas abajo desde el primer nucleótido de la primera repetición de CRISPR, en el extremo 5' de un locus CRISPR o al menos aproximadamente 44 nucleótidos aguas arriba desde el 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 de aproximadamente 24 nucleótidos y el tamaño mínimo de la secuencia marcadora es de aproximadamente 20 nucleótidos. Por lo tanto, en algunas realizaciones preferidas, la expresión "una porción 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 de aproximadamente 40 nucleótidos y el tamaño máximo de la secuencia marcadora es de aproximadamente 58 nucleótidos. Por lo tanto, en algunas realizaciones, la expresión "una porción del mismo" cuando se emplea en el contexto de un locus CRISPR, significa al menos aproximadamente 98 nucleótidos. En algunas realizaciones preferidas, la expresión "una porción del mismo" en el contexto de un locus CRISPR, significa al menos aproximadamente 44-98 nucleótidos.

La presente invención también proporciona variantes de CRISPR, así como métodos para generar variantes de CRISPR. En realizaciones adicionales, las variantes de CRISPR se aíslan, se clonan y/o se secuencian empleando métodos conocidos en la técnica. En algunas realizaciones, las variantes de CRISPR se emplean como dianas para fines de detección y/o de identificación, mientras que en realizaciones alternativas, las variantes de CRISPR se emplean en modificar genéticamente la resistencia contra moléculas de ácido nucleico.

C. Extremo de un locus CRISPR

Cuando se compara el locus CRISPR o una porción del mismo procedente de la bacteria original y de una bacteria marcada, se compara 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 (*p. ej.*, al menos aproximadamente 44-98 nucleótidos) de un locus CRISPR. En algunas realizaciones preferidas, se compara 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 (*p. ej.*, al menos aproximadamente 44-98 nucleótidos) en uno o en ambos extremos de un locus CRISPR.

En algunas realizaciones preferidas, se compara al menos aproximadamente los 10 primeros 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 (*p. ej.*, 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 compara al menos aproximadamente los 44 primeros nucleótidos en el extremo 5' de un locus CRISPR o aproximadamente los 44 últimos nucleótidos en el extremo 3' de un locus CRISPR.

En algunas realizaciones, se compara al menos aproximadamente los 10 primeros 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 (*p. ej.*, al menos aproximadamente 44-98 nucleótidos) aguas abajo desde el primer nucleótido de la primera repetición de CRISPR, en el extremo 5' de un locus CRISPR, o aguas arriba desde el último nucleótido de la última repetición de CRISPR en el extremo 3' de un locus CRISPR. En algunas realizaciones preferidas, se compara al menos aproximadamente los 44 primeros nucleótidos aguas abajo desde el primer nucleótido de la primera repetición de CRISPR, en el extremo 5' de un locus CRISPR o aproximadamente al menos 44 nucleótidos aguas arriba desde el 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 de aproximadamente 24 nucleótidos y el tamaño mínimo de la secuencia marcadora es de aproximadamente 20 nucleótidos. En algunas realizaciones preferidas, se comparan al menos 44 nucleótidos. En algunas realizaciones alternativas, el tamaño máximo de la secuencia duplicada es de aproximadamente 40 nucleótidos y el tamaño máximo de la secuencia marcadora es de aproximadamente 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

Tal y como se emplea en esta memoria, la expresión "repetición de CRISPR" tiene el significado convencional que se emplea en la técnica (es decir, múltiples repeticiones directas cortas, que no muestran una variación de la secuencia, o muy ligera variación dentro de un locus CRISPR dado). Tal y como se emplea en esta memoria, en contexto, "repetición de CRISPR" es un sinónimo del término "CRISPR".

Un locus CRISPR comprende una o varias repeticiones más de CRISPR que los espaciadores de CRISPR existentes. Por lo tanto, la repetición de CRISPR se corresponde con la secuencia repetida dentro de un locus

CRISPR. Por ejemplo, a excepción de la repetición terminal, la secuencia típica de una repetición de la secuencia de CRISPR1 de *S. thermophilus* es:

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

5 Variaciones puntuales de esta secuencia de la repetición se han observado en secuencias de repeticiones en un locus CRISPR de una cepa dada y en secuencias de repeticiones en un locus CRISPR o en cepas procedentes de especies dadas, pero son muy raras. Comparada con esta secuencia típica de la repetición, la secuencia de la repetición terminal de un locus CRISPR dado, siempre muestra la misma variación en su extremo 3'. Variaciones puntuales de esta secuencia de la repetición terminal, también se han observado, pero son raras. Las repeticiones de CRISPR pueden aparecer de forma natural en la bacteria original. Los números de orden de GenBank para las secuencias de CRISPR1 incluyen: CP000023, CP000024, 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.

15 Tal y como se describe con más detalle en esta memoria, una secuencia duplicada se deriva, se puede derivar, se obtiene o se puede obtener a partir de una bacteria original. En algunas realizaciones preferidas, la secuencia comprende el ADN genómico de una bacteria original. En algunas realizaciones particularmente preferidas, la repetición de CRISPR duplicada (p. ej., en el mismo locus CRISPR) se integra de forma reiterativa, secuencial, simultánea o sustancialmente simultánea, junto con la secuencia marcadora, en la bacteria original, para dar lugar a una bacteria marcada.

20 El número de nucleótidos en una repetición es generalmente de aproximadamente 20 a aproximadamente 40 pares de bases (p. ej., aproximadamente 36 pares de bases). Sin embargo, no se pretende que la presente invención esté limitada por ningún intervalo particular dentro de aproximadamente 20 y aproximadamente 40 pares de bases. De hecho, se pretende que toda limitación numérica máxima dada a lo largo de esta memoria descriptiva, incluya todas las limitaciones numéricas inferiores, como si tales limitaciones numéricas inferiores estuvieran expresamente escritas en esta memoria. Cada limitación numérica inferior dada a lo largo de esta memoria descriptiva, incluirá todas las limitaciones numéricas superiores, como si tales limitaciones numéricas superiores, estuvieran expresamente escritas en esta memoria. Cada intervalo numérico dado a lo largo de esta memoria descriptiva, incluirá cada intervalo numérico más estrecho que esté incluido dentro de ese intervalo numérico más extenso, como si tales intervalos numéricos más estrechos estuvieran todos expresamente escritos en esta memoria.

30 En realizaciones adicionales, el número de repeticiones está en el intervalo desde aproximadamente 1 hasta aproximadamente 250. Sin embargo, no se pretende que la presente invención esté limitada por ningún intervalo particular dentro de aproximadamente 1 y aproximadamente 250 repeticiones. En efecto, tal y como se ha indicado anteriormente, se pretende que cada limitación numérica máxima dada a lo largo de esta memoria descriptiva, incluya cada limitación numérica inferior, como si cada limitación numérica inferior estuviera expresamente escrita en esta memoria. Cada limitación numérica inferior dada a lo largo de esta memoria descriptiva, incluirá todas las limitaciones numéricas superiores, como si tales limitaciones numéricas superiores estuvieran expresamente escritas en esta memoria. Cada intervalo numérico dado a lo largo de esta memoria descriptiva, incluirá cada intervalo numérico más estrecho que esté incluido dentro de ese intervalo numérico más extenso, como si tales intervalos numéricos más estrechos estuvieran todos expresamente escritos en esta memoria. Realmente se pretende que esto se aplique a todos los intervalos numéricos proporcionados en esta memoria.

45 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 tiene origen genómico, mientras que en otras realizaciones, tiene origen sintético o recombinante. En algunas realizaciones, los genes de las repeticiones de CRISPR son bicatenarios o monocatenarios, representando la cadena codificadora o la no codificadora, o combinaciones de las mismas. En algunas realizaciones, los genes de las repeticiones de CRISPR se preparan empleando técnicas de ADN recombinante (p. ej., ADN recombinante), tal y como se ha descrito en esta memoria.

50 En algunas realizaciones, una o varias de las repeticiones de CRISPR se emplean para modificar genéticamente una célula (p. ej., una célula receptora). Por ejemplo, en algunas realizaciones, la(s) repetición(es) de CRISPR se inserta(n) en el ADN de una célula (p. ej., el ADN plasmídico y/o genómico de una célula receptora), empleando cualquier método adecuado conocido en la técnica. En realizaciones adicionales, la(s) repetición(es) de CRISPR se emplea(n) como molde según el cual se modifica (p. ej., se muta) el ADN de una célula (p. ej., ADN plasmídico y/o genómico de una célula receptora), de modo que se crean o se modifican genéticamente las repeticiones de CRISPR en el ADN de la célula. En realizaciones adicionales, la(s) repetición(es) de CRISPR está(n) presente(s) en al menos una estructura artificial, en al menos un plásmido y/o en al menos un vector, etc. En realizaciones adicionales, las repeticiones de CRISPR se introducen en la célula empleando cualquier método adecuado, conocido en la técnica.

En algunas realizaciones adicionales, uno o varios genes o proteínas *cas* se emplean, junto con o en combinación, con una o varias, 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, forman una combinación funcional, tal y como se describe a continuación.

E. Espaciador de CRISPR

5 Tal y como se emplea en esta memoria, un "espaciador de CRISPR" incluye secuencias de espaciador no repetitivas que se localizan entre las repeticiones (es decir, las repeticiones de CRISPR) de los loci CRISPR. En algunas realizaciones de la presente invención, un "espaciador de CRISPR" se refiere al segmento de ácido nucleico que está flanqueado por dos repeticiones de CRISPR. Se ha observado que las secuencias de espaciadores de CRISPR tienen frecuentemente similitudes significativas con una variedad de moléculas de ADN móvil (*p. ej.*, bacteriófagos y plásmidos). En algunas realizaciones preferidas, los espaciadores de CRISPR se localizan entre dos repeticiones de
10 CRISPR idénticas. En algunas realizaciones, los espaciadores de CRISPR se identifican mediante análisis de la secuencia de tramos de ADN localizados entre dos repeticiones de CRISPR.

Curiosamente, se ha observado que las células que son portadoras de estos espaciadores de CRISPR, no se pueden infectar con moléculas de ADN que contienen secuencias homólogas a estos espaciadores (Mojica y *col.* [2005], *supra*). En la mayoría de las realizaciones, el espaciador de CRISPR está presente de forma natural entre
15 dos repeticiones directas cortas, múltiples e 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 a 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 la secuencia. En realizaciones preferidas, el análisis de las
20 secuencias homólogas incluye un espaciador de CRISPR, que en algunas realizaciones tiene una identidad de 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% con la secuencia de ácido nucleico diana o con un producto de la transcripción del mismo o con una secuencia identificada (*p. ej.*, una secuencia de interés). En algunas realizaciones, el espaciador de CRISPR tiene una
25 identidad del 100% con la secuencia de ácido nucleico de la diana.

El número de espaciadores de CRISPR en los loci o en un locus CRISPR dado, puede variar entre las especies. En algunas realizaciones preferidas, el número de espaciadores está en el intervalo desde aproximadamente 1 hasta aproximadamente 248. Sin embargo, no se pretende que la presente invención esté limitada por ningún intervalo particular entre aproximadamente 1 y aproximadamente 140 espaciadores. En efecto, tal y como se ha indicado
30 anteriormente, se pretende que cada limitación numérica máxima dada a lo largo de esta memoria descriptiva, incluya cada limitación numérica inferior, como si cada limitación numérica inferior estuviera expresamente escrita en esta memoria. Cada limitación numérica inferior dada a lo largo de esta memoria descriptiva, incluirá todas las limitaciones numéricas superiores, como si tales limitaciones numéricas superiores estuvieran expresamente escritas en esta memoria. Cada intervalo numérico dado a lo largo de esta memoria descriptiva, incluirá cada
35 intervalo numérico más estrecho que está incluido dentro de ese intervalo numérico más extenso, como si tales intervalos numéricos más estrechos estuvieran todos expresamente escritos en esta memoria. Realmente se pretende que esto se aplique a todos los intervalos numéricos proporcionados en esta memoria.

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

Tal y como se describe en esta memoria, la presente invención proporciona métodos y composiciones que facilitan el uso de uno o de varios genes o proteínas *cas*, en combinación con una o varias repeticiones de CRISPR, preferentemente dos, adecuados para conferir una especificidad de la inmunidad a al menos un espaciador de
45 CRISPR en una célula receptora. En algunas realizaciones preferidas, al menos un gen o proteína *cas* y al menos una repetición de CRISPR, se emplean en combinaciones funcionales para conferir una especificidad de la inmunidad a al menos un espaciador de CRISPR en una célula.

Tal y como se emplea en esta memoria, la expresión "especificidad de la inmunidad" significa que la inmunidad se confiere frente a una secuencia de ácido nucleico específica o un producto de la transcripción de la misma, empleando un espaciador de CRISPR específico o una secuencia espaciadora de pseudo-CRISPR. Tal y como se indica en esta memoria, un espaciador de CRISPR dado, no confiere resistencia frente a ninguna secuencia de
50 ácido nucleico o ningún producto de la transcripción de la misma, sino solo frente a las secuencias cuyo espaciador de CRISPR o espaciador de pseudo-CRISPR es homólogo (*p. ej.*, las que tienen una identidad del 100%).

En algunas realizaciones, el(los) espaciador(es) de CRISPR se obtiene(n) a partir 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, al menos un gen o proteína *cas*
55 y/o al menos una repetición de CRISPR, se obtienen a partir de un organismo diferente 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 obtiene al menos un gen y/o proteína *cas*, y/o al menos una repetición de CRISPR. En algunas realizaciones alternativas preferidas, los espaciadores de CRISPR se obtienen a partir de un organismo que es homólogo al receptor o a otra célula donante, a partir del cual se obtiene al menos un gen y/o proteína *cas*, y/o al menos una repetición de CRISPR. En algunas realizaciones preferidas, el(los) espaciador(es) de CRISPR se diseña(n) y se produce(n) empleando métodos recombinantes conocidos en la técnica. De hecho, se pretende que los espaciadores de CRISPR se produzcan empleando cualquier método adecuado conocido en la técnica.

En algunas realizaciones, los espaciadores de CRISPR son heterólogos con la célula receptora a partir de la cual se obtienen al menos un gen o proteína *cas*, y/o al menos una repetición de CRISPR (en algunas realizaciones, preferentemente dos o más). En algunas realizaciones alternativas, los espaciadores de CRISPR son homólogos con la célula receptora a partir de la cual se obtiene al menos un gen o proteína *cas* y/o al menos una repetición de CRISPR (en algunas realizaciones, preferentemente dos o más). De hecho, se pretende que cualquiera de los elementos utilizados en los métodos, sea heterólogo u homólogo. En algunas realizaciones, en las que se emplean múltiples elementos (*p. ej.*, cualquier combinación de espaciador(es) de CRISPR, de repetición(es) de CRISPR, de gen(es) *cas* y proteína(s) *Cas*), algunos elementos son homólogos entre sí y algunos elementos son heterólogos entre sí (*p. ej.*, en algunas realizaciones, los espaciador(es) de CRISPR y los genes *cas* son homólogos, pero la(s) repetición(es) de CRISPR(s) es/son homóloga(s)). Por lo tanto, en algunas realizaciones, el espaciador de CRISPR no está asociado de forma natural con la repetición de CRISPR y/o los genes *cas* y/o una combinación funcional de repetición de CRISPR-gen *cas*. De hecho, se pretende que cualquier combinación de elementos heterólogos y homólogos se emplee en la presente invención. En todavía realizaciones adicionales, las células donantes y receptoras son heterólogas, mientras que en realizaciones adicionales, son homólogas. También se pretende que los elementos contenidos en las células donantes y receptoras sean homólogos y/o heterólogos. Los elementos (*p. ej.*, los espaciadores de CRISPR) se introducen en el ADN plasmídico y/o genómico de la célula receptora, utilizando cualquier método adecuado conocido en la técnica.

En algunas realizaciones preferidas, al menos un espaciador de CRISPR se utiliza para modificar genéticamente una célula (*p. ej.*, una célula receptora). En otras realizaciones, los espaciadores de CRISPR se emplean como molde según el cual se modifica (*p. ej.*, se muta) el ADN genómico y/o plasmídico de una célula (*p. ej.*, una célula receptora), de modo que se crean espaciadores de CRISPR en el ADN de la célula. En algunas realizaciones, el(los) espaciador(es) de CRISPR se clona(n) en al menos una estructura artificial, plásmido u otro vector, con el que se transforma a continuación la célula receptora, 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 por ningún mecanismo, teoría o hipótesis particular, se contempla que cuanto más alejado está un espaciador de CRISPR dado, del extremo 5' del locus CRISPR que comprende el(los) gen(es) *cas* y/o la secuencia líder, menor será la resistencia conferida por el espaciador de CRISPR. Por lo tanto, en algunas realizaciones de la presente invención, uno o varios de los 100 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en otras realizaciones, uno o varios de los primeros 50 espaciadores de CRISPR, desde el extremo 5' del locus CRISPR están modificados, en realizaciones adicionales, uno o varios de los 40 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en aún otras realizaciones, uno o varios de los 30 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en todavía otras realizaciones, uno o varios de los 20 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, en aún otras realizaciones, uno o varios de los 15 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados, y en algunas realizaciones preferidas, uno o varios de los 10 primeros espaciadores de CRISPR desde el extremo 5' del locus CRISPR están modificados. Tal y como se indica en esta memoria, bacterias diferentes tienen cantidades diferentes de espaciadores de CRISPR, de modo que en algunas realizaciones se modifican varios espaciadores.

F. Agrupación de espaciadores de CRISPR

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

Tal y como se utiliza en esta memoria, la expresión "agrupación de espaciadores de CRISPR" se refiere a la longitud más corta observada de un espaciador en un tipo de CRISPR. Por tanto, por ejemplo, en CRISPR1 de *S. thermophilus*, la longitud dominante del espaciador es de 30 pb, con una minoría de espaciadores con un tamaño entre 28 pb y 32 pb. Por ello, en CRISPR1 de *S. thermophilus*, la agrupación de espaciadores de CRISPR se define como un tramo continuo de 28 pb.

En algunas realizaciones preferidas de la presente invención, la agrupación de espaciadores de CRISPR es homóloga al ácido nucleico diana, a un producto de la transcripción del mismo o a una secuencia identificada en toda la extensión de la secuencia de la agrupación. Aunque la homología también se puede considerar en términos

de similitud, en algunas realizaciones preferidas de la presente invención, la homología se expresa en términos de identidad de la secuencia. Por ello, en algunas realizaciones, una secuencia homóloga incluye una agrupación de espaciadores de CRISPR, que puede tener una identidad de al menos aproximadamente 90%; o de al menos aproximadamente 91, aproximadamente 92, aproximadamente 93, aproximadamente 94, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98 o aproximadamente 99% con la secuencia de ácido nucleico diana, un producto de la transcripción de la misma o una secuencia identificada en toda la extensión de la secuencia de la agrupación. En algunas realizaciones particularmente preferidas, la agrupación del espaciador de CRISPR tiene una identidad del 100% con la secuencia de ácido nucleico diana, un producto de la transcripción de la misma o una secuencia identificada en toda la extensión de la secuencia de la agrupación.

Durante el desarrollo de la presente invención, se analizaron secuencias de CRISPR de diversas cepas de *S. thermophilus*, que incluyen cepas industriales estrechamente relacionadas y variantes resistentes al fago. Diferencias en el número y en el tipo de espaciadores, se observaron principalmente en el locus CRISPR1. Sobre todo, la sensibilidad del fago parecía estar correlacionada con el contenido del espaciador de CRISPR1. Específicamente, el contenido del espaciador era casi idéntico entre las cepas originales y los derivados resistentes al fago, excepto en los espaciadores adicionales presentes en los últimos. Estos descubrimientos sugerían una relación potencial entre la presencia de espaciadores adicionales y las diferencias observadas en la sensibilidad del fago de una cepa dada. Esta observación dio lugar a la investigación del origen y la función de espaciadores adicionales, presentes en mutantes resistentes al fago.

G. Seudo-espaciador de CRISPR

Tal y como se emplea en esta memoria, las expresiones "seudo-espaciador de CRISPR", "proespaciador" y "proto-espaciador", se refieren a una secuencia de ácido nucleico presente en un organismo (*p. ej.*, un organismo donante, que incluye un bacteriófago pero no está limitado al mismo), que es preferentemente esencial para la función y/o supervivencia y/o replicación y/o infecciosidad, etc., y que forma una secuencia de espaciador de CRISPR. En algunas realizaciones, los pseudo-espaciadores de CRISPR se emplean para producir secuencias de espaciadores de CRISPR que son complementarias u homólogas a un pseudo-espaciador de CRISPR.

En algunas realizaciones, al menos un pseudo-espaciador de CRISPR y espaciador(es) de CRISPR que es/son complementarios u homólogos a al menos un pseudo-espaciador de CRISPR se emplean para modificar genéticamente una célula receptora. En algunas realizaciones, los pseudo-espaciadores de CRISPR o el(los) espaciador(es) de CRISPR que es/son complementarios u homólogos a uno o a varios pseudo-espaciadores de CRISPR, se insertan en el ADN plasmídico y/o genómico de una célula receptora, empleando cualquier método adecuado conocido en la técnica.

En algunas realizaciones adicionales, los pseudo-espaciadores de CRISPR se utilizan como un molde a partir del cual se modifica (*p. ej.*, se muta) el ADN plasmídico y/o genómico de una célula receptora, de modo que se crean espaciadores de CRISPR en el ADN plasmídico y/o genómico de la célula. En algunas realizaciones adicionales, los pseudo-espaciadores de CRISPR o espaciadores de CRISPR que es/son complementarios u homólogos a uno o a varios pseudo-espaciadores de CRISPR, se clonan en una estructura artificial, en un plásmido y/o en un vector, etc., se introduce(n) en la célula hospedadora empleando cualquier método adecuado, conocido en la técnica.

H. Proteínas Cas y genes cas

Tal y como se utiliza en esta memoria, la expresión "gen cas" tiene el significado convencional, tal y como se emplea en la técnica y se refiere a uno o a varios genes cas que están generalmente acoplados, asociados o próximos o en la proximidad de los loci CRISPR flanqueantes.

Una revisión completa de la familia de las proteínas Cas, es presentada por Haft y col. (Haft y col., *Comput. Biol.*, 1, 6 e60 [2005]). Tal y como se describe en ese artículo, se describen 41 familias de genes "CRISPR-associated" (*cas*), además de las cuatro familias del gen, previamente conocidas. Tal y como se ha indicado, los sistemas de CRISPR pertenecen a diferentes clases, con diferentes patrones de repetición, grupos de genes y distribución en las especies. De hecho, el número de genes cas en un locus CRISPR dado, puede variar entre las especies.

En algunas realizaciones, uno o varios de los genes y/o proteínas cas y/o proteínas presentes en la naturaleza en una célula receptora y uno o varios espaciadores heterólogos, se integran o se insertan en los loci CRISPR, de forma adyacente a uno o varios de los genes o proteínas cas.

En algunas realizaciones, uno o varios de los genes y/o proteínas cas son heterólogos con la célula receptora y uno o varios de los espaciadores son homólogos o heterólogos. En algunas realizaciones preferidas, los espaciadores se integran o se insertan en los loci CRISPR de forma adyacente a uno o a varios de los genes o proteínas cas.

Los loci CRISPR se encuentran típicamente en la proximidad de cuatro genes denominados *cas1* a *cas4*. La disposición más común de estos genes es *cas3-cas4-cas1-cas2*. La proteína Cas3 parece ser una helicasa, mientras que Cas4 se parece a la familia de exonucleasas RecB y contiene un motivo rico en cisteínas, lo que sugiere la unión al ADN. Cas1 es generalmente muy básica y es la única proteína Cas que se encuentra sistemáticamente en todas las especies que contienen loci CRISPR. Cas2 aún no se ha caracterizado. *cas 1-4* se

caracterizan típicamente por su proximidad a los loci CRISPR y su amplia distribución entre las especies bacterianas y arqueobacterianas. Aunque no todos los genes *cas1-4* se asocian con todos los loci CRISPR, todos se encuentran en subtipos múltiples.

5 Además, hay otra agrupación de tres genes asociada con estructuras de CRISPR en muchas especies bacterianas, denominados en esta memoria, *cas1B*, *cas5* y *cas6* (véase, Bolotin y col., [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 aún otras realizaciones, el gen *cas* se selecciona entre fragmentos, variantes, homólogos y/o sus derivados de *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* y/o *cas6*. En algunas realizaciones adicionales, se emplea una combinación de dos o más genes *cas*, en cualquier combinación adecuada. Es de señalar que la nomenclatura de los genes *cas* está en cambio continuo. Por lo tanto, el texto en esta memoria se tiene que tomar dentro del contexto.

15 La expresión "proteína Cas" también incluye una pluralidad de proteínas Cas (p. ej., 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, más preferentemente, entre aproximadamente 4 y aproximadamente 8 proteínas Cas, y más preferentemente, entre aproximadamente 4 y aproximadamente 7 proteínas Cas; tales como 4, 5, 6 o 7 proteínas Cas).

20 En algunas realizaciones, las proteínas Cas están codificadas por genes *cas* que comprenden ADN, mientras que en otras realizaciones, *cas* comprende ARN. En algunas realizaciones, el ácido nucleico tiene origen genómico, mientras que en otras realizaciones, tiene origen sintético o recombinante. En algunas realizaciones, los genes *cas* que codifican las proteínas Cas son bicatenarios o monocatenarios, según representen la cadena codificadora o la cadena no codificadora o combinaciones de las mismas. En algunas realizaciones, los genes *cas* se preparan mediante el uso de técnicas de ADN recombinante (p. ej., ADN recombinante), tal y como se describe en esta memoria.

25 I. Bacteriófago

Tal y como se emplea en esta memoria, el término "bacteriófago" (o "fago") tiene su significado convencional, tal y como se entiende en la técnica (es decir, un virus que infecta selectivamente una o varias especies bacterianas). Muchos bacteriófagos son específicos de un género o una especie o una cepa bacteriana en particular. En algunas realizaciones preferidas, los fagos son capaces de infectar las bacterias originales y/o células hospedadoras. En algunas realizaciones, los bacteriófagos son virulentos frente a la bacteria original. En algunas realizaciones, los fagos son líticos, mientras que en otras realizaciones, los fagos son lisogénicos.

35 Un bacteriófago lítico es uno que sigue la vía lítica hasta la finalización del ciclo lítico, en lugar de entrar en la vía lisogénica. Un bacteriófago lítico se somete a una replicación vírica que conduce a la lisis de la membrana celular, la destrucción de la célula y la liberación de las partículas de bacteriófago de la progenie, capaces de infectar otras células.

Un bacteriófago lisogénico es uno capaz de entrar en la vía lisogénica, en la que el bacteriófago se convierte en una parte inactiva, pasiva del genoma de la célula, antes de la finalización de su ciclo lítico.

40 Los bacteriófagos que se emplean en la presente invención incluyen bacteriófagos que pertenecen a cualquiera de las siguientes familias de virus, pero no están limitados a los mismos: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae o Tectiviridae. En algunas realizaciones, se emplea particularmente un bacteriófago que infecta bacterias que son patógenas para vegetales y/o animales (incluyendo los seres humanos).

45 En algunas realizaciones particularmente preferidas, el bacteriófago de la presente invención incluye bacteriófagos pero no está limitado a los mismos, capaces de infectar una bacteria que comprende de forma natural uno o varios loci CRISPR. Los loci CRISPR se han identificado en más de 40 procariontes (véase, p. ej., Jansen y col., Mol. Microbiol., 43:1565-1575 [2002]; y Mojica y col., [2005]) que incluyen sin estar limitados a los mismos, *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*.

55 En algunas realizaciones, el bacteriófago incluye pero no está limitado a los mismos, aquellos bacteriófagos capaces de infectar bacterias pertenecientes a los 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 aún otras realizaciones adicionales, el bacteriófago incluye los bacteriófagos capaces de infectar (o transducir) bacterias lácticas pero no está limitado a los mismos, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (p. ej., *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*.

5 En aún otras realizaciones, el bacteriófago incluye pero no está limitado a los mismos, aquellos bacteriófagos capaces de infectar a *Lactococcus lactis* (p. ej., *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*,
10 *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 están limitados a los mismos, aquellos bacteriófagos capaces de infectar bacterias que comprenden uno o varios loci CRISPR heterólogos. En algunas realizaciones, las bacterias comprenden uno o varios loci CRISPR heterólogos y/o uno o varios genes *cas* heterólogos y/o una o varias repeticiones heterólogas de CRISPR, y/o uno o varios espaciadores heterólogos de CRISPR.
15

La infección de bacterias a través de fagos es el resultado de la inyección o la transferencia del ADN del fago dentro de las células. En algunas realizaciones, la infección conduce a la expresión (es decir, a la transcripción y traducción) del ácido nucleico del bacteriófago dentro de la célula y la continuación del ciclo vital del bacteriófago. En algunas realizaciones que implican un bacteriófago recombinante, secuencias recombinantes también se expresan dentro del genoma del fago (p. ej., ácidos nucleicos informadores).
20

Se ha encontrado que las secuencias espaciadoras de CRISPR en procariontes, tienen frecuentemente similitudes significativas con una variedad de moléculas de ADN, incluyendo elementos genéticos tales como cromosomas, bacteriófagos y plásmidos conjugativos. Se ha descrito que células portadoras de esos espaciadores de CRISPR son incapaces de ser infectadas por moléculas de ADN que contienen secuencias homólogas a los espaciadores (véase, Mojica y col., [2005]).
25

En algunas realizaciones de la presente invención, las bacterias originales se exponen (p. ej., de forma reiterativa, secuencial, simultánea o sustancialmente simultánea) a más de un bacteriófago. En algunas realizaciones preferidas, las bacterias se exponen a mezclas de uno o más (p. ej., varios) fagos diferentes. En algunas realizaciones alternativas preferidas, las bacterias originales son sensibles a cada uno de los bacteriófagos a los que se exponen.
30

En algunas realizaciones, cada una de las secuencias marcadoras procedentes de cada uno de los bacteriófagos y/o de cada una de las secuencias duplicadas (p. ej., la repetición duplicada de CRISPR) procedente de la bacteria original, se integran 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 en ambos extremos del mismo locus CRISPR. En aún otras realizaciones, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran en el extremo 5' y/o en 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 integran en el extremo 5' del mismo locus CRISPR.
35

En algunas realizaciones, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas procedentes de las bacterias originales se integran de forma reiterativa, simultánea o sustancialmente simultánea. En realizaciones en las que cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran de forma secuencial, la primera secuencia marcadora y/o la primera secuencia duplicada se integran en las bacterias originales. Una segunda secuencia marcadora procedente de un segundo bacteriófago y/o otra secuencia duplicada, se integran a continuación en la bacteria original. En algunas realizaciones preferidas, la secuencia marcadora y/o la secuencia duplicada se integran en el ADN cromosómico de las bacterias originales.
40
45

En algunas realizaciones, cada una de las secuencias marcadoras y/o cada una de las secuencias duplicadas se integran en un extremo (p. ej., el extremo 5') del mismo locus CRISPR adyacentes entre sí (es decir, próximas). Por lo tanto, en algunas realizaciones, cada una de las secuencias marcadoras y/o de las secuencias duplicadas se integran de forma secuencial, en donde las primeras secuencias se integran en la bacteria original en un extremo (p. ej., en el extremo 5' y/o 3' o próximo a los mismos) del locus CRISPR. En algunas realizaciones preferidas, una segunda secuencia marcadora y/o una secuencia duplicada se integran a continuación en las bacterias originales, de forma adyacente (p. ej., de forma directamente contigua) a la primera pareja de secuencias. En algunas realizaciones, las segundas secuencias se integran en la bacteria original de forma adyacente (p. ej., de forma directamente contigua) al extremo 5' o 3' de las primeras secuencias. En algunas realizaciones preferidas, las segundas secuencias se integran en la bacteria original de forma adyacente (p. ej., de forma directamente contigua) al extremo 3' de las primeras secuencias. En realizaciones en las que se proporcionan secuencias adicionales, estas se integran a continuación.
50
55

En algunas realizaciones, cada una de las secuencias se integra de forma adyacente (es decir, de forma contigua) entre sí, dentro o en el extremo 3' y/o en el extremo 5' del mismo locus CRISPR de las bacterias originales. En algunas realizaciones preferidas, cada una de las secuencias se integran de forma adyacente (es decir, de forma contigua) entre sí en el extremo 5' del mismo locus CRISPR de las bacterias originales. En algunas realizaciones particularmente preferidas, cada una de las secuencias se integra de forma adyacente (es decir, de forma contigua) entre sí, aguas arriba del extremo 5' del locus CRISPR de las bacterias originales. En algunas realizaciones alternativamente preferidas, cada una de las secuencias se integra de forma adyacente (es decir, de forma contigua) entre sí, en una posición que está aguas arriba de la repetición de CRISPR 5' del locus CRISPR de las bacterias originales. En algunas realizaciones preferidas más particularmente, cada una de las secuencias se integran de forma adyacente (es decir, de forma contigua) entre sí, aguas arriba de la primera repetición de CRISPR 5' del locus CRISPR de la bacteria original.

J. Bacterias originales

Tal y como se emplea en esta memoria, las expresiones "bacteria original" "bacterias originales" y "cepa original" se refieren a cualquier bacteria/bacterias/cepas que se expone a uno o a varios bacteriófagos virulentos. En algunas realizaciones particularmente preferidas, las bacterias originales son sensibles al fago virulento. En algunas realizaciones preferidas, la cepa original se infecta con el bacteriófago. En algunas realizaciones particularmente preferidas, la infección con un fago vuelve la bacteria/bacterias/cepa originales o una subpoblación de las mismas, insensible a una infección adicional con el bacteriófago. En algunas realizaciones preferidas, la infección de una "bacteria original" con uno o varios bacteriófagos da como resultado la creación de una cepa marcada, que se puede seleccionar basándose en su insensibilidad frente al bacteriófago. En algunas realizaciones preferidas, "mutantes resistentes al bacteriófago" son bacterias que están marcadas o que se marcan de acuerdo con los métodos de la presente invención. En algunas realizaciones, las bacterias originales son cepas bacterianas de tipo silvestre. En algunas realizaciones preferidas, las bacterias originales son cepas de tipo silvestre de bacterias que no se han infectado previamente con ningún bacteriófago. En algunas realizaciones preferidas, las bacterias originales son cepas de tipo silvestre de bacterias que no se han marcado previamente, mientras que en algunas realizaciones alternativas, las bacterias originales son mutantes resistentes al bacteriófago que se han marcado previamente.

En algunas realizaciones particularmente preferidas, la bacteria original se selecciona a partir de cualquier bacteria que comprende de forma natural uno o varios loci CRISPR. Los loci CRISPR se han identificado en más de 40 procariontes (Jansen y col. [2002] *supra*; Mojica y col., [2005], *supra*; y Haft y col., [2005], *supra*), que incluyen pero no están limitados a los mismos, *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoarrobacter*, *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 original comprende uno o varios espaciadores heterólogos de CRISPR, una o varias repeticiones heterólogas de CRISPR, y/o uno o varios genes *cas* heterólogos. En algunas realizaciones alternativas, la bacteria original comprende uno o varios loci CRISPR heterólogos, preferentemente, uno o varios loci CRISPR completos. En algunas realizaciones adicionales, la bacteria original comprende de forma natural uno o varios loci CRISPR y también comprende uno o varios espaciadores heterólogos de CRISPR, una o varias repeticiones heterólogas de CRISPR, y/o uno o varios genes *cas* heterólogos. En algunas realizaciones adicionales, la bacteria original comprende de forma natural uno o varios loci CRISPR y también comprende uno o varios loci CRISPR heterólogos, preferentemente, uno o varios loci CRISPR completos.

En algunas realizaciones preferidas, la subpoblación resistente al fago creada por exposición de las bacterias originales 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 comprenda cultivos mixtos de células y fagos. En algunas realizaciones, el cultivo mixto es una mezcla de diferentes mutantes que se corresponden con diferentes casos de integración en el mismo loci CRISPR y/o en loci diferentes.

Aunque no se pretende que la presente invención esté limitada de este modo, los géneros bacterianos originales preferidos son *Streptococcus* y *Lactobacillus*. De hecho, se pretende que cualquier especie bacteriana se emplee en la presente invención, incluyendo pero no limitando a las mismas, *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*, *Enterococcus*, *Pediococcus*, *Leuconostoc*, *Oenococcus* y/o *Xanthomonas*. En algunas realizaciones, las bacterias originales son o se obtienen a partir de bacterias lácticas, que incluyen pero no están limitadas a las mismas, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (p. ej., *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc* y/o *Oenococcus*. En realizaciones adicionales, las bacterias originales son o se obtienen a partir de *Lactococcus lactis* (p. ej., *L. lactis* subsp. *lactis* y *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 realizaciones de la presente invención, la bacteria original es una "bacteria de calidad alimentaria" (es decir, una bacteria que se emplea y se considera generalmente segura para el uso en la preparación y/o producción de alimentos para humanos y/o animales). En algunas realizaciones preferidas, la bacteria original es adecuada para el uso como cultivo iniciador, cultivo probiótico y/o suplemento dietético. En algunas realizaciones adicionales, la bacteria original se emplea en la fermentación de carne (*p. ej.*, vaca, cerdo, cordero y aves de corral), incluyendo pero sin estar limitadas a las mismas, bacterias lácticas, *Pediococcus cerevisiae*, *Lactobacillus plantarum*, *L. brevis*, *L. sakei*, *L. curvatus*, especies de *Micrococcus*, *Pediococcus pentosaceus*, *Staphylococcus xylosus*, *S. vitulinus* y mezclas de las mismas (véase, *p. ej.*, Knorr (compilador), Food Biotechnology en 538-39 [1987]; y Pederson, Microbiology of Fermented Foods, en 210-34, 2ª ed., [1979]; y el documento de Patente de EE.UU. nº 2.225.783). En todavía otras realizaciones adicionales, la bacteria original se emplea en la fermentación de vegetales (*p. ej.*, zanahorias, pepinos, tomates, pimientos y col), incluyendo pero sin estar limitadas a las mismas, *L. plantarum*, *L. brevis*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, y mezclas de las mismas (véase *p. ej.*, Knorr, *supra*; Pederson, *supra*; y los documentos de Patente de EE.UU. nº 3.024.116, 3.403.032, 3.932.674 y 3.897.307). En aún otras realizaciones adicionales, la bacteria original se emplea en la fermentación de masas formadas a partir de cereales (*p. ej.*, trigo, centeno, arroz, avena, cebada y maíz). En aún otras realizaciones, la bacteria original se emplea en la producción de vino a través de la fermentación de jugo de frutas (*p. ej.*, mosto). En algunas realizaciones adicionales, la bacteria original se emplea en la fermentación de la leche (*p. ej.*, *L. delbrueckii subsp. bulgaricus*, *L. acidophilus*, *S. thermophilus*, y mezclas de las mismas (véase, Knorr, *supra*; y Pederson *supra*, en las páginas 105-35). En algunas realizaciones preferidas, la bacteria original se emplea en la producción de queso, incluyendo pero sin estar limitadas a las mismas, *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 mezclas de las mismas (véase *p. ej.*, Knorr, *supra*, y Pederson, *supra*, en 135-51). En aún otras realizaciones, la bacteria original se emplea en la fermentación de huevos, incluyendo pero sin estar limitadas a las mismas, *Pediococcus pentosaceus*, *Lactobacillus plantarum* y mezclas de las mismas (véase, Knorr, *supra*). En algunas realizaciones, la bacteria original se emplea en la fermentación para producir diversos productos, que incluyen sin estar limitados a los mismos, queso Cheddar y requesón (*p. ej.*, *L. lactis subsp. lactis*, *L. lactis subsp. cremoris*), yogur (*L. delbrueckii subsp. bulgaricus* y *S. thermophilus*), queso suizo (*p. ej.*, *S. thermophilus*, *L. lactis*, y *L. helveticus*), queso azul (*Leuconostoc cremoris*), queso italiano (*L. bulgaricus* y *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 saccharoperbutylaceticum*). En algunas realizaciones preferidas, las especies de bacterias originales se seleccionan entre *S. thermophilus*, *L. delbrueckii subsp. bulgaricus* y/o *L. acidophilus*.

En aún otras realizaciones adicionales, las bacterias originales se emplean en métodos que incluyen pero no están limitados a los mismos, la producción de antibióticos, la producción de aminoácidos, la producción de disolventes y la producción de otros materiales económicamente útiles. En aún otras realizaciones, las bacterias originales se emplean en cosmética, terapéutica y/o composiciones farmacéuticas. En algunas realizaciones las composiciones tienen actividades particulares, que incluyen sin estar limitadas a las mismas, la regeneración de la piel, que incluyen sin estar limitadas a las mismas, propiedades antiarrugas, desaparición de antiguas cicatrices, reparación de tejidos dañados por quemaduras, promover la cicatrización de la piel, eliminación de manchas de pigmentación, etc. En algunas realizaciones, las composiciones favorecen o inhiben el crecimiento de las uñas, el pelo o el vello. En algunas realizaciones adicionales, las composiciones comprenden al menos un cultivo microbiano y/o una bacteria marcada y/o un cultivo celular producidos empleando los métodos y las composiciones de la presente invención.

En realizaciones adicionales, las bacterias originales son mutantes insensibles a bacteriófagos. Por lo tanto, en algunas realizaciones, las bacterias originales son insensibles a uno o a varios bacteriófagos. En algunas realizaciones preferidas, la bacteria original no es un mutante insensible a los bacteriófagos, frente al bacteriófago al que está expuesto durante el uso de la presente invención.

K. Secuencia marcadora

5 Tal y como se emplea en esta memoria, la expresión "secuencia marcadora" se refiere a la porción de una "unidad adicional de repetición-espaciador" que proviene, puede provenir, se obtiene o se puede obtener a partir del genoma de uno o varios bacteriófagos a los que se expone la bacteria original, de acuerdo con los métodos de la presente invención y que se emplea como marcador o etiqueta (*p. ej.*, un marcador único o una etiqueta única). 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 desde aproximadamente 20 a aproximadamente 58 nucleótidos de longitud. Sin embargo, en algunas realizaciones alternativas, un "marcador" se genera empleando elementos genéticos procedentes de fuentes distintas a las del fago. Por ejemplo, en algunas realizaciones, el marcador es proporcionado por plásmidos, elementos transponibles, ácido nucleico aislado, etc. De hecho, en algunas realizaciones, el marcador es una secuencia única sintética, no funcional. Por tanto, no se pretende que la presente invención esté limitada por marcadores de ácido nucleico que se generan únicamente a partir de ácido nucleico de fagos.

15 La secuencia marcadora es típicamente una secuencia que está presente en la naturaleza en el bacteriófago. Preferentemente, la secuencia marcadora tiene una identidad de al menos aproximadamente 90%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99%, con la secuencia presente en la naturaleza en el bacteriófago (*p. ej.*, el genoma del bacteriófago a partir del cual proviene, puede provenir, se obtiene o se puede obtener la secuencia). En algunas realizaciones muy preferidas, la secuencia marcadora tiene una identidad del 100% con la secuencia presente en la naturaleza en el bacteriófago (*p. ej.*, el genoma del bacteriófago a partir del cual proviene, puede provenir, se obtiene o se puede obtener).

20 En algunas realizaciones, la secuencia marcadora tiene una identidad menor de aproximadamente 40%, aproximadamente 30%, aproximadamente 20%, aproximadamente 10%, aproximadamente 5%, aproximadamente 4%, aproximadamente 3%, aproximadamente 2%, aproximadamente 1% o aproximadamente 0% con cualquier otro espaciador de CRISPR o agrupación de espaciadores de CRISPR, en uno o en varios loci CRISPR de la bacteria marcada.

25 En algunas realizaciones, la secuencia marcadora tiene menos de aproximadamente 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 uno o varios loci CRISPR de la bacteria marcada.

30 En algunas realizaciones alternativas, la secuencia marcadora es una secuencia que es idéntica a una secuencia (*p. ej.*, 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 (*p. ej.*, un espaciador de CRISPR) en el locus CRISPR de la bacteria porque contiene uno o varios polimorfismos de nucleótidos aislados (*p. ej.*, uno o dos polimorfismos de nucleótidos aislados).

35 En algunas realizaciones, al menos una secuencia marcadora se integra en la bacteria original. En algunas realizaciones alternativas, también se integra al menos una secuencia duplicada (*p. ej.*, una secuencia de una repetición duplicada de CRISPR) que proviene, puede provenir, se obtiene o se puede obtener a partir del genoma de la bacteria original o de uno o varios plásmidos de la bacteria original (*p. ej.*, megaplásmidos). En algunas realizaciones particularmente preferidas, al menos una secuencia duplicada se copia o se replica a partir del genoma de la bacteria original. En algunas realizaciones, la secuencia de la repetición de CRISPR en un locus CRISPR se duplica y la secuencia marcadora se integra en el genoma de la bacteria, inmediatamente después (*p. ej.*, aguas abajo) de la nueva repetición de CRISPR duplicada. Sin embargo, no se pretende que la presente invención esté limitada por ningún mecanismo específico o ninguna teoría sobre la actuación.

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

45 En algunas realizaciones preferidas, al menos una secuencia marcadora y al menos una secuencia duplicada, se integran en la bacteria original. En algunas realizaciones, cada vez que una secuencia marcadora se integra en el genoma de la bacteria original, se acompaña por una integración reiterativa, secuencial, simultánea o sustancialmente simultánea de al menos una secuencia duplicada. Por lo tanto, al menos un par de secuencias que comprenden la secuencia marcadora y la secuencia duplicadas, se integra en la bacteria original, dando como

resultado una bacteria marcada. Sin embargo, no se pretende que la presente invención se limite a ningún mecanismo específico o a ninguna teoría sobre la actuación.

5 En algunas realizaciones preferidas, la secuencia marcadora, al menos una, y, al menos una secuencia duplicada, se integran de forma adyacente entre sí. En algunas realizaciones particularmente preferidas, la, al menos una, secuencia marcadora, y la, al menos una, secuencia duplicada, se integran de forma directamente adyacente entre sí, de modo que no hay nucleótidos intercalados entre las secuencias.

10 En algunas realizaciones, la secuencia duplicada se fija, se une o se fusiona con un extremo (*p. ej.*, el extremo 5' o 3') de la secuencia marcadora. En algunas realizaciones preferidas, la secuencia duplicada se fija, se une o se 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 de repetición espaciador de CRISPR. Por lo tanto, en algunas realizaciones, después de la integración de la unidad de repetición espaciador 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), en el locus CRISPR, aguas abajo de la secuencia duplicada. En aún
15 otras realizaciones preferidas, las secuencias dentro de una unidad de repetición espaciador de CRISPR, están fijadas directamente, unidas directamente o fusionadas directamente, de modo que no hay nucleótidos intercalados entre la secuencia duplicada y la secuencia marcadora.

20 En algunas realizaciones particularmente preferidas, una unidad de repetición espaciador de CRISPR se integra en el genoma de la bacteria original para producir una bacteria marcada. En algunas realizaciones preferidas, la secuencia duplicada proviene, puede provenir, se obtiene o se puede obtener a partir del genoma de la bacteria original. En algunas realizaciones adicionales, la secuencia marcadora proviene, puede provenir, se obtiene o se puede obtener a partir del genoma del bacteriófago que se emplea para infectar la bacteria original.

25 En algunas realizaciones adicionales, las unidades múltiples de espaciador y repetición de CRISPR, se integran en el genoma de la bacteria original. En algunas realizaciones, las unidades múltiples de espaciador y repetición de CRISPR comprenden una primera unidad de repetición espaciador de CRISPR que comprende una secuencia duplicada y una secuencia marcadora y una segunda unidad de espaciador y 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 (*p. ej.*, superior a aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98%, aproximadamente 99% o aproximadamente 100% de identidad) que la primera secuencia duplicada. En algunas realizaciones, la secuencia marcadora tiene típicamente una secuencia diferente (*p. ej.*, menos de aproximadamente 40%, aproximadamente 30%, aproximadamente 20%, aproximadamente 10%, aproximadamente 5%, aproximadamente 4%, aproximadamente 3%, aproximadamente 2%, aproximadamente 1% o aproximadamente 0% de identidad) de la primera secuencia marcadora. Esto es también el caso en realizaciones que contienen secuencias adicionales de unidades de espaciador y repetición de CRISPR integradas.

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

[secuencia duplicada-secuencia marcadora]_n

en donde $n = 2, 3, 4, 5$ o ≥ 6 .

40 En algunas realizaciones particularmente preferidas, la configuración de las unidades múltiples de espaciador y repetición de CRISPR, es típicamente:

[repetición de CRISPR-secuencia marcadora]_n

en donde $n = 2, 3, 4, 5$ o ≥ 6 .

En algunas realizaciones preferidas, la configuración de las unidades múltiples de espaciador y repetición de CRISPR, es:

45 5'-[secuencia duplicada-secuencia marcadora]_n-3'

en donde $n = 2, 3, 4, 5$ o ≥ 6 .

En algunas realizaciones particularmente preferidas, la configuración de las unidades múltiples de espaciador y repetición de CRISPR, es:

50 5'-[repetición de CRISPR-secuencia marcadora]_n-3'

en donde $n = 2, 3, 4, 5$ o ≥ 6 .

En algunas realizaciones preferidas, las unidades múltiples de espaciador y repetición de CRISPR, se integran en la bacteria original.

En algunas realizaciones, la porción de la secuencia marcadora de la unidad de espaciador y repetición de CRISPR, se integra de forma adyacente a: (i) una secuencia duplicada que es homóloga (*p. ej.*, idéntica) a una secuencia presente en la naturaleza en la bacteria original; (ii) una secuencia duplicada que es homóloga (*p. ej.*, idéntica) a una secuencia presente en la naturaleza en el locus CRISPR de la bacteria original; o (iii) lo más preferido, a una secuencia duplicada que es homóloga (*p. ej.*, idéntica) a una repetición de CRISPR presente en la naturaleza, en el locus CRISPR de la bacteria original.

Después de cada exposición de una bacteria original a un bacteriófago dado, en experimentos independientes, la secuencia marcadora en cada una de las bacterias marcadas, presenta diferentes secuencias de nucleótidos, creando de este modo una secuencia que es única para cada bacteria. De este modo, después de la exposición de una bacteria original a un bacteriófago dado, la secuencia marcadora que se integra en la bacteria original, se selecciona a partir del genoma del bacteriófago. Tal y como se ha indicado anteriormente, no se pretende que la presente invención esté limitada por eventos de integración al azar, por ningún mecanismo particular, ni por medios de actuación.

Este descubrimiento sorprendente se empleó para el desarrollo de la presente invención, ya que la secuencia marcadora seleccionada proporciona un marcador o una etiqueta única en la bacteria marcada. Se ha encontrado sorprendentemente que cuando se expone la misma bacteria original al mismo bacteriófago, la secuencia marcadora que se integra en experimentos independientes/distintos, tiene una secuencia diferente, dando por ello como resultado 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 gracias a una o a varias de las siguientes propiedades de la secuencia marcadora: (1) la posición de la secuencia marcadora en uno o en varios loci CRISPR del mutante insensible a los bacteriófagos (tal y como se indica en esta memoria, la secuencia marcadora se localiza típicamente en uno y/o en ambos extremos 5' y 3' (más preferentemente, el extremo 5') del locus CRISPR de la bacteria marcada; (2) la secuencia marcadora tiene un grado de homología o de identidad elevado (*p. ej.*, 100% de identidad) con una secuencia en el genoma del bacteriófago, a la que estaba expuesta la bacteria original; y/o (3) la secuencia marcadora se fusiona, se une o se fija (*p. ej.*, fusionada, unida o fijada directamente) con al menos una secuencia (*p. ej.*, una repetición de CRISPR; es decir, una "unidad de espaciador y repetición de CRISPR") que está duplicada, procedente del genoma de la bacteria original. Típicamente, tal y como se describe en esta memoria, esta unidad de espaciador y repetición de CRISPR, se localiza en uno y/o en ambos extremos (*p. ej.*, el extremo 5' y/o 3'; más preferentemente, el extremo 5') del locus CRISPR de la bacteria marcada. Por tanto, en algunas realizaciones, las unidades de espaciador y repetición de CRISPR se integran en ambos extremos del locus CRISPR de la bacteria original, de modo que las secuencias están en el extremo 5' y 3' del locus CRISPR. En algunas realizaciones adicionales, una de las secuencias duplicadas es la primera secuencia en el extremo 5' del locus CRISPR y la secuencia marcadora se localiza inmediatamente aguas abajo de la secuencia duplicada. En algunas realizaciones, la otra secuencia duplicada es la última secuencia en el extremo 3' del locus CRISPR y la secuencia marcadora está inmediatamente aguas arriba de la secuencia duplicada.

En algunas realizaciones preferidas, la(s) secuencia(s) marcadora(s) y/o la(s) secuencia(s) duplicada(s) de la unidad de espaciador y repetición de CRISPR, se integran en un extremo del locus CRISPR de la bacteria original, de modo que la(s) secuencia(s) están en el extremo 3' del locus CRISPR. En algunas realizaciones adicionales, la secuencia duplicada es la última secuencia en el extremo 3' del locus CRISPR y la secuencia marcadora está localizada inmediatamente aguas arriba de la secuencia duplicada. En algunas realizaciones preferidas, la(s) secuencia(s) marcadora(s) y/o la(s) secuencia(s) duplicada(s) se integran en un extremo del locus CRISPR de la bacteria original, de modo que las secuencias están en el extremo 5' del locus CRISPR. En algunas realizaciones, la secuencia duplicada es la primera secuencia en el extremo 5' del locus CRISPR y la secuencia marcadora está inmediatamente aguas abajo de la secuencia duplicada.

Tal y como se describe en esta memoria, la(s) secuencia(s) marcadora(s) es un marcador específico de una cepa, en el sentido de que la secuencia marcadora que está integrada o insertada desde el bacteriófago en la bacteria original, es diferente cada vez que la bacteria original (*p. ej.*, la misma bacteria original) se expone al bacteriófago (*p. ej.*, el mismo bacteriófago). Por consiguiente, la secuencia marcadora se emplea como un marcador único para una cepa bacteriana dada.

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

L. Loci CRISPR marcados

El genoma de *Streptococcus thermophilus* LMG 18311 contiene 3 loci CRISPR; las secuencias repetidas de 36 pb son diferentes en CRISPR1 (34 repeticiones), CRISPR2 (5 repeticiones), y CRISPR3 (una secuencia única). Sin

embargo, están perfectamente conservadas en cada locus. Las repeticiones CRISPR1 y CRISPR2 están intercaladas respectivamente con 33 y 4 secuencias de 30 pb de longitud. Todas estas secuencias que están intercaladas son diferentes entre sí. También son diferentes de las encontradas en la cepa CNRZ1066 (41 secuencias intercaladas dentro de CRISPR1) y en la cepa LMD-9 (16 dentro de CRISPR1 y 8 dentro de CRISPR3), que son ambas *S. thermophilus*.

La cepa DGCC7710 de *Streptococcus thermophilus* (depositada en la "Collection Nationale de Cultures de Microorganismes" francesa con el número CNCM I-2423) posee al menos 3 loci CRISPR: CRISPR1, CRISPR2 y CRISPR3. En las cepas CNRZ1066 y LMG18311 de las que se conoce la secuencia completa del genoma (Bolotin y col., [2004] *supra*), CRISPR1 se localiza en el mismo locus cromosómico: entre str0660 (o stu0660) y str0661 (o stu0661). En la cepa DGCC7710, CRISPR1 está localizado 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 es diferente de los demás. Aunque la mayoría de estos espaciadores son nuevos (es decir, que no se han identificado previamente dentro de los loci CRISPR), cuatro espaciadores próximos al remolque de CRISPR1, son idénticos a espaciadores ya conocidos de CRISPR1. Estos cuatro incluyen: el 28° espaciador de DGCC7710, que tiene una identidad del 100% con el 31° espaciador de CRISPR1 de la cepa CNRZ1575 (número de orden de GenBank DQ072991); el 30° espaciador de DGCC7710, que tiene una identidad del 100% con el 27° espaciador de CRISPR1 de la cepa CNRZ703 (número de orden de GenBank DQ072990); el 31° espaciador de DGCC7710, que tiene una identidad del 100% con el 28° espaciador de CRISPR1 de la cepa CNRZ703 (número de orden de GenBank DQ072990); y el 32° espaciador de DGCC7710, que tiene una identidad del 100% con el 30° espaciador de CRISPR1 de la cepa CNRZ703 (número de orden de GenBank DQ072990). La secuencia de CRISPR1 (5'-3') de la cepa DGCC7710, se muestra en SEQ ID NO:1.

La cepa de *Streptococcus thermophilus* DGCC7778 se aisló como un mutante natural resistente al fago, empleando DGCC7710 como cepa original, y el fago D858 como el fago virulento. CRISPR1 de la cepa DGCC7778 contiene 35 repeticiones (incluyendo la repetición terminal), y por ello, 34 espaciadores. Cuando se compara con la secuencia de CRISPR1 de DGCC7710, la secuencia de CRISPR1 de DGCC7778 posee dos espaciadores nuevos, adyacentes, así como dos repeticiones adicionales que flanquean los nuevos espaciadores, en un extremo del locus CRISPR (es decir, próximo al líder). Todos los demás espaciadores del locus CRISPR1 permanecen invariables. La secuencia de CRISPR1 (5'-3') de la cepa DGCC7778 se muestra en SEQ ID NO:2.

Por tanto, en el caso de DGCC7778, el primer espaciador (5'-caacacattcaacagattaatgaagaatac-3' [SEQ ID NO:3] y el segundo espaciador (5'-tccactcagctacaataatagtgagtactc-3' [SEQ ID NO:4]) constituyen el marcador específico de la cepa que identifica esta cepa marcada. Durante el desarrollo de la presente invención, se mostró que la secuencia de ambos espaciadores nuevos, existe dentro del genoma del fago D858. La secuencia del segundo nuevo espaciador está localizada entre las posiciones 25471 y 25442 pb (es decir, en la hebra negativa) del genoma de D858, con un emparejamiento incorrecto (96,7% de nucleótidos idénticos a lo largo de 30 nucleótidos). La secuencia del primer espaciador está localizada entre las posiciones 31481 y 31410 pb (es decir, en la hebra positiva) del genoma de D858 (100% de nucleótidos idénticos a lo largo de 30 nucleótidos). Aunque no se pretende que la presente invención se limite a ningún mecanismo ni teoría particulares, el hecho de que dos nuevos espaciadores presentes en el locus CRISPR1 de DGCC7778, sean necesarios para conferir resistencia a la cepa DGCC7778, frente al fago D858, se contempla que el espaciador "2" se inserta en primer lugar en el locus CRISPR1 de DGCC7710 (33 repeticiones y 32 espaciadores), en un extremo de este locus CRISPR, junto con una repetición. Esta inserción da lugar a un mutante insensible a los bacteriófagos (cepa intermedia), marcado con este nuevo espaciador adicional (es decir, que ahora tiene 34 repeticiones y 33 espaciadores). Este espaciador se obtiene a partir del genoma de D858, pero un error de la replicación o un error de la transcripción inversa, ocurrido probablemente durante el proceso de inserción, lleva a una mutación puntual. Debido a un emparejamiento erróneo (es decir, 1 emparejamiento erróneo) entre este nuevo espaciador adquirido y la secuencia diana del fago, la eficacia de la resistencia de esta cepa intermedia frente al fago D858, era baja.

Sin embargo, un segundo evento de inserción de espaciador, ocurrido en esta cepa intermedia (es decir, la cepa que es más resistente frente al fago D858 que la cepa original DGCC7710, pero no "totalmente" resistente debido al emparejamiento incorrecto), conduce a la inserción de un segundo espaciador nuevo (el espaciador "1", tal y como se encontró en DGCC7778) en el mismo extremo del locus CRISPR1, junto con una repetición. Esta segunda inserción produjo un nuevo mutante insensible a los bacteriófagos, que se aisló y se denominó DGCC7778. DGCC7778 es más resistente a D858 que la cepa intermedia, y mucho más resistente que la cepa original DGCC7710, debido a la presencia del espaciador "1", que tiene una identidad del 100% con la secuencia del fago diana.

La cepa DGCC7710-RH de *Streptococcus thermophilus* se aisló como un mutante natural resistente al fago, empleando DGCC7710 como cepa original y el fago D858 como fago virulento. CRISPR1 de la cepa DGCC7710-RH 1 contiene 34 repeticiones (incluyendo la repetición terminal) y 33 espaciadores. Cuando se compara con la secuencia de CRISPR1 de *Streptococcus thermophilus*, cepa DGCC7710, la secuencia de CRISPR1 de *Streptococcus thermophilus*, cepa DGCC7710-RH posee un nuevo espaciador adicional (es decir, una secuencia marcadora) y una repetición adicional que flanquea el nuevo espaciador, en un extremo del locus CRISPR (es decir, próximo al líder, en el extremo 5' del locus CRISPR). Todos los demás espaciadores del locus CRISPR1 permanecieron invariables. La secuencia de CRISPR1 (5'-3') de la cepa DGCC7710-RH1 es:

caaggacagtattgatttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatctataaccactt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcaagattcgaattgt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttacc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggccttactc aacagcga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttcctgttcttttgttatctttc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctattctccgttttgttgcaatcct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaggcctcaaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaacagattcaa
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgcccgttgaattacacggcaaggca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatcttaattacaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtgtaacgtatta
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatcctgattaacttactcg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggagatgatgctatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaactgatgcaagagttcaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattctcatccggtaactgtcaagtg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaaggccatagaaggagacaacatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaatcatttccataacttcat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagatcagcaagcaagcgttagttact
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataattttaaga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggtatagcttaataatcattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagttaccgtttc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggcaactaacaattatgct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaatctgttttaggtacatt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatagacaagagttaaatggctt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaactgggatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctacttttacagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgaagctt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgctacatcacaaggatgatccaga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggttat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctaagaagcattggagcgtatattgattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgcccttctttgcccttgaclag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttggccattgagtg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTtgattcaacataaaaaggccagttcaattgaactggctt
(SEQ ID NO:5)

La secuencia líder de este CRISPR1 es:

5' caaggacagtattgatttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' (SEQ ID NO:6). La secuencia integrada que
 comprende repeticiones de CRISPR se muestra en mayúsculas, mientras que los espaciadores de CRISPR se
 muestran con minúsculas. 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' ttgattcaacataaaaagccagttcaattgaactggctt 3' (SEQ ID NO:8). Por tanto, para la cepa DGCC7710-RH1 de
S. thermophilus, el espaciador (5'-tcaacaattgcaacatctataaccactt-3' [SEQ ID NO:9]) constituye la secuencia
 10 marcadora específica de la 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, en la hebra positiva) del
 genoma de D858 (y tiene una identidad del 100% con la secuencia genómica de D858 a lo largo de 30 nucleótidos).
 El nuevo espaciador (es decir, la secuencia marcadora) que se integra en el locus CRISPR1 de la cepa DGCC7710-
 RH1 de *S. thermophilus*, confiere resistencia a esta cepa frente al fago D858.

15 La cepa DGCC7710-RH2 de *S. thermophilus*, se aisló como un mutante natural resistente al fago, empleando la
 cepa DGCC7710 de *S. thermophilus* como cepa original, y el fago D858 como fago virulento. El CRISPR1 de *S.*
thermophilus, cepa DGCC7710-RH2, contiene 34 repeticiones (incluyendo la repetición terminal) y 33 espaciadores.
 Cuando se compara con la secuencia de CRISPR1 de *S. thermophilus*, cepa DGCC7710, la secuencia de CRISPR1
 de *S. thermophilus*, cepa DGCC7710-RH2, posee un nuevo espaciador adicional (es decir, una secuencia
 20 marcadora) y una repetición adicional que flanquea el nuevo espaciador en un extremo del locus CRISPR (es decir,
 próximo al líder, en el extremo 5' del locus CRISPR). Todos los demás espaciadores del locus CPISPR1
 permanecen invariables. La secuencia de CRISPR1 (5'-3') de la cepa DGCC7710-RH2, es:

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaatcaatga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcaagattcgaattgt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttcttctttgttatcttttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttgttgcgaatcct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagatggaaaactagaacagattcaa
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaaggca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatctaattacaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtgtaacgtattha
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatcctgattaaactactcgcg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggagatgatgctatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacagaacgatgcaagagtcaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattctcatccgtaactgctcaagtg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatttaaggcatagaaaggagacaacatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaaatcatttcaacttcat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact

GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggtatagcttaatatcattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagttaccgtttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtaactaacaattatgct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgttcaaattctgttttaggtacatt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatagacaagagttaaaatggtcct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctacttttacagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgaagtct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgtacatctcaaggatgatccaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgaccttacaatggtttat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctaagaagcatttgagcgtatattgattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatttggccccttcttggccccttgactag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttggcccattgagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTtgattcaacataaaaagccagttcaattgaactggcttt
 (SEQ ID NO:10)

5 La secuencia líder es 5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' (SEQ ID NO:6). Las
 secuencias integradas que comprenden repeticiones de CRISPR se muestran con letras mayúsculas, mientras que
 el espaciador de CRISPR (es decir, la secuencia marcadora) se muestra con letras minúsculas. La repetición
 terminal tiene la secuencia 5' gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO:7), y la secuencia remolque tiene la
 secuencia 5' ttgattcaacataaaaagccagttcaattgaactggcttt 3' (SEQ ID NO:8). Por lo tanto, en el caso de *S. thermophilus*,
 cepa DGCC7710-RH2, el espaciador (5'-ttacgtttgaaaagaatatcaaatcaatga-3' [SEQ ID NO:11]) constituye el marcador
 10 específico de la cepa que identifica esta cepa mutante (es decir, la bacteria marcada). También se ha mostrado que
 la secuencia del nuevo espaciador existe dentro del genoma del fago D858. La secuencia del espaciador está
 localizada entre las posiciones 17215 y 17244 pb (es decir, en la hebra positiva) del genoma de D858 (y tiene una
 identidad del 100% con la secuencia genómica de D858 a lo largo de 30 nucleótidos). El nuevo espaciador que se
 integra en el locus CRISPR1 de *S. thermophilus*, cepa DGCC7710-RH2, confiere a la cepa resistencia frente al fago
 D858.

15 Además de los loci CRISPR marcados, desarrollados de forma natural, descritos anteriormente, en algunas
 realizaciones, las bacterias marcadas se producen empleando métodos de ADN recombinante, tal y como se
 conocen en la técnica. Por ejemplo, en algunas realizaciones, los oligonucleótidos sintéticos se producen y se
 insertan en un cultivo de bacterias originales, para producir bacterias marcadas. No se pretende que la presente
 invención esté limitada a loci CRISPR marcados, ya que loci adicionales se emplean en realizaciones de marcación.

20 M. Tipificación

La presente invención también proporciona métodos y composiciones para identificar (p. ej., tipificar) una bacteria
 marcada. En algunas realizaciones, la identificación implica la amplificación (p. ej., usando PCR) del locus CRISPR o
 de una porción del mismo. En algunas realizaciones, se diseña un primer cebador para hibridarlo con una secuencia
 que está localizada aguas arriba de la primera repetición de CRISPR de un locus CRISPR. Por ejemplo, en algunas
 25 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 vecino que está localizado aguas arriba del locus
 CRISPR. En algunas realizaciones, un segundo cebador se hibrida aguas abajo con el primer espaciador de

CRISPR o la primera agrupación de espaciadores de CRISPR. En algunas realizaciones, el segundo cebador se hibrida con el remolque o incluso con un gen vecino aguas abajo. En algunas realizaciones preferidas, el segundo cebador se hibrida dentro del locus CRISPR. En algunas realizaciones alternativas preferidas, el segundo cebador se hibrida al menos parcialmente, con un espaciador de CRISPR aguas abajo o con una agrupación de espaciadores de CRISPR.

En algunas realizaciones particularmente preferidas, después de la amplificación, la secuencia marcadora se identifica empleando cualquier método adecuado, conocido en la técnica. Por ejemplo, en algunas realizaciones, la secuencia marcadora se identifica determinando el producto de la amplificación, mediante un patrón de restricción. Por lo tanto, en algunas realizaciones, una vez que se ha amplificado el ADN que comprende el locus CRISPR o una porción del mismo, se digiere con una o varias enzimas de restricción.

En algunas realizaciones adicionales, las secuencias marcadoras se identifican empleando métodos de secuenciación tal y como se conocen en la técnica. En aún otras realizaciones, los métodos de hibridación bien conocidos en la técnica, se emplean en la presente invención. En algunas realizaciones, se emplean métodos que incluyen técnicas de hibridación conocidas en la técnica para la detección y/o la diferenciación de cepas bacterianas, incluyendo pero sin estar limitados a las mismas, la transferencia de tipo Southern, ensayos de cambio de la movilidad, ensayos de secuenciación empleando matrices de oligonucleótidos, tipificación de oligonucleótidos espaciadores, hibridación fluorescente in situ (FISH), ensayos de rastreo de heteroduplos y análisis de la movilidad de heteroduplos.

En algunas realizaciones adicionales preferidas, la secuencia marcadora identificada se compara con secuencias en al menos una base de datos de secuencias de fagos y/o al menos una base de datos de secuencias bacterianas. En algunas realizaciones, la secuencia marcadora se empareja con una o varias secuencias en la base de datos de secuencias de fagos, pero no con secuencias en la base de datos de secuencias bacterianas. Se contempla que a medida que se preparan nuevas bacterias marcadas, empleando los métodos proporcionados en esta memoria, bases de datos adicionales de marcadores se emplean en la presente invención.

N. Bacterias marcadas

Tal y como se emplean en esta memoria, las expresiones "bacterias marcadas", "bacteria marcada", "bacterias etiquetadas" y "bacteria etiquetada", se emplean todas indistintamente, en referencia a una bacteria original o a bacterias originales, en las que se ha modificado uno o varios loci CRISPR o una porción de los mismos se ha modificado (p. ej., mutado) de modo que son insensibles frente a uno o varios bacteriófagos, a los que estaban expuestas.

Tal y como se describe con más detalle en esta memoria, en algunas realizaciones, la bacteria marcada se expone a más de un bacteriófago (p. ej., de forma reiterativa, secuencial o simultánea), de modo que se acumula una o varias modificaciones genómicas en uno o en varios loci CRISPR, de tal manera que se vuelve insensible frente a cada uno de los bacteriófagos a los que se ha expuesto.

Para infectar las células, se inyecta un bacteriófago o se transfiere su ácido nucleico en la célula, con el ácido nucleico del fago que se encuentra de forma independiente del genoma de la célula. En algunas realizaciones, la infección da como resultado la expresión (es decir, la transcripción o la traducción) del ácido nucleico del bacteriófago dentro de la célula y la continuación del ciclo vital del bacteriófago.

En algunas realizaciones de la presente invención, después de la exposición al bacteriófago, la bacteria marcada tiene una susceptibilidad reducida o no tiene ninguna, frente a la infección y/o la multiplicación del bacteriófago, cuando se compara con la bacteria original. Tal y como se emplea en esta memoria, la expresión "susceptibilidad reducida frente a la infección y/o la multiplicación del bacteriófago", significa que el nivel de infección y/o multiplicación del bacteriófago en la bacteria marcada, no produce un efecto perjudicial sobre la bacteria marcada.

Por lo tanto, en algunas realizaciones de la presente invención, no se destruye una bacteria original después de la exposición al bacteriófago, debido a una mutación de la bacteria original de modo que se vuelve insensible frente al bacteriófago.

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

Tal y como se ha indicado anteriormente, la adición de los loci CRISPR marcados desarrollados de forma natural, descritos anteriormente, en algunas realizaciones, las bacterias marcadas se producen empleando técnicas de ADN

recombinante, tal y como se conocen en la técnica. Por ejemplo, en algunas realizaciones, los oligonucleótidos sintéticos se producen y se insertan en un cultivo de bacterias originales, para producir bacterias marcadas. No se pretende que la presente invención esté limitada a loci CRISPR marcados, ya que se emplean loci adicionales en realizaciones de marcación.

5 O. Cultivos

Los cultivos, en particular los cultivos iniciadores, se emplean de forma extensa en la industria alimentaria en la obtención de productos fermentados, que incluyen los productos lácticos (p. ej., yogur y queso), productos cárnicos, productos de panadería, vino y productos del reino vegetal. En particular, los cultivos iniciadores se emplean ampliamente en la obtención de muchos productos de leche fermentada, queso y mantequilla. Estos cultivos
10 iniciadores bacterianos imparten unas características específicas a diversos productos lácteos, mediante la realización de una serie de funciones. En algunas realizaciones particularmente preferidas, los cultivos empleados en la presente invención son cultivos "útiles industrialmente". Tal y como se emplea en esta memoria, esta expresión se refiere a cualquier cultivo bacteriano que se emplea en cualquier industria, incluyendo pero sin estar limitadas a las mismas, la producción de alimentos, piensos, cosméticos, productos farmacéuticos, nutracéuticos, probióticos,
15 enzimas, metabolitos, etc. De hecho, no se pretende que la presente invención esté limitada por ningún cultivo o industria particular, ya que la presente invención se emplea en numerosas aplicaciones.

Los cultivos comerciales, sin concentrar de bacterias, se denominan en la industria "cultivos madre", y se propagan en el sitio de la producción (p. ej., productos lácteos), antes de añadirlos a un producto comestible iniciador (p. ej.,
20 leche), para la fermentación. El cultivo iniciador propagado en el sitio de producción para inocularlo en un producto iniciador comestible, se denomina la "masa de iniciación".

Cultivos iniciadores adecuados para el uso en la presente invención incluyen cualquier organismo que sea adecuado para el uso en la industria alimentaria, cosmética y/o farmacéutica. En algunas realizaciones preferidas, el cultivo iniciador se emplea en la industria láctea. De hecho, los cultivos de bacterias lácticas se emplean generalmente en la preparación de productos lácteos fermentados (p. ej., leche agria, yogur y crema agria), y en la preparación de de
25 mantequilla y queso (p. ej., queso brie, havarti, Cheddar, Monterey jack, etc.).

Tal y como se emplea en esta memoria, la expresión "bacteria láctica" se refiere a bacterias Gram positivas, microaerófilas o anaerobias que fermentan el azúcar, con la producción de ácidos, que incluyen el ácido láctico como el ácido producido de forma predominante, ácido acético, ácido fórmico y ácido propiónico. Las bacterias lácticas más útiles en la industria incluyen especies de *Lactococcus* (p. ej., *Lactococcus lactis*), *Lactobacillus*,
30 *Bifidobacterium*, *Streptococcus*, *Leuconostoc*, *Pediococcus* y *Propionibacterium*. En algunas realizaciones, la presente invención proporciona cultivos iniciadores que comprenden al menos una especie de bacteria láctica, tal como, *L. lactis*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Streptococcus thermophilus* o combinaciones de las mismas. Los cultivos iniciadores de bacterias lácticas se emplean generalmente en la industria alimentaria como cultivos de cepas mixtas que comprenden una o varias especies. En algunas realizaciones que comprenden cultivos
35 de cepas mixtas (p. ej., cultivos iniciadores para el yogur), que comprenden cepas de *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Streptococcus thermophilus*, hay una relación simbiótica entre las especies en las que la producción de ácido láctico es superior, comparadas con cultivos de bacterias lácticas de una sola cepa (véase, p. ej., Rajagopal y col., J. Dairy Sci., 73:894-899 [1990]).

En algunas realizaciones particularmente preferidas, el cultivo iniciador es una especie de bacteria láctica, que incluye pero no está limitada a las cepas de *Bifidobacterium*, *Brevibacterium* o *Propionibacterium*. Cultivos iniciadores adecuados del grupo de las bacterias lácticas incluyen, pero no están limitados a los mismos, cepas empleadas generalmente, *Lactococcus*, *Streptococcus*, *Lactobacillus* (p. ej., *Lactobacillus acidophilus*),
40 *Enterococcus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*. La especie *Lactococcus* incluye, pero no está limitada a *Lactococcus lactis*, empleada ampliamente, que incluye *Lactococcus lactis* subsp. *lactis* y *Lactococcus lactis* subsp. *cremoris*, y *Lactococcus lactis* subsp. *lactis biovar*. Otras especies de bacterias ácidas incluyen *Leuconostoc*, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Lactobacillus helveticus*. Además, cepas probióticas tales como *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*, se emplean en la potenciación del sabor y proporcionan beneficios para la salud. Los cultivos termofílicos de bacterias lácticas
45 empleados generalmente 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* (p. ej., *L. helveticus*) durante la fabricación, para obtener un sabor deseado.

En algunas realizaciones, el cultivo iniciador comprende o consiste en una cepa modificada genéticamente (preparada de acuerdo con los métodos deseados en esta memoria), con una o varias cepas de bacterias lácticas o cualquier otra cepa adecuada como cultivo iniciador. Tal y como es conocido por los expertos en la técnica, la
50 selección de organismos para los cultivos iniciadores empleados en la presente invención, depende del tipo particular de productos que se van a preparar o a tratar. Por ello, por ejemplo, para la fabricación de queso y de mantequilla, se emplea ampliamente cultivos mesofílicos de especies de *Lactococcus*, de *Leuconostoc* y de *Lactobacillus*, mientras que las cepas termofílicas de especies de *Streptococcus* y de *Lactobacillus* se emplean ampliamente para el yogur y otros productos lácteos fermentados.

En algunas realizaciones, el producto iniciador es un iniciador seco, mientras que en otras realizaciones es un iniciador concentrado, y en otras realizaciones es un cultivo congelado. En algunas realizaciones preferidas, los cultivos iniciadores secos comprenden al menos una bacteria láctica. En algunas realizaciones, el cultivo iniciador se emplea para la inoculación directa. En algunas realizaciones preferidas, el cultivo es un cultivo iniciador concentrado, empleado para la inoculación directa.

En algunas realizaciones, el cultivo bacteriano iniciador comprende una cepa o especie bacteriana (es decir, es un cultivo puro). Por tanto, en estas realizaciones, todo el cultivo bacteriano iniciador o al menos una porción significativa del mismo, 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, es un cultivo mixto, tal como una mezcla definida)

Los cultivos iniciadores preparados empleando cualquier técnica adecuada, conocida en la técnica, se emplean en la presente invención (véase, *p. ej.*, el documento de patente de EE.UU. nº 4.621.058). Por ejemplo, los cultivos iniciadores preparados mediante la introducción de un inóculo (*p. ej.*, un cultivo bacteriano) en un medio de crecimiento, 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 por ningún método particular para preparar cultivos iniciadores.

Cultivos iniciadores secos preparados empleando cualquier método adecuado conocido en la técnica, se emplean en la presente invención (véase, *p. ej.*, los documentos de patente de EE.UU. nº 4.423.079 y 4.140.800). En algunas realizaciones, los cultivos iniciadores secos empleados en la presente invención son preparaciones sólidas (*p. ej.*, comprimidos, pastillas, cápsulas, partículas, gránulos, polvos, cualquiera de los cuales es humectable, se ha secado por atomización o está liofilizado, en algunas realizaciones). En algunas realizaciones alternativas, los cultivos iniciadores secos de la presente invención son una pastilla muy congelada o un polvo liofilizado. Estos cultivos iniciadores secos se preparan empleando cualquier método adecuado, conocido en la técnica.

En algunas realizaciones, los cultivos iniciadores empleados en la presente invención comprenden material concentrado que tiene concentraciones sustancialmente elevadas de al menos una especie bacteriana. En algunas realizaciones preferidas, el material concentrado se diluye en agua o se resuspende en agua o en otro diluyente adecuado (*p. ej.*, un medio de crecimiento adecuado, aceite mineral o aceite vegetal), para el empleo en la presente invención. En algunas realizaciones, los cultivos iniciadores secos y concentrados de la presente invención se preparan empleando métodos bien conocidos en la técnica, centrifugación, filtración o una combinación de estas técnicas, pero no limitados a los mismos.

P. Productos

La presente invención se emplea en la producción de diversos productos, incluyendo pero sin estar limitada a los mismos, productos alimentarios, piensos, productos cosméticos y/o productos farmacéuticos. De hecho, se contempla que cualquier producto preparado a partir de un cultivo bacteriano o que comprenda un cultivo bacteriano, se emplee en la presente invención. Estos incluyen, sin estar limitados a los mismos, frutas, legumbres, plantas forrajeras y vegetales que incluyen productos derivados, grano y productos derivados del grano, productos lácteos y productos derivados de lácteos, carnes, aves, mariscos, productos cosméticos, enzimas, metabolitos y productos farmacéuticos.

Tal y como se emplea en esta memoria, el término "alimento" se utiliza en un amplio sentido e incluye piensos, productos alimenticios, ingredientes alimenticios, suplementos alimenticios y alimentos funcionales. Aunque el término incluye alimento para humanos, se entiende que el término también comprende alimentos para animales (es decir, "piensos"). Sin embargo, en algunas realizaciones preferidas, la presente invención proporciona alimento para el consumo humano. Tal y como se emplea en esta memoria, la expresión "ingrediente alimenticio" incluye una formulación adecuada para la adición a los alimentos. En algunas realizaciones, las formulaciones se emplean en pequeñas cantidades en una variedad de productos que las requieren, por ejemplo, la acidificación o la emulsión. Tal y como se emplea en esta memoria, la expresión "alimento funcional" se refiere a alimentos que son capaces de proporcionar, no solo un efecto nutritivo y/o una satisfacción del gusto, sino que también son capaces de proporcionar un efecto beneficioso adicional al consumidor. En algunas realizaciones las bacterias de la presente invención comprenden o se añaden a un ingrediente alimentario, un suplemento o un alimento funcional. Se contempla que el alimento se proporcione en cualquier forma adecuada, que incluye, sin estar limitada a las mismas, soluciones, sólidos, geles, emulsiones, etc., dependiendo del uso y/o del modo de aplicación y/o del modo de administración. De hecho, no se pretende que la presente invención esté limitada por alimentos en cualquier forma particular. En algunas realizaciones, las bacterias de la presente invención se emplean en la preparación de numerosos productos alimenticios, que incluyen sin estar limitados a los mismos, los productos pasteleros, los productos lácteos, los productos cárnicos, los productos de aves de corral, los pescados y los productos de panadería. En algunas realizaciones, las bacterias se emplean como ingredientes en bebidas gaseosas, zumos de frutas, bebidas que comprenden proteína del suero, té para la salud, bebidas con cacao, bebidas lácteas, bebidas con bacterias lácticas, queso, yogur, yogur líquido y vino. En algunas realizaciones adicionales, la presente invención proporciona métodos para preparar alimentos, que incluyen métodos que comprenden la mezcla por adición de bacterias de acuerdo con la presente invención, con un ingrediente alimenticio (*p. ej.*, un material iniciador

para un alimento). En algunas realizaciones preferidas, el alimento proporcionado en esta memoria en un producto lácteo. En algunas realizaciones particularmente preferidas, el producto lácteo se selecciona entre yogur, queso (p. ej., un queso fresco ácido, un queso curado, un queso semicurado, un queso fresco), leche agria, requesón, crema agria, kefir, una bebida a base de suero de leche fermentado, kumis, una bebida a base de leche y una bebida de yogur.

Tal y como se emplea en esta memoria, los términos "alimento" y "pienso" incluyen, pero no están limitados a, materias vegetales primas y procesadas, tal y como el material no vegetal. Se pretende que el alimento/pienso sea adecuado para el consumo de cualquier animal, humano o no humano. En algunas realizaciones preferidas, el alimento/pienso se emplea con ganado (p. ej., vacas, ovejas, cerdos, etc.), aves de corral (p. ej., pollos y pavos), peces, reptiles o crustáceos. De hecho, no se pretende que la presente invención esté limitada por el alimento/pienso para ningún organismo en particular.

EXPERIMENTAL

Los siguientes ejemplos se proporcionan para mostrar e ilustrar adicionalmente ciertas realizaciones y aspectos preferidos de la presente invención y no deben interpretarse como una limitación del alcance de la misma.

En la descripción experimental que sigue a continuación, las siguientes abreviaturas se aplican a: °C (grados Centígrados); H₂O (agua); aa (aminoácido); pb (par de bases); kb (kilopar de bases); kD (kilodaltons); g y gm (gramos); µg y ug (microgramos); mg (miligramos); ng (nanogramos); µl y µl (microlitros); ml (mililitros); mm (milímetros); nm (nanómetros); µm y um (micrómetro); M (molar); mM (milimolar); µM y uM (micromolar); seg y s (segundos); min(s) (minuto/minutos); h(s) (hora/horas); MOI (multiplicidad de la infección); EOP (eficiencia de la extensión en placas); UFP (unidades formadoras de placas); MgCl₂ (cloruro de magnesio); NaCl (cloruro sódico); PAGE (electroforesis en gel de poliacrilamida); PBS (solución salina tamponada con fosfato [NaCl 150 mM, tampón de fosfato sódico 10 mM, pH 7,2]); SDS (dodecilsulfato sódico); Tris (tris(hidroximetil)aminometano); p/v (peso a volumen); v/v (volumen a 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 emplea, a no ser que se indique de otro modo, técnicas convencionales de química, biología molecular, microbiología, ADN recombinante e inmunología, que se incluyen en las capacidades de una persona con conocimientos en la materia. Tales métodos son bien conocidos por los expertos en la técnica.

Tal y como se emplea en esta memoria, DGCC7710 también se denomina "WT"; DGCC7710RH1 también se denomina "DGCC7710-RH1" y "RH1"; DGCC7710RH2 también se denomina "DGCC7710-RH2" y "RH-2."

EJEMPLO 1

Marcación de DGCC7710 de *Streptococcus thermophilus* empleando el bacteriófago D2972, mediante la inserción de una unidad aislada de repetición-espaciador dentro de CRISPR1

En este ejemplo, la cepa DGCC7710 de *S. thermophilus* (depositada en la "Collection Nationale de Cultures de Microorganismes" francesa, con el número CNCM 1-2423), se marcó mediante la inserción "natural" dentro de su locus CRISPR1, de una unidad adicional de repetición-espaciador, siendo originario el espaciador del bacteriófago D2972. El locus CRISPR1 de DGCC7710 contiene 33 repeticiones (incluyendo la repetición terminal) y 32 espaciadores (véase, el número de orden de GenBank: EF434469). El bacteriófago D2972 se aisló a partir de un producto lácteo fermentado, empleando la cepa DGCC7710. Su genoma se ha secuenciado totalmente (véase, el número de orden de GenBank: AY699705). La secuencia del locus CRISPR1 de DGCC7710 es:

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgtttgacagcaaatcagattcgaattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgattgacaatctgctgaccactgttacc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacactggcaggcttattactcaacagcga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttcctgtttgttatctttc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctcattctccgtttgttgcgaatcct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaacagattcaa
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgtgaattacaggcaagggtca
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatctaatacaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgtagcgtcatgtgtaactgattta
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtccaatcctgattaacttactctg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggatgatgctatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatcgaaggttcaag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattctcatccgtaactgctcaagtg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaggagacaacatg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattaaatcatttcataactcat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaatttataattttaaga
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataattatggtatagcttaatatcattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcattcagcagcttcgagttaccgttcc
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatctgagggtcaactaacaattatgct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgttcaattctgttttaggtacatt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatcagacaagagttaaatggtctt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaactgtagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctacttttacagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgaaggtct
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaggatgatccaga

GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtgtgatgacctctacaatggtttat
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatttgccccttcttggccccttgactag
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcattgtgccattgagt
GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTtgattcaacataaaaagccagttcaattg
aacttggctt (SEQ ID NO: 1)

5 En la primera etapa, la cepa original (DGCC7710) se expuso al bacteriófago donante (D2972). DGCC7710 se cultivó
 previamente en medio a base de leche estéril (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 a base de leche estéril,
 aproximadamente 0,05% (p/v). Aproximadamente 10⁷ bacteriófagos D2972 se añadieron al medio a base de leche
 10 inoculado (número final de bacteriófagos, aproximadamente 10⁶ ufp/ml). La mezcla se cultivó después a 42°C
 durante 16 horas. Después de la incubación, diluciones del cultivo se extendieron en placas sobre medio
 M17-glucosa (0,5% p/v), para obtener colonias aisladas después de incubar a 42°C durante 24 horas. Una cantidad
 de colonias se eligieron y se dejaron crecer de forma separada en medio a base de leche estéril a 42°C durante
 18 horas, para proporcionar una reserva de material aislado.

15 A continuación, se identificó cada cepa marcada mediante un análisis de la secuencia de sus loci CRISPR1. En este
 Ejemplo, una cepa marcada es una variante de la cepa original, en la que la variante contiene una unidad adicional
 de repetición-espaciador dentro del locus CRISPR1 (véase la Figura 3). Típicamente, la parte del espaciador de la
 unidad adicional tiene un tamaño de aproximadamente 30 nucleótidos, y su secuencia es idéntica a una
 subsecuencia del fago donante. El material aislado se cultivó por separado a 42°C durante 18 horas en medio
 M17-glucosa. Las células se eligieron y se extrajo su ADN. En cada material aislado, se amplificó la región del
 20 cromosoma correspondiente al extremo del líder del locus CRISPR1, mediante PCR empleando los siguientes
 cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:15]; véase, Bolotin y col., Microbiol.,
 151:2551-2561 [2005]); y cebador inverso CR1-89R5 (5'-acaaaaacggaagaatgaagttg-3' [SEQ ID NO:16]). La mezcla
 de la reacción PCR (volumen final de 25 µL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno
 de los cuatro dNTP 2 mM, 10 a 100 ng de ADN, cada cebador 0,2 µM, 1,25 U de polimerasa Taq (Promega). Las
 25 condiciones de la ciclación de la PCR eran las siguientes: desnaturalización previa 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 de una etapa de elongación final a 72°C durante 4 min. La secuencia nucleica de cada
 fragmento de la PCR se determinó a continuación empleando el "cebador directo" y una metodología clásica de
 secuenciación, tal y como se conoce en la técnica. Cada secuencia se comparó con la obtenida para la cepa
 original, para identificar la presencia de secuencias adicionales.

30 A través de experimentos independientes, se crearon múltiples cepas marcadas a partir de DGCC7710, empleando
 D2972 como donante. Algunas de estas cepas marcadas se describen en la Tabla 1-1. Todas difieren de la cepa

original en una sola unidad adicional de repetición-espaciador en el extremo del líder del locus CRISPR1. En todos los casos, la parte del espaciador de la unidad adicional tiene una identidad del 100% con una subsecuencia del bacteriófago donante D2972 (esta subsecuencia del bacteriófago donante también se denomina "pro-espaciador"). Típicamente, las cepas marcadas descritas en este Ejemplo, poseen un locus CRISPR1 compuesto por 34 repeticiones y 33 espaciadores; 32 espaciadores que proceden de DGCC7710, el espaciador suplementario originario de D2972.

5

Tabla 1-1. Descripción de cepas marcadas en el locus CRISPR1 obtenidas a partir de DGCC7710 empleando D2972 como donante				
Cepa marcada	Cepa original	Fago donante	ADN del donante insertado	
			Secuencia	Posición *
DGCC7710 _{phi2972} ^{+S40}	DGCC7710	D2972	TCTGGAAAGCATATTGAGGGA GCTACTCTT (SEQ ID NO:17)	27974-28003 (+)
DGCC7710 _{phi2972} ^{+S41}	DGCC7710	D2972	TCTAATCCCCTAGGAATAGT GGGTAGTAA (SEQ ID NO:18)	25693-25722 (+)
DGCC7710 _{phi2972} ^{+S43}	DGCC7710	D2972	TTATAACATAACGGTTAGTTG GCCTCTAT (SEQ ID NO:19)	23410-23382 (-)
DGCC7710 _{phi2972} ^{+S44}	DGCC7710	D2972	AAGGAGCTAGCCACATTTCCG CAATTGATA (SEQ ID NO:20)	23334-23363 (+)
DGCC7710 _{phi2972} ^{+S45}	DGCC7710	D2972	CAGCTTGAAATGTTTATTGAA GCAGCAGTG (SEQ ID NO:21)	24624-24653 (+)
DGCC7710 _{phi2972} ^{+S46}	DGCC7710	D2972	AAATCAGTTTTTTGTTTCAGAA ACTTGTCT (SEQ ID NO:22)	25582-25611 (+)

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

EJEMPLO 2

Marcación de DGCC7710 de *Streptococcus thermophilus* utilizando el bacteriófago D858 mediante la inserción de una unidad aislada de repetición-espaciador dentro de CRISPR1

10 En este Ejemplo, la cepa DGCC7710 de *S. thermophilus* se marcó mediante la inserción "natural" dentro de su locus CRISPR1, con una unidad adicional de repetición-espaciador, en donde el espaciador provenía 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 los virus *Siphoviridae*. La secuencia de su genoma se ha determinado completamente (número de orden de GenBank: EF529515). Este fago es virulento para la cepa DGCC7710 de *S. thermophilus*.

15 Su genoma se ha secuenciado totalmente

20 En primer lugar, la cepa original (DGCC7710) se expuso al bacteriófago donante (D858). DGCC7710 se cultivó previamente en medio a base de leche estéril (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 estéril a base de leche, aproximadamente 0,05% (p/v). Aproximadamente 10⁷ bacteriófagos D858 se añadieron al medio a base de leche inoculado (número final de bacteriófagos, aproximadamente 10⁶ ufp/ml). La mezcla se cultivó después a 42°C durante 16 horas. Después de la incubación, diluciones del cultivo se extendieron en placas sobre medio de M17-glucosa (0,5% p/v) para obtener colonias aisladas después de la incubación a 42°C durante 24 horas. Una cantidad de colonias se recogieron y se dejaron crecer de forma separada en medio a base de leche estéril, a 42°C durante 18 horas, para proporcionar una reserva de material aislado.

25 A continuación, cada cepa marcada se identificó mediante análisis de la secuencia de su locus CRISPR1. En este Ejemplo, una cepa marcada es una variante de la cepa original que contiene una unidad adicional de repetición-espaciador dentro de su locus CRISPR1 (véase la Figura 3). Típicamente, la parte del espaciador de la unidad adicional tenía aproximadamente un tamaño de 30 nucleótidos y su secuencia era idéntica a una subsecuencia del fago donante. El material aislado se cultivó por separado a 42°C durante 18 horas en medio M 17-glucosa. Las células se eligieron y se extrajo su ADN. En cada material aislado, se amplificó la región del cromosoma correspondiente al extremo del líder del locus CRISPR1, mediante PCR empleando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:1 5], Bolotin y col., [2005], *supra*); y cebador inverso CR1-89R5 (5'-acaaaaacggaagaatgaagttg-3' [SEQ ID NO:20]). La mezcla de la reacción PCR (volumen final de 30 35 25 µL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno de los cuatro dNTP 2 mM, 10 a

100 ng de ADN, cada cebador 0,2 μM, 1,25 U de polimerasa *Taq* (Promega). Las condiciones de la ciclación de la PCR eran las siguientes: desnaturalización previa 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 de una etapa de elongación final a 72°C durante 4 min. La secuencia nucleica de cada fragmento de la PCR se determinó a continuación, empleando el "cebador directo" y una metodología clásica de secuenciación, tal y como se conoce en la técnica. Cada secuencia se comparó con la obtenida para la cepa original, para identificar la presencia de secuencias adicionales.

Una cepa marcada creada a partir de DGCC7710, empleando D858 como donante, se describe en la Tabla 2-1. Difiere de la cepa original por una sola unidad adicional de repetición-espaciador en el extremo del líder del locus CRISPR1. La parte del espaciador de la unidad adicional tiene una identidad del 100% con una subsecuencia del bacteriófago donante D858.

Tabla 2-1. Descripción de cepas marcadas en el locus CRISPR1, obtenidas a partir de DGCC7710 y utilizando D858 como donante

Cepa marcada	Cepa original	Fago donante	ADN del donante insertado	
			Secuencia	Posición *
DGCC7710 _{phi858} ^{+S42}	DGCC7710	D858	TCGATAAATCAGCCAAAGTATT AAGTGGTT (SEO ID NO:23)	27560-27589 (+)

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

EJEMPLO 3

Marcación de DGCC7710 de *Streptococcus thermophilus* empleando el bacteriófago D2972 mediante la inserción de múltiples unidades de repetición-espaciador dentro de CRISPR1

En este Ejemplo, la cepa DGCC7710 de *S. thermophilus*, se marcó con una inserción "natural" dentro del locus CRISPR1, mediante la adición de múltiples unidades de repetición-espaciador, proviniendo los espaciadores del bacteriófago D2972.

En primer lugar, la cepa original (DGCC7710) se expuso al bacteriófago donante (D2972). DGCC7710 se cultivó previamente en medio a base de leche estéril (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 estéril a base de leche, aproximadamente 0,05% (p/v). Aproximadamente 10⁷ bacteriófagos D2972 se añadieron al medio a base de leche inoculado (número final de bacteriófagos aproximadamente 10⁶ ufp/ml). La mezcla se cultivó después a 42°C durante 16 horas. Después de la incubación, diluciones del cultivo se extendieron en placas sobre medio M17-glucosa (0,5% p/v), para obtener colonias aisladas después de incubar a 42°C durante 24 horas. Una cantidad de colonias se eligieron y se dejaron crecer de forma separada en medio a base de leche estéril a 42°C durante 18 horas, para proporcionar una reserva de material aislado.

A continuación, cada cepa marcada se identificó a través de un análisis de la secuencia de su locus CRISPR1. En este Ejemplo, una cepa marcada es una variante de la cepa original que contiene múltiples unidades adicionales de repetición-espaciador dentro de su locus CRISPR1 (véase la Figura 3). Típicamente, la parte del espaciador de cada unidad adicional tenía aproximadamente un tamaño de 30 nucleótidos y su secuencia era idéntica a una subsecuencia del fago donante. El material aislado se cultivó por separado a 42°C durante 18 horas en medio M17-glucosa. Las células se eligieron y se extrajo su ADN. En cada material aislado, se amplificó la región del cromosoma correspondiente al extremo del líder del locus CRISPR1, mediante PCR empleando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctagtctctc-3' [SEQ ID NO:15], Bolotin y col., 2005, *supra*); cebador inverso CR1-89R5 (5'-acaaaaacggaagaatgaagttg-3' [SEQ ID NO:16]). La mezcla de la reacción PCR (volumen final de 25 μL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno de los cuatro dNTP 2 mM, 10 a 100 ng de ADN, cada cebador 0,2 μM, 1,25 U de polimerasa *Taq* (Promega). Las condiciones de la ciclación de la PCR eran las siguientes: desnaturalización previa 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 de una etapa de elongación final a 72°C durante 4 min. La secuencia nucleica de cada fragmento de la PCR se determinó a continuación empleando el "cebador directo" y una metodología clásica de secuenciación, tal y como se conoce en la técnica. Cada secuencia se comparó con la obtenida para la cepa original, para identificar la presencia de secuencias adicionales.

A través de experimentos independientes, se crearon múltiples cepas marcadas a partir de DGCC7710, empleando D2972 como donante. Algunas de las cepas marcadas se describen en la Tabla 3-1. Todas ellas difieren de la cepa

original en múltiples unidades adicionales de repetición-espaciador en el extremo del líder del locus CRISPR1. En todos los casos, la parte del espaciador de cada unidad adicional, tiene una identidad del 100% con una subsecuencia del bacteriófago donante D2972.

Tabla 3-1. Descripción de cepas marcadas en el locus CRISPR1 procedentes de DGCC7710, empleando D2972 como donante

Cepa marcada	Cepa original	Fago donante	ADN del donante insertado	
			Secuencia	Posición
DGCC7710 _{phi2972} ^{+S46+S47}	DGCC7710	D2972	AAATCAGTTTTTTGTTTCAGAAACTTG TTCT (SEQ ID NO:24) TTGTCTATTACGACAACATGGAAGA TGAT (SEQ ID NO:25)	33045-33073 (+) 25582-25611 (+)
DGCC7710 _{phi2972} ^{+S48+S49}	DGCC7710	D2972	TTTTGAGAAAAGTCTTTAACGATGCA GTAGC (SEQ ID NO:26) TAATAGTTTACCAAAATCATCTTTATT CCAA (SEQ ID NO:27)	25967-25938 (-) 6008-6037 (+)
DGCC7710 _{phi2972} ^{+S50+S51}	DGCC7710	D2972	GAAGTTGAAATAATTCGAGAAATAG AACTC (SEQ ID NO:28) TGGAAACCAAGAAAATGCAATAGAAT GGAAG (SEQ ID NO:29)	34105-34134 (+) 29246-29275 (+)
DGCC7710 _{phi2972} ^{+S52+S54}	DGCC7710	D2972	CTGATTGTTAATGTACGAGGGCTCC AGCCA (SEQ ID NO:30) CTCAGTCGTTACTGGTGAACCAAGTTT CAAT (SEQ ID NO:31)	31582-31611 (+) 21732-21703 (-)
DGCC7710 _{phi2972} ^{+S53+S54}	DGCC7710	D2972	TGTTTCAAAGTTTTCGGGTCCAAGTAT CAAT (SEQ ID NO:32) TTTTCCGTCTCTTTTTAGCAAAGA TAGC (SEQ ID NO:33)	29647-29618 (-) 16681-16652 (-)
DGCC7710 _{phi2972} ^{+S61+S62}	DGCC7710	D2972	GAATCGTGGGATATTCGTCTTACGT TTGA (SEQ ID NO:34) ACATATCGACGTATCGTGATTATCCC ATT (SEQ ID NO:35)	31709-31737 (+) 17182-17211 (+)
DGCC7710 _{phi2972} ^{+S55+S63}	DGCC7710	D2972	CTGGAAGCATATTGAGGGAGCTAC TCTT (SEQ ID NO:36) GTATATCGAAGAACGACTGAAAGAG CTTGA (SEQ ID NO:37) TCTAATCCCACTAGGAATAGTGGGT AGTAA (SEQ ID NO:38)	25693-25722 (+) 1114-1142 (+) 27381-27409 (+)

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

EJEMPLO 4

Marcación de DGCC7710 de *Streptococcus thermophilus*, empleando el bacteriófago D2972 mediante la inserción reiterativa de unidades de repetición-espaciador dentro de CRISPR1

5 En este Ejemplo, la cepa DGCC7710 de *S. thermophilus*, se marcó por medios "naturales" a través de la inserción reiterativa dentro de su locus CRISPR1, de unidades adicionales de repetición-espaciador, procediendo los espaciadores del bacteriófago D2972 y de bacteriófagos obtenidos a partir de D2972.

10 En la primera reiteración, la cepa original (DGCC7710) se expuso al bacteriófago donante (D2972) y una cepa marcada se aisló y se caracterizó empleando la misma metodología que se ha descrito en el Ejemplo 1. Comparada con DGCC7710, esta cepa marcada (denominada DGCC7710_{phi2972}^{S6}) posee una unidad adicional de repetición-espaciador, tal y como se describe en la Tabla 4, en su locus CRISPR1.

15 Debido a la inserción de una unidad adicional de repetición-espaciador en el locus CRISPR1 de la cepa DGCC7710_{phi2972}^{S6}, el bacteriófago donante D2972 ya no era virulento contra DGCC7710_{phi2972}^{S6}, y no se podía utilizar como bacteriófago donante para esta cepa. Este problema se superó mediante el uso de un fago donante mutado, obtenido a partir de D2972 que incluye al menos una modificación específica dentro de su genoma (es decir, un "fago mutado"). Este fago mutado se seleccionó mediante la exposición del bacteriófago donante a la cepa marcada, de modo que una modificación (es decir, mutación) del fago original lo volvía virulento frente a la cepa marcada.

20 DGCC7710_{phi2972}^{S6} se cultivó previamente en medio a base de leche a 42°C durante 18 horas. Un medio a base de leche se inoculó a continuación con el precultivo de DGCC7710_{phi2972}^{S6}, con una concentración de aproximadamente 10⁶ ufp/ml y con una suspensión de D2972 con una MOI (multiplicidad de la infección) superior a 100. El cultivo se incubó a 42°C durante 18 horas, y a continuación se centrifugó durante 10 min a 10.000 x g. El material sobrenadante se recogió y se filtró empleando un filtro de 0,45 µm. Se emplearon diluciones adecuadas del material sobrenadante filtrado para inocular medios de agar y M17-glucosa, sembrados con un césped de DGCC7710_{phi2972}^{S6}, empleando métodos bien conocidos en la técnica. Las placas de agar sembradas, se incubaron durante 24 horas a 42°C. Una placa aislada se eligió y se cultivó sobre DGCC1710_{phi2972}^{S6} en medio M17-glucosa, durante 6 horas a 42°C. Una suspensión de este nuevo bacteriófago denominado D4724, se obtuvo por filtración del cultivo a través de un filtro de 0,45 µm. Se verificó la virulencia de D4724 frente a DGCC7710_{phi2972}^{S6}.

30 A continuación, la cepa DGCC7710_{phi2972}^{S6} se expuso al bacteriófago donante D4724 y se aisló y se caracterizó una cepa marcada empleando la metodología descrita en el Ejemplo 1. Comparada con DGCC7710, esta cepa marcada (denominada DGCC7710_{phi2972}^{S6}_{phi4724}^{S15}), posee en su locus CRISPR1, 2 unidades adicionales de repetición-espaciador, tal y como se describe en la Tabla 4-1.

35 Con el fin de una tercera reiteración, se seleccionó un segundo bacteriófago mutado, denominado D4733, mediante la estimulación de DGCC7710_{phi2972}^{S6}_{phi4724}^{S15} con D4724, usando la misma metodología que para obtener D4724. El bacteriófago D4733 es virulento frente a DGCC7710_{phi2972}^{S6}_{phi4724}^{S15}. Después de la exposición a la virulencia de DGCC7710_{phi2972}^{S6}_{phi4724}^{S15} con el bacteriófago donante D4733, se aisló y se caracterizó una cepa marcada, empleando la misma metodología que se ha descrito en el Ejemplo 1. Comparada con DGCC7710, esta cepa marcada, denominada DGCC7710_{phi2972}^{S6}_{phi4724}^{S15}_{phi4733}^{S29}, posee 3 unidades adicionales de repetición-espaciador en su locus CRISPR1, tal y como se describe en la Tabla 4-1.

Tabla 4-1. Descripción de las cepas marcadas de forma reiterativa en el locus CRISPR1 procedentes de DGCC7710, empleando D2972 y los fagos mutados D4724 y D4733 como bacteriófagos donantes				
Cepa marcada	Cepa original	Fago donante	ADN del donante insertado	
			Secuencia	Posición
DGCC7710 ^{S6} _{phi2972}	DGCC7710	D2972	GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:39)	34521-34492 (-)
DGCC7710 ^{S6} _{phi2972} ^{S20} _{phi4724}	DGCC7710 ^{S6} _{phi2972}	D4724	GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:40) TTATATCGAAGAACGACTGAA AGAGCTTGA (SEQ ID NO:41)	34521-34492 (-) 1113-1142 (+)
DGCC7710 ^{S6} _{phi2972} ^{S20} _{phi4724} ^{S29} _{phi4733}	DGCC7710 ^{S6} _{phi2972} ^{S20} _{phi4724}	D4733	GCCCTTCTAAATTGGATTACCTT CCGAGGTG (SEQ ID NO:42) TTATATCGAAGAACGACTGAA AGAGCTTGA (SEQ ID NO:43) ATTGCCATGATTTCAAATTTTAA TTGGGAT (SEQ ID NO:44)	34521-34492 (-) 1113-1142 (+) 32136-32164 (+)

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

EJEMPLO 5

Marcación de DGCC3198 de *Streptococcus thermophilus*, empleando el bacteriófago D4241 mediante inserción de una unidad aislada de repetición-espaciador dentro de CRISPR1

5 En este Ejemplo, la cepa DGCC3198 de *S. thermophilus*, (también conocida como LMD-9 y depositada ante la "American Type Culture Collection" como ATCC BAA-365), se marcó mediante la inserción "natural" dentro de su locus CRISPR1 de una unidad adicional de repetición-espaciador, procediendo el espaciador del bacteriófago D4241. El locus CRISPR1 de DGCC3198 contiene 17 repeticiones (incluyendo la repetición terminal) y 16 espaciadores (número de orden de GenBank: CP000419). El bacteriófago D4241 se aisló a partir de un producto lácteo fermentado, empleando la cepa DGCC3198. La secuencia del locus CRISPR 1 de DGCC3198 es:

caagaacagttattgattttataatcactatgtgggtatgaaaactcaaaaalcatttgag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACatgatgatgaagatcgcctactaac
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcttcacctcaaatcttagagctggactaaa
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACatgctgtaaaaataaccgacctactact
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgaagctcatcatgtaaggctaaaacctat
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtagtctaatagatttctgcaccattgta
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACattcgtgaaaaaatatcgtgaaataggcaa
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctaggctcatcaagataaatcagtagc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaaaaacatggggcggcgtaaatgtgtaag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaaccagcaagagagcgcggacaacatt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtataacacaggttagaggatgttataact
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctagaagctcaagcggtaaaagttagggcg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcttgaggggcaagccctcggcgtccatt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaactaccaagcaaatcagcaatcaataagt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctataagtgacaatcagcgtagggaatcgc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACatcagtgcggtatattaccctagacgcta

10

GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacagttactattaatcacgattccaacgg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTtgattcaacataaaaagccgggtcaattg
aactggcttt (SEQ ID NO:45)

15

La cepa original (DGCC3198) se expuso al bacteriófago donante (D4241), tal y como se ha descrito en el Ejemplo 1. El material aislado se obtuvo y se analizó adicionalmente, empleando los métodos descritos en el Ejemplo 1. Las reacciones de PCR y la determinación de la secuencia del ADN se realizaron tal y como se ha descrito en el Ejemplo 1, excepto que el cebador inverso empleado para la PCR era CR122-R3, con la siguiente secuencia: 5'-gctctaagattgaggtgaagg-3' (SEQ ID NO:46).

20

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 difieren de la cepa original en una unidad adicional de repetición-espaciador aislada, en el extremo del líder del locus CRISPR1. Para las 3 nuevas secuencias de espaciadores descritas en la Tabla 5-1, una búsqueda de homología con secuencias disponibles en bases de datos públicas, mostraba las mejores puntuaciones de la homología, con las subsecuencias del bacteriófago de *S. thermophilus*, DT1 (número de orden de GenBank AF085222), lo que confirma que la nueva secuencia espaciadora se originó a partir del bacteriófago.

Tabla 5-1. Descripción de las cepas marcadas en el locus CRISPR1 de DGCC3198, empleando D4241 como bacteriófago donante			
Cepa marcada	Cepa original	Fago donante	Secuencia de ADN del donante insertada
DGCC3198 _{phi4241} ^{+S64}	DGCC3198	D4241	ACCAAGTAGCATTGAGCAAAGATAGATTG (SEQ ID NO:47)
DGCC3198 _{phi4241} ^{+S65}	DGCC3198	D4241	TAGATCTCATGAGTGGCGACAGTGAGCTT (SEQ ID NO:48)
DGCC3198 _{phi4241} ^{+S66}	DGCC3198	D4241	TACCATCTTGGGATAGGTACTGGTCATGCC (SEQ ID NO:49)

EJEMPLO 6

Marcación de DGCC3198 de *Streptococcus thermophilus* empleando el bacteriófago D4241 mediante la inserción de una sola unidad de repetición-espaciador dentro de CRISPR3

5 En este Ejemplo, la cepa DGCC3198 de *S. thermophilus* se marcó mediante una inserción "natural" dentro del locus CRISPR3 de una unidad adicional de repetición-espaciador, procediendo el espaciador del bacteriófago D4241. El locus CRISPR3 de DGCC3198 contiene 9 repeticiones (incluyendo la repetición terminal) y 8 espaciadores (número de orden de GenBank: CP000419). La secuencia del locus CRISPR3 de DGCC3198 es:

**taaattgtaataagtatagatagcttgagttattcaagactatcttttagtatttagtctgtatgaagtgatgggataatcatttgtagagagtaga
ttataaggattgatagagggaattaagtgcttgacatatgatta**agaaataatctaataatggtgacagtcacatctgtctaaacgttgatataaag**

**gattttaaggataataataataaaatggaattatttgaagctgaagctgctgagattaatagtgcgattacgaaatctggtgaaagatacctacga
gTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACggtgaaaagggtcactgtacgagtacttaTTTTAG
AGCTGTGTTGTTTCGAATGGTTCCAAAACtcaatgagtggtatccaagacgaaaacttaTTTTAGAGCTGTGT
TGTTCGAATGGTTCCAAAACccttgctggtgctccatacgccatataTTTTAGAGCTGTGTTGTTTCGAA
TGGTTCCAAAACtgttgggaaaccgagtagccatgattaaTTTTAGAGCTGTGTTGTTTCGAATGGTTCCA
AAAACacagagtacaatattgctctcattggagacacTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACctcat
attcgttagttgctttgtcataaaTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACagaactttatcaagataaaac
tactttaaTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAACatagtaataattcattgaaaataattgtTTTT
AGAGCTGTGTTGTTTCGAATGGTTCCAAAACtttggatcacaatttcggtgacatctctagaactcatctatcataaagg
agtctagattgaaatgtagaaggac (SEQ ID NO:50)**

10 La cepa original (DGCC3198) se expuso al bacteriófago donante (D4241), tal y como se ha descrito en el Ejemplo 1. El material aislado se obtuvo y se analizó adicionalmente, empleando los métodos descritos en el Ejemplo 1. Las reacciones de la PCR y la determinación de la secuencia del ADN se realizaron tal y como se ha descrito en el Ejemplo 1, excepto que se emplearon CR3Icad-FI (5'-ctgagattaatagtgcgattacg-3'; SEQ ID NO:51) y CR3trail-R2 (5'-gctggatattcgataacatgctc-3'; SEQ ID NO:52).

15 Una cepa marcada se creó a partir de DGCC3198 empleando D4241 como donante. Esta cepa marcada se describe en la Tabla 6-1. Esta cepa marcada difiere de la cepa parental en una sola unidad adicional de repetición-espaciador, en el extremo del líder del locus CRISPR3. Para la nueva secuencia del espaciador descrita en la Tabla 6, una búsqueda de homología con secuencias disponibles procedentes de bases de datos públicas, mostraba las mejores puntuaciones de homología con subsecuencias de los bacteriófagos DT1 y Sfi19 de *S. thermophilus*
20 (números de orden de GenBank AF085222 y AF115102), lo que confirma que la nueva secuencia espaciadora se origina a partir del bacteriófago.

Tabla 6-1. Descripción de la cepa marcada en el locus CRISPR3 procedente de DGCC3198 empleando D4241 como bacteriófago donante			
Cepa marcada	Cepa original	Fago donante	Secuencia de ADN del donante insertada
DGCC319 8 _{phi4241} ^{S67}	DGCC3198	D4241	5'- tgcaatttcattagttcttgacgccttt -3' (SEQ ID NO:53)

EJEMPLO 7

Método de la PCR para la detección específica de cepas marcadas

25 En este Ejemplo, se describen los métodos de la PCR para la detección específica de cepas marcadas. Cuando una cepa se marca de forma natural por la adición de una o varias secuencias de oligonucleótidos, constituidas por la adición de espaciadores en uno o varios de los loci CRISPR1, es necesario poder detectar las cepas marcadas. Los métodos se basan en la presencia de secuencias específicas para las cepas marcadas que se insertan dentro de una región identificada con precisión, del cromosoma de la cepa. Por ello, se diseña un método de amplificación con
30 PCR que es específico para la cepa marcada. Una cepa que carece de esta secuencia única de oligonucleótidos no da como resultado un ADN amplificado con PCR, mientras que la PCR que emplea el ADN de la cepa marcada, da como resultado la amplificación de un fragmento de ADN de una longitud definida.

35 Para configurar el método de detección específica con PCR, se diseñaron 2 cebadores. Uno de los cebadores es el cebador directo y no es específico de la cepa marcada, pero es específico de un locus CRISPR; Se denomina "cebador de CRISPR". En *S. thermophilus*, y dependiendo de la cepa marcada, el "cebador de CRISPR" es idéntico a una secuencia dentro del locus CRISPR1 ("cebador de CRISPR1") o dentro del locus CRISPR3 ("cebador de CRISPR3") o dentro del locus CRISPR2 ("cebador de CRISPR2"). Los cebadores de CRISPR se escogen entre

5 secuencias que se conservan entre cepas de las especies de interés. Para *S. thermophilus*, se recomiendan los siguientes cebadores: cebador de CRISPR1, 5'-tgctgagacaacctagtctctc-3' (yc70, Bolotin y col., [2005], *supra* [SEQ ID NO:15]); cebador de CRISPR3, 5'-ctgagattaatagtcgattacg-3' (CR3lead-F1; SEQ ID NO:51). El segundo cebador es el cebador inverso y es específico de la cepa marcada; se denomina "cebador de TAG". Los cebadores de TAG son complementarios a uno de los espaciadores de las unidades de repetición-espaciador añadidas en las cepas marcadas. Preferentemente, el cebador de TAG es complementario al espaciador de la unidad de repetición-espaciador añadida, que es el más distal de la secuencia líder del locus CRISPR. La Figura 4 ilustra la posición del cebador de CRISPR y del cebador de TAG.

10 La Tabla 7-1 proporciona ejemplos de cebadores de TAG para la detección de cepas marcadas enumeradas en el Ejemplo 1 a 6.

Tabla 7-1. Cebadores empleados para la detección de cepas marcadas descritas en los Ejemplos 1 a 6.		
Cepa marcada	Cebador de CRISPR	Cebador de TAG
DGCC7710 _{phi2972} ^{+S40}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-aagagtagctccctcaatagc-3' (SEQ ID NO:54)
DGCC7710 _{phi2972} ^{+S41}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-ttactaccactattcctagtg-3' (SEQ ID NO:55)
DGCC7710 _{phi2972} ^{+S43}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-atagaggccaactaacggtat-3' (SEQ ID NO:56)
DGCC7710 _{phi2972} ^{+S44}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-tatcaattgcggaatgtggct-3' (SEQ ID NO:57)
DGCC7710 _{phi2972} ^{+S43}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-cactgctgctcaataaacatt-3' (SEQ ID NO:58)
DGCC7710 _{phi2972} ^{+S46}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-agaacaagtttgaacaaaa-3' (SEQ ID NO:59)
DGCC7710 _{phi858} ^{+S42}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-aacctaataactttggctga-3' (SEQ ID NO:60)
DGCC7710 _{phi2972} ^{+S46+S47}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-agaacaagtttgaacaaaa-3' (SEQ ID NO:61)
DGCC7710 _{phi2972} ^{+S48+S49}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-gctactgcatcgtaaagactt-3' (SEQ ID NO:62)
DGCC7710 _{phi2972} ^{+S50+S51}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-gagttctatttctgaattatt-3' (SEQ ID NO:63)
DGCC7710 _{phi2972} ^{+S52+S54}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-tggtggagccctgtacatta-3' (SEQ ID NO:64)
DGCC7710 _{phi2972} ^{+S53+S54}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-aatgatacttgaccgaaacc-3' (SEQ ID NO:65)
DGCC7710 _{phi2972} ^{+S61+S62}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-tcaaacgtaagacgaatcg-3' (SEQ ID NO:66)
DGCC7710 _{phi2972} ^{+S55+S63+S41}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-aagagtagctccctcaatagc-3' (SEQ ID NO:54)
DGCC7710 _{phi2972} ^{+S56}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC7710 _{phi2972} ^{+S6} _{phi4724} ^{+S20} _{phi4733} ^{+S29}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-cacctcggaacctaataccaatt-3' (SEQ ID NO:67)
DGCC3198 _{phi4241} ^{+S64}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-caatctatcttctgctcaatgc-3' (SEQ ID NO:67)
DGCC3198 _{phi4241} ^{+S65}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-aagctcactgtcgccacttag-3' (SEQ ID NO:68)
DGCC3198 _{phi4241} ^{+S66}	5'-tgctgagacaacctagtctctc-3' (SEQ ID NO:15)	5'-ggcatgccagtacatcccaa-3' (SEQ ID NO:69)
DGCC3198 _{phi4241} ^{+S67}	5'-ctgagattaatagtcgattacg-3'	5'-tcaagaactaatggaaattgcag-3'

Tabla 7-1. Cebadores empleados para la detección de cepas marcadas descritas en los Ejemplos 1 a 6.

Cepa marcada	Cebador de CRISPR	Cebador de TAG
	(SEQ ID NO:51)	(SEQ ID NO:70)

En estos experimentos, la muestra que contiene la cepa que se va a detectar, se trató empleando cualquier método adecuado conocido en la técnica, para separar las bacterias del resto de la muestra. A modo de ejemplo en el caso de yogur o de muestras lácteas frescas que contienen *S. thermophilus*, la muestra se trató tal y como describen Lick y col. (Lick y col., *Milchwissenschaft* 50:183-186 [1996]).

Dependiendo de la cantidad de bacterias contenidas dentro del material resultante, las células de *S. thermophilus* se amplificaron mediante el cultivo en medio M17-glucosa, durante 18 horas a 42°C. El ADN se extrajo a continuación de las bacterias y se sometió a la PCR específica. Los cebadores de la PCR (cebador de CRISPR y cebador de TAG) se escogieron de forma adecuada, tal y como se describe en la Tabla 7-1. La mezcla de la reacción PCR (volumen final de 25 µL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno de los cuatro dNTP 2 mM, 10 a 100 ng de ADN, cada cebador 0,2 µM, 1,25 U de polimerasa *Taq* (Promega). Las condiciones de la ciclación con PCR eran: desnaturalización previa a 98°C durante 5 min, seguida de 33 ciclos alternando desnaturalización a 94°C durante 30 s e hibridación a 56°C durante 30 s, y elongación final a 72°C durante 1 min; seguido de una etapa de elongación a 72°C durante 4 min. En una PCR testigo, el ADN extraído se sometió a una segunda PCR, dirigida a genes de ARN 16S, empleando los siguientes cebadores universales: BSF8-20, 5'-agagtttgatcctggctcag-3' (SEQ ID NO:71) y BSR1541-20, 5'-aaggagggtgccagccgca-3' (SEQ ID NO:72; véase, Wilmotte y col., *FEBS Lett.*, 317:96-100 [1993]).

La mezcla de la reacción PCR (volumen final de 25 µL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno de los cuatro dNTP 2 mM, 10 a 100 ng de ADN, cada cebador 0,2 µM, 1,25 U de polimerasa *Taq* (Promega). Las condiciones de la ciclación de la PCR eran las siguientes: desnaturalización previa a 95°C durante 7 min, seguida 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 elongación final a 72°C durante 5 min. Los productos de la amplificación con PCR se analizaron a continuación empleando electroforesis en gel de agarosa (1%, p/v) y se registró el tamaño de los fragmentos de ADN amplificados. Para cada PCR específica, se prepararon testigos empleando ADN extraído de la cepa original (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 para la cepa marcada sometida a ensayo, ya que los productos de la PCR de tamaño adecuado se obtuvieron siempre en PCRs testigo y en PCRs específicas, dando como resultado el fragmento de ADN amplificado solo cuando se empleaba ADN específico de la cepa marcada.

Tabla 7-2. Tamaño de los productos de la PCR obtenidos mediante PCR específica y PCR testigo sobre ADN de la cepa marcadas, ADN de la cepa original y ADN de DGCC_{phi2972}^{S6}

PCR ^c específica	ADN específico de la cepa marcada		ADN de DGCC _{phi2972} ^{+S6}		ADN de la cepa original	
	PCR ^a específica	PCR ^a testigo	PCR ^a específica	PCR ^a testigo	PCR ^a específica	PCR ^a testigo
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 _{phi2972} ^{+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+S511}	310	1530	0	1530	0	1530
DGCC7710 _{phi2972} ^{+S52+S4}	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

Tabla 7-2. Tamaño de los productos de la PCR obtenidos mediante PCR específica y PCR testigo sobre ADN de la cepa marcadas, ADN de la cepa original y ADN de DGCC_{phi2972}^{S6}

PCR ^c específica	ADN específico de la cepa marcada		ADN de DGCC _{phi2972} ^{+S6}		ADN de la cepa original	
	PCR ^a específica	PCR ^a testigo	PCR ^a específica	PCR ^a testigo	PCR ^a específica	PCR ^a testigo
DGCC7710 _{phi2972} ^{+S55+S63+S41}	370	1530	0	1530	0	1530
DGCC7710 _{phi2972} ^{+S6}	240	1530	240	1530	0	1530
DGCC7710 _{phi2972} ^{+S6} _{phi2724} ^{+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 un fragmento de PCR; ^c empleando el cebador específico de PCR, tal y como se ha mencionado en la Tabla 7-1.

EJEMPLO 8

Método para la identificación de cepas marcadas

5 En este Ejemplo, se describe un método para detectar la presencia de cepas marcadas en una muestra y para identificar su naturaleza. Esto se realiza mediante la amplificación con PCR de uno o varios loci CRISPR y la determinación de la secuencia parcial. Este método se emplea en varias realizaciones, mientras que el método descrito en el Ejemplo 7 es útil para la detección de una cepa marcada, pero puede no ser suficiente para su identificación formal. Además, en muchos casos, la naturaleza de la cepa marcada contenida dentro de la muestra, no es conocida. Por tanto, el método de PCR específica no se puede utilizar para su detección. Eventualmente, la cepa marcada se puede detectar e identificar a través del análisis del locus CRISPR modificado.

10 Las células de *S. thermophilus* contenidas en una muestra, se extrajeron de la muestra empleando un método adecuado, conocido en la técnica, y se extendieron en placas con diluciones adecuadas sobre agar M17-glucosa, a continuación se incubaron durante 24 horas a 42°C, para obtener colonias aisladas. Una o varias colonias aisladas se recogieron después y se dejaron crecer en medio líquido M17-glucosa, durante 18 horas a 42°C. De cada cultivo celular se eligieron células y se extrajo su ADN. Para cada material aislado, la región del cromosoma correspondiente al locus CRISPR1, se amplificó mediante PCR empleando los siguientes cebadores: cebador directo yc70 (5'-tgctgagacaacctgtctc-3' [SEQ ID NO:15], Bolotin y col., [2005]; *supra*); y cebador inverso SPIDR-dws (5'-taacagagcctccctatcc-3' [SEQ ID NO:73]). La mezcla de la reacción PCR (volumen final de 25 µL) contenía: tampón exento de Mg 1x (Promega), MgCl₂ 2,5 mM, cada uno de los cuatro dNTP 2 mM, 10 a 100 ng de ADN, cada cebador 0,2 µM, 1,25 U de polimerasa *Taq* (Promega). Las condiciones de ciclación de la PCR eran las siguientes: desnaturalización previa a 98°C durante 5 min, seguida de 33 ciclos alternando desnaturalización a 94°C durante 30 s e hibridación a 56°C durante 30 s y elongación a 72°C durante 1 min; seguido de una etapa final de elongación a 72°C durante 4 min. En algunos casos, la región correspondiente al locus CRISPR3, se amplificó empleando los siguientes cebadores: CR3lead-F1, 5'-ctgagattaatgtgcgattacg-3' (SEQ ID NO:51) y CR3trail-R2, 5'-gctggatattcgataacatgtc-3' (SEQ ID NO:52). La secuencia de ácido nucleico de cada fragmento de PCR se determinó a continuación, empleando el "cebador directo" y la metodología clásica de secuenciación, tal y como se conoce en la técnica. A continuación, se comparó cada secuencia con secuencias de loci CRISPR, disponibles en bases de datos.

30 En un experimento, el locus CRISPR1 procedente de un material aislado a partir de un producto lácteo fermentado, se sometió a PCR y el amplicón resultante se sometió a secuenciación. La secuencia se comparó con las secuencias disponibles en las bases de datos. La Figura 5 proporciona los resultados de la comparación. Al parecer, la secuencia tenía una identidad del 100% con la de la cepa DGCC7710, con una secuencia adicional de 66 nucleótidos. Esta secuencia de 66 nucleótidos tiene 36 nucleótidos en su extremo 5' que son idénticos a los de las repeticiones en el locus CRISPR1 de *S. thermophilus* y la secuencia de 30 nucleótidos restante, es idéntica a una subsecuencia del bacteriófago D2972. Además, esta secuencia adicional de 66 nucleótidos, estaba localizada inmediatamente aguas abajo de la secuencia líder de CRISPR1. Por consiguiente, el locus CRISPR1 del material aislado contiene el locus CR1SPR1 de DGCC7710, con una unidad adicional de repetición-espaciador, tal y como se describe en la Figura 3. Además, la secuencia restante de 30 nucleótidos también era idéntica a la secuencia del espaciador adicional de la cepa marcada DGCC_{phi2972}^{S41}. Esto indicó definitivamente que el material aislado obtenido a partir del producto lácteo fermentado era la cepa marcada DGCC_{phi2972}^{S41}.

REIVINDICACIONES

1. Un método para marcar una bacteria que comprende las etapas de:
 - (a) exponer una bacteria original que comprende un locus CRISPR a un bacteriófago para introducir una unidad adicional de repetición espaciador dentro del locus CRISPR, para producir una bacteria marcada, en donde la unidad adicional de repetición-espaciador proporciona el marcador;
 - (b) seleccionar un mutante insensible a los bacteriófagos;
 - (c) comparar un locus CRISPR o una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos
 - (d) seleccionar una bacteria marcada que comprende dicha unidad adicional de repetición-espaciador en el locus CRISPR que no está presente en la bacteria original.
2. El método de acuerdo con la reivindicación 1, en donde se compara el extremo 5' y/o el extremo 3' del locus CRISPR de la bacteria original.
3. El método de acuerdo con la reivindicación 1 o la reivindicación 2, en el que se compara al menos la primera repetición de CRISPR o el primer espaciador de CRISPR en el extremo 5' del locus CRISPR.
4. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde se compara al menos la última repetición de CRISPR o el último espaciador de CRISPR en el extremo 3' del locus CRISPR.
5. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el método comprende la etapa de seleccionar una bacteria marcada que comprende una unidad adicional de repetición-espaciador en el extremo 5' y/o en el extremo 3' del locus CRISPR que no está presente en la bacteria original.
6. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde dicho método comprende exponer la bacteria original a dos o varios bacteriófagos, de forma simultánea o sucesiva.
7. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el locus CRISPR o una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos se comparan mediante la amplificación de dicho locus CRISPR o de una porción del mismo procedente de dicha bacteria original y/o de dicho mutante insensible a los bacteriófagos.
8. El método de acuerdo con la reivindicación 7, en donde la amplificación se realiza empleando PCR.
9. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el locus CRISPR o una porción del mismo procedente de la bacteria original y el mutante insensible a los bacteriófagos se comparan mediante la secuenciación de dicho locus CRISPR o de una porción del mismo procedente de dicha bacteria original y/o de dicho mutante insensible a los bacteriófagos.
10. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el locus CRISPR o a una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos se comparan amplificando y secuenciando después dicho locus CRISPR o a una porción del mismo procedente de dicha bacteria original y/o de dicho mutante insensible a los bacteriófagos.
11. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la unidad adicional de repetición-espaciador tiene al menos 44 nucleótidos de longitud.
12. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde en la etapa (d) se selecciona una bacteria marcada que comprende dos, tres o más unidades adicionales de repetición-espaciador.
13. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene una identidad de al menos 95%, preferentemente una identidad del 100% con una repetición de CRISPR en el locus CRISPR de la bacteria original.
14. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la unidad adicional de repetición-espaciador comprende al menos una secuencia de nucleótidos que tiene una identidad de al menos 95%, preferentemente, una identidad del 100% con una secuencia de nucleótidos en el genoma del bacteriófago utilizado para la selección de la bacteria marcada.
15. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la unidad adicional de repetición-espaciador comprende una primera secuencia de nucleótidos que tiene al menos una identidad del 95%, preferentemente, una identidad del 100% con una repetición de CRISPR en el locus CRISPR de la bacteria original y una segunda secuencia de nucleótidos que tiene al menos una identidad del 95%, preferentemente, una identidad

del 100% con una secuencia de nucleótidos en el genoma del bacteriófago empleado para la selección de la bacteria marcada.

16. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la bacteria original es adecuada para el uso como cultivo iniciador, cultivo probiótico o suplemento dietético.

5 **17.** El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde la bacteria original se selecciona entre el grupo de géneros consistente 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*
10 y *Oenococcus*.

18. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en donde el bacteriófago se selecciona entre el grupo de familias de virus consistente en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae y Tectiviridae.

19. Un método para generar una variante de CRISPR que comprende las etapas de:

15 (a) exponer una bacteria original que comprende un locus CRISPR a un bacteriófago para introducir una unidad adicional de repetición espaciador dentro del locus CRISPR, para producir una bacteria marcada, en donde la unidad adicional de repetición-espaciador proporciona el marcador;

(b) seleccionar una bacteria resistente a los bacteriófagos;

20 (c) comparar el locus CRISPR o una porción del mismo procedente de la bacteria original y del mutante insensible a los bacteriófagos;

(d) seleccionar una bacteria marcada que comprende una unidad adicional de repetición-espaciador en el locus CRISPR que no está presente en la bacteria original; y

(e) aislar y/o clonar y/o secuenciar la unidad adicional de repetición-espaciador.

FIGURA 1

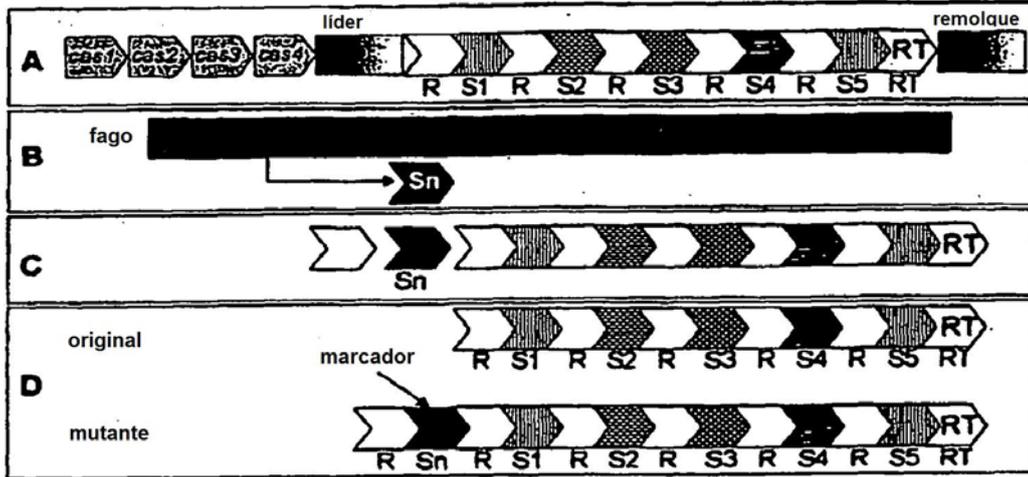


FIGURA 2

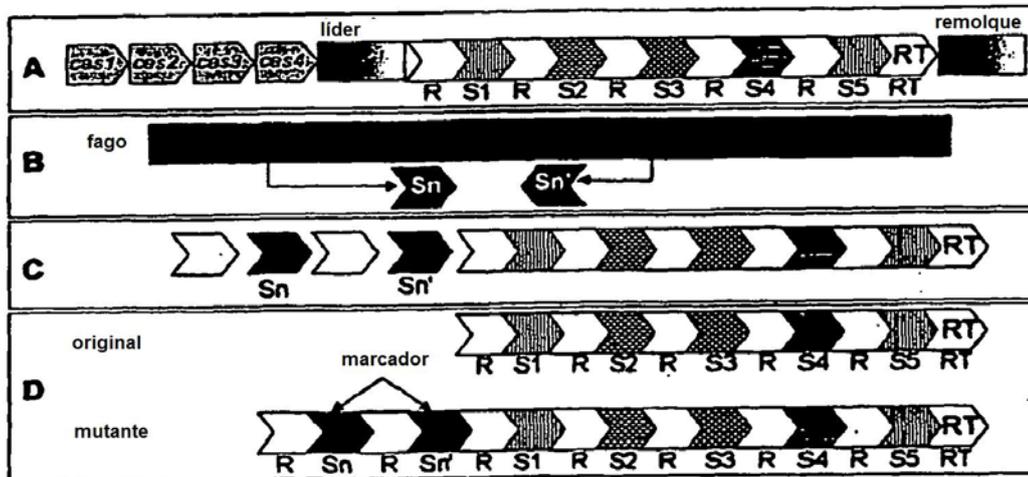


FIGURA 3

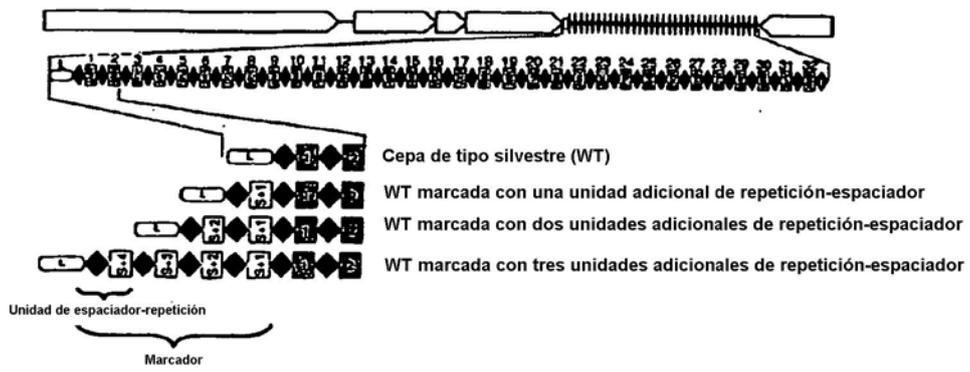


FIGURA 4

