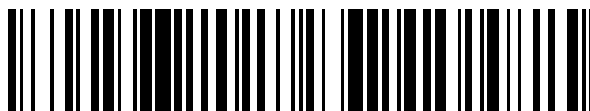


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 541 693**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.02.2008** **E 08726285 (3)**

97 Fecha y número de publicación de la concesión europea: **25.03.2015** **EP 2126130**

54 Título: **Cultivos con resistencia mejorada a fagos**

30 Prioridad:

02.03.2007 US 904721 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
23.07.2015

73 Titular/es:

**DUPONT NUTRITION BIOSCIENCES APS
(100.0%)**

**Langebrogade 1, Postboks 17
1001 Copenhagen K, DK**

72 Inventor/es:

**BARRANGOU, RODOLPHE;
FREMAUX, CHRISTOPHE;
HORVATH, PHILIPPE;
ROMERO, DENNIS y
BOYAVAL, PATRICK**

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 541 693 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

Cultivos con resistencia mejorada a fagos

5 Campo de la Invención

La presente invención proporciona métodos y composiciones que encuentran uso en el desarrollo y el uso de combinaciones de cepas y rotaciones de cultivo de inicio.

10 Antecedentes de la Invención

Cultivos, y cultivos de inicio, en particular, se utilizan ampliamente en la industria alimentaria en la fabricación de productos fermentados incluidos los productos lácteos (por ejemplo, yogur, mantequilla y queso), productos cárnicos, productos de panadería, vino y productos vegetales. La preparación de los cultivos es una labor intensiva, que ocupa mucho espacio y equipo, y existe un riesgo considerable de contaminación con bacterias de putrefacción y/o fagos durante las etapas de propagación. El fracaso de los cultivos bacterianos, debido a la infección por un bacteriófago (fago) y a la multiplicación es un problema importante con el uso industrial de cultivos bacterianos. Hay muchos tipos diferentes de fagos y nuevas cepas siguen apareciendo. Además, existe la necesidad de métodos y composiciones para el seguimiento de las bacterias utilizadas en dichos cultivos. De hecho, a pesar de los avances en el desarrollo de cultivos, existe una continua necesidad de mejorar los cultivos para su uso en la industria.

Breve Descripción de la Invención

La presente invención proporciona métodos y composiciones que encuentran uso en el desarrollo y el uso de combinaciones de cepas y rotaciones de cultivo de inicio.

De acuerdo con un aspecto, la presente invención proporciona un método para la generación de un cultivo de inicio que comprende al menos dos variantes de una cepa resistente a bacteriófagos, que comprende las etapas de: (a) la exposición de una cepa bacteriana precursora, que comprende al menos una porción de una ubicación CRISPR, a un bacteriófago para producir una mezcla de bacterias que comprende una variante de la cepa resistente a bacteriófagos que comprende una ubicación CRISPR modificada que comprende al menos un espaciador adicional en dicha ubicación CRISPR modificada; (b) la exposición independiente de la misma cepa bacteriana precursora que en el paso (a) que comprende al menos una porción de una ubicación CRISPR al mismo bacteriófago que en el paso (a) para producir una mezcla de bacterias que comprende otra variante de la cepa resistente a bacteriófagos que comprende una ubicación CRISPR modificada que comprende al menos un espaciador adicional en dicha ubicación CRISPR modificada; (c) la selección de dichas variantes de las cepas resistentes a bacteriófagos de dichas mezclas de bacterias; (d) la selección de dichas variantes de las cepas resistentes a bacteriófagos que comprenden un espaciador adicional en dicha ubicación CRISPR modificada de dichas cepas resistentes a bacteriófagos seleccionadas en el paso (c); y (e) el aislamiento de dichas variantes de las cepas resistentes a bacteriófagos, donde dichas cepas comprenden un espaciador adicional en dicha ubicación CRISPR modificada; donde la secuencia de al menos un espaciador adicional en la variante de la cepa resistente a bacteriófagos es diferente de la secuencia de al menos un espaciador adicional en la otra cepa resistente a bacteriófagos.

De acuerdo con un segundo aspecto, la invención proporciona un cultivo de inicio obtenido mediante el método de la presente invención.

De acuerdo con un tercer aspecto, la invención proporciona un método de fermentación que comprende la adición del cultivo de inicio de la invención a un medio de fermentación, en condiciones tales que tenga lugar la fermentación de los componentes de dicho medio de fermentación.

En algunas modalidades preferidas, el bacteriófago se selecciona de un grupo de familias de virus que consiste en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, y Tectiviridae. En algunas modalidades preferidas adicionales, el bacteriófago es un bacteriófago natural, mientras que en otras modalidades preferidas, el bacteriófago es un bacteriófago mutado obtenido mediante presión selectiva utilizando una cepa bacteriana resistente a bacteriófagos. Aún en otras modalidades, la cepa resistente a bacteriófagos es una mutante insensible a bacteriófagos. Todavía en otras modalidades, la cepa bacteriana precursora es una mutante insensible a bacteriófagos. En algunas otras modalidades, las porciones terminales 5' y/o 3' de la ubicación CRISPR de la cepa bacteriana precursora se comparan con la ubicación CRISPR modificada de la variante de la cepa resistente a bacteriófagos. En algunas modalidades adicionales más, las porciones terminales 5' y/o 3' del primer espaciador CRISPR de la ubicación CRISPR de la cepa bacteriana precursora se comparan con la ubicación CRISPR modificada de la variante de la cepa resistente a bacteriófagos. En algunas modalidades adicionales, por lo menos una parte de la ubicación CRISPR de la cepa bacteriana precursora y, por lo menos una parte de la ubicación CRISPR modificada de la variante de la cepa resistente a bacteriófagos, se comparan amplificando por lo menos una parte de la ubicación CRISPR y al menos, una parte de la ubicación CRISPR modificada, para producir una secuencia amplificada de la ubicación CRISPR y una amplificación de la secuencia de la ubicación CRISPR modificada. En aún más modalidades, la amplificación se realiza mediante la reacción en

cadena de la polimerasa. En algunas modalidades preferidas, por lo menos una parte de la ubicación CRISPR de la cepa bacteriana precursora y por lo menos una parte de la ubicación CRISPR modificada de la variante de la cepa resistente a bacteriófagos, se comparan secuenciando por lo menos una parte de la ubicación CRISPR y al menos una parte de la ubicación CRISPR modificada. En algunas modalidades preferidas en particular, los métodos además comprenden la etapa de secuenciar la secuencia amplificada de la ubicación CRISPR y la secuencia amplificada de la ubicación CRISPR modificada. En algunas modalidades adicionales, el espaciador adicional en la ubicación CRISPR modificada, forma una unidad de espaciador repetido adicional. En algunas modalidades preferidas, la unidad de espaciador repetido adicional comprende al menos alrededor de 44 nucleótidos. En algunas modalidades alternativas preferidas, la unidad de espaciador repetido adicional está compuesta por alrededor de entre 44 y alrededor de 119 nucleótidos. Sin embargo, no es la intención que la presente invención se limite a estos intervalos de tamaño específicos, ya que se ha encontrado uso para otros tamaños en la presente invención, tal como se describe en este documento. En algunas modalidades, la unidad de espaciador repetido adicional, comprende al menos una secuencia de nucleótidos que tiene por lo menos alrededor del 95% de identidad de una repetición CRISPR en la ubicación CRISPR de la cepa bacteriana precursora. En algunas otras modalidades, la unidad espaciador repetido adicional, comprende al menos una secuencia de nucleótidos que tiene por lo menos alrededor del 95% de identidad de una secuencia de nucleótidos en el genoma de al menos un bacteriófago. En algunas modalidades preferidas en particular, la cepa bacteriana precursora, es una cepa industrialmente útil. En algunas modalidades adicionales preferidas, la cepa bacteriana precursora es susceptible a la infección por lo menos por un bacteriófago. En algunas modalidades preferidas adicionales, la cepa bacteriana precursora, comprende un cultivo seleccionado de cultivos de inicio, cultivos probióticos y cultivos de suplemento dietético. En algunas modalidades preferidas, la cepa bacteriana precursora comprende una cepa obtenida de un cultivo. En algunas modalidades preferidas en particular, el cultivo es un cultivo de inicio, un cultivo probiótico, y/o un cultivo de suplemento dietético. Más aún en modalidades adicionales, la cepa bacteriana precursora se selecciona de *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Campylobacter*, *Klebsiella*, *Frankia*, *Bartonella*, *Rickettsia*, *Shewanella*, *Serratia*, *Enterobacter*, *Proteus*, *Providencia*, *Brochothrix*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc*, y *Oenococcus*.

Se describe al menos una variante de la cepa resistente a bacteriófagos obtenida mediante los métodos establecidos en el presente. Se describen variantes de las cepas resistentes a bacteriófagos, en las que la variante de la cepa resistente a bacteriófagos es una cepa industrialmente útil. En algunas modalidades preferidas, la variante de la cepa resistente a bacteriófagos comprende una cepa industrialmente útil que es al menos un componente de un cultivo de inicio, cultivos probióticos, cultivos de suplemento dietético, y/u otros cultivos útiles.

La presente invención también proporciona composiciones que comprenden una variante de la cepa resistente a bacteriófagos producida utilizando los métodos establecidos en el presente. La presente invención proporciona composiciones que comprenden al menos dos variantes de las cepas resistentes a bacteriófagos producidas utilizando los métodos establecidos en el presente. Se proporcionan los alimentos y/o piensos que comprenden al menos una de estas composiciones. Se proporcionan métodos para la preparación de alimentos y/o piensos que comprenden la adición de al menos una de estas composiciones para el alimento o el pienso. La presente invención también proporciona cultivos de inicio, cultivos probióticos, cultivos de suplemento dietético y otros cultivos útiles que componen por lo menos una de estas composiciones. La presente invención también proporciona métodos de fermentación que comprenden la adición de al menos una de estas composiciones a un cultivo de inicio. La presente invención proporciona métodos de fermentación que comprenden la adición de al menos una de estas composiciones a un medio de fermentación, en condiciones tales que la fermentación de los componentes del medio de fermentación ocurra. En algunas modalidades, la fermentación no se ve afectada por la presencia de bacteriófagos. En algunas modalidades, el medio de fermentación es un producto alimenticio. En algunas modalidades preferidas, el producto alimenticio es un producto lácteo. En algunas modalidades preferidas en particular, el producto lácteo es la leche. En algunas modalidades adicionales, al menos dos composiciones diferentes compuestas de dos o más variantes de las cepas resistentes a bacteriófagos son secuencialmente expuestas al medio de fermentación.

Se proporcionan métodos para reducir la población de bacteriófagos perjudiciales en un medio de fermentación que comprende la exposición de un medio de fermentación a al menos una variante de la cepa resistente a bacteriófagos, producida utilizando los métodos establecidos en el presente, en condiciones tales que la población de bacteriófagos se reduce.

Se proporcionan métodos para la generación de al menos una variante de la cepa resistente a bacteriófagos, que comprenden las etapas de: (a) la exposición de una cepa bacteriana precursora que comprende por lo menos una porción de una ubicación CRISPR a por lo menos una secuencia de ácido nucleico para producir una mezcla de bacterias que comprenda al menos una variante de la cepa resistente a bacteriófagos que comprenda una ubicación CRISPR modificada; (b) la selección de una variante de la cepa resistente a bacteriófagos de una mezcla de bacterias; (c) la comparación de la ubicación CRISPR o una parte de la misma de la cepa bacteriana precursora y la ubicación CRISPR modificada de la variante de la cepa resistente a bacteriófagos para identificar variantes de las cepas resistentes a bacteriófagos que comprendan al menos otro fragmento de ácido nucleico en la ubicación

CRISPR modificada, que no se encuentre en la ubicación CRISPR de la cepa bacteriana precursora; (d) la selección de la variante de cepas resistentes a bacteriófagos que comprendan un fragmento de ácido nucleico adicional en la ubicación CRISPR modificada; (e) el análisis de por lo menos un fragmento de ácido nucleico adicional en la ubicación CRISPR modificada, para identificar al menos una variante de la cepa resistente a bacteriófagos; y (f) el aislamiento de al menos una variante de la cepa resistente a bacteriófagos.

Se proporcionan métodos para la generación de fagos mutantes de escape a CRISPR, que comprenden: (a) la obtención de: al menos uno de los fagos precursores y una cepa bacteriana resistente a fagos que comprenda al menos una ubicación CRISPR, donde la ubicación CRISPR comprenda una secuencia de ácidos nucleicos que sea por lo menos alrededor del 95% idéntica a al menos una secuencia proto-espaciadora en el genoma de por lo menos uno de los fagos precursores; (b) la exposición de al menos uno de los fagos precursores a una cepa bacteriana resistente a fagos, en condiciones tales que por lo menos una variante de fagos se produzca; y (c) la selección de al menos una variante de fagos, en la que por lo menos una variante de fagos exhiba la capacidad de infectar las cepas de bacterias resistentes a fagos y sea un fago mutante de escape a CRISPR. La cepa de bacterias resistentes a fagos, puede ser una variante de la cepa resistente a bacteriófagos obtenida mediante los métodos establecidos en el presente. Los métodos además pueden comprender la etapa de la comparación de al menos una parte de por lo menos una secuencia proto-espaciadora y una porción CRISPR posicionada cerca de al menos una secuencia proto-espaciadora en la variante del fago con la al menos una secuencia proto-espaciadora y una porción CRISPR del fago precursor. Los métodos pueden además comprender la etapa de la selección de la variante de fagos que infectan la cepa de bacterias resistentes a fagos, en la que la variante de fagos comprende los fagos mutantes de escape a CRISPR, y donde los fagos de escape a CRISPR comprenden al menos una mutación en la al menos una secuencia proto-espaciadora y/o en la porción CRISPR de los fagos mutantes de escape a CRISPR. Los métodos pueden ser iterativamente repetidos una o más veces mediante los fagos mutantes de escape a CRISPR y diferentes cepas de bacterias resistentes a fagos CRISPR que incluyen al menos una ubicación CRISPR, en la que la ubicación CRISPR comprende una secuencia de ácidos nucleicos que sea por lo menos alrededor del 95% idéntica a al menos una secuencia proto-espaciadora en el genoma de fagos mutantes de escape a CRISPR. Todavía en modalidades adicionales, por lo menos un bacteriófago se selecciona del grupo de familias de virus que consiste en: *Corticoviridae*, *Cystoviridae*, *Inoviridae*, *Leviviridae*, *Microviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, y *Tectiviridae*. En algunas modalidades preferidas, la cepa de bacterias resistentes a fagos se selecciona de *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Enterococcus*, *Clostridium*, *Campylobacter*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Klebsiella*, *Frankia*, *Bartonella*, *Rickettsia*, *Shewanella*, *Serratia*, *Enterobacter*, *Proteus*, *Providencia*, *Brochothrix*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc*, *Streptococcus*, y *Oenococcus*.

Se proporcionan fagos mutantes de escape a CRISPR obtenidos mediante los métodos establecidos en el presente. Los fagos mutantes de escape a CRISPR pueden comprender dos o más mutaciones presentes en al menos dos secuencias proto-espaciadoras y/o en la porción CRISPR.

Se proporcionan fagos mutantes de escape a CRISPR, en los que el genoma de los fagos mutantes de escape a CRISPR es genéticamente modificado para incluir mutaciones en al menos un proto-espaciador y/o una porción CRISPR. Por lo menos una porción CRISPR puede ser mutada en los fagos mutantes de escape a CRISPR, mientras que en algunas modalidades alternativas, al menos una porción CRISPR, se suprime en los fagos mutantes de escape a CRISPR. También se proporcionan composiciones que comprenden al menos un fago mutante de escape a CRISPR.

La presente invención también proporciona métodos para el control de las poblaciones de bacterias en un producto que comprende la exposición de composiciones que incluyen al menos un fago mutante de escape a CRISPR en un medio de fermentación, donde el medio de fermentación contiene al menos una población de bacterias indeseables, en condiciones tales que la población de estas bacterias no deseadas se reduce, y el medio de fermentación se utiliza para generar el producto. En algunas modalidades, el producto se selecciona de los alimentos, los aditivos, cosméticos, productos para el cuidado personal, productos sanitarios, productos veterinarios y suplementos alimenticios. En algunas otras modalidades, los métodos se repiten por lo menos una vez y las diferentes composiciones y/o composiciones que comprenden diferentes fagos mutantes de escape a CRISPR, se utilizan en rotación.

En algunas modalidades, la presente invención proporciona métodos y composiciones para el uso de uno o más genes o proteínas *cas* para la modulación de la resistencia en una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo. En algunas modalidades adicionales, la presente invención proporciona composiciones y métodos para el uso de una secuencia recombinante de ácidos nucleicos recombinantes que comprende al menos un gen *cas* y por lo menos dos repeticiones CRISPR junto con al menos un espaciador de CRISPR, donde al menos un espaciador de CRISPR es heterólogo a por lo menos un gen *cas* y/o por lo menos dos repeticiones CRISPR para modular la resistencia contra un ácido nucleico objetivo o de un producto de la transcripción del mismo. Todavía en modalidades adicionales, la presente invención proporciona al menos una secuencia de ácidos nucleicos que comprende al menos un gen *cas*.

Todavía en modalidades adicionales, la presente invención proporciona al menos una secuencia de ácido nucleico que comprende al menos un gen *cas* y por lo menos dos repeticiones CRISPR. En algunas modalidades, la presente invención proporciona una secuencia de ácido nucleico que comprende al menos un gen *cas* y por lo menos un espaciador de CRISPR. Todavía en modalidades adicionales, la presente invención proporciona una secuencia de

5 ácido nucleico que comprende al menos un gen *cas*, por lo menos un espaciador de CRISPR y al menos dos repeticiones CRISPR. En otras modalidades, la presente invención proporciona una secuencia recombinante de ácido nucleico que comprende al menos un gen *cas* y por lo menos dos repeticiones CRISPR junto con al menos un espaciador de CRISPR, en el que el espaciador de CRISPR es heterólogo para al menos un gen *cas* y/o por lo menos dos repeticiones CRISPR.

10 La presente invención también proporciona constructos que comprenden una o varias de las secuencias de ácidos nucleicos que se describen en este documento. Todavía en modalidades adicionales, la presente invención proporciona vectores que comprenden una o varias de las secuencias de ácidos nucleicos o una o varias de las construcciones descritas en este documento. Todavía en modalidades adicionales, la presente invención

15 proporciona células que incluyen la secuencia de ácido nucleico o el constructo o el vector descritos en el presente.

La presente invención también proporciona métodos para modular (por ejemplo, conferir o aumentar) la resistencia de una célula en contra de un ácido nucleico objetivo o un producto de la transcripción del mismo que comprende las etapas de: (i) la identificación de una secuencia (por ejemplo, una secuencia conservada) en un organismo (en

20 algunas modalidades, se trata de una secuencia que es esencial para la función o la supervivencia del organismo); (ii) la preparación de un espaciador de CRISPR que es homóloga a la secuencia identificada; (iii) la preparación de un ácido nucleico (por ejemplo, un ácido nucleico recombinante) que comprenda al menos un gen *cas* y por lo menos dos repeticiones CRISPR junto con el espaciador de CRISPR.

25 La presente invención también proporciona métodos para la identificación de un espaciador de CRISPR o un pseudoespaciador de CRISPR para su uso en la modulación de la resistencia de una célula en contra de un ácido nucleico objetivo o un producto de la transcripción del mismo que comprende las etapas de: (i) la elaboración de una célula compuesta por al menos dos repeticiones CRISPR y al menos un gen *cas* o una proteína; (ii) la identificación de al menos un espaciador de CRISPR o pseudoespaciadores de CRISPR en un organismo que es

30 considerablemente resistente al ácido nucleico objetivo o al producto de la transcripción del mismo; (iii) la modificación de la secuencia del espaciador de CRISPR en la célula de tal modo que el espaciador de CRISPR tenga homología con el espaciador del organismo; y (iv) determinar si la célula modula la resistencia contra el ácido nucleico objetivo o producto de la transcripción del mismo, en donde la modulación de la resistencia de la célula contra el ácido nucleico objetivo o producto de la transcripción del mismo, es indicativo de que el espaciador de

35 CRISPR modula la resistencia de la célula.

Se proporcionan métodos para la identificación de un gen *cas* para su uso en la modulación de la resistencia de una célula en contra de un ácido nucleico objetivo o un producto de la transcripción del mismo, que comprende las etapas de: (i) la elaboración de una célula compuesta por al menos un espaciador de CRISPR y al menos dos

40 repeticiones CRISPR; (ii) la ingeniería de la célula de manera que incluya al menos un gen *cas*; y (iii) determinar si la célula modula la resistencia contra el ácido nucleico objetivo o producto de la transcripción del mismo, en donde la modulación de la resistencia de la célula en contra del ácido nucleico objetivo o producto de la transcripción del mismo es indicativo de que el gen *cas* puede ser utilizado para modular la resistencia de la célula.

45 Se proporcionan métodos para la identificación de una repetición de CRISPR para su uso en la modulación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo que comprende las etapas de: (i) la elaboración de una célula compuesta por al menos un espaciador de CRISPR y al menos un gen *cas*; (ii) la ingeniería de una célula de forma que contenga la repetición CRISPR; y (iii) determinar si la célula modula la resistencia contra el ácido nucleico objetivo o el producto de la transcripción del mismo, en el que la

50 modulación de la resistencia de la célula contra el ácido nucleico objetivo o producto de la transcripción del mismo es indicativo de que la repetición de CRISPR se puede utilizar para modular la resistencia.

Se proporcionan métodos para la identificación de una combinación funcional de un gen *cas* y una repetición CRISPR que comprende las etapas de: (a) determinar las secuencias del gen *cas* y la repetición CRISPR; (b) la

55 identificación de uno o más grupos de genes *cas* según lo determinado por el análisis de comparación de secuencias; (c) la identificación de uno o más grupos de repeticiones CRISPR; y (d) la combinación de aquellos genes *cas* y secuencias repetidas de CRISPR que caigan dentro del mismo grupo, donde la combinación del gen *cas* y las secuencias repetidas de CRISPR dentro del mismo grupo es indicativo de que la combinación es una combinación funcional.

60 Se proporcionan métodos para modular el lisotipo de una célula bacteriana que comprende uno o más genes o proteínas *cas* y dos o más repeticiones CRISPR que comprenden las etapas de: (i) la identificación de uno o varios pseudoespaciadores CRISPR en la secuencia genómica de un bacteriófago contra el que la resistencia debe ser modulada; y (ii) modificar la secuencia de uno o más espaciadores CRISPR de la célula bacteriana de tal manera

65 que el espaciador o espaciadores de CRISPR de la célula bacteriana tenga(n) homología con el pseudoespaciador o espaciadores de CRISPR del bacteriófago contra el que la resistencia debe ser modulada.

Se proporcionan métodos para modular (por ejemplo, conferir o aumentar) la resistencia de una célula bacteriana contra un bacteriófago que comprende las etapas de: (i) la identificación de una secuencia (por ejemplo, una secuencia conservada) en un bacteriófago (preferiblemente, una secuencia esencial para la función o la supervivencia del bacteriófago); (ii) la preparación de un espaciador de CRISPR que sea homólogo a la secuencia identificada; (iii) la preparación de un ácido nucleico que comprenda al menos un gen *cas* y por lo menos dos repeticiones CRISPR junto con el espaciador de CRISPR; y (iv) la introducción de un ácido nucleico en la célula bacteriana, para hacer a la célula bacteriana resistente al ácido nucleico objetivo o producto de la transcripción del mismo.

Se proporcionan métodos para modular (por ejemplo, conferir o aumentar) la resistencia de una célula bacteriana contra un ácido nucleico objetivo o producto de la transcripción, en un bacteriófago del mismo que comprende las etapas de: (i) la identificación de uno o varios pseudoespaciadores CRISPR en el genoma de un bacteriófago, que sea capaz de proporcionar resistencia a un ácido nucleico objetivo o a un producto de la transcripción del mismo; (ii) la preparación de un ácido nucleico recombinante, que comprenda al menos un gen *cas* y por lo menos dos repeticiones CRISPR junto con uno o más pseudoespaciadores de CRISPR identificados; y (iii) la introducción del ácido nucleico recombinante en la célula bacteriana, para hacer así que la célula bacteriana sea resistente al ácido nucleico objetivo o producto de la transcripción del mismo.

Se proporcionan métodos para modular la resistencia de una célula bacteriana que comprende uno o más genes o proteínas *cas* y dos o más repeticiones CRISPR, contra un ácido nucleico objetivo o un producto de la transcripción del mismo en un bacteriófago, que comprende las etapas de: (i) la identificación de uno o más pseudoespaciadores CRISPR en un bacteriófago que es capaz de proporcionar resistencia a un ácido nucleico objetivo o producto de la transcripción del mismo; (ii) la identificación de uno o más espaciadores CRISPR en una célula bacteriana, en la que la resistencia debe ser modulada; y (iii) la modificación de la secuencia del espaciador o espaciadores de CRISPR en la célula bacteriana en la que la resistencia debe ser modulada de tal manera que el espaciador o espaciadores de CRISPR tengan un mayor grado de homología a los pseudoespaciadores de CRISPR del bacteriófago contra los cuales la resistencia debe de ser modulada.

Se proporcionan métodos para la determinación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo, que comprende la identificación de una o más combinaciones funcionales de repeticiones CRISPR-*cas* y uno o más espaciadores CRISPR en la célula.

Se proporcionan células obtenidas u obtenibles mediante el método o los métodos proporcionados en el presente. En algunas modalidades, la presente invención proporciona espaciadores CRISPR o pseudoespaciadores CRISPR obtenidos u obtenibles por el método o los métodos descritos en el presente.

Se proporcionan genes *cas* obtenidos u obtenibles por el método o los métodos descritos en el presente. Se proporcionan repeticiones CRISPR obtenidas u obtenibles por el método o los métodos descritos en el presente. Se proporcionan combinaciones funcionales obtenidas por el método o los métodos descritos en el presente. Se proporciona una ubicación recombinante de CRISPR que comprende al menos un espaciador de CRISPR o un pseudoespaciador de CRISPR, y/o por lo menos un gen *cas* y/o por lo menos una repetición CRISPR y/o una combinación funcional.

Se proporcionan métodos para el uso de células, por lo menos un espaciador de CRISPR o pseudoespaciador de CRISPR, al menos un gen *cas*, por lo menos una repetición de CRISPR, o una combinación funcional del mismo, para modular la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

Se proporcionan cultivos celulares que incluyen por lo menos una célula, al menos un espaciador de CRISPR o pseudoespaciador de CRISPR, por lo menos un gen *cas*, por lo menos una repetición de CRISPR o una combinación funcional para la modulación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

Se proporcionan productos alimenticios y/o aditivos que comprenden cultivos proporcionados en el presente. Se proporcionan procedimientos para la preparación de productos alimenticios y/o aditivos que incluyen cultivos proporcionados en la presente. En aún modalidades adicionales, la presente invención proporciona productos alimenticios y/o alimentos obtenidos u obtenibles por los métodos proporcionados en la presente. En algunas modalidades preferidas, la presente invención proporciona métodos para el uso de los cultivos proporcionados en la presente para la elaboración de productos alimenticios y/o alimentos.

Además, la presente invención proporciona las secuencias de nucleótidos que comprenden o que consisten en las secuencias establecidas en cualquiera de SEQ ID NOS: 7-10 y SEQ ID NOS: 359-405, así como variantes, fragmentos, homólogos y sus derivados. La presente invención también proporciona los aminoácidos codificados por las secuencias de nucleótidos establecidas en la presente. Se proporcionan constructos y/o vectores que

comprenden una o más de las secuencias de nucleótidos del presente. Se proporcionan células hospedadoras que comprenden al menos uno de los constructos y/o secuencias de nucleótidos proporcionados en la presente.

En algunas modalidades, uno o más genes *cas* o proteínas, se utilizan en combinación con dos o más repeticiones CRISPR. En algunas otras modalidades, uno o más genes o proteínas *cas* y/o las dos o más repeticiones de CRISPR se derivan de la misma célula. En algunas modalidades adicionales, uno o más genes o proteínas *cas*, y las dos o más repeticiones de CRISPR se co-presentan naturalmente en la misma célula. Todavía en algunas modalidades adicionales, uno o más genes o proteínas *cas*, son utilizados en combinación con uno o más espaciadores de CRISPR.

En algunas modalidades, el espaciador o los espaciadores de CRISPR se derivan de un organismo diferente a la célula de la que uno o más genes o proteínas *cas* y/o las dos o más repeticiones CRISPR se derivan.

En algunas modalidades, el espaciador se obtiene a partir de una célula que es resistente a un ácido nucleico objetivo. En algunas modalidades, el espaciador de CRISPR es una secuencia de ácido nucleico sintética. En algunas otras modalidades, el espaciador o los espaciadores CRISPR tienen homología con el ácido nucleico objetivo. En algunas modalidades, el espaciador o los espaciadores CRISPR tienen el 100% de identidad con el ácido nucleico objetivo o por lo menos la longitud del núcleo espaciador de CRISPR.

En algunas modalidades, uno o más genes o proteínas *cas* se utilizan en combinación con al menos uno o más espaciadores CRISPR y por lo menos dos o más repeticiones CRISPR. El ácido nucleico objetivo o el producto de la transcripción del mismo, se obtienen a partir de ADN de un bacteriófago. En algunas modalidades alternativas, el ácido nucleico objetivo o el producto de la transcripción del mismo se derivan de un gen de resistencia a antibiótico/antimicrobiano. En algunas otras modalidades, el ácido nucleico objetivo o el producto de la transcripción del mismo se derivan de un ácido nucleico que codifica al menos un factor de virulencia. En algunas modalidades preferidas, el factor de virulencia incluye toxinas, internalinas, hemolisinas, y/u otros factores de virulencia.

En algunas modalidades de la presente invención, uno o más genes *cas* y las dos o más repeticiones CRISPR se derivan de la misma célula. En algunas modalidades alternativas, uno o más genes *cas* y las dos o más repeticiones de CRISPR se co-presentan naturalmente en la misma célula. Aún en más modalidades, los espaciadores de CRISPR se derivan de un organismo diferente a célula de la que uno o más genes *cas* y/o las dos o más repeticiones CRISPR se derivan. En algunas modalidades, la célula es una célula receptora o una célula hospedadora.

En algunas modalidades, uno o más genes *cas* o las proteínas y las dos o más repeticiones de CRISPR se co-presentan naturalmente en la misma célula.

En algunas modalidades, la modificación comprende la inserción de uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR en la célula. En algunas modalidades, la modificación comprende producir por ingeniería genética un espaciador de CRISPR de la célula. En algunas modalidades, el espaciador de la célula tiene 100% de homología al espaciador CRISPR o pseudoespaciador CRISPR del organismo. En algunas modalidades, la totalidad o parte del espaciador en la célula se modifica. En algunas modalidades, la modificación comprende la modificación de un espaciador recombinante. En algunas modalidades, la modificación se produce a través de mutación espontánea o mutagénesis. En algunas modalidades, por lo menos uno o más espaciadores CRISPR en la célula se eliminan. En algunas modalidades, por lo menos una o más repeticiones CRISPR en la célula, se eliminan. En algunas modalidades, uno o más genes *cas* se eliminan. En algunas modalidades, CRISPR y/o uno o más genes *cas* se eliminan. En algunas modalidades, una o más proteínas o genes *cas* y/o dos o más repeticiones CRISPR en la célula se eliminan. En algunas modalidades, las secuencias de nucleótidos del gen *cas* y la repetición CRISPR se derivan de la misma o de diferentes cepas. En algunas modalidades, las secuencias de nucleótidos del gen *cas* y la repetición CRISPR se derivan de la misma o de diferentes especies.

En algunas modalidades, las secuencias de nucleótidos del gen *cas* y la repetición CRISPR se derivan del mismo o de diferentes géneros. En algunas modalidades, las secuencias de nucleótidos del gen *cas* y la repetición CRISPR se derivan del mismo o de diferentes organismos.

En algunas modalidades de la presente invención, el ácido nucleico objetivo en el bacteriófago, es una secuencia de ácido nucleico altamente conservada. En algunas modalidades, el ácido nucleico objetivo en el bacteriófago codifica una proteína de la especificidad de la hospedadora. En algunas otras modalidades, el ácido nucleico objetivo en el bacteriófago, codifica una proteína que es esencial para la supervivencia, la reproducción o el crecimiento del bacteriófago. En algunas modalidades, el ácido nucleico objetivo en el bacteriófago codifica una helicasa, una primasa, una cabeza o una cola de proteínas estructurales, una proteína con un dominio conservado (por ejemplo, holina, lisina, etc) o por lo menos una secuencia conservada entre importantes genes de fagos.

El método para determinar la resistencia de una célula a un ácido nucleico objetivo o producto de la transcripción del mismo, puede comprender la etapa adicional de la comparación de la secuencia de uno o más espaciadores de CRISPR en la célula, con la secuencia de ácido nucleico objetivo. El método para determinar la resistencia de una

célula a un ácido nucleico objetivo o un producto de la transcripción del mismo, puede comprender la etapa adicional de la determinación del perfil de resistencia de la célula.

El cultivo de la invención es un cultivo de inicio. Se describe un cultivo probiótico.

También se proporcionan “bacterias etiquetadas” que son resistentes a fagos (es decir, “mutantes insensibles a bacteriófagos”, “BIMS”). Se proporcionan bacterias que comprenden una o varias secuencias procedentes de al menos un bacteriófago del genoma, que son integradas en la ubicación CRISPR de la bacteria. Esta secuencia derivada del fago proporciona una etiqueta, que es identificable por su ubicación y/o secuencia y/o secuencia adyacente.

En algunas modalidades alternativas, la presente invención proporciona la duplicación de secuencias (por ejemplo, duplicar repeticiones de CRISPR) que se originan a partir de una bacteria precursora y que también están integradas iterativamente, secuencialmente, simultáneamente o casi simultáneamente junto con la secuencia procedente del genoma del bacteriófago.

Además, la presente invención utiliza métodos que facilitan la integración de una o varias secuencias de bacteriófagos diferentes en la ubicación CRISPR de la cepa bacteriana. La integración de las secuencias diferentes del bacteriófago en la ubicación CRISPR de la cepa bacteriana es un acontecimiento aleatorio. La integración de las diferentes secuencias del bacteriófago en la ubicación CRISPR de la cepa bacteriana puede no ser un acontecimiento aleatorio. Por lo tanto, no es siempre la misma ubicación del genoma del bacteriófago la que se integra en la ubicación CRISPR de la bacteria. Sin embargo, una vez que se integra y se mantiene se convierte por lo tanto en una potente marca o etiqueta y/o seguimiento de la bacteria. En consecuencia, la secuencia o secuencias procedentes del genoma del bacteriófago, no solo son nuevas para la ubicación CRISPR de la bacteria precursora, sino que también son una etiqueta que es única para cada bacteria. En el presente se incluyen por tanto métodos de etiquetado (por ejemplo, marcado) y/o la identificación de bacterias.

En algunas modalidades, los métodos de la presente invención son “naturales” y no resultan en la producción de organismos modificados genéticamente. Se proporcionan métodos para el etiquetado de bacterias que comprenden las etapas de: (a) la exposición de una bacteria precursora a un bacteriófago; (b) la selección de una mutante insensible a bacteriófagos; (c) la comparación de una ubicación CRISPR o una parte de la misma de una bacteria precursora y el mutante insensible a bacteriófagos; y (d) la selección de una bacteria etiquetada que incluya un fragmento de ADN adicional en la ubicación CRISPR que no está presente en la bacteria precursora.

También se proporcionan bacterias etiquetadas obtenidas mediante los métodos de la presente invención. En algunas modalidades, la presente invención proporciona cultivos celulares que comprenden al menos una cepa bacteriana etiquetada. Aún en más modalidades, la presente invención proporciona alimentos y/o piensos que comprenden bacterias etiquetadas, incluyendo pero no limitado a cultivos de células que comprenden dichas bacterias etiquetadas. La presente invención también proporciona métodos para la preparación de alimentos y/o piensos que comprenden al menos una cepa bacteriana etiquetada. En algunas modalidades, los métodos incluyen la adición de al menos una cepa bacteriana etiquetada o cultivos celulares al alimento y/o pienso.

La presente invención también proporciona métodos para la generación de variantes de CRISPR que comprenden las etapas de: (a) la exposición de una bacteria precursora a un bacteriófago; (b) la selección de una bacteria resistente a bacteriófagos (es decir, una “mutante insensible a bacteriófagos”); (c) comparar la ubicación CRISPR o una parte de la misma, de una bacteria precursora y una mutante insensible a bacteriófagos; (d) la selección de una bacteria etiquetada que comprenda un fragmento adicional de ADN en la ubicación CRISPR, que no esté presente en la bacteria precursora; y (e) el aislamiento y/o la clonación y/o la secuenciación del fragmento de ADN adicional. Se proporcionan variantes de CRISPR producidas utilizando los métodos establecidos en el presente. Las variantes de CRISPR son cepas mutantes resistentes a fagos que tienen una ubicación CRISPR modificada con un espaciador adicional.

En algunas modalidades, la presente invención proporciona bacterias precursoras que son aptas para su uso como iniciadoras de cultivos, cultivos probióticos y/o suplementos dietéticos. En algunas modalidades, la bacteria precursora se selecciona de cualquier género adecuado, incluyendo pero no limitado a *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 modalidades, el bacteriófago es seleccionado de una familia de virus adecuada, incluyendo pero no limitado a *Corticoviridae*, *Cystoviridae*, *Inoviridae*, *Leviviridae*, *Microviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, y *Tectiviridae*. En algunas modalidades, la presente invención proporciona cultivos celulares que son seleccionados de cultivos de inicio, cultivos probióticos y/o suplementos dietéticos. Se proporcionan métodos para la identificación de bacterias etiquetadas, que comprenden la etapa de la comparación de al menos un fragmento adicional de ADN, con una secuencia de la base de datos de bacteriófagos y/o una secuencia de la base de datos bacteriana.

La presente invención también proporciona cepas de *S. thermophilus* que comprenden una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, en donde la secuencia comprenda un espaciador de CRISPR de la cepa de *Streptococcus thermophilus* DGCC7778, referida aquí en lo sucesivo como SEQ ID NO: 680 (caacacattcaacagattaatgaagaatac; SEQ ID NO: 680).

En aún más modalidades, la presente invención puede proporcionar cepas de *S. thermophiles* que incluyen una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia comprende la SEQ ID NO: 680 corriente abajo (por ejemplo, directamente corriente abajo) de la primera repetición CRISPR en al menos una ubicación CRISPR.

En otras modalidades, la presente invención puede proporcionar cepas de *S. thermophiles* que incluyen una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia incluye el espaciador de CRISPR (5'-3') de la cepa *Streptococcus thermophilus* DGCC7778 (tccactcacgtacaaatagtgagtgtactc; SEQ ID NO: 681). Aún en modalidades adicionales, la presente invención puede proporcionar cepas de *S. thermophiles* que comprenden una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia comprende la SEQ ID NO: 681 corriente abajo (por ejemplo, directamente corriente abajo) de la primera repetición CRISPR en al menos una ubicación CRISPR.

Todavía en modalidades adicionales, la presente invención puede proporcionar cepas de *S. thermophiles* que comprenden una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia incluye la SEQ ID NO: 683:

Secuencia CRISPR1 (5'-3') de la cepa de *Streptococcus thermophilus* DGCC7710-RH1

(Continúa en página siguiente)

caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataaccactt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattgacgaggagctattggcacaacttaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgttggaatcct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaaataacacggcaaggtca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggtaacgtattta
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcggtcccaatcctgattaatacttactcg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggttaactgctcaagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggtcatagaaaggagacaacatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattttaaatacttttcataacttcat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggatatagcttaatatcattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagttaccgtttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtcaactaacaattatgct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatggctct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggttaacagctactttttacagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatcccaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGTTtgattcaacataaaaagccagttcaattg
 aacttggcttt (SEQ ID NO: 683)

- 5 Aún en modalidades adicionales, la presente invención puede proporcionar cepas de *S. thermophiles* que comprenden una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia incluye la SEQ ID NO: 683 corriente abajo (por ejemplo, directamente corriente abajo) de la primera repetición de CRISPR en por lo menos una ubicación CRISPR. Las cepas de *S. thermophilus* incluyen una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia incluye la SEQ ID NO: 685 (es decir, 5'-TACGTTTGAAGAATATCAATCAATGA-3').

10 Aún en modalidades adicionales, la presente invención proporciona cepas de *S. thermophiles* que comprenden una secuencia que se haya obtenido o pueda obtenerse de un bacteriófago, donde la secuencia incluye la SEQ ID NO: 685 corriente abajo (por ejemplo, directamente corriente abajo) de la primera repetición CRISPR en al menos una ubicación CRISPR.

15 Se proporcionan métodos y composiciones que encuentran uso en el desarrollo y el uso de combinaciones de cepas y rotaciones de cultivo de inicio. En algunas modalidades, se proporciona el uso de uno o más CRISPR BIMS simultáneamente en un cultivo de inicio (es decir, una combinación de BIMS). En algunas otras modalidades, se proporciona el uso de uno o más CRISPR BIMS en un esquema de rotación. En algunas otras modalidades, se proporciona el uso de una o más combinaciones CRISPR BIMS en un esquema de rotación.

20 La presente invención también proporciona métodos y composiciones para utilizar CRISPR-cas (es decir, mutagénesis natural, en algunas modalidades preferidas) para construir al menos una variante resistente a fagos CRISPR, de al menos un microorganismo objetivo, que luego es utilizada para generar fagos mutantes que eludan la

resistencia a CRISPR-cas a través de la mutación en el fago que corresponde a una secuencia seleccionada de las secuencias espaciadoras, las secuencias pseudoespaciadoras, las secuencias proximales, las porciones de reconocimiento, etc., para aumentar la virulencia de los fagos.

En algunas modalidades, la presente invención proporciona composiciones y métodos adecuados para la producción de fagos que tengan una mayor virulencia, en comparación con el fago precursor. En algunas modalidades, por lo menos un espaciador de clonado se introduce en una ubicación activa CRISPR-cas, para producir una variante de la célula resistente a fagos para su uso en la generación de fagos mutantes. En algunas modalidades preferidas, los métodos comprenden la introducción de una secuencia que sirve como objetivo específico de una secuencia del genoma del fago (por ejemplo, una región que es altamente susceptible a un objetivo espaciador o una secuencia de reconocimiento para la incorporación de un hospedador espaciador). En algunas modalidades adicionales, la presente invención proporciona métodos y composiciones para la producción por ingeniería directa de fagos, de modo que la secuencia del genoma correspondiente al espaciador es mutada en consecuencia.

La presente invención también proporciona métodos para dirigir la evolución de un determinado fago, utilizando la resistencia a fagos CRISPR adquirida de una cepa hospedadora correspondiente, para crear un agente de biocontrol más virulento y, por tanto, más efectivo.

Descripción de los dibujos

La Figura 1 proporciona un esquema que muestra que la integración de un espaciador CRISPR en la ubicación CRISPR del *S. thermophilus* proporciona resistencia contra un bacteriófago al cual el espaciador CRISPR muestra identidad. El DGCC7710 precursor es sensible al fago, y el BIM DGCC7710RH1 es resistente al fago. El BIM DGCC7710RH1 tiene un espaciador nuevo (Sn) en la ubicación CRISPR, el cual muestra identidad al 100% para la secuencia de fago. Como se muestra en la etapa (B), la cepa se estimula con el fago 858 y se selecciona un fago mutante resistente. Como se muestra en la etapa (C), la ubicación CRISPR I del mutante tiene un espaciador adicional el cual comparte identidad al 100% con la región 31.921-31.950bp del fago.

La Figura 2 proporciona un esquema que muestra que la integración de un espaciador CRISPR en la ubicación CRISPR de *S. thermophilus* proporciona resistencia contra un bacteriófago para el cual el espaciador CRISPR muestra identidad. El DGCC7710 precursor es sensible al fago, y el BIM DGCC7710RH2 es resistente al fago. El BIM DGCC7710RH2 tiene un espaciador nuevo (Sn) en la ubicación CRISPR, el cual muestra identidad al 100% para la secuencia de fago. Como se muestra en la etapa (B), la cepa se estimula con el fago 858 y un fago mutante resistente se selecciona. Como se muestra en la etapa (C), el experimento se repitió independientemente y se seleccionó otro mutante. La ubicación CRISPR I del mutante tiene un espaciador adicional (diferente de aquel en RH1) el cual comparte identidad al 100% con la región 17.125-17.244bp del fago.

La Figura 3 proporciona una representación gráfica que ilustra la preparación del constructo CAS1KO en el cual el gen *cas1* se fracciona por la recombinación homóloga.

La Figura 4 proporciona una representación gráfica de la preparación del constructo RT usando una enzima de restricción para generar el constructo RT del constructo S1S2. Existen sitios de restricción *Bgl*I dentro de las repeticiones que permiten que la parte "media" del constructo sea cortada. Después de la digestión enzimática, se usó una ligasa para unir las dos piezas de los extremos, generando así un nuevo constructo que tiene RT, pero no espaciadores.

La Figura 5 proporciona una representación gráfica de la integración del constructo RT.

La Figura 6 proporciona una representación gráfica que ilustra el constructo S1S2 producido por ingeniería usando cebadores específicos y reacciones PCR iterativas. El primer panel ilustra todos los cebadores usados y el inicio de las dos primeras reacciones PCR (reacción #1 con los cebadores P1 y P2 y la reacción #2 con los cebadores P2 y P3). El segundo panel muestra los productos PCR obtenidos de las dos primeras reacciones PCR, con el producto de reacción #1 en el lado izquierdo y el producto de reacción #2 en el lado derecho. El tercer panel muestra la tercera reacción PCR, usando una combinación de los productos de los dos primeros PCRs como la plantilla para la tercera reacción PCR, y el cebador P1 de la primera reacción junto con el cebador P4 de la segunda reacción. El cuarto panel muestra el producto de PCR#3, el cual genera técnicamente el constructo S1S2.

La Figura 7 proporciona una representación gráfica de los detalles para el diseño del cebador para los cebadores 2 y 3, los cuales contienen secuencias claves para el experimento, derivadas de los espaciadores idénticos a las secuencias de fago (los productos PCR derivados de estos cebadores PCR generarán los espaciadores que al final proporcionarán resistencia a los fagos).

La Figura 8 proporciona una representación gráfica de la integración del constructo S1S2.

La Figura 9 proporciona una representación gráfica que muestra una vista general de la ubicación CRISPR1 de *S. thermophilus*, los espaciadores recientemente adquiridos en las mutantes resistentes al fago, y sensibilidad de fago

correspondiente. La ubicación CRISPR1 del DGCC7710 (WT) está en la parte superior. La región repetición/espaciador del WT está a la mitad: repeticiones (diamante negro), espaciadores (recuadros grises enumerados), líder (recuadro blanco L) y repetición terminal (diamante negro T). En el fondo, el contenido de espaciador en el lado líder de la ubicación en mutantes resistentes al fago se detalla en el lado izquierdo, con espaciadores recientemente adquiridos (recuadros blancos, S1-S14). En el lado derecho, la sensibilidad de cada cepa a los fagos 858 y 2972 se representa como un histograma de la eficiencia de plaqueo (EOP), la cual es la relación de cuentas de placa de una cepa mutante a aquella de tipo natural.

La Figura 10 proporciona el espaciador CRISPR producido por ingeniería, inactivación del gen *cas* y la sensibilidad de fago correspondiente. I, WT_{Φ858}^{+S1S2} mutante; II, WT_{Φ858}^{+S1S2}ΔCRISPR1 mutante donde CRISPR1 se eliminó; III, WT_{Φ858}^{+S1S2}::pR mutante donde CRISPR1 se colocó y reemplazó con una repetición única; IV, WT₀₂₉₇₂^{+S4}::pS1S2, mutante de la cepa WT₀₂₉₇₂^{+S4} donde el CRISPR1 se colocó y reemplazó con una versión que contenía S1 y S2; V, WT_{Φ858}^{+S1S2}::pcas5- con el *cas5* inactivado; VI, WT_{Φ858}^{+S1S2}::pcas7- con el *cas7* inactivado. El pORI indica el plásmido integrado (12). La sensibilidad del fago de cada cepa a los fagos 858 y 2972 se representa en el fondo como un histograma de la eficiencia de plaqueo (EOP).

La Figura 11 proporciona un esquema que muestra la construcción del constructo S1S2.

La Figura 12 proporciona un esquema que muestra la construcción del WT_{Φ858}^{+S1S2}ΔCRISPR1.

La Figura 13 proporciona una alineación del espaciador CRISPR S1 con la región genómica correspondiente del fago 858 y los dos fagos mutantes que sortean la resistencia CRISPR de la cepa WT_{Φ858}^{+S1S2}.

La Figura 14 proporciona una representación esquemática de la construcción de un primer nivel de la variante resistente al fago. Cada variante tiene un espaciador adicional sencillo dentro de su CRISPR. Los espaciadores adicionales no se relacionan cada uno con los otros (por ejemplo, cada uno tiene una secuencia diferente). Todos los espaciadores se originan del fago P.

La Figura 15 proporciona una representación esquemática del segundo nivel de variantes resistentes al fago que presentan un incremento en la resistencia a los fagos. Las variantes finales (A1.n y A2.n) son originadas de la cepa A y tienen una integración secuencial de los espaciadores adicionales dentro de su CRISPR, con todos los espaciadores siendo diferentes unos de otros y originados del fago P.

La Figura 16 proporciona una representación esquemática del segundo nivel de variantes resistentes al fago que presentan resistencia ampliada a los fagos. La variante final (A1^{pqr}) es originada de la cepa A y tiene una integración secuencial de espaciadores adicionales dentro de su CRISPR originado de 3 fagos diferentes (esto es, del fago P, Q y R).

La Figura 17 proporciona una representación esquemática de la ubicación CRISPR1 (sección A) y de la ubicación CRISPR3 (sección B) de las cepas *S. thermophilus* descritas en los Ejemplos 7 hasta Ejemplo 16. Los nombres de las cepas se dan en el lado izquierdo de la Figura. Las flechas negras representan las CRISPR repetidas, "R" se establece para Repetición y "RT" se establece para Repetición de Terminal. Las flechas grises numeradas desde 1 hasta 32 en la sección A y desde 1 hasta 12 en la sección B representan los espaciadores CRISPR1 y espaciadores CRISPR3, respectivamente, como están en DGCC7710. Las flechas blancas numeradas desde S4 hasta S35 representan los espaciadores CRISPR adicionales específicos para describir las cepas.

La Figura 18 proporciona un esquema que muestra una modalidad de la presente invención, en la cual se indican una secuencia de marcado y una repetición CRISPR se integra en un extremo de la ubicación CRISPR. En la sección A, la ubicación y elementos CRISPR, incluyendo las repeticiones (R), espaciadores (S), el líder corriente arriba y el remolque corriente abajo, con la repetición de terminal (RT) adyacente al remolque, y se indican los genes *cas* en la vecindad (los 4 genes *cas* nombrados *cas1* hasta *cas4* en la presente, no se muestran a escala). Los genes *cas* pueden estar en cualquier extremo, o dividirse y presentarse en ambos extremos. Además, los genes *cas* pueden localizarse en cualquiera de las dos hebras de ADN. La sección B muestra la secuencia del fago, con un fragmento de la secuencia (Sn) siendo usado como un espaciador adicional (esto es, secuencia de marcado). La figura 18C muestra la inserción de un nuevo espaciador (Sn) (esto es, secuencia de marcado) en un extremo de la ubicación CRISPR (cerca del líder en este ejemplo en el extremo 5' de la ubicación CRISPR), entre las dos repeticiones. La figura 18D proporciona una comparación de la ubicación CRISPR contenida entre el precursor y la bacteria mutante (esto es, bacteria etiquetada), con un espaciador nuevo (Sn) (esto es, secuencia de marcado) integrada en un extremo de la ubicación CRISPR (cerca al líder en este ejemplo), entre las repeticiones. El espaciador nuevo (Sn) constituye la secuencia de marcado la cual es específica para la bacteria mutante (esto es, bacteria etiquetada). En algunas modalidades, el uso de este método resulta en la adición de uno o más espaciadores de la secuencia fago.

La Figura 19 proporciona un esquema que muestra una modalidad de la presente invención, en la cual dos secuencias de marcado y dos repeticiones CRISPR se integran en un extremo de la ubicación CRISPR. En la sección A, se indican la ubicación CRISPR (A) y los elementos, incluyendo las repeticiones (R), espaciadores (S), el

líder corriente arriba y el remolque corriente abajo, con la terminal repetida (RT) adyacente al remolque, y los genes *cas* en la vecindad (4 genes *cas* nombrados como *cas1* hasta *cas4* en la presente, no se muestran a escala). Los genes *cas* pueden estar en cualquier extremo o dividirse y presentarse en ambos extremos. Los genes *cas* pueden localizarse en cualquiera de las dos hebras de ADN. La sección B muestra la secuencia fago en negro, con dos fragmentos de la secuencia (Sn y Sn') siendo usados como espaciadores adicionales (esto es, secuencias de marcado). La sección C muestra la inserción de los nuevos espaciadores (esto es, secuencias de marcado) (Sn y Sn') en el mismo extremo de la ubicación CRISPR (cercano al líder en este ejemplo en la terminal 5'), cada uno de los cuales está entre dos repeticiones. La sección B proporciona una comparación de la ubicación CRISPR contenida entre el precursor y la bacteria mutante (esto es, bacteria etiquetada), con dos espaciadores nuevos (Sn y Sn') integrados en el mismo extremo de la ubicación CRISPR (cercano al líder en este ejemplo en la terminal 5'), con cada uno localizado entre las repeticiones. Los espaciadores nuevos Sn y Sn' constituyen la secuencia de marcado la cual es específica del mutante. En algunas modalidades, este método resulta en la adición de uno o más espaciadores de la secuencia fago.

La Figura 20 proporciona una gráfica que muestra la evolución del conteo de fago en leche que contiene 10^7 pfu/ml de D2972 durante la fermentación con DGCC7710 (diamantes negros) o con DGCC9836 (cuadros abiertos). La leche fue leche en polvo al 10% (p/v) en agua. La temperatura de incubación fue 42°C.

La Figura 21 proporciona una gráfica que muestra la evolución del conteo de fago acumulado en WT_{phi2972}^{+S20} y WT_{phi2972}^{+S21} en leche que contiene 10^7 pfu/ml de D2972 durante la fermentación inoculada con WT_{phi2972}^{+S20} (punteado) o con WT_{phi2972}^{+S21} (gris claro) o con ambos WT_{phi2972}^{+S20} y WT_{phi2972}^{+S21} (gris oscuro). La leche fue leche en polvo al 10% (p/v) en agua. La temperatura de incubación fue 42°C.

La Figura 22 proporciona un Logo Web para la porción CRISPR1 NNAGAAW (SEQ ID NO:696).

La Figura 23 proporciona una alineación de los protoespaciadores CRISPR3 seleccionados y las regiones de flanqueo y el logo Web para la porción CRISPR3 NGGNG (SEQ ID NO:723). En esta figura, se proporcionan S42 DGCC7710_{phi2972}^{+S40}_{phi3821}^{+S41}_{S42} (SEQ ID NO:724), S41 DGCC7710_{phi2972}^{+S40}_{phi3821}^{+S41}_{S42}, (SEQ ID NO:699) S41 DGCC7710_{phi858}^{+S1}_{S2delta}_{CRISPR1}_{phi848}^{+S43} (SEQ ID NO:700), y S78 LMD-9_{phi4241}^{+S78} (SEQ ID NO:701).

Descripción Detallada de la Invención

La presente invención proporciona métodos y composiciones que encuentran uso en el desarrollo y el uso de combinaciones de cepas y rotaciones de cultivo de inicio.

La *Streptococcus thermophilus* es una especie de bacteria Gram positiva G+C baja que es una especie clave explotada en la formulación de sistemas de cultivos lácteos para la producción de yogur y queso. Los análisis genómicos comparativos de las cepas *S. thermophilus* cercanamente relacionadas revelan previamente que principalmente el polimorfismo genético se presenta en las ubicaciones hipervariables, tales como los operones *eps* y *rps*, así como dos repeticiones palindrómicas cortas interespaciadas regularmente agrupadas en las ubicaciones (CRISPR) (Ver, por ejemplo, Jansen et al., Mol. Microbiol., 43: 1565 [2002]; Bolotin et al., Microbiol., 151:2551 [2005]; y Bolotin et al., Nat. Biotechnol., 22: 1554 [2004]). Como se describe en la presente en mayor detalle, las ubicaciones CRISPR consisten típicamente de diversas repeticiones directas no contiguas separadas por extensiones de las secuencias variables llamadas espaciadores, y son muchas veces adyacentes a los genes *cas* (CRISPR asociados). Aunque la función de las ubicaciones CRISPR no se ha establecido biológicamente, en los análisis *in silico* de los espaciadores se revela la homología de secuencia con los elementos externos, incluyendo el bacteriófago y secuencias de plásmido (ver por ejemplo, Bolotin et al., Microbiol., supra; Mojica et al., supra; y Pourcel et al., supra). Con base exclusivamente en análisis *in silico*, diversas hipótesis se han presentado por adelantado que proponen el papel del CRISPR y los genes *cas*, que incluyen proporcionar inmunidad contra los elementos genéticos externos por medio de un mecanismo basado en la interferencia del ARN (ver, Makarova et al., Biol. Direct., 1:7 [2006]). Sin embargo, no se intenta que la presente invención se limite a ningún mecanismo particular y/o medios de acción.

Las estrategias actuales usadas en la industria para minimizar la infección del bacteriófago y la falla resultante de los cultivos de la bacteria, incluyen el uso de: (i) cultivos de inicio mezclados; y (ii) el uso de cepas alternas que tienen diferentes perfiles de susceptibilidad al fago (esto es, rotación de cepa). Tradicionalmente, los cultivos de inicio usados en la industria láctea son mezclas de cepas de bacterias de ácido láctico. La composición compleja de los cultivos de inicio mezclados asegura que se proporciona un cierto nivel de resistencia al ataque del fago. Sin embargo, los sub-cultivos repetidos de los cultivos de cepas mezclados llevan a cambios imprevisibles en la distribución de las cepas individuales y eventualmente a dominios de cepas no deseadas. Esto a su vez puede llevar a incrementar la susceptibilidad al ataque del fago y riesgo de fallas en la fermentación.

La rotación de las cepas bacterianas seleccionadas las cuales son sensibles a diferentes fagos es otro enfoque actualmente usado para limitar el desarrollo del fago. Sin embargo, es difícil e incómodo identificar y seleccionar un número suficiente de las cepas que tienen diferentes tipos de perfiles para proporcionar un programa de rotación eficiente y confiable. Además, el uso continuo de las cepas requiere el monitoreo cuidadoso para los nuevos fagos

infecciosos y la necesidad de substituir rápidamente una cepa infectada con una cepa bacteriana resistente. En las plantas de fabricación donde se preparan grandes cantidades de cultivos de inicio en volumen antes de usarse, usualmente no es posible tal respuesta rápida. Así, diversos intentos se han hecho para mejorar la resistencia de los cultivos para uso en la industria.

Además, aunque sería útil tener cultivos de inicio que estén etiquetados de tal manera que su origen pudiera determinarse, esto no se ha hecho. Efectivamente, aunque es posible insertar un oligonucleótido sintético en una cepa para etiquetarla o marcarla, usando tecnologías de ADN recombinante, la cepa etiquetada podría considerarse que es un organismo genéticamente modificado y por lo tanto se enfrentan a cuestiones reglamentarias en aplicaciones comerciales. Así, existe una necesidad en el arte para los métodos naturales y composiciones adecuadas para introducir una secuencia única en la bacteria que pudiera usarse para identificar y/o rastrear la bacteria.

Los bacteriófagos posiblemente son la entidad biológica más abundante en el planeta (Ver, Breitbart y Rohwer, Trends Microbiol., 13:278 [2005]). Su distribución y abundancia ubicua tienen un impacto importante en la ecología microbiana y la evolución de los genomas bacterianos (ver, Chibani-Chennoufi et al, J. Bacteriol., 186:3677 [2004]). En consecuencia, la bacteria ha desarrollado una variedad de mecanismos de defensa naturales que dirigen diversas etapas objetivo del ciclo de vida fago, absorción del bloqueo notablemente, prevención de la inyección de ADN, restricción de la entrada de ADN y sistemas de infección abortivos. Estas barreras antiviricas pueden también producirse por ingeniería y manipularse para mejorar el control de las poblaciones fago (ver por ejemplo, Chibani-Chennoufi et al., supra; y Sturino y Klaenhammer, Nat. Rev. Microbiol., 4:395 [2006]).

Numerosas bacterias se han seleccionado por los humanos y usado ampliamente para la fermentación y procesos de biotecnología. Desafortunadamente, las bacterias domesticadas usadas en las aplicaciones industriales son frecuentemente susceptibles al ataque de fago, incluyendo aquellos generos y especies ampliamente usados como cultivos lácteos (ver, Brussow, Ann. Rev. Microbiol., 55:283 [2001]). En consecuencia, la industria ha ideado diversas estrategias para combatir al fago basadas en la diversidad en la cepa, mutantes insensibles al bacteriófago, y plásmidos que portan mecanismos de resistencia al fago.

Definiciones

A menos de que se defina de otra manera en la presente, todos los términos técnicos y científicos usados en la presente tienen el mismo significado como se entiende comúnmente por alguien de habilidad ordinaria en el arte para la cual esta invención concierne. Aunque ninguno de los métodos y materiales similares o equivalentes a aquellos descritos en la presente encuentran uso en la práctica de lo que se describe en la presente, se describen en la presente métodos y materiales ilustrativos. Como se usa en la presente, los términos singulares "un", "una," y "el" "la" incluyen la referencia plural a menos de que el contexto indique claramente lo contrario. A menos de que se indique de otra manera, los ácidos nucleicos se escriben de izquierda a derecha en la orientación 5' hasta 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en la orientación de amino a carboxi, respectivamente. Se entiende que esta invención no se limita a la metodología, protocolos y reactivos particulares descritos, ya que estos pueden variar, dependiendo del contexto en que se usan por aquellos de habilidad en el arte.

Se intenta que cada limitación numérica máxima dada a través de esta especificación incluya cada limitación numérica inferior, como si tales limitaciones numéricas inferiores se expresaron por escrito en la presente. Cada limitación numérica mínima dada a través de esta especificación incluirá cada limitación numérica más alta, como si tales limitaciones numéricas más altas se escribieron expresamente en la presente. Cada intervalo numérico dado a través de esta especificación incluye cada intervalo numérico más estrecho que cae dentro de tal intervalo numérico más amplio, como si tales intervalos numéricos más estrechos fueron todos escritos expresamente en la presente.

Como se usa en la presente, el término "que se presenta naturalmente" se refiere a elementos y/o procesos que se presentan en la naturaleza.

Como se usa en la presente, los términos "constructo," "conjugado," "casete," e "híbrido," incluyen una secuencia de nucleótido directa o indirectamente enlazada a otra secuencia (por ejemplo, una secuencia reguladora, tal como un promotor). En algunas modalidades, la presente invención proporciona constructos que comprenden una secuencia de nucleótido ligada operablemente a tal secuencia reguladora. El término "ligado operablemente" se refiere a una yuxtaposición en donde los componentes descritos están en una relación que les permite funcionar en su manera pretendida. Una secuencia reguladora "ligada operablemente" a una secuencia codificadora se liga de tal manera que la expresión de la secuencia codificadora se realiza bajo condición compatible con las secuencias de control. Como se usa en la presente, el término "secuencias reguladoras" incluyen promotores y aumentadores y otras señales de regulación de expresión. Como se usa en la presente, el término "promotor" se usa en el sentido normal del arte, por ejemplo, un sitio de enlace de la polimerasa de ARN. En algunas modalidades, los constructos comprenden o expresan un marcador, el cual permite la selección del constructo de la secuencia de nucleótido en, por ejemplo, una bacteria. Existen diversos marcadores que pueden usarse, por ejemplo, aquellos marcadores que proporcionan resistencia antibiótica/antimicrobiana.

En algunas modalidades, el constructo comprende un vector (por ejemplo, un plásmido). En algunas modalidades adicionales, la presente invención proporciona vectores que comprenden uno o más de los constructos o secuencias descritos en la presente. Como se usa en la presente, el término “vector” incluye los vectores de expresión, vectores de transformación, y vectores de transferencia. El término “vector de transformación” se refiere a un constructo capaz de transferirse de una entidad a otra entidad, la cual puede ser de la misma especie o puede ser de especies diferentes. Los constructos que son capaces de transferirse de una especie a otra se refieren algunas veces a “vectores de transferencia”. En algunas modalidades, los vectores se transforman en una célula hospedadora adecuada como se describe en la presente. En algunas modalidades, los vectores son vectores de plásmido o fago proporcionados con un origen de replicación, opcionalmente un promotor para la expresión del polinucleótido, y opcionalmente un regulador del promotor. En algunas modalidades, los vectores que contienen una o más secuencias de nucleótidos marcadas seleccionadas. Los sistemas de selección más adecuados para los microorganismos industriales son aquellos formados por el grupo de marcadores de selección los cuales no requieren una mutación en el organismo hospedador. En algunas modalidades, los vectores se usan *in vitro* (por ejemplo, para la producción del ARN o usado para transfecta o transformar una célula hospedadora). En algunas modalidades, los polinucleótidos se incorporan en un vector recombinante (típicamente un vector replicable), tal como un vector de clonación o expresión. El vector encuentra uso en la replicación del ácido nucleico en una célula hospedadora compatible.

La introducción de un ácido nucleico (por ejemplo, un fago, constructo o vector) en una célula puede efectuarse por diversos métodos. Por ejemplo, en algunas modalidades, la transducción, transformación, transfección de fosfato de calcio, transfección mediada por DEAE-dextrano, transfección mediada por lípidos catiónicos, electroporación, transducción o infección pueden encontrar uso. Efectivamente, cualquier método adecuado conocido en el arte encuentra uso en la presente invención. En algunas modalidades, las células que contienen ácido nucleico exógeno (introducidas por medio de fago, constructo, o un vector) se seleccionan para usar cualquier método adecuado conocido en el arte.

Las enseñanzas sobre la transformación de las células están bien documentadas en el arte, por ejemplo, ver Sambrook et al (Molecular Cloning: A Laboratory Manual, 2ª edición, 1989, Cold Spring Harbor Laboratory Press) y Ausubel et al., Current Protocols in Molecular Biology (1995), John Wiley & Sons, Inc.

En el contexto de introducir un ácido nucleico en una célula, en algunas modalidades, se prefiere que el término “introducir” signifique uno o más de entre transformar, transfectar, conjugar o transducir. En algunas modalidades particularmente preferidas, las cepas bacterianas (por ejemplo, cepas bacterianas precursoras, cepas bacterianas variantes, etc.) se “exponen” a al menos un fago, de tal manera que el ácido nucleico del fago se introduce en las células de la cepa bacteriana.

Como se usa en la presente, los términos “secuencia de ácido nucleico,” “secuencia de nucleótido,” y “ácido nucleico,” se refieren a cualquier secuencia de ácido nucleico, que incluye ADN, ARN, genómico, sintético, recombinante (por ejemplo, cADN). Se pretende que los términos abarquen las secuencias de hebra doble y/o hebra sencilla, ya sea que representen la hebra sentido o antisentido o combinaciones de las mismas. Las secuencias de ácidos nucleicos recombinantes se preparan por el uso de cualquier técnica de ADN recombinante adecuada. En algunas modalidades, como se describe en la presente, las secuencias de ácido nucleico proporcionadas incluyen secuencias de gen que codifican CRISPR, Cas, y otras secuencias. Efectivamente, como se usa en el contexto, la presente invención abarca secuencias de ácidos nucleicos que codifican diversas secuencias CRISPR, que incluyen pero no se limitan a espaciadores, pseudoespaciadores, líderes, etc., así como secuencias cas, y otras secuencias de ácidos nucleicos bacterianas y de fago (“bacteriófago”).

Los términos “molécula de ácido nucleico codificada”, “secuencia de ácido nucleico codificada”, “Secuencia de ADN codificada”, y “ADN codificado”, se refieren al orden o secuencia de desoxirribonucleótidos a lo largo de una hebra de ácido desoxirribonucleico. El orden de estos desoxirribonucleótidos determina el orden de los aminoácidos a lo largo de la cadena del polipéptido (proteína). La secuencia de ADN codifica de esta manera la secuencia de aminoácido.

Como se usa en la presente en el contexto de introducir una secuencia de ácido nucleico en una célula, el término “introducido” se refiere a cualquier método adecuado para transferir la secuencia de ácido nucleico en la célula. Tales métodos para la introducción incluyen pero no se limitan a fusión de protoplastos, transfección, transformación, conjugación, y transducción. En algunas modalidades particularmente preferidas, el ácido nucleico se introduce en las células receptoras durante la infección de las células por el o los bacteriófagos.

En algunas modalidades, las secuencias de ácidos nucleicos y los ácidos nucleicos proporcionados en la presente se aíslan o purifican substancialmente. Por “aislado” o “purificado substancialmente” se pretende que las moléculas de ácido nucleico, o fragmentos biológicamente activos o variantes, homólogos, o derivados de los mismos estén substancialmente o esencialmente libres de componentes normalmente encontrados en asociación con el ácido nucleico en su estado natural. Tales componentes incluyen, pero no se limitan a otro material celular, medios de cultivo, materiales de producción recombinante, y diversos productos químicos usados en sintetizar químicamente los ácidos nucleicos.

En algunas modalidades, una secuencia de ácido nucleico o ácido nucleico “aislado” está típicamente libre de secuencias de ácidos nucleicos que flanquean el ácido nucleico de interés en el ADN genómico del organismo a partir del cual el ácido nucleico se derivó (por ejemplo, secuencias codificadoras presentes en los extremos 5' o 3'). Sin embargo, la molécula puede incluir algunas bases o restos adicionales que no afectan perjudicialmente las características básicas de la composición.

Como se usa en la presente, el término “modificación” se refiere a cambios hechos dentro del ácido nucleico y/o secuencias de aminoácidos. En algunas modalidades, las modificaciones se logran usando métodos producidos por ingeniería genética (por ejemplo, recombinante), mientras en otras modalidades, las modificaciones se hacen usando mecanismos genéticos que se presentan naturalmente. Se entiende que la totalidad o parte de una secuencia se modificará usando los métodos de la presente invención. En algunas modalidades preferidas, los ácidos nucleicos modificados incluyen uno o más espaciadores CRISPR producidos recombinantemente o que se presentan naturalmente, genes o proteínas cas, repeticiones CRISPR, ubicaciones CRISPR, así como ácido nucleicos de bacteriófagos. Cualquier método adecuado conocido en el arte encuentra uso en la presente invención, incluyendo pero no limitando al uso del PCR, clonación, mutagénesis dirigida al sitio, etc. Efectivamente, los kits comercialmente disponibles encuentran uso en la presente invención. En algunas modalidades, los oligonucleótidos sintéticos se usan. En algunas modalidades, los métodos tales como recombinación homóloga encuentran uso (por ejemplo, para la inserción o eliminación de los espaciadores CRISPR). En algunas modalidades, la producción por ingeniería genética incluye la activación de una o más secuencias de ácidos nucleicos (por ejemplo, ubicaciones CRISPR, repeticiones CRISPR, espaciadores CRISPR, genes o proteínas cas, combinaciones funcionales de los genes o proteínas cas y repeticiones CRISPR, o combinaciones de los mismos).

En algunas modalidades, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR se insertan en al menos una ubicación CRISPR. En algunas modalidades adicionales, la modificación no interrumpe uno o más genes cas de al menos una ubicación CRISPR. En otras modalidades, uno o más genes cas permanecen intactos. En algunas modalidades adicionales, la modificación no interrumpe una o más repeticiones CRISPR de al menos una ubicación CRISPR. En algunas modalidades, una o más repeticiones CRISPR permanecen intactas. En algunas modalidades adicionales, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR se insertan en o dentro de al menos una ubicación CRISPR. En algunas modalidades adicionales, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR se insertan en la terminal 5' de al menos una ubicación CRISPR.

En algunas modalidades, la modificación comprende insertar al menos un espaciador CRISPR o pseudoespaciadores CRISPR en una célula (por ejemplo, una célula receptora). En algunas otras modalidades, la modificación comprende insertar uno o más espaciadores CRISPR o pseudoespaciadores CRISPR en (por ejemplo, para modificar o reemplazar) uno o más espaciadores CRISPR de una célula receptora. En algunas modalidades, los espaciadores CRISPR de la célula son los mismos, mientras en otras modalidades, son diferentes. En algunas modalidades, la modificación comprende insertar al menos un espaciador CRISPR o pseudoespaciador CRISPR de un organismo donador en una célula receptora. En algunas modalidades adicionales, la modificación comprende insertar uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador en una célula receptora bajo condiciones adecuadas para modificar o reemplazar uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de la célula receptora. En algunas modalidades, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador se insertan en uno o más, preferiblemente, dos o más repeticiones CRISPR de la célula. En algunas modalidades preferidas, al menos una combinación cas repetida CRISPR funcional permanece intacta en la célula.

En algunas modalidades adicionales, la inserción se presenta adyacente a uno o más (preferiblemente dos o más) espaciadores o pseudoespaciadores CRISPR. Como se usa en la presente, el término “adyacente” significa “próximo a” en su sentido más amplio e incluye “directamente adyacente.” Así, en algunas modalidades, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo se insertan “directamente adyacente” a uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de la célula receptora, (esto es, el o los espaciadores CRISPR o pseudoespaciadores CRISPR se insertan de manera que no hay intervención de nucleótidos entre los espaciadores).

En algunas modalidades adicionales, el o los espaciadores CRISPR o pseudoespaciadores CRISPR se insertan de manera que existen al menos alrededor de 2, alrededor de 3, alrededor de 4, alrededor de 5, alrededor de 6, alrededor de 7, alrededor de 8, alrededor de 9, alrededor de 10, alrededor de 15, alrededor de 20, alrededor de 25, alrededor de 30, alrededor de 35, alrededor de 40, alrededor de 45, alrededor de 50, alrededor de 55, alrededor de 60, alrededor de 65, alrededor de 70, alrededor de 75, alrededor de 80, alrededor de 85, alrededor de 90, alrededor de 95, alrededor de 100, alrededor de 200, alrededor de 300, alrededor de 400, alrededor de 500, alrededor de 600, alrededor de 700, alrededor de 800, alrededor de 900, alrededor de 1000, alrededor de 10.000, 1 alrededor de 00.000, o alrededor de 1.000.000 o más nucleótidos que intervienen entre los espaciadores.

En algunas modalidades adicionales, el nucleótido que interviene hace referencia a una “secuencia líder”. Estos términos se usan intercambiabilmente en la presente. La secuencia líder puede ser de una longitud diferente en bacterias diferentes. En algunas modalidades, la secuencia líder es de al menos alrededor de 20, alrededor de 25,

alrededor de 30, alrededor de 35, alrededor de 40, alrededor de 45, alrededor de 50, alrededor de 55, alrededor de 60, alrededor de 65, alrededor de 70, alrededor de 75, alrededor de 80, alrededor de 85, alrededor de 90, alrededor de 95, alrededor de 100, alrededor de 200, alrededor de 300, alrededor de 400, o alrededor de 500 o más nucleótidos en longitud. En algunas modalidades preferidas, la secuencia líder está entre el último gen *cas* (en el extremo 3') y la primera repetición CRISPR (en el extremo 5') de la ubicación CRISPR. En algunas modalidades, la secuencia líder está entre alrededor de 20-500 nucleótidos en longitud.

En algunas modalidades, uno o más de los espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador se insertan adyacentes a uno o más genes *cas* de una célula receptora, en donde los genes *cas* son los mismos o diferentes. En algunas modalidades adicionales, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador se insertan adyacentes al mismo o diferentes espaciadores de la célula receptora.

En otra modalidad, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR tales como uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador, cada uno se inserta adyacente a las mismas o diferentes repeticiones CRISPR de la célula. En otra modalidad, uno o más espaciadores CRISPR o pseudoespaciadores CRISPR, tales como uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador, son cada uno insertados adyacentes a los mismos o diferentes genes *cas* de la célula receptora.

En algunas modalidades adicionales, la secuencia de uno o más espaciadores CRISPR de un organismo donador se proporcionan bajo condiciones en las que la célula receptora se modifica de tal manera que el espaciador CRISPR tiene homología al espaciador CRISPR o pseudoespaciador CRISPR del organismo donador. En algunas modalidades, el espaciador CRISPR tiene homología al 100% al espaciador CRISPR del organismo donador.

En algunas modalidades, el o los espaciadores CRISPR o pseudoespaciadores CRISPR comprenden un ADN o ARN de origen genómico, sintético o recombinante. En algunas modalidades, el o los espaciadores CRISPR o pseudoespaciadores CRISPR son de hebra doble, mientras en otras modalidades, son de hebra sencilla, si representan la hebra sentido o antisentido o combinaciones de las mismas. Se contempla que el o los espaciadores CRISPR o pseudoespaciadores CRISPR se preparen por el uso de las técnicas de ADN recombinantes (por ejemplo, ADN recombinante), como se describe en la presente.

En algunas modalidades, la modificación comprende insertar uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de un organismo donador que son substancialmente resistentes a un ácido nucleico objetivo o un producto de transcripción del mismo en una o más ubicaciones CRISPR de una célula substancialmente sensible. En algunas modalidades, la inserción se presenta en o entre una combinación funcional de al menos dos repeticiones de CRISPR y al menos un gen *cas* en una célula substancialmente sensible. En algunas modalidades, la modificación comprende modificar (por ejemplo, mutar) el ADN de una célula receptora (por ejemplo, ADN plásmido o ADN genómico), de tal manera que uno o más genes *cas* se crean en el ADN de la célula. En algunas modalidades, los genes *cas* se clonan en un constructo, un plásmido o un vector, etc., el cual luego se transforma en la célula, usando cualquier método adecuado.

En algunas modalidades, la modificación comprende modificar (por ejemplo, mutar) el ADN de una célula receptora (por ejemplo, tal como ADN plásmido o ADN genómico), de tal manera que uno o más, preferiblemente, dos o más repeticiones de CRISPR se crean en el ADN de la célula. En algunas modalidades, las repeticiones de CRISPR se clonan en un constructo, un plásmido o un vector, etc., el cual luego se transforma en la célula, usando cualquier método adecuado.

En algunas modalidades adicionales, la modificación comprende modificar (por ejemplo, mutar) el ADN de una célula receptora (por ejemplo, ADN plásmido o ADN genómico), de tal manera que una o más combinaciones funcionales repetidas del *cas*-CRISPR se crean en el ADN de la célula. En algunas modalidades, las combinaciones funcionales repetidas del *cas*-CRISPR pueden clonarse en un constructo, un plásmido o un vector, el cual luego se transforma en la célula, usando cualquier método adecuado.

En algunas modalidades, la modificación comprende modificar (por ejemplo, mutar) el ADN de una célula receptora (por ejemplo, ADN plásmido o ADN genómico), de tal manera que uno o más espaciadores CRISPR se crean en el ADN de la célula. En algunas modalidades, los espaciadores CRISPR pueden clonarse en un constructo, un plásmido o un vector, el cual luego se transforma en la célula, usando cualquier método adecuado. En algunas modalidades preferidas, un espaciador CRISPR se flanquea por dos repeticiones de CRISPR (esto es, un espaciador CRISPR que tiene al menos una repetición CRISPR en cada lado).

En algunas modalidades, la modificación comprende insertar uno o más espaciadores CRISPR (por ejemplo, espaciadores CRISPR heterólogos) en la vecindad de (por ejemplo, adyacente a/directamente adyacente a) uno o más genes *cas* y/o la secuencia líder. Así, en algunas modalidades, la organización de la ubicación del CRISPR que se presenta naturalmente se mantiene siguiendo la inserción de uno o más espaciadores CRISPR.

Como se usa en la presente, el término “ácido nucleico objetivo” se refiere a cualquier secuencia de ácido nucleico o producto de transcripción del mismo, contra cuya resistencia en una célula (por ejemplo, una célula receptora) se modula. En algunas modalidades, la resistencia se dirige contra la secuencia de ácido nucleico objetivo per se. Ventajosamente, esto confiere resistencia a una célula contra un organismo donador a partir del cual el o los ácidos nucleicos objetivos se derivan. Así, en algunas modalidades, la inserción de un pseudo-espaciador CRISPR derivado de un bacteriófago o un o unos espaciadores CRISPR los cuales son complementarios u homólogos a uno o más pseudoespaciadores CRISPR en una célula receptora que confiere resistencia al bacteriófago. Así, en algunas modalidades preferidas, la inserción entre dos repeticiones de CRISPR de un pseudo-espaciador CRISPR derivado de un bacteriófago o un o unos espaciadores CRISPR son complementarios u homólogos a uno o más pseudoespaciadores CRISPR en una célula receptora que confiere resistencia al bacteriófago. En un aspecto adicional, se proporciona un método para modular la resistencia de una célula receptora contra un ácido nucleico objetivo o un producto de transcripción del mismo.

La presente invención también proporciona métodos para determinar el perfil de resistencia de una célula contra un ácido nucleico objetivo. Como se usa en la presente, el término “perfil de resistencia” significa una o más entidades contra las cuales la célula es sensible o resistente. En consecuencia, en algunas modalidades, el perfil de resistencia de una célula refleja que la célula es resistente a un primer bacteriófago, sensible a un segundo bacteriófago, resistente a un primer elemento genético móvil y sensible a un primer gen de resistencia al antibiótico, etc.

En algunas modalidades, uno o más genes o proteínas *cas*, una o más repeticiones de CRISPR, uno o más genes *cas*, una o más combinaciones funcionales repetidas del *cas*-CRISPR, uno o más espaciadores CRISPR, y/o uno o más espaciadores CRISPR etc., dentro de una célula se detectan y/o se procesan por secuenciado con el fin de predecir/determinar el perfil de resistencia probable de una célula particular. En algunas otras modalidades, uno o más espaciadores CRISPR dentro de una célula se detectan o se procesan por secuenciado con el fin de predecir/determinar el perfil de resistencia probable de una célula particular. Los métodos de detección adecuados incluyen, pero no se limitan a PCR, hibridización de ADN-ADN, hibridización de ADN-ARN, microconfiguraciones de ADN, etc. Efectivamente, se pretende que cualquier método adecuado encuentre uso en la presente invención. En las modalidades adicionales, el perfil de resistencia probable de una célula bacteriana particular a uno o más bacteriófagos se usa como un predictor de lisotipo para selección microbiana. En algunas modalidades adicionales, uno o más genes *cas* y/o una o más repeticiones CRISPR se procesan en secuencia además de uno o más espaciadores CRISPR, con objeto de verificar la compatibilidad de la combinación repetida del gen *cas*-CRISPR o para identificar nuevos pares de combinaciones *cas*/repetición compatibles.

Como se usa en la presente, el término “resistencia modulada” se refiere a suprimir, reducir, disminuir, inducir, conferir, restaurar, elevar, incrementar o de otra manera afectar la resistencia de una célula a un ácido nucleico objetivo, como se toma en el contexto.

Como se usa en la presente, el término “resistencia” no está destinado a implicar que una célula es resistente al 100% a un ácido nucleico objetivo o un producto de transcripción del mismo, pero incluye células que son tolerantes al ácido nucleico objetivo o un producto de transcripción del mismo.

Como se usa en la presente el término “resistencia al ácido nucleico objetivo o producto de transcripción del mismo” significa que se confiere resistencia contra una célula o un organismo (por ejemplo, fago) que comprende o produce el ácido nucleico objetivo o producto de transcripción del mismo. En algunas modalidades, el componente mínimo requerido para conferir inmunidad o resistencia contra un ácido nucleico objetivo o producto de expresión del mismo es al menos un gen *cas* (o una proteína *Cas*) y al menos dos repeticiones de CRISPR que flanquean un espaciador.

En algunas modalidades, la presente invención proporciona métodos para modular (por ejemplo, conferir o incrementar) la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo que comprende las etapas de: identificar una secuencia (por ejemplo, una secuencia conservada) en un organismo (preferiblemente, una secuencia esencial para la función o supervivencia del organismo); preparar un espaciador CRISPR el cual comprende una secuencia homóloga (por ejemplo, idéntica al 100%), a la secuencia identificada; preparar un ácido nucleico que comprende al menos un gen *cas* y al menos dos repeticiones de CRISPR junto con el espaciador CRISPR; y (iv) transformar una célula con el ácido nucleico para de esta manera volver a la célula resistente al ácido nucleico objetivo o producto de transcripción del mismo.

Como se usa en la presente, el término “secuencia conservada” en el contexto de identificar una secuencia conservada en un organismo no necesariamente tiene que ser conservada en su sentido estricto, como el conocimiento de una secuencia de un organismo dado es suficiente. Además, la secuencia no es necesariamente parte de una entidad esencial. Sin embargo, en algunas modalidades, la secuencia conservada es una secuencia que es esencial para la función y/o supervivencia y/o replicación e/o infectividad y similares de un organismo o una célula. En algunas modalidades, la secuencia conservada comprende una helicasa, una proteína estructural de cabeza o cola de primasa, una proteína con un dominio conservado (por ejemplo, holina, lisina, y otros), o secuencias conservadas entre los genes fago importantes.

En algunas modalidades adicionales, la presente invención proporciona métodos para modular (por ejemplo, conferir o incrementar) la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo que comprende las etapas de: identificar uno o más espaciadores CRISPR en un organismo resistente al ácido nucleico objetivo o producto de transcripción del mismo; preparar un ácido nucleico recombinante que comprende al menos un gen o proteína cas y al menos dos repeticiones de CRISPR junto con uno o más espaciadores identificados; y transformar una célula con el ácido nucleico recombinante de esta manera haciendo a la célula receptora resistente al ácido nucleico objetivo o producto de transcripción del mismo.

En algunas modalidades, la presente invención proporciona métodos para modular (por ejemplo, conferir o incrementar) la resistencia de una célula que comprende al menos uno o más genes o proteínas cas y una o más, preferiblemente, dos o más repeticiones de CRISPR contra un ácido nucleico objetivo o un producto de transcripción del mismo que comprende las etapas de: identificar uno o más espaciadores CRISPR en un organismo resistente al ácido nucleico objetivo o producto de transcripción del mismo; y modificar la secuencia de uno o más espaciadores CRISPR en la célula de tal manera que el o los espaciadores CRISPR tienen homología al o los espaciadores CRISPR en el organismo. En algunas modalidades, uno o más espaciadores CRISPR en una célula receptora se modifican (por ejemplo, producidos por ingeniería genéticamente) de tal manera que el o los espaciadores CRISPR tienen homología a uno o más espaciadores CRISPR en un organismo donador que es substancialmente resistente a un ácido nucleico objetivo o un producto de transcripción del mismo, con el fin de hacer a la célula resistente al ácido nucleico objetivo. En algunas modalidades preferidas, uno o más genes o proteínas cas y una o más, preferiblemente, dos o más repeticiones de CRISPR en la célula son una combinación funcional como se describe en la presente.

Los métodos producidos por ingeniería genética incluyen cualquiera de los métodos adecuados conocidos en el arte, que incluyen pero no se limitan a, agregar (por ejemplo, insertar), eliminar (por ejemplo, remover) o modificar (por ejemplo, mutar) la secuencia de uno o más espaciadores CRISPR y/o uno o más pseudoespaciadores CRISPR en una célula, de tal manera que el espaciador CRISPR tiene homología (por ejemplo, incremento de la homología después de producirse por ingeniería genética) a uno o más espaciadores CRISPR de un organismo donador. Esta etapa producida por ingeniería resulta en una célula que fue substancialmente sensible a un ácido nucleico objetivo o un producto de transcripción del mismo que es substancialmente resistente al ácido nucleico objetivo o un producto de transcripción del mismo.

En algunas modalidades adicionales, la presente invención proporciona métodos para disminuir o reducir la resistencia de una célula receptora que comprende al menos uno o más genes o proteínas cas y una o más, preferiblemente, dos o más repeticiones de CRISPR contra un ácido nucleico objetivo o un producto de transcripción del mismo.

En algunas modalidades, los métodos comprenden las etapas de: identificar uno o más espaciadores CRISPR en un organismo que es substancialmente resistente al ácido nucleico objetivo o un producto de transcripción del mismo; y modificar la secuencia de uno o más espaciadores CRISPR en la célula de tal manera que el o los espaciadores CRISPR tienen un grado reducido de homología al o los espaciadores CRISPR en el organismo.

En otras modalidades, los métodos para modular (por ejemplo, disminuir) la resistencia de una célula que comprende uno o más genes o proteínas cas y una o más, preferiblemente, dos o más repeticiones de CRISPR contra un ácido nucleico objetivo o producto de transcripción del mismo que comprende las etapas de: identificar un espaciador CRISPR o un pseudoespaciador CRISPR en un organismo que comprende un ácido nucleico objetivo o producto de transcripción del mismo contra el cual la resistencia se modula; e identificar el espaciador CRISPR en el organismo en el cual la resistencia se modula; y (iii) adaptar la secuencia del espaciador CRISPR en el organismo en el cual la resistencia se modula de tal manera que el espaciador CRISPR tiene un grado inferior de homología al espaciador CRISPR o pseudoespaciador CRISPR del organismo que comprende el ácido nucleico objetivo o producto de transcripción del mismo contra el cual la resistencia se modula.

Uno o más espaciadores CRISPR en una célula substancialmente resistente se producen por ingeniería con el fin de hacer la célula sensible a un ácido nucleico objetivo. Los métodos producidos por ingeniería genética que encuentran uso incluyen, pero no se limitan a, la adición (por ejemplo, inserción), eliminación (por ejemplo, remoción) o modificación de una o más combinaciones repetición CRISPR-cas o porciones o fragmentos funcionales de los mismos en la célula substancialmente resistente y/o la adición (por ejemplo, inserción), eliminación (por ejemplo, remoción) o modificación de uno o más espaciadores CRISPR o porciones o fragmentos de los mismos en la célula substancialmente resistente. Esta etapa producida por ingeniería resulta en una célula que fue substancialmente resistente a un ácido nucleico objetivo o un producto de transcripción del mismo que llega a ser substancialmente sensible a un ácido nucleico objetivo o un producto de transcripción del mismo.

En algunas modalidades, con objeto de conferir sensibilidad a una célula, se contempla que uno o más espaciadores CRISPR, uno o más genes o proteínas cas, una o más, preferiblemente, dos o más repeticiones de CRISPR, y/o una o más combinaciones repeticiones CRISPR-cas funcionales de una célula substancialmente resistente se removerán, eliminarán o modificarán de tal manera que la resistencia no se confiere más. En algunas modalidades, las células que son sensibles a un ácido nucleico objetivo o un producto de transcripción del mismo se preparan de

tal manera que sus niveles dentro de un cultivo dado (por ejemplo, un cultivo de inicio) pueden modularse (por ejemplo, disminuirse) como se desee. Así, en algunas modalidades, los cultivos de inicio que comprenden dos o más cepas bacterianas se desarrollan de tal manera que todos los miembros del cultivo son sensibles al mismo agente (por ejemplo, el mismo bacteriófago). Así, cuando el tiempo ya no es el deseado para el cultivo vivo, el cultivo se pone en contacto con el mismo agente sencillo con el fin de eliminar todos los miembros del cultivo. En algunas modalidades, la sensibilidad de las células se modula para uno o más agentes (por ejemplo, fagos), de tal manera que el agente elimina solamente una cierta proporción de las células en un cultivo dado (por ejemplo, alrededor de 10, alrededor de 20, alrededor de 30, alrededor de 40, alrededor de 50, alrededor de 60, alrededor de 70, alrededor de 80, alrededor de 90, o alrededor de 95% de las células en el cultivo).

En algunas modalidades, una célula receptora se produce por ingeniería de tal manera que comprende un espaciador CRISPR o una secuencia correspondiente a un pseudoespaciador CRISPR, por ello la célula se hace resistente a un ácido nucleico objetivo o producto de transcripción del mismo. Adecuadamente, la célula se produce por ingeniería de tal manera que el espaciador CRISPR o secuencia correspondiente al pseudoespaciador CRISPR se usa junto con una combinación gen *cas*-repetición CRISPR funcional, como se describe en la presente.

En algunas modalidades, una célula que es resistente a un ácido nucleico objetivo o producto de transcripción del mismo se produce por ingeniería de tal manera que el espaciador CRISPR confiere la inmunidad contra el ácido nucleico objetivo o producto de transcripción del mismo insertado en una célula que comprende una combinación gen *cas*-repetición CRISPR funcional, por ello la célula se hace resistente al ácido nucleico objetivo o producto de transcripción del mismo.

En algunas modalidades adicionales, la secuencia de uno o más espaciadores CRISPR o pseudoespaciadores CRISPR de una célula que es resistente a un ácido nucleico objetivo o producto de transcripción del mismo se determina. Una célula receptora luego se produce por ingeniería de tal manera que comprende la secuencia del espaciador CRISPR y una combinación gen *cas*-repetición CRISPR funcional, por ello la célula se hace resistente al ácido nucleico objetivo o producto de transcripción del mismo.

En algunas modalidades adicionales, un espaciador CRISPR de una célula receptora y una combinación gen *cas*-repetición CRISPR funcional de la misma o diferente célula (por ejemplo, la misma o diferente célula receptora) se preparan. Una célula receptora adicional luego se produce por ingeniería de tal manera que comprende la secuencia del espaciador CRISPR y la combinación gen *cas*-repetición CRISPR funcional, por ello la célula se hace resistente al ácido nucleico objetivo o producto de transcripción del mismo.

En algunas modalidades, la resistencia se dirige contra un producto de transcripción de la secuencia de ácido nucleico objetivo (por ejemplo, un transcripto de la secuencia de ácido nucleico objetivo, en particular un ARN o mRNA), un transcripto (por ejemplo, un transcripto de ARN sentido o antisentido), o un producto de transcripción del polipéptido. En algunas modalidades, esto concede resistencia a una célula contra un organismo donador a partir del cual el producto de transcripción se deriva.

En algunas modalidades, la secuencia del nucleótido objetivo comprende ADN o ARN de origen genómico, sintético o recombinante. En algunas modalidades adicionales, la secuencia de nucleótido es de hebra doble, mientras en otra modalidad es de hebra sencilla, si se presenta la hebra sentido o antisentido o combinaciones de las mismas. Aún en modalidades adicionales, la secuencia de nucleótido se prepara por el uso de técnicas de ADN recombinante (por ejemplo, ADN recombinante). En todavía modalidades adicionales, la secuencia de nucleótido es la misma que una forma que se presenta naturalmente, mientras en otras modalidades se deriva de los mismos. Aún en modalidades adicionales, la secuencia de ácido nucleico objetivo se deriva de un gen. En algunas otras modalidades, la secuencia de ácido nucleico objetivo se deriva de una variante, homólogo, fragmento o derivado de un gen. En algunas modalidades preferidas, la secuencia nucleica objetivo es o se deriva del bacteriófago. En algunas modalidades, la secuencia nucleica objetivo se deriva del ADN plásmido. En algunas modalidades, la secuencia nucleica objetivo se deriva de un elemento genético móvil. En algunas modalidades adicionales, la secuencia nucleica objetivo se deriva de un elemento transponible o una secuencia de inserción. Aún en modalidades adicionales, la secuencia nucleica objetivo se deriva de un gen que concede resistencia. En algunas modalidades adicionales, la secuencia nucleica objetivo se deriva de un gen que concede resistencia a un antibiótico o antimicrobiano. En algunas modalidades, la secuencia nucleica objetivo se deriva de un factor de virulencia. En algunas modalidades adicionales, la secuencia nucleica objetivo se deriva de una toxina, una internalina o una hemolisina.

En algunas modalidades, la secuencia nucleica objetivo o un producto de transcripción de la misma se deriva de una o más bacterias. Así, en algunas modalidades preferidas la resistencia de las células bacterianas se modula usando los métodos y composiciones de la presente invención. En algunas modalidades preferidas, la secuencia del nucleótido objetivo se deriva de un gen asociado con la resistencia al plásmido transferido en la bacteria. En algunas modalidades, uno o más espaciadores CRISPR en la célula se modifican de tal manera que el espaciador CRISPR de la célula tiene homología al espaciador CRISPR y/o pseudoespaciador CRISPR contenido en el ADN plásmido de la célula bacteriana, por ello se proporciona resistencia contra el o los plásmidos particulares. Así, la transferencia del ADN externo en la célula se previene. En algunas modalidades preferidas, las regiones particulares dentro del

ADN plásmido se dirigen, de manera que proporcionan inmunidad contra el ADN plásmido. Por ejemplo, en algunas modalidades, se dirigen las secuencias dentro del origen plásmido de replicación o secuencias dentro de los genes que se codifican para las proteínas de replicación.

En algunas modalidades, la presente invención proporciona métodos que comprenden las etapas de: identificar un espaciador CRISPR y/o pseudoespaciador CRISPR derivado del ADN plásmido de una célula bacteriana contra la cual la resistencia se modula; y modificar la secuencia de un espaciador CRISPR en la célula en la cual la resistencia se modula, de tal manera que el espaciador CRISPR de la célula tiene homología al espaciador CRISPR y/o pseudoespaciador CRISPR contenido en el ADN plásmido de la célula bacteriana.

En todavía modalidades adicionales, la presente invención proporciona métodos para conceder resistencia a una célula contra la transferencia del plásmido, que comprende las etapas de: identificar un espaciador, CRISPR y/o pseudoespaciador CRISPR derivado del ADN plásmido; identificar una a más combinaciones gen *cas*-repeticón CRISPR funcional en una célula que es sustancialmente sensible al plásmido; y producir por ingeniería una o más ubicaciones CRISPR en la célula sustancialmente sensible de tal manera que comprendan uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR del plásmido, por ello la célula se hace resistente.

En algunas modalidades, la secuencia del nucleótido objetivo se deriva de un gen asociado con resistencia a uno o más elementos genéticos móviles. En algunas modalidades, los espaciadores CRISPR particulares y/o pseudoespaciadores CRISPR derivados de uno o más elementos genéticos móviles se agregan dentro de una ubicación CRISPR de una célula de manera que proporcionan resistencia contra los elementos genéticos móviles (por ejemplo, elementos de transposición y secuencias de inserción), previniendo de esta manera la transferencia del ADN externo y la desviación genética. En algunas modalidades, las regiones particulares dentro de los transposones y las secuencias de inserción se dirigen de manera que proporcionan inmunidad contra los elementos genéticos móviles. Por ejemplo, en algunas modalidades objetivo incluyen, pero no se limitan a transposones conjugados (Tn916), transposones de clase II (Tn501), secuencias de inserción (IS26), y genes de transposasa.

En algunas modalidades, la presente invención proporciona métodos que comprenden las etapas de: identificar un espaciador CRISPR y/o pseudoespaciador CRISPR derivado de uno o más elementos genéticos móviles de una célula contra la resistencia a ser modulada; y modificar la secuencia de un espaciador CRISPR en una célula en la cual la resistencia se modula de manera que el espaciador CRISPR y/o pseudoespaciador CRISPR de la célula tiene homología al espaciador CRISPR contenido en los elementos genéticos móviles de la célula.

En todavía modalidades adicionales, la presente invención proporciona métodos para conceder resistencia a una célula contra uno o más elementos genéticos móviles que comprenden las etapas de: identificar un espaciador CRISPR y/o pseudoespaciador CRISPR derivado de uno o más elementos genéticos móviles; identificar una o más combinaciones *cas*-repeticón CRISPR funcionales en una célula que es sustancialmente sensible a uno o más elementos genéticos móviles; y producir por ingeniería una o más las ubicaciones CRISPR en la célula sustancialmente sensible de tal manera que comprende o tiene homología a uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR de uno o más elementos genéticos móviles para volver a la célula resistente.

En algunas modalidades, la secuencia del nucleótido objetivo se deriva de un gen asociado con resistencia a antibióticos y/o antimicrobianos. Como se usa en la presente, el término "antimicrobiano" se refiere a cualquier composición que elimina o inhibe el crecimiento o reproducción de los microorganismos. Se entiende que el término abarca los antibióticos (esto es, composiciones producidas por otros microorganismos), así como composiciones sintéticamente producidas. Los genes de resistencia antimicrobiana incluyen pero no se limitan a *bla_{tem}*, *bla_{rob}*, *bla_{shv}*, *aadB*, *aacC1*, *aacC2*, *aacC3*, *aacA4*, *mecA*, *vanA*, *vanH*, *vanX*, *satA*, *aacA-aphH*, *vat*, *vga*, *msrA* *sul*, y/o *int*. Los genes de resistencia antimicrobiana incluyen aquellos que se obtienen de especies bacterianas que incluyen pero no se limitan al género *Escherichia*, *Klebsiella*, *Pseudomonas*, *Proteus*, *Streptococcus*, *Staphylococcus*, *Enterococcus*, *Haemophilus*, y *Moraxella*. Los genes de resistencia antimicrobiana también incluyen aquellos que se obtienen de especies bacterianas que incluyen pero no se limitan a *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Enterococcus faecalis*, *Staphylococcus saprophyticus*, *Streptococcus pyogenes*, *Haemophilus influenzae*, y *Moraxella catarrhalis*. En algunas modalidades, los espaciadores CRISPR particulares y/o pseudoespaciadores CRISPR derivados de genes que codifican resistencia antimicrobiana se agregan dentro de una ubicación CRISPR de una célula receptora, bajo condiciones tales que se previene la transferencia de los genes de resistencia. Así, el riesgo de adquirir genes de resistencia antimicrobiana (esto es, marcadores) se reduce. En algunas modalidades, los objetivos también incluyen *vanR*, (esto es, resistencia a la vancomicina), *tetR* (esto es, resistencia a la tetraciclina), y/o factores de resistencia que proporcionan resistencia a la beta-lactamasa.

En algunas modalidades, la presente invención proporciona métodos que comprenden las etapas de: identificar uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR derivados de una célula que comprende uno o más genes de resistencia antimicrobiana o marcadores; y modificar la secuencia del espaciador CRISPR en una célula que no comprende o no expresa los genes de resistencia antimicrobianos o marcadores de tal manera que el espaciador CRISPR de la célula tiene homología a uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR contenidos en la célula que comprende uno o más genes de resistencia antimicrobianos o marcadores.

En todavía modalidades adicionales, la presente invención proporciona métodos para modular la adquisición de los marcadores de resistencia antimicrobianos en una célula que comprende las etapas de: identificar uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR derivados de una célula que comprende uno o más genes de resistencia antimicrobianos o marcadores; identificar una o más ubicaciones CRISPR en una célula que no comprende o no expresa los genes de resistencia antimicrobianos o marcadores; y modificar la secuencia del espaciador CRISPR en la célula que no comprende o no expresa los genes de resistencia antimicrobianos o marcadores de tal manera que el espaciador CRISPR y/o pseudoespaciadores CRISPR tienen homología al espaciador CRISPR contenido en la célula resistente a la transferencia de genes que conceden resistencia a uno o más antimicrobianos.

En algunas modalidades, la secuencia del nucleótido objetivo se deriva de al menos un gen asociado con los factores de virulencia. En algunas modalidades, los espaciadores CRISPR particulares y/o pseudoespaciadores CRISPR derivados de los genes que codifican los factores de virulencias se agregan dentro de una ubicación CRISPR bacteriana para proporcionar resistencia contra la transferencia de genes que conceden virulencia en la bacteria. En algunas modalidades, se dirigen los factores que comúnmente contribuyen a la virulencia microbiana (por ejemplo, en los patógenos), tales como toxinas, internalinas, hemolisinas y otros factores de virulencias.

La presente invención también proporciona métodos que comprenden las etapas de: identificar uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR derivados de una célula que comprende uno o más factores de virulencia; y modificar la secuencia del espaciador CRISPR en una célula que no comprende o no expresa los factores de virulencia o marcadores de tal manera que el espaciador CRISPR de la célula tiene homología a uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR contenidos en la célula que comprende uno o más factores de virulencia.

En todavía modalidades adicionales, la presente invención proporciona métodos para conceder resistencia a una célula contra uno o más factores de virulencia o marcadores que comprenden las etapas de: identificar un espaciador CRISPR y/o pseudoespaciador CRISPR derivado de uno o más factores de virulencia o marcadores; identificar una o más combinaciones cas-repetición CRISPR funcionales en una célula que es substancialmente sensible a uno o más factores de virulencia o marcadores; y producir por ingeniería una o más las ubicaciones CRISPR en la célula substancialmente sensible de tal manera que comprenden uno o más espaciadores CRISPR y/o pseudoespaciadores CRISPR de uno o más factores de virulencia o marcadores para volver a la célula resistente.

La presente invención abarca el uso de variantes, homólogos, derivados y fragmentos de los mismos, incluyendo variantes, homólogos, derivados y fragmentos de las ubicaciones CRISPR, espaciadores CRISPR, pseudoespaciadores CRISPR, genes o proteínas cas, repeticiones de CRISPR, combinaciones gen cas-repetición CRISPR funcionales y secuencias de ácido nucleico objetivo o productos de transcripción de los mismos.

El término "variante" se usa para significar un polipéptido que se presenta naturalmente o secuencias de nucleótidos las cuales difieren de una secuencia tipo natural.

El término "fragmento" indica que un polipéptido o secuencia de nucleótido comprende una fracción de una secuencia tipo natural. Puede comprender una o más secciones grandes contiguas de la secuencia o una pluralidad de las secciones pequeñas. La secuencia también puede comprender otros elementos de la secuencia, por ejemplo, puede ser cualquier proteína de fusión con otra proteína. Preferiblemente la secuencia comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural.

Preferiblemente, el fragmento conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, un espaciador CRISPR o pseudoespaciador CRISPR comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, un espaciador CRISPR conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, un gen *cas* comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, un gen *cas* conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, una proteína *cas* comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, una proteína *cas* conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, una repetición CRISPR comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, una repetición CRISPR conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de la actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, una combinación *cas*-repetición CRISPR funcional comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, la combinación *cas*-repetición CRISPR funcional conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de la actividad del polipéptido tipo natural o secuencia de nucleótido.

Preferiblemente, una secuencia de ácido nucleico objetivo comprende al menos 50%, más preferiblemente al menos 65%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, más preferiblemente al menos 95%, más preferiblemente al menos 96%, más preferiblemente al menos 97%, más preferiblemente al menos 98%, más preferiblemente al menos 99% de la secuencia tipo natural. Preferiblemente, una secuencia de ácido nucleico objetivo conserva 50%, más preferiblemente 60%, más preferiblemente 70%, más preferiblemente 80%, más preferiblemente 85%, más preferiblemente 90%, más preferiblemente 95%, más preferiblemente 96%, más preferiblemente 97%, más preferiblemente 98%, o más preferiblemente 99% de actividad del polipéptido tipo natural o secuencia de nucleótido.

En algunas modalidades, el fragmento es un fragmento funcional. Por “fragmento funcional” de una molécula se entiende que un fragmento que retiene o posee sustancialmente la misma actividad biológica que la molécula intacta. En todos los casos, un fragmento funcional de una molécula conserva al menos 10% y al menos alrededor de 25%, alrededor de 50%, alrededor de 75%, alrededor de 80%, alrededor de 85%, alrededor de 90%, alrededor de 95%, alrededor de 96%, alrededor de 97%, alrededor de 98%, o alrededor de 99% de la actividad biológica de la molécula intacta.

El término “homólogo” significa una entidad que tiene cierta homología con las secuencias de aminoácidos objeto y las secuencias de nucleótidos objeto. En la presente, el término “homología” puede equipararse con la “identidad”.

En el contexto presente, una secuencia homóloga se toma para incluir una secuencia de aminoácido, la cual puede ser al menos idéntica al 75, 85 ó 90%, preferiblemente al menos idéntica al 95%, 96%, 97%, 98% ó 99% a la secuencia objeto. Aunque la homología puede también considerarse en los términos de la similitud (esto es, residuos de aminoácidos que tienen propiedades químicas/funciones similares), en el contexto de la presente invención se prefiere para expresar la homología en los términos de la identidad de secuencia.

En el contexto presente, una secuencia homóloga se toma para incluir una secuencia de nucleótido, la cual puede ser al menos idéntica al 75, 85 ó 90%, preferiblemente al menos idéntica al 95%, 96%, 97%, 98% ó 99% a la secuencia objeto. Aunque la homología puede también considerarse en los términos de la similitud (esto es, residuos de aminoácidos que tienen propiedades químicas/funciones similares), en el contexto de la presente invención se prefieren para expresar homología en los términos de la identidad de secuencia.

Las comparaciones de homología pueden realizarse por el ojo, o más usualmente, con la ayuda de programas de comparación de secuencia comercialmente disponibles. Estos programas de ordenador comercialmente disponibles pueden calcular el % de homología entre dos o más secuencias.

5 El porcentaje (%) de homología puede calcularse sobre las secuencias contiguas (esto es, una secuencia se liga con otra secuencia y cada aminoácido en una secuencia se compara directamente con el aminoácido correspondiente en la otra secuencia, un residuo cada vez). Esto se llama una alineación "sin espacios". Típicamente, tales alineaciones sin espacios se realizan solamente sobre un número relativamente corto de residuos.

10 Aunque este es un método muy simple y consistente, no tiene en cuenta que, por ejemplo, en un par de secuencias que de otro modo sería idéntico, una inserción o eliminación causará que se pongan los siguientes residuos de aminoácidos fuera de alineación, así potencialmente resultando en una reducción grande en el % de homología cuando una alineación global se realiza. Posteriormente, más métodos de comparación de secuencia se diseñaron para producir las alineaciones óptimas que se toman en las inserciones y eliminaciones posibles en consideración
15 sin penalizar indebidamente el registro de homología general. Esto se realiza por insertar los "espacios" en la alineación de secuencia para intentar de maximizar la homología local.

Sin embargo, estos métodos más complejos asignan "penalizaciones del espacio" a cada espacio que se presenta en la alineación de manera que, para el mismo número de aminoácidos idénticos, una secuencia se alinea con el
20 menos número de espacios posible, reflejando mayor relación entre las dos secuencias comparadas, que lograrán un mayor registro que uno con muchos espacios. "Costos de espacios continuos" se usa típicamente de manera que carga un costo relativamente alto para la existencia de un espacio y una penalización más pequeña para cada residuo posterior en el espacio. Este es el sistema de registro de espacio más comúnmente usado. La alta penalización de espacios naturalmente producirá alineaciones optimizadas con un número menor de espacios. La
25 mayoría de los programas de alineación permiten que las penalizaciones del espacio se modifiquen. Sin embargo, se prefieren usar los valores por omisión cuando se usa un software de este tipo para las comparaciones de secuencia. Por ejemplo, cuando se usa el paquete GCG Wisconsin Bestfit, la penalización del espacio por omisión para las secuencias de aminoácidos es -12 para un espacio y -4 para cada extensión.

30 El cálculo del % de homología máximo, por lo tanto, en primer lugar requiere la producción de una alineación óptima, tomando en cuenta las penalizaciones del espacio. Un programa de ordenador adecuado para realizar tal alineación es el paquete GCG Wisconsin Bestfit (University of Wisconsin, U.S.A.; Devereux et al., 1984, Nucleic Acids Research 12:387). Los ejemplos de otros softwares que pueden realizar comparaciones de secuencia incluyen pero no se limitan a, el paquete BLAST (ver Ausubel et al., 1999 ibid - Capítulo 18), FASTA (Atschul et al., 1990, J. Mol.
35 Biol., 403-410), la serie GENEWORKS de herramientas de comparación y CLUSTAL. Tanto BLAST como FASTA están disponibles para búsqueda en línea y fuera de línea (ver Ausubel et al., 1999 ibid, páginas 7-58 hasta 7-60). Sin embargo, para algunas aplicaciones, se prefiere usar el programa GCG Bestfit. Una nueva herramienta, llamada Secuencias BLAST 2 también está disponible para comparar secuencias de proteínas y nucleótidos (ver FEMS Microbiol Lett 1999 174(2): 247-50; FEMS Microbiol Lett 1999 177(1): 187-8).

40 Aunque el % de homología final puede medirse en los términos de identidad, los procesos de alineación en sí no se basan típicamente en una comparación de pares todo o nada. Al contrario, una matriz de registro de similitud en escala se usa generalmente de modo que asigna registro a cada comparación de pares basándose en la similitud química o distancia evolutiva. Un ejemplo de tal matriz comúnmente usado es la matriz BLOSUM62, la matriz de
45 omisión para la serie BLAST de programas. Los programas GCG Wisconsin generalmente usan ya sea los valores de omisión públicos o una tabla de comparación de símbolo personalizado si se suministra (ver manual de usuario para detalles adicionales). Para algunas aplicaciones, se prefiere el uso de los valores de omisión públicos para el paquete GCG, o en el caso de otro software, la matriz de omisión tal como BLOSUM62.

50 Una vez que el software ha producido una alineación óptima, es posible calcular el % de homología, preferiblemente el % de identidad de secuencia. El software típicamente lo hace como parte de la comparación de secuencia y genera un resultado numérico.

Las penalizaciones del espacio deberán usarse cuando se determina la identidad de secuencia, a continuación se usan adecuadamente los siguientes parámetros:

Para BLAST			
Abertura del espacio		0	
Extensión del espacio		0	
Para CLUSTAL		ADN	PROTEÍNA
Tamaño de la palabra	2	1	Triple K
Penalización del espacio	10	10	
Extensión del espacio	0.1	0.1	

Para la comparación de secuencia del polipéptido, los siguientes ajustes pueden usarse: la penalización de la creación del ESPACIO de 3.0 y la penalización de la extensión del ESPACIO de 0.1. Adecuadamente, el grado de identidad con respecto a una secuencia de aminoácido se determina sobre al menos 5 aminoácidos contiguos, se determina sobre al menos 10 aminoácidos contiguos, sobre al menos 15 aminoácidos contiguos, sobre al menos 20 aminoácidos contiguos, sobre al menos 30 aminoácidos contiguos, sobre al menos 40 aminoácidos contiguos, sobre al menos 50 aminoácidos contiguos, o sobre al menos 60 aminoácidos contiguos.

Las secuencias pueden también tener eliminaciones, inserciones o sustituciones de los residuos de aminoácidos, los cuales producen un cambio silencioso y resultan en una sustancia funcionalmente equivalente. Las sustituciones de aminoácidos deliberadas pueden hacerse con base a la similitud en la polaridad, carga, solubilidad, hidrofobicidad, hidrofiliidad y/o la naturaleza anfipática de los residuos siempre y cuando la actividad de enlace secundaria de la sustancia se mantenga. Por ejemplo, los aminoácidos cargados negativamente incluyen el ácido aspártico y ácido glutámico; los aminoácidos cargados positivamente incluyen lisina y arginina; y los aminoácidos con grupos de cadena polar no cargados tienen valores de hidrofiliidad similares que incluyen leucina, isoleucina, valina, glicina, alanina, asparagina, glutamina, serina, treonina, fenilalanina, y tirosina.

Las sustituciones conservadoras pueden hacerse, por ejemplo, de acuerdo a la tabla a continuación. Los aminoácidos en el mismo bloque en la segunda columna y preferiblemente en la misma línea en la tercera columna pueden substituirse cada uno por el otro

ALIFÁTICO	No polar	ESPACIO
		I L V
	Polar-no cargado	C S T M
		N Q
	Polar-cargado	D E
AROMÁTICO		K R
		H F W Y

La presente invención también abarca la sustitución homóloga (sustitución y reemplazo se usan ambos en la presente para significar el intercambio de un residuo de aminoácido existente, con un residuo alternativo) que puede presentarse, esto es, la sustitución bajo condiciones iguales tales como básica para básica, ácida para ácida, polar para polar, etc. La sustitución no homóloga también puede presentarse esto es, de una clase del residuo a otro o involucrando alternativamente la inclusión de los aminoácidos no naturales tales como ornitina (en la presente de aquí en adelante se refiere como Z), ornitina de ácido diaminobutírico (en la presente de aquí en adelante se refiere como B), ornitina de norleucina (en la presente de aquí en adelante se refiere como O), piriilalanina, tienilalanina, naftilalanina y fenilglicina.

Los reemplazos pueden también hacerse por los aminoácidos no naturales que incluyen; aminoácidos alfa* y alfa-disustituidos*, aminoácidos N-alquilo*, ácido láctico*, haluros derivados de aminoácidos naturales tales como trifluorotirosina*, p-Cl-fenilalanina*, p-Br-fenilalanina*, p-I-fenilalanina*, L-alil-glicina*, β -alanina*, ácido L- α -amino butírico*, ácido L- γ -amino butírico*, ácido L- α -amino isobutírico*, ácido L- ϵ -amino caproico#, ácido 7-amino heptanoico*, L-metionina sulfona#, L-norleucina*, L-norvalina*, p-nitro-L-fenilalanina*, L-hidroxirolina#, L-tioprolina*, derivados de metilo de la fenilalanina (Fe) - tales como 4-metil-Fe*, pentametil-Fe*, L-Fe (4-amino)#, L-Tyr (metilo)*, L-Fe (4-isopropilo)*, ácido L-Tic (1,2,3,4-tetrahidroisoquinolina-3-carboxílico)*, ácido L-diaminopropionico # y L-Fe (4-bencilo)*. La anotación * se utiliza para el propósito de la discusión de arriba (refiriéndose a la sustitución homóloga o no homóloga), para indicar la naturaleza hidrofóbica del derivado mientras que # se utiliza para indica la naturaleza hidrofílica del derivado, #* indica las características anfipáticas.

Las secuencias de aminoácidos variantes incluyen los grupos espaciadores adecuados que son adecuados para la inserción insertada entre cualquiera de dos residuos de aminoácidos de la secuencia que incluye los grupos alquilo tales como los grupos metilo, etilo o propilo en adición a los aminoácidos espaciadores tales como los residuos de glicina o β -alanina. Una forma adicional de la variación involucra la presencia de uno o más residuos de aminoácidos en la forma de peptido, que se entenderá por aquellos de habilidad en el arte. Para evitar dudas, "la forma de peptido" se usa para referirse a residuos de aminoácidos variantes en donde el grupo substituyente de α -carbono está en los residuos del átomo de nitrógeno en lugar del α -carbono. Los procesos para preparar los péptidos en la forma de peptido son bien conocidos en el arte.

Las secuencias de nucleótidos para uso en la presente invención pueden incluirse dentro de los nucleótidos sintéticos o modificados. Un número de diferentes tipos de modificación a los oligonucleótidos es conocido en el arte. Estos incluyen estructuras de metilfosfonato y fosforotioato y/o la adición de acridina o cadenas de polilisina en las terminales 3' y/o 5' de la molécula. Para los propósitos de la presente invención, se entenderá que las secuencias de nucleótidos pueden modificarse por cualquiera de los métodos disponibles en el arte. Tales modificaciones pueden realizarse para aumentar la actividad in vivo o vida útil de las secuencias de nucleótidos útiles en la presente invención.

CRISPRs

Los CRISPR (Repeticiones palindrómicas cortas interespaciadas regularmente agrupadas); también conocidas como SPIDR (Repeticiones directas intercaladas del espaciador) constituyen una familia de las ubicaciones de ADN recientemente descritas que son usualmente específicas para una especie bacteriana particular. La ubicación CRISPR es una clase distinta de las repeticiones de las secuencias cortas intercaladas (SSRs) que se reconocieron primero en *E. coli* (Ishino et al, J. Bacteriol., 169:5429-5433 [1987]; y Nakata et al., J. Bacteriol., 171:3553-3556 [1989]). Las SSR intercaladas similares se identifican en *Haloflex mediterranei*, *Streptococcus pyogenes*, *Anabaena*, y *Mycobacterium tuberculosis* (ver, Groenen et al., Mol. Microbiol., 10: 1057-1065 [1993]; Hoe et al, Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al, Biochim. Biophys. Acta 1307:26-30 [1996]; y Mojica et al, Mol. Microbiol., 17:85-93 [1995]). Las ubicaciones CRISPR difieren de otras SSRs por la estructura de las repeticiones, la cuales se denominan repeticiones cortas regularmente espaciadas (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; y Mojica et al, Mol. Microbiol., 36:244-246 [2000]). Las repeticiones son elementos cortos que se presentan en las agrupaciones que son siempre regularmente espaciadas por secuencias de intervención únicas con una longitud constante (Mojica et al, [2000], supra). Aunque las secuencias repetidas se conservan altamente entre las cepas, el número de las repeticiones interespaciadas y las secuencias de las regiones del espaciador difieren de cepa a cepa (van Embden et al, J. Bacteriol., 182:2393-2401 [2000]).

Las ubicaciones CRISPR consisten en repeticiones de ADN parcialmente palindrómicas altamente conservadas típicamente de 24 a 40 pb, que contienen repeticiones invertidas terminales e interiores de hasta 11 pb. Se ha reportado que estas repeticiones se presentan desde 1 hasta 140 veces. Aunque los elementos aislados se han detectado, estos generalmente se configuran en grupos (hasta alrededor de 20 o más por genoma) de unidades repetidas espaciadas por una intervención única de secuencias de 20-58 pb. Hasta la fecha, hasta 20 distintas ubicaciones CRISPR se han encontrado dentro de un cromosoma sencillo.

Las CRISPRs son generalmente homogéneas dentro de un genoma dado siendo la mayoría de ellos idénticos. Sin embargo, hay ejemplos de heterogeneidad en, por ejemplo, las arqueobacterias (Mojica et al, [2000], supra).

Como se usa en la presente, el término "ubicación CRISPR" se refiere al segmento de ADN el cual incluye todas las repeticiones de CRISPR, partiendo con el primer nucleótido de la primera repetición CRISPR y que termina con el último nucleótido de la última repetición CRISPR (terminal).

Aunque la función biológica de las ubicaciones CRISPR es desconocida, algunas hipótesis se han propuesto. Por ejemplo, se propone que estas pueden estar involucradas en el enlace del cromosoma a una estructura celular, o en la replicación del cromosoma y división del replicón (Jansen et al., OMICS 6:23-33 [2002]; Jansen et al, Mol. Microbiol., 43: 1565-1575 [2002]; y Pourcel et al., Microbiol., 151:653-663 [2005]). Mojica et al. (Mojica et al., J. Mol. Evol., 60: 174-182 [2005]) tiene la hipótesis de que CRISPR puede involucrarse en conceder la inmunidad específica contra el ADN externo y Pourcel et al. (supra), la hipótesis de que las CRISPR son estructuras que son capaces de ocuparse de las piezas del ADN externo como parte de un mecanismo de defensa. Bolotin et al. (supra) sugiere que los elementos del espaciador CRISPR son los rastros de las invasiones pasadas por los elementos extracromosomales, y tiene la hipótesis de que se proporciona una célula con inmunidad contra la infección fago y más generalmente la expresión de ADN exterior, al codificar un ARN anti-sentido. Bolotin et al. (supra) también sugiere que los genes *cas* son necesarios para la formación de CRISPR. Sin embargo, no se pretende que la presente invención se limite por cualquier mecanismo, función, teoría, ni medios de acción particulares.

El genoma de *Streptococcus thermophilus* LMG18311 contiene 3 ubicaciones CRISPR; los 36 pb de las secuencias repetidas son diferentes en el CRISPR1 (34 repeticiones), CRISPR2 (5 repeticiones), y CRISPR3 (una secuencia sencilla). Sin embargo, están perfectamente conservadas dentro de cada ubicación. Las repeticiones de CRISPR1 y CRISPR2 son respectivamente interespaciadas por 33 y 4 secuencias de 30 pb en longitud. Todas estas secuencias interespaciadoras son diferentes una de la otra. Estas también son diferentes de aquellas encontradas en la cepa CNRZ1066 (41 secuencias interespaciadoras dentro de CRISPR1) y en la cepa LMD-9 (16 dentro del CRISPR1 y 8 dentro del CRISPR3), que ambos son *S. thermophilus*.

Diversos métodos para identificar las ubicaciones CRISPR se conocen en el arte. Por ejemplo, Jensen et al. (Jensen et al., [2002], supra) describen un enfoque basado en el ordenador en el cual las secuencias de nucleótidos se buscan para las porciones CRISPR usando el programa PATSCAN en el servidor de Mathematics and Computer Science Division en el Argonne National Laboratory, Argonne, IL, EE.UU. El algoritmo que se usó para identificar las porciones CRISPR fue $pl = a...b\ c...d\ plc...d\ plc...dpl$, donde *a* y *b* fueron el límite de tamaño inferior y superior de la repetición y *pl* y *c* y *d* fueron el límite del tamaño inferior y superior de las secuencias del espaciador. Los valores de *a*, *b*, *c* y *d* pueden variar desde alrededor de 15 hasta alrededor de 70 pb en incrementos desde alrededor de 5 pb. En algunas modalidades preferidas, las ubicaciones CRISPR se identifican usando el dotplots (por ejemplo, al usar el programa de ordenador Dotter).

Cualquier método adecuado conocido en el arte encuentra uso en analizar la similitud de la secuencia. Por ejemplo, el análisis puede realizarse usando el NCBI BLAST con una base de datos de los genomas microbianos y el GenBank, como se conoce en el arte. Además, las secuencias de nucleótidos, incluyen aquellas proporcionadas en

la presente, se incluyen en las bases de datos (por ejemplo, GenBank o la página web del genoma JGI). Como se usa en la presente, “corriente arriba” significa en la dirección 5' y “corriente abajo” significa en la dirección 3'.

En las modalidades adicionales, los métodos de la presente invención utilizan los procedimientos de amplificación (ver por ejemplo, Mojica et al., [2005], supra; y Pourcel et al., [2005], supra). La amplificación de la región deseada del ADN puede realizarse por cualquier método conocido en el arte, incluyendo la reacción de la cadena de polimerasa (PCR). La “Amplificación” se refiere a la producción de las copias adicionales de una secuencia de ácido nucleico. Esta se realiza generalmente usando las tecnologías PCR bien conocidas en el arte. La “reacción de la cadena de polimerasa” (“PCR”) es bien conocida por aquellos de habilidad en el arte. En la presente invención, los cebadores de oligonucleótidos se diseñan para el uso en las reacciones PCR para amplificar todo o parte de una ubicación CRISPR.

El término “cebador” se refiere a un oligonucleótido, ya sea si se presenta naturalmente como en una restricción purificada digerida o producida sintéticamente, la cual es capaz de actuar como un punto de iniciación de la síntesis cuando se coloca bajo condiciones en las cuales la síntesis de un producto de extensión del cebador el cual es complementario a una hebra de ácido nucleico se induce (esto es, en la presencia de nucleótidos y un agente inductor tal como polimerasa de ADN y en una temperatura adecuada y pH). En algunas modalidades, el cebador es la hebra sencilla para máxima eficacia en la amplificación, aunque en otras modalidades, el cebador es la hebra doble. En algunas modalidades, el cebador es un oligodesoxirribonucleótido. El cebador debe ser lo suficientemente largo para la síntesis de los productos de extensión en la presencia del agente inductor. La longitud exacta de los cebadores depende de muchos factores, incluyendo la temperatura, fuente del cebador, y el uso del método. Los cebadores PCR son típicamente al menos alrededor de 10 nucleótidos en longitud, y más típicamente al menos alrededor de 20 nucleótidos en longitud. Los métodos para el diseño y realización del PCR son bien conocidos en el arte, e incluyen pero no se limitan a métodos que usan cebadores pares, cebadores anidados, cebadores específicos sencillos, cebadores degenerados, cebadores específicos del gen, cebadores específicos del vector, cebadores parcialmente mal alineados, etc.

En algunas modalidades preferidas de la presente invención, una ubicación CRISPR o una porción de la misma de una bacteria precursora y una bacteria etiquetada se comparan usando cualquier método adecuado conocido en el arte. En algunas modalidades preferidas de la presente invención, la ubicación CRISPR o una porción de la misma de la bacteria precursora y la bacteria etiquetada se comparan al amplificar la ubicación CRISPR o una porción de la misma. Además, los métodos de amplificación cíclicos bien conocidos (por ejemplo, PCR, reacción de cadena de ligasa, etc.), otros métodos, que incluyen pero no se limitan a métodos de amplificación isotérmica que encuentran uso en la presente invención. Los métodos de amplificación isotérmica bien conocidos que encuentran uso en la presente invención incluyen pero no se limitan a amplificación de desplazamiento de hebra (SDA), Q-beta-replicasa, amplificación de secuencia basada en ácido nucleico (NASBA), y replicación de la secuencia auto-sostenida.

En algunas otras modalidades preferidas de la presente invención, la ubicación CRISPR o una porción de la misma de la bacteria precursora y la bacteria etiquetada se comparan por el proceso de secuencia de la presente invención, la ubicación CRISPR o una porción de la misma de la bacteria precursora y la bacteria etiquetada se comparan al amplificar y luego procesar por secuencia las ubicaciones CRISPR o una porción de la misma. En algunas modalidades, una terminal de las ubicaciones CRISPR se comparan, mientras en otras modalidades, ambos extremos 5' y 3' de las ubicaciones se comparan. En algunas modalidades preferidas, una terminal (por ejemplo, el extremo 5') de las ubicaciones CRISPR se comparan. En aún otras modalidades, al menos la última repetición CRISPR en el extremo 3' de la ubicación CRISPR y/o al menos el último espaciador CRISPR (por ejemplo, el último núcleo del espaciador CRISPR) en el extremo 3' de la ubicación CRISPR y/o al menos la primera repetición CRISPR en el extremo 5' de la ubicación CRISPR y/o al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' de la ubicación CRISPR se comparan. En algunas modalidades preferidas, al menos la primera repetición CRISPR en el extremo 5' de la ubicación CRISPR y/o al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' de la ubicación CRISPR se comparan. En algunas modalidades adicionales preferidas, al menos el último espaciador CRISPR (por ejemplo, el último núcleo del espaciador CRISPR) en el extremo 3' de la ubicación CRISPR y/o al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' de la ubicación CRISPR se comparan. En algunas otras modalidades preferidas, al menos el primer espaciador CRISPR (por ejemplo, el primer núcleo del espaciador CRISPR) en el extremo 5' de las ubicaciones CRISPR se comparan.

En algunas modalidades, las ubicaciones CRISPR comprenden el ADN, mientras en otras modalidades, las ubicaciones CRISPR comprenden el ARN. En algunas modalidades, el ácido nucleico es de origen genómico, mientras en otras modalidades, es de origen sintético o recombinante. En algunas modalidades, las ubicaciones CRISPR son de hebra doble, mientras en otras modalidades, son de hebra sencilla, si se representa la hebra sentido o antisentido o combinaciones de las mismas. En algunas modalidades, las ubicaciones CRISPR se preparan por el uso de las técnicas de ADN recombinante (por ejemplo, ADN recombinante), como se describe en la presente.

La presente invención también proporciona métodos para generar las variantes CRISPR. Estas variantes se expresan, aíslan, clonan y/o procesan por secuencia usando cualquier método adecuado conocido en el arte. Las variantes CRISPR son cepas mutantes resistentes al fago que tienen una ubicación CRISPR modificada con un

espaciador adicional. Estas variantes encuentran uso como propósitos de detección/identificación, o por la resistencia producida por ingeniería contra las moléculas de ácido nucleico. Estas variantes encuentran uso en el desarrollo de los agentes de biocontrol.

5 En el contexto de la presente invención, la ubicación CRISPR se orienta como se describe a continuación. El CRISPR líder es un segmento de ADN conservado de tamaño definido. La orientación de la ubicación CRISPR1 *S. thermophilus* se establece usando las siguientes características:

10 La posición relativa del CRISPR a los genes *cas* cercanos (Secuencias asociadas con CRISPR); el CRISPR1 se localiza corriente abajo de 4 genes *cas* (genes str0657, str0658, str0659, y str0660 dentro de la secuencia del cromosoma CNRZ1066);

15 La secuencia repetida tiene el potencial para formar una estructura secundaria de horquilla, aunque no es completamente palindrómica, y la secuencia complementaria inversa;

20 (5'-GTTGTACAGTTACTTAAATCTTGAGAGTACAAAAAC-3'; SEQ ID NO:695) es diferente de la secuencia directa (5'-GTTTTTGTACTCTCAAGATTTAAGTAAGTGTACAAC-3'; SEQ ID NO: 1). Generalmente, el extremo 5' de la secuencia directa es más rico en los nucleótidos G y T del extremo 5' de la secuencia complementaria inversa. Además, como el emparejamiento base G-T es mejor que el emparejamiento base A-C, la estructura de la horquilla es generalmente más fuerte en la hebra directa; y

Como se usa en la presente, la posición de la terminal repetida es la terminal repetida la cual muestra variación en la secuencia en su extremo 3' es generalmente la repetición de terminal.

25 El CRISPR líder es un segmento de ADN conservado de tamaño definido el cual se localiza inmediatamente en la corriente arriba de la primera repetición. Por ejemplo, la secuencia líder del CRISPR1 *S. thermophilus* es el segmento de ADN partiendo inmediatamente después del codón de detección del gen str0660, y finalizando justo antes de la primera repetición. El CRISPR líder se localiza en el extremo 5' de la ubicación CRISPR. El CRISPR líder se localiza inmediatamente corriente arriba de la primera repetición CRISPR de la ubicación CRISPR.

30 El remolque de CRISPR es un segmento de ADN conservado de tamaño definido, el cual se localiza inmediatamente corriente abajo de la terminal repetida. Por ejemplo, la secuencia de remolque del CRISPR1 *S. thermophilus* es el segmento de ADN partiendo inmediatamente después de la terminal repetida, y finalizando justo antes del codón de detención del gen str0661 (localizado en la hebra de ADN opuesta). El remolque CRISPR se localiza en el extremo 3' de la ubicación CRISPR. El remolque de CRISPR se localiza inmediatamente en la corriente abajo de la terminal repetida.

Por ejemplo, el CRISPR líder y las secuencias de remolque de CRISPR en la ubicación CRISPR1 de la cepa *Streptococcus thermophilus* CNRZ1066 son:

40 CRISPR líder: 5'-CAAGGACAGTTATTGATTTTATAATCACTATGTGGGTATAAAAACGTCAAATTTTCATTTGA G-3' (SEQ ID NO:688)

CRISPR remolque:

45 5'-TTGATTCAACATAAAAAGCCAGTTCAATTGAACTTGGCTTT-3' (SEQ ID NO:691)

El CRISPR líder corresponde a las posiciones 625038 hasta 625100, y el CRISPR remolque corresponde a las posiciones 627845 hasta 627885 en el genoma completo (CP000024) de *S. thermophilus*.

50 Como se usa en la presente, el término "corriente arriba" significa la dirección 5' y "corriente abajo" significa la dirección 3'. Como se usa en la presente el término "porción del mismo" en el conexto de una ubicación CRISPR significa al menos alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos o aún alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) de una ubicación CRISPR. En algunas modalidades preferidas, el término "porción del mismo" significa al menos alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos o alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) de una o ambas terminaciones (esto es, las terminaciones 5' y/o 3') de una ubicación CRISPR. En algunas modalidades preferidas, el término "porción del mismo" se refiere a al menos alrededor de los primeros 44 nucleótidos en el extremo 5' de una ubicación CRISPR o alrededor de los últimos 44 nucleótidos en el extremo 3' de una ubicación CRISPR.

65

En algunas modalidades adicionales, el término “porción del mismo” en el contexto de una ubicación CRISPR significa al menos el primero de alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos, o alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) en corriente abajo del primer nucleótido de la primera repetición CRISPR en el extremo 5' de una ubicación CRISPR o corriente arriba del último nucleótido de la última repetición CRISPR en el extremo 3' de una ubicación CRISPR. En algunas modalidades preferidas, el término “porción del mismo” se refiere a al menos alrededor de los primeros 44 nucleótidos en corriente abajo del primer nucleótido de la primera repetición CRISPR en el extremo 5' de una ubicación CRISPR o al menos alrededor de 44 nucleótidos en corriente arriba del último nucleótido de la última repetición CRISPR en el extremo 3' de una ubicación CRISPR.

En algunas modalidades, el tamaño mínimo de la secuencia duplicada es alrededor de 24 nucleótidos y el tamaño mínimo de la secuencia de marcado es alrededor de 20 nucleótidos. Así, en algunas modalidades preferidas, el término “porción del mismo” en el contexto de una ubicación CRISPR, significa al menos 44 nucleótidos.

En algunas modalidades, el tamaño máximo de la secuencia duplicada es alrededor de 40 nucleótidos y el tamaño máximo de la secuencia de marcado es alrededor de 58 nucleótidos. Así, en algunas modalidades, el término “porción del mismo” cuando se usa en el contexto de una ubicación CRISPR significa al menos alrededor de 98 nucleótidos. En algunas modalidades preferidas, el término “porción del mismo” en el contexto de una ubicación CRISPR significa al menos alrededor de 44-98 nucleótidos.

Cuando se compara la ubicación CRISPR o una porción de la misma de la bacteria precursora y una bacteria etiquetada, al menos alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos, o alrededor de 100 nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) de una ubicación CRISPR se comparan. En algunas modalidades preferidas, al menos alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos, o alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) en uno o ambos extremos de una ubicación CRISPR se comparan.

En algunas modalidades preferidas, al menos el primero de alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos o alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) en el extremo 5' de una ubicación CRISPR o en el extremo 3' de una ubicación CRISPR se comparan. En algunas modalidades preferidas, al menos el primero de alrededor de 44 nucleótidos en el extremo 5' de una ubicación CRISPR o el último de alrededor de 44 nucleótidos en el extremo 3' de una ubicación CRISPR se comparan.

En algunas modalidades, al menos el primero de alrededor de 10 nucleótidos, alrededor de 20 nucleótidos, alrededor de 24 nucleótidos, alrededor de 30 nucleótidos, alrededor de 40 nucleótidos, alrededor de 44 nucleótidos, alrededor de 50 nucleótidos, alrededor de 60 nucleótidos, alrededor de 70 nucleótidos, alrededor de 80 nucleótidos, alrededor de 90 nucleótidos, alrededor de 98 nucleótidos, o alrededor de 100 o más nucleótidos (por ejemplo, al menos alrededor de 44-98 nucleótidos) en corriente abajo del primer nucleótido de la primera repetición CRISPR en el extremo 5' de una ubicación CRISPR o corriente arriba del último nucleótido de la última repetición CRISPR en el extremo 3' de una ubicación CRISPR se comparan. En algunas modalidades preferidas, al menos alrededor de los primeros 44 nucleótidos en corriente abajo del primer nucleótido de la primera repetición CRISPR en el extremo 5' de una ubicación CRISPR o alrededor de al menos 44 nucleótidos en corriente arriba del último nucleótido de la última repetición CRISPR en el extremo 3' de una ubicación CRISPR se comparan.

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

Como se usa en la presente, el término “Repetición CRISPR” tiene el significado convencional como se usa en el arte (esto es, repeticiones directas cortas múltiples, las cuales no muestran o tienen muy poca variación de secuencia dentro de una ubicación CRISPR dada). Como se usa en la presente, en el contexto, “Repetición CRISPR” es sinónimo del término “CRISPR”.

Una ubicación CRISPR comprende una o más repeticiones CRISPR que existen en los espaciadores CRISPR. Así, la repetición CRISPR corresponde a la secuencia repetida dentro de una ubicación CRISPR. Por ejemplo, excepto por la terminal repetida, la secuencia repetida típica de la secuencia CRISPR1 *S. thermophilus* es:

5 5'-GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC-3' (SEQ ID NO: 1)

Puntos de variaciones de esta secuencia repetida se observan pero son muy raros. En comparación a esta secuencia de repetición típica, la secuencia repetida terminal siempre muestra la misma variación en el extremo 3'. Los puntos de variaciones de esta secuencia repetida terminal también se observan pero son muy raros. Las repeticiones de CRISPR pueden presentarse de forma natural en la bacteria precursora. Los números de acceso al GenBank de las secuencias 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.

Como se describe en detalle adicional en la presente, una secuencia duplicada se deriva, puede derivarse, se obtiene o puede obtenerse de una bacteria precursora. En algunas modalidades preferidas, la secuencia comprende el ADN genómico de una bacteria precursora. En algunas modalidades particularmente preferidas, la repetición CRISPR duplicada (por ejemplo, en la misma ubicación CRISPR) se integra iterativamente, secuencialmente, simultáneamente o substancialmente simultáneamente junto con la secuencia de marcado en la bacteria precursora para dar una bacteria etiquetada.

El número de los nucleótidos en una repetición es generalmente alrededor de 20 hasta alrededor de 40 pares base (por ejemplo, 36 pares base), pero en otras modalidades es alrededor de 20 hasta alrededor de 39 pares base, alrededor de 20 hasta alrededor de 37 pares base, alrededor de 20 hasta alrededor de 35 pares base, alrededor de 20 hasta alrededor de 33 pares base, alrededor de 20 hasta alrededor de 30 pares base, alrededor de 21 hasta alrededor de 40 pares base, alrededor de 21 hasta alrededor de 39 pares base, alrededor de 21 hasta alrededor de 37 pares base, alrededor de 23 hasta alrededor de 40 pares base, alrededor de 23 hasta alrededor de 39 pares base, alrededor de 23 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 40 pares base, alrededor de 25 hasta alrededor de 39 pares base, alrededor de 25 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 35 pares base, o alrededor de 28 ó 29 pares base.

El número de nucleótidos en una repetición es generalmente alrededor de 20 hasta alrededor de 40 pares base, pero puede ser alrededor de 20 hasta alrededor de 39 pares base, alrededor de 20 hasta alrededor de 37 pares base, alrededor de 20 hasta alrededor de 35 pares base, alrededor de 20 hasta alrededor de 33 pares base, alrededor de 20 hasta alrededor de 30 pares base, alrededor de 21 hasta alrededor de 40 pares base, alrededor de 21 hasta alrededor de 39 pares base, alrededor de 21 hasta alrededor de 37 pares base, alrededor de 23 hasta alrededor de 40 pares base, alrededor de 23 hasta alrededor de 39 pares base, alrededor de 23 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 40 pares base, alrededor de 25 hasta alrededor de 39 pares base, alrededor de 25 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 35 pares base, or alrededor de 28 ó 29 pares base. El número de repeticiones puede estar en el intervalo de alrededor de 1 hasta alrededor de 140, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 5 hasta alrededor de 100, desde alrededor de 10 hasta alrededor de 100, desde alrededor de 15 hasta alrededor de 100, desde alrededor de 20 hasta alrededor de 100, desde alrededor de 25 hasta alrededor de 100, desde alrededor de 30 hasta alrededor de 100, desde alrededor de 35 hasta alrededor de 100, desde alrededor de 40 hasta alrededor de 100, desde alrededor de 45 hasta alrededor de 100, desde alrededor de 50 hasta alrededor de 100, desde alrededor de 1 hasta alrededor de 135, desde alrededor de 1 hasta alrededor de 130, desde alrededor de 1 hasta alrededor de 125, desde alrededor de 1 hasta alrededor de 120, desde alrededor de 1 hasta alrededor de 115, desde alrededor de 1 hasta alrededor de 110, desde alrededor de 1 hasta alrededor de 105, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 1 hasta alrededor de 95, desde alrededor de 1 hasta alrededor de 90, desde alrededor de 1 hasta alrededor de 80, desde alrededor de 1 hasta alrededor de 70, desde alrededor de 1 hasta alrededor de 60, desde alrededor de 1 hasta alrededor de 50, desde alrededor de 10 hasta alrededor de 140, desde alrededor de 10 hasta alrededor de 130, desde alrededor de 10 hasta alrededor de 120, desde alrededor de 10 hasta alrededor de 110, desde alrededor de 10 hasta alrededor de 95, desde alrededor de 10 hasta alrededor de 90, desde alrededor de 20 hasta alrededor de 80, desde alrededor de 30 hasta alrededor de 70, desde alrededor de 30 hasta alrededor de 60, desde alrededor de 30 hasta alrededor de 50, desde alrededor de 30 hasta alrededor de 40, o alrededor de 32.

En algunas otras modalidades, el número de nucleótidos en una repetición es alrededor de 20 hasta alrededor de 39 pares base, alrededor de 20 hasta alrededor de 37 pares base, alrededor de 20 hasta alrededor de 35 pares base, alrededor de 20 hasta alrededor de 33 pares base, alrededor de 20 hasta alrededor de 30 pares base, alrededor de 21 hasta alrededor de 40 pares base, alrededor de 21 hasta alrededor de 39 pares base, alrededor de 21 hasta alrededor de 37 pares base, alrededor de 23 hasta alrededor de 40 pares base, alrededor de 23 hasta alrededor de 39 pares base, alrededor de 23 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 40 pares base, alrededor de 25 hasta alrededor de 39 pares base, alrededor de 25 hasta alrededor de 37 pares base, alrededor de 25 hasta alrededor de 35 pares base, o alrededor de 28 ó 29 pares base.

En algunas modalidades, el número de las repeticiones está en el intervalo desde alrededor de 1 hasta alrededor de 144, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 5 hasta alrededor de 100, desde alrededor de 10 hasta alrededor de 100, desde alrededor de 15 hasta alrededor de 100, desde alrededor de 20 hasta alrededor de 100, desde alrededor de 25 hasta alrededor de 100, desde alrededor de 30 hasta alrededor de 100, desde alrededor de 35 hasta alrededor de 100, desde alrededor de 40 hasta alrededor de 100, desde alrededor de 45 hasta alrededor de 100, desde alrededor de 50 hasta alrededor de 100, desde alrededor de 1 hasta alrededor de 135, desde alrededor de 1 hasta alrededor de 130, desde alrededor de 1 hasta alrededor de 125, desde alrededor de 1 hasta alrededor de 120, desde alrededor de 1 hasta alrededor de 115, desde alrededor de 1 hasta alrededor de 110, desde alrededor de 1 hasta alrededor de 105, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 1 hasta alrededor de 95, desde alrededor de 1 hasta alrededor de 90, desde alrededor de 1 hasta alrededor de 80, desde alrededor de 1 hasta alrededor de 70, desde alrededor de 1 hasta alrededor de 60, desde alrededor de 1 hasta alrededor de 50, desde alrededor de 10 hasta alrededor de 140, desde alrededor de 10 hasta alrededor de 130, desde alrededor de 10 hasta alrededor de 120, desde alrededor de 10 hasta alrededor de 110, desde alrededor de 10 hasta alrededor de 95, desde alrededor de 10 hasta alrededor de 90, desde alrededor de 20 hasta alrededor de 80, desde alrededor de 30 hasta alrededor de 70, desde alrededor de 30 hasta alrededor de 60, desde alrededor de 30 hasta alrededor de 50, desde alrededor de 30 hasta alrededor de 40, o alrededor de 30, 31, 32, 33, 34 ó 35 repeticiones.

En algunas modalidades, el número de las repeticiones está en el intervalo desde alrededor de 2 hasta alrededor de 140, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 5 hasta alrededor de 100, desde alrededor de 10 hasta alrededor de 100, desde alrededor de 15 hasta alrededor de 100, desde alrededor de 20 hasta alrededor de 100, desde alrededor de 25 hasta alrededor de 100, desde alrededor de 30 hasta alrededor de 100, desde alrededor de 35 hasta alrededor de 100, desde alrededor de 40 hasta alrededor de 100, desde alrededor de 45 hasta alrededor de 100, desde alrededor de 50 hasta alrededor de 100.

En algunas modalidades adicionales, el número de repeticiones está en el intervalo desde alrededor de 2 hasta alrededor de 135, desde alrededor de 2 hasta alrededor de 130, desde alrededor de 2 hasta alrededor de 125, desde alrededor de 2 hasta alrededor de 120, desde alrededor de 2 hasta alrededor de 115, desde alrededor de 2 hasta alrededor de 110, desde alrededor de 2 hasta alrededor de 105, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 2 hasta alrededor de 95, desde alrededor de 2 hasta alrededor de 90, desde alrededor de 2 hasta alrededor de 80, desde alrededor de 2 hasta alrededor de 70, desde alrededor de 2 hasta alrededor de 60, desde alrededor de 2 hasta alrededor de 50, desde alrededor de 2 hasta alrededor de 40, desde alrededor de 2 hasta alrededor de 30, desde alrededor de 2 hasta alrededor de 20, desde alrededor de 2 hasta alrededor de 10, desde alrededor de 2 hasta alrededor de 9, desde alrededor de 2 hasta alrededor de 8, desde alrededor de 2 hasta alrededor de 7, desde alrededor de 2 hasta alrededor de 6, desde alrededor de 2 hasta alrededor de 5, desde alrededor de 2 hasta alrededor de 4, o desde alrededor de 2 hasta alrededor de 3.

En algunas modalidades, las repeticiones de CRISPR comprenden ADN, mientras en otras modalidades, las repeticiones de CRISPR comprenden el ARN. En algunas modalidades, el ácido nucleico es de origen genómico, mientras en otras modalidades, es de origen sintético o recombinante. En algunas modalidades, los genes de repetición CRISPR son de hebra doble o hebra sencilla, se representan en la hebra sentido o antisentido o combinaciones de los mismos. En algunas modalidades, los genes de repetición CRISPR se preparan por el uso de las técnicas de ADN recombinante (por ejemplo, ADN recombinante), como se describe en la presente.

En algunas modalidades, una o más de las repeticiones de CRISPR se usan para producir por ingeniería una célula (por ejemplo, una célula receptora). En algunas modalidades preferidas, una o más, preferiblemente, dos o más repeticiones de CRISPR se usan para producir por ingeniería una célula (por ejemplo, una célula receptora), que en combinación con uno o más genes o proteínas *cas* y uno o más espaciadores CRISPR modulan la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo. Por ejemplo, en algunas modalidades, las repeticiones de CRISPR se insertan en el ADN de una célula (por ejemplo, plásmido y/o ADN genómico de una célula receptora), usando cualquier método adecuado conocido en el arte. En las modalidades adicionales, las repeticiones de CRISPR encuentran uso como una plantilla para modificar (por ejemplo, mutar) el ADN de una célula (por ejemplo, ADN plásmido y/o genómico de una célula receptora), tal que las repeticiones de CRISPR se crean o producen por ingeniería en el ADN de la célula. En las modalidades adicionales, las repeticiones de CRISPR se presentan en al menos un constructo, al menos un plásmido, y/o al menos un vector, etc. En modalidades adicionales, las repeticiones CRISPR se introducen en la célula usando cualquier método adecuado conocido en el arte.

En las modalidades adicionales, la presente invención proporciona métodos para identificar una repetición CRISPR para uso en modular la resistencia de una célula contra un ácido nucleico objetivo o producto de transcripción del mismo que comprende las etapas de: (i) preparar una célula que comprende al menos un espaciador CRISPR y al menos un gen *cas*; (ii) producir por ingeniería la célula de tal manera que contiene una repetición CRISPR; y (iii) determinar si la célula presenta resistencia contra el ácido nucleico objetivo o producto de transcripción del mismo,

en donde la modulación de la resistencia de la célula contra el ácido nucleico objetivo o producto de transcripción del mismo es indicativo de que la repetición CRISPR puede usarse para modular la resistencia.

En algunas modalidades adicionales, uno o más genes o proteínas *cas* se usan junto con o en combinación con una o más, preferiblemente, dos o más repeticiones de CRISPR y opcionalmente uno o más espaciadores CRISPR. En algunas modalidades particularmente preferidas, los genes o proteínas *cas* y repeticiones de CRISPR forman una combinación funcional como se describe a continuación. En algunas modalidades, las repeticiones CRISPR comprenden cualquiera de los nucleótidos establecidos en la SEQ ID NOS: 1-22. Las SEQ ID NOS: 1-12 son de *S. thermophilus*, mientras las SEQ ID NOS: 13-16 son de *Streptococcus agalactiae*, la SEQ NO: 17 es de *S. mutans*, y las SEQ ID NOS: 18-22 son de *S. pyogenes*.

GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAAC (SEQ ID NO:1)
 GTTTTTGTATTCTCAAGATTTAAGTAACTGTACAGT (SEQ ID NO:2)
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAGT (SEQ ID NO:3)
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAGT (SEQ ID NO:4)
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTGCAAC (SEQ ID NO:5)
 GTTTTTGTA CTCTCAAGATTTAAGTAGCTGTACAGT (SEQ ID NO:6)
 GTTTTTGTA CTCTCAAGATATAAGTAACTGTACAAC (SEQ ID NO:7)
 GTTTTTGTA CTCTCAAGATCTAAGTAACTGTACAAC (SEQ ID NO:8)
 GTTTTTGTA CTCTCAAGATGTAAGTAACTGTACAAC (SEQ ID NO:9)
 GTCTTTGTA CTCTCAAGATTTAAGTAACTGTACAAC (SEQ ID NO:10)
 AAAAAAGTCCCCTCTCGAGGTAATTAGGTTTATATC (SEQ ID NO:11)
 GTTTCGTCCTCTCTCGAGGTAATTAGGTTTATATC (SEQ ID NO:12)
 GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC (SEQ ID NO:13)
 GTTTTAAAGCTGTGCTGTTATTATGCTAGGGCACCA (SEQ ID NO:14)
 GTTTTAGAGCTGTGCTGTTTCGAATGGTTCCAAAAC (SEQ ID NO:15)
 GTTTTAGAGCTGTGCTGTTATTATGCTAGGACATCA (SEQ ID NO:16)
 GTTTTAGAGCCATGTTAGTTACTGATTTACTAAAAT (SEQ ID NO:17)
 GTTTTAGAGCTATGCTGTTTTGAATGGTCCAAAAC (SEQ ID NO:18)
 GTTTTAGAGCTATGCTGTTTTGAATGGTCTCCATTC (SEQ ID NO:19)
 CTTTCAATCCACTCACCCATGAAGGGTGAGACG (SEQ ID NO:20)
 ATTTCAATCCACTCACCCATGAAGGGTGAGACT (SEQ ID NO:21)
 ATTTCAATCCACTCACCCATGAAGGGTGAGACC (SEQ ID NO:22)

Espaciador CRISPR

Como se usa en la presente, el “espaciador CRISPR” abarca secuencias espaciadoras no repetitivas que se encuentran entre repeticiones directas cortas múltiples (esto es, repeticiones de CRISPR) de las ubicaciones CRISPR. En algunas modalidades de la presente invención, un “espaciador CRISPR” se refiere al segmento de ácido nucleico que se flanquea por dos repeticiones de CRISPR. Se ha encontrado que las secuencias del espaciador CRISPR frecuentemente tienen similitudes importantes a una variedad de moléculas de ADN móviles (por ejemplo, bacteriófagos y plásmidos). En algunas modalidades preferidas, los espaciadores CRISPR se localizan entre dos repeticiones de CRISPR idénticas. En algunas modalidades, los espaciadores CRISPR se identifican por el análisis de secuencia en los extremos de ADN localizados entre dos repeticiones de CRISPR. En algunas modalidades preferidas, el espaciador CRISPR se presenta naturalmente entre dos repeticiones directas cortas múltiples idénticas que son palindrómicas.

De forma interesante, las células que llevan estos espaciadores CRISPR no pueden ser infectadas por las moléculas de ADN que contienen las secuencias homólogas a los espaciadores (Mojica et al. 2005). En algunas modalidades preferidas, el espaciador CRISPR es homólogo al ácido nucleico objetivo o un producto de transcripción del mismo o una secuencia identificada. Aunque la homología puede también considerarse en los términos de la similitud, en el contexto de la presente invención se prefiere expresar la homología en los términos de identidad de secuencia. Una secuencia homóloga se toma para incluir un espaciador CRISPR, el cual puede ser al menos idéntico a alrededor de 70, alrededor de 75, alrededor de 80, alrededor de 85, o alrededor de 90%, o al menos idéntico a alrededor de 91, alrededor de 92, alrededor de 93, alrededor de 94, alrededor de 95, alrededor de 96, alrededor de 97, alrededor de 98, o alrededor de 99% a la secuencia de ácido nucleico objetivo o un producto de transcripción del mismo o una secuencia identificada. En algunas modalidades preferidas, el espaciador CRISPR es idéntico alrededor del 100% a la secuencia de ácido nucleico objetivo. También se nota que el número de los espaciadores CRISPR en una de las ubicaciones o ubicación CRISPR dada puede variar entre las especies. Además, el número de los espaciadores está en el intervalo desde alrededor de 1 hasta alrededor de 140, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 2 hasta alrededor de 100, desde alrededor de 5 hasta alrededor de 100, desde alrededor de 10 hasta alrededor de 100, desde alrededor de 15 hasta alrededor de 100, desde alrededor de 20 hasta alrededor de 100, desde alrededor de 25 hasta alrededor de 100, desde alrededor de 30 hasta alrededor de 100, desde alrededor de

35 hasta alrededor de 100, desde alrededor de 40 hasta alrededor de 100, desde alrededor de 45 hasta alrededor de 100, o desde alrededor de 50 hasta alrededor de 100. En algunas modalidades preferidas, el número de los espaciadores está en el intervalo desde alrededor de 1 hasta alrededor de 135, desde alrededor de 1 hasta alrededor de 130, desde alrededor de 1 hasta alrededor de 125, desde alrededor de 1 hasta alrededor de 120, desde alrededor de 1 hasta alrededor de 115, desde alrededor de 1 hasta alrededor de 110, desde alrededor de 1 hasta alrededor de 105, desde alrededor de 1 hasta alrededor de 100, desde alrededor de 1 hasta alrededor de 95, desde alrededor de 1 hasta alrededor de 90, desde alrededor de 1 hasta alrededor de 80, desde alrededor de 1 hasta alrededor de 70, desde alrededor de 1 hasta alrededor de 60, desde alrededor de 1 hasta alrededor de 50, desde alrededor de 1 hasta alrededor de 40, desde alrededor de 1 hasta alrededor de 30, desde alrededor de 1 hasta alrededor de 20, desde alrededor de 1 hasta alrededor de 10, desde alrededor de 1 hasta alrededor de 9, desde alrededor de 1 hasta alrededor de 8, desde alrededor de 1 hasta alrededor de 7, desde alrededor de 1 hasta alrededor de 6, desde alrededor de 1 hasta alrededor de 5, desde alrededor de 1 hasta alrededor de 4, desde alrededor de 1 hasta alrededor de 3, o desde alrededor de 1 hasta alrededor de 2. En algunas modalidades preferidas, los espaciadores CRISPR se identifican por el análisis de secuencia como los extremos del ADN localizados entre las dos repeticiones.

Como se describe en la presente, la presente invención proporciona métodos y composiciones que facilitan el uso de uno o más genes o proteínas *cas* en combinación con una o más, preferiblemente, dos o más repeticiones de CRISPR adecuadas para conceder especificidad de inmunidad a al menos un espaciador CRISPR en una célula receptora. En algunas modalidades preferidas, al menos un gen o proteína *cas* y al menos una repetición CRISPR se usan en combinaciones funcionales para conceder la especificidad de la inmunidad a al menos un espaciador CRISPR en una célula.

Como se usa en la presente, el término “especificidad de inmunidad” significa que la inmunidad se confiere contra una secuencia de ácido nucleico específica o producto de transcripción del mismo, usando un espaciador CRISPR específico o secuencia del pseudo-espaciador CRISPR. Como se indica en la presente, un espaciador CRISPR dado no concede resistencia contra cualquier secuencia de ácido nucleico o producto de transcripción del mismo pero solamente aquellas secuencias contra las cuales el espaciador CRISPR o pseudo-espaciador CRISPR es homólogo (por ejemplo, aquellos que son idénticos al 100%).

En algunas modalidades, los espaciadores CRISPR se obtienen de un organismo donador que es diferente de la célula receptora. En algunas modalidades preferidas, el donador y las células receptoras son diferentes cepas bacterianas, especies y/o géneros. En algunas modalidades preferidas, al menos un gen o proteína *cas* y/o al menos unas repeticiones de CRISPR se obtienen de un organismo diferente al del organismo receptor. En algunas modalidades preferidas, al menos dos repeticiones de CRISPR se transfieren. En algunas modalidades adicionales preferidas, los espaciadores CRISPR se obtienen de un organismo que es heterólogo al receptor o una célula donadora adicional de la cual al menos un gen y/o proteína *cas*, y/o al menos una repetición CRISPR se obtienen. En algunas modalidades preferidas alternativas, los espaciadores CRISPR se obtienen de un organismo que es homólogo al receptor o una célula donadora adicional de la cual al menos un gen y/o proteína *cas*, y/o al menos una repetición CRISPR se obtienen. En algunas modalidades preferidas, los espaciadores CRISPR se diseñan y producen usando los métodos recombinantes conocidos en el arte. Efectivamente, se entiende que los espaciadores CRISPR se producen usando cualquier método adecuado conocido en el arte.

En algunas modalidades, los espaciadores CRISPR son heterólogos a la célula receptora de la cual al menos un gen o proteína *cas* y/o al menos una, y en algunas modalidades, preferiblemente, dos o más, repeticiones de CRISPR se obtienen. En algunas modalidades alternativas, los espaciadores CRISPR son homólogos a la célula receptora de la cual al menos un gen o proteína *cas* y/o al menos una, y en algunas modalidades, preferiblemente, dos o más, repeticiones de CRISPR se obtienen. Efectivamente, se entiende que cualquiera de los elementos utilizados en los métodos son heterólogos u homólogos. En algunas modalidades, donde los elementos múltiples se usan (por ejemplo, cualquier combinación de los espaciadores CRISPR, repeticiones de CRISPR, genes *cas*, y proteínas *cas*), algunos elementos son homólogos unos de otros y algunos elementos son heterólogos unos de otros (por ejemplo, en algunas modalidades, los espaciadores CRISPR y genes *cas* son homólogos, pero las repeticiones de CRISPR son heterólogas). Así, en algunas modalidades, el espaciador CRISPR no se asocia naturalmente con la repetición CRISPR y/o genes *cas* y/o combinación gen *cas*-repetición CRISPR funcional. Efectivamente, se entiende que cualquier combinación de los elementos heterólogos y homólogos encuentra uso en la presente invención. En modalidades aún adicionales, las células donadoras y receptoras son heterólogas, mientras en modalidades adicionales, son homólogas. También se entiende que los elementos contenidos dentro de las células donadoras y receptoras son homólogos y/o heterólogos. Los elementos (por ejemplo, espaciadores CRISPR) se introducen en el plásmido y/o ADN genómico de la célula receptora utilizando cualquier método adecuado conocido en el arte.

Al menos un espaciador CRISPR se usa para producir por ingeniería una célula (por ejemplo, una célula receptora). En algunas modalidades adicionales, uno o más espaciadores CRISPR se usan en combinación con uno o más genes o proteínas *cas* y/o una o más, preferiblemente, dos o más repeticiones de CRISPR (en algunas modalidades preferidas, una o más combinaciones funcionales de los mismos se usan) para modular la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo, para producir por ingeniería una célula. En algunas modalidades adicionales, los espaciadores CRISPR se usan como una plantilla para modificar (por

ejemplo, mutar) el plásmido y/o ADN genómico de una célula (por ejemplo, una célula receptora), de tal manera que los espaciadores CRISPR se crean en el ADN de la célula. En algunas modalidades, los espaciadores CRISPR se clonan en al menos un constructo, plásmido u otro vector, con el cual la célula receptora luego se transforma a continuación, usando cualquier método adecuado conocido en el arte.

5 En algunas modalidades adicionales, la presente invención proporciona métodos para identificar un espaciador CRISPR para uso en modular la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo, que comprende las etapas de: preparar una célula que comprende al menos dos repeticiones CRISPR y al menos un gen o proteína cas; identificar al menos un espaciador CRISPR en un
10 organismo (por ejemplo, un organismo donador); modificar la secuencia del espaciador CRISPR de la célula de tal manera que tiene homología al espaciador CRISPR del organismo donador que comprende el ácido nucleico objetivo; y determinar si la célula modula la resistencia contra el ácido nucleico objetivo, en donde la modulación de la resistencia de la célula contra el ácido nucleico objetivo o producto de transcripción del mismo es indicativa del espaciador CRISPR que modula la resistencia de la célula contra el ácido nucleico objetivo.

15 En algunas modalidades preferidas, los espaciadores CRISPR comprenden o consisten en la secuencia de nucleótido establecida para cualquiera de una o más de cualquiera de las SEQ ID NO: 23-460 y/o SEQ ID NOS: 522-665. Las SEQ ID NOS: 23-339, 359-408, 522-665 son de *S. thermophilus*, mientras las SEQ ID NOS: 340-358 son de *S. vestibularis*, las SEQ ID NOS: 409-446 son de *S. agalactiae*, las SEQ ID NOS: 447-452 son de *S. mutans*,
20 y las SEQ ID NOS: 453-460 son de *S. pyogenes*.

(Continúa en página siguiente)

AGAACGTATTCCAAAACCTCTTTACGATTA (SEQ ID NO:23)
 TTAAGTGTATCAAAATGATAAGATAGTCT (SEQ ID NO:24)
 CGTTGATGTTTATTCAAGTAAAATAATTAA (SEQ ID NO:25)
 TCCTTTCACGGGTAGCACACTAACATACAC (SEQ ID NO:26)
 GTTGGCAATGCAAACAACCTTTATGAACCG (SEQ ID NO:27)
 TTTATTTCTTTCGCGATAACGTTCCACCTTT (SEQ ID NO:28)
 AGATTATAAGGAACACAACCAACTATATAG (SEQ ID NO:29)
 ACGACATCAAGCTGATTGTCTTCTACATAA (SEQ ID NO:30)
 TTTGGAATACTGAATGTTTTACTGAAAATC (SEQ ID NO:31)
 ACACCACTATCTTTTCTCCTGAAAATGAA (SEQ ID NO:32)
 GTAATTCCACGAAATTATCAACCTTATGCA (SEQ ID NO:33)
 TTGGAGGATTGCCCCATATTCCCAAGAGT (SEQ ID NO:34)
 GAGAGGCGTTAAATATAGAAATGCAAGATT (SEQ ID NO:35)
 TTTTAACGTCATCAGTCCACCGCCTTAAAT (SEQ ID NO:36)
 CACCTCTTTCGATGGAAAGGTATCCTTCTA (SEQ ID NO:37)
 GACCAAAGTTTGATTATAGAGCTATACACC (SEQ ID NO:38)
 ACCATCATTCTTACCATTACAACCTGTAATG (SEQ ID NO:39)
 ATACGAATTCGGTTCGCACAATTACAATTC (SEQ ID NO:40)
 TATCAACGCAATCATTACAACAACCTCAAACA (SEQ ID NO:41)
 ATCTACGTGTCAATACATATCACAAAACAG (SEQ ID NO:42)
 ATTTTATAGAAATTTCTGATATAATAATGA (SEQ ID NO:43)
 TTGTTGGAACAAGGACGACTTGGTAAACTA (SEQ ID NO:44)
 CATATTAAGCTGACTGGGCCTAATGCTTTT (SEQ ID NO:45)
 TTCATAGCATACCGTAGTTGTAAAATCTAT (SEQ ID NO:46)
 AACATTTAGGGAATGAAATTGATAAGACTG (SEQ ID NO:47)
 AACATGAGAACTGTAGAAAACAAGCAATA (SEQ ID NO:48)
 TGGTGAAGATGGCAGTCATAAATGGCACATT (SEQ ID NO:49)
 AAGGGTTGAAAAATGTTGGTATATCAAACG (SEQ ID NO:50)
 TTCTGGTAGTGGATTTAGTCAAACAGATGT (SEQ ID NO:51)
 TCCATAGAGCGTCTTAAACAAAGAATAGTC (SEQ ID NO:52)
 TTATGATTGAATGACATGGTTGTATAAGTA (SEQ ID NO:53)
 TTTCTTTAGGAATACCAGGGAGTTCAGCTT (SEQ ID NO:54)
 TGGCAGAGATTACACAGCAACGGAAACAGC (SEQ ID NO:55)
 GGGTATCATTGTATCTAGTGATGGACCTGA (SEQ ID NO:56)
 ATTTGAAAAATGCACAACAGCGTTTGATAG (SEQ ID NO:57)
 GAGCTACCAGCTACCCCGTATGTCAGAGAG (SEQ ID NO:58)
 CGTTCCTTTTTTCAAGGTAATCTTTGAAAG (SEQ ID NO:59)
 AAGTCCGTAAGCACCAGTTCCAATCGTCAT (SEQ ID NO:60)
 TTGAATACCAATGCCAGCTTCTTTTAAGGC (SEQ ID NO:61)
 AACCTCATACATGGGGAAAATTGGTAAGTA (SEQ ID NO:62)
 TAACCTCATTAGTGTAGTTGTAATTAGCAT (SEQ ID NO:63)
 TTAGCTACCCAAATATCTTCTGTTTTCCAA (SEQ ID NO:64)

GAGTTTTCAATATTGGCACAGGAGACAATT (SEQ ID NO:65)
 TGATACTATTTTAGTCAGATATGAAATATC (SEQ ID NO:66)
 TCATCAATGTTTAAAGCCCAACAATACATGA (SEQ ID NO:67)
 TAGATTTAATCAGTAATGAGTTAGGCATAA (SEQ ID NO:68)
 AGGAAAATAGCATGAGCGTACAACAATCTA (SEQ ID NO:69)
 TGTCTATCACGCTTCCTAAGTGCATGAAAA (SEQ ID NO:70)
 ATGTCACCAATCACTAAAGAACCTACGCTG (SEQ ID NO:71)
 AACATCTTCCTCTCCGATTGCAAATAGTGC (SEQ ID NO:72)
 CATATTTGGTGCCCGTTTCGATAAAGAGTA (SEQ ID NO:73)
 CATTAAATCGCTTGAAGCAGACATTGAAGC (SEQ ID NO:74)
 GACTTATCTTGAAGGTAGTGAAGGCACTT (SEQ ID NO:75)
 TCCTTGCCATCTGCACTGTAAGCCCAAGCA (SEQ ID NO:76)
 TAGTACGCATAATCAATTCATCAAGCTTGA (SEQ ID NO:77)
 GTAGTGACCCAAAATTCTATGACCTTGAAA (SEQ ID NO:78)
 AGATTGTGGTGCTTACGGAAAAATTCCTTGT (SEQ ID NO:79)
 TGGCAAGAAGTGTAAGAGATGCAATGGATA (SEQ ID NO:80)
 TTTATTATCATTATTCTTCTTCCCAAGCGT (SEQ ID NO:81)
 TTTTATAGAATTTGGTGGTGAACCTTTTCA (SEQ ID NO:82)
 AATGGGTCACAGATTGCCATAATAAGGAG (SEQ ID NO:83)
 CCGAGGTCACTTTAGAACCCACAAAATAAG (SEQ ID NO:84)
 ATGAGAGAACACAGTATAGACCCTGATACA (SEQ ID NO:85)
 CAGTATTAATGAGGTTTGGGTGGTCATTCC (SEQ ID NO:86)
 CCATACTCTCTATCAGTTCATTTAATTCTTC (SEQ ID NO:87)
 TAATATGTCGCTCTACTGATTCCAAAACGG (SEQ ID NO:88)
 ATGAATTACATTCATGATTTTATCGAGTTT (SEQ ID NO:89)
 CGTGCCATTGTTTCGGTCGGACGTGGGCA (SEQ ID NO:90)
 CTTTCTAAGTTGAATTAATTCAAGTTTGT (SEQ ID NO:91)
 TCGCTACTATGGTTAACGATGAGGAACTCT (SEQ ID NO:92)
 AGCAACTTTAAAACTAAAAGAGCTACTTGA (SEQ ID NO:93)
 AAAACCCTACACAGTGTGTGAGATGTGTCA (SEQ ID NO:94)
 AATGGGTCACAGATTGCCATAATAAGGAGG (SEQ ID NO:95)
 TTTTTTAAAAATCCGTCATGCTATACTATAT (SEQ ID NO:96)
 AATTCAAACCTTCTCCAATAATACCCTCCA (SEQ ID NO:97)
 CATGCTTTCAGTTAATAAGACGTGGGACTA (SEQ ID NO:98)
 TGGAAGGGGTGTCTAGTGAAGAAATTGTCTG (SEQ ID NO:99)
 CTCGAAGCGCTTCATTGCCCTATTCCTTTC (SEQ ID NO:100)
 ATGTCTAAGGTATCCACTCGTGAAATCAT (SEQ ID NO:101)
 ATATTAATGGAAATTTCAATTCAAACGCAGT (SEQ ID NO:102)
 TAGAGAGTTTATATCCTGATGGAATCGATG (SEQ ID NO:103)
 TGGCGAATTAGAGAGCCAATGGCAAGCAAG (SEQ ID NO:104)
 AGAAGACCAATAAACTTGAGAAAAAGCAAG (SEQ ID NO:105)
 AAATGGTTCGTTTAAATTGTTAATGTCAAAGC (SEQ ID NO:106)
 CAATTGATTCTAAAATGCTTGGTACACGTA (SEQ ID NO:107)
 TCTTCGTGTTATCACAGCTTCTACACGTTG (SEQ ID NO:108)
 GAAATCTCATTGAAACCAACTTCAAGACCA (SEQ ID NO:109)
 TGCTTGGTAGTTGATGCACTGCATTAGTAA (SEQ ID NO:110)
 AATGTACCGGAATAGCGTTACATTGCACAT (SEQ ID NO:111)
 TTCATAAATTCTCACTTTTCCTTGCTATTC (SEQ ID NO:112)
 TGTCGAAAAAATTACCTAGTCACGACAGAC (SEQ ID NO:113)
 CAACAATTACTTATGCATTAGGAACATCTG (SEQ ID NO:114)
 AATTCGTGAAAAACAATAAAAAACAAAAAA (SEQ ID NO:115)
 TAACATTTCTGTCCATTTCTTCCTTGATGC (SEQ ID NO:116)
 CAAGGCAACTCAACCAACCAATTGACC (SEQ ID NO:117)
 CTAATAATCGTAAATGGTAAGTTGCACGATG (SEQ ID NO:118)
 AACGTAAGGAGTTTTTTTATTCTTTGTGA (SEQ ID NO:119)

GTGGAAAATTTACACCCTACATATATCAA (SEQ ID NO:120)
 CCTCTGCTAATGACTTAAACGGCTCGTTTT (SEQ ID NO:121)
 AAAATCAAAGTTTTGGGTTTGTCTACGTTG (SEQ ID NO:122)
 ATATGTACATACCTAAAGAAAACACGGGCA (SEQ ID NO:123)
 CGTTGTCAAAATATGTGATTACTTTGTATT (SEQ ID NO:124)
 CCATAGCTGTAATGTTGTTTGTGACTGCTT (SEQ ID NO:125)
 CGCTAAGTTTGGCTTTAAGTATAACAAGCT (SEQ ID NO:126)
 AAAGTACGCTTCAAGGCACGTTGAAGACAT (SEQ ID NO:127)
 CTTTTTAACGTGTTAGCGTCTTAGCTTTG (SEQ ID NO:128)
 TTGGCTTCGTGAATAATTTTTAAAACGCAT (SEQ ID NO:129)
 TGTTGAATCAATACGCTGAAACACACTCCC (SEQ ID NO:130)
 CGTTATCAGTTGAAAGTTTCAACTCGTAAG (SEQ ID NO:131)
 TAAACTAGTTGGCATCTATGCTCCAGGAAG (SEQ ID NO:132)
 TAGACCACCATAGCCGAGTTGTCTTTTTCG (SEQ ID NO:133)
 ACATCCCACCTTCTGGGTTTTTTAGCCATG (SEQ ID NO:134)
 AGTATGGCTATTGTCCTGATACTCATCCAC (SEQ ID NO:135)
 CGCTCTTGACGTGGCTGGTGACATCTACGC (SEQ ID NO:136)
 GAGTACATGGAGTTTCTGCTAGATACTA (SEQ ID NO:137)
 TAAGTTATGAAATATAAAGTTATTGTCTA (SEQ ID NO:138)
 AACGTTATGACATTTAGGAGCTTCCAAATT (SEQ ID NO:139)
 AACACAGCAAGACAAAAGGATGACACTTT (SEQ ID NO:140)
 CAACCATAACTTACGCATCAGGTACATCTG (SEQ ID NO:141)
 ACACGCGCTTACCTCGTATATCAAATTCA (SEQ ID NO:142)
 TGCCCCGAAACTAGCGATACACAACAGCAT (SEQ ID NO:143)
 CTCAAGCTCTTCATCTGTGATAGGTGTTTTG (SEQ ID NO:144)
 ATCACTCTTTGATAGTATCTCAAACGCTGG (SEQ ID NO:145)
 GAAACAGTCAGACCAGCTAATTGCGCAATT (SEQ ID NO:146)
 ATATTTGCGAAAGATACAAGGACACTTACAC (SEQ ID NO:147)
 GCGGATGAAACACAACCTTCAATTGTATTCA (SEQ ID NO:148)
 TAATGCTACATCTCAAAGGATGATCCCAGA (SEQ ID NO:149)
 ACGTCTGTCTAACTGGAAAGTACCTGCTAAT (SEQ ID NO:150)
 CTGTTCTCTAATCGAGAGGCGCGTGATTGA (SEQ ID NO:151)
 AAACCTCACTAGTCACTTAGTGCGGTTAGG (SEQ ID NO:152)
 TATTAAGTTTAGTCCCAGGTTTCTTATCGT (SEQ ID NO:153)
 AAACCAATAAACATACCGATTGCTGCCAAT (SEQ ID NO:154)
 GCAAACGTTAGCCCAGGAAAGCATCATGAA (SEQ ID NO:155)
 AAGAGCAAAAAATAACTCTAGCTCTCGTCC (SEQ ID NO:156)
 AAGAAACCTCTAAGTTGAGCATTTAATGAT (SEQ ID NO:157)
 ATATAGTTTTAACTTTCTTGACCTTCTG (SEQ ID NO:158)
 ACGTTGATGAATATTGTTGATAAACTTTA (SEQ ID NO:159)
 CAAGAAGTGAACAAAGTACACGCTGGAAGT (SEQ ID NO:160)
 GACAGCAAGATACACGTAGTTGATGAATTG (SEQ ID NO:161)
 TAAGAAATCAACGCAGATTTTTAGCCAACA (SEQ ID NO:162)
 TAACCAATAATTACAGTGAAGCACAATAG (SEQ ID NO:163)
 CAGGCGTAAGGTATGCTAATTATAACGAT (SEQ ID NO:164)
 GCTATCGAACTAATAGCTTAGAGGAACTCA (SEQ ID NO:165)
 GTGGAATATTAAGCCCGAATTGTTGCAGCA (SEQ ID NO:166)
 TATTGCAATATTTGCGTTTGGGAAACCTTC (SEQ ID NO:167)
 CGTCTGTCTAACTGGAAAGTACCGGCTAAT (SEQ ID NO:168)
 AAAGAGATGTACCCATCCATTCTAACAGGT (SEQ ID NO:169)
 GGGGAGTTGATTTCTTACATCAAAACAATG (SEQ ID NO:170)
 CATCAAAGTTGAAAGGACTACAACAGCCC (SEQ ID NO:171)
 CTTAAATTTAGAGCGTGGGATCTTGAATAT (SEQ ID NO:172)
 ATATACCGATGGCACATCTGAAACTGGCTG (SEQ ID NO:173)
 TAACTCATATGTATCTTGACCAACTATTTT (SEQ ID NO:174)

AAATAGCACCTCTAAGCGTTAATGGTATTC (SEQ ID NO:175)
 AATATCTACAGGTCACTACAAAGCTACGCT (SEQ ID NO:176)
 GTTGGGGTGTGTTTGTAAACGGCGTATGCTA (SEQ ID NO:177)
 TCAATCAGGTGACGGTGATGCTTATATTAA (SEQ ID NO:178)
 CATACATGATAGTTTGTCAACACTTTTGAT (SEQ ID NO:179)
 TCAGCATTTGGTTTACATGACCCACGTCTG (SEQ ID NO:180)
 CAATCAACAGGTTTGACTGATTATAACGGT (SEQ ID NO:181)
 TAGCTACACATGAATTTTATTACAATGGTG (SEQ ID NO:182)
 CTTACGTTTGAAAAGAATATCAAATCAATG (SEQ ID NO:183)
 TTAAAAAAGGGCCTTTCTCTAAATCAAGTA (SEQ ID NO:184)
 TGCTGAACGTATCTGTCCACTGTGTGGCCA (SEQ ID NO:185)
 CCGTTCTTCAAACGTTAAATTCCAAGGTGT (SEQ ID NO:186)
 GCTGCGATTATGACAATGCTGTCTGTAAGG (SEQ ID NO:187)
 GAAGAATTTATTAATAAAGATGGTTCTGCT (SEQ ID NO:188)
 AGGCAGAAAAGAAGTATTTTGGTAAGTATG (SEQ ID NO:189)
 AAATGGTTTATCGACAAGAAAATGAAGCT (SEQ ID NO:190)
 CCAAATTTGCATTATACAAAACGCTCCTTC (SEQ ID NO:191)
 ATCCTAACTGCTTTGCTAACTACATCATGG (SEQ ID NO:192)
 TAACAAGATAAGATTAGCGTCTTCAACAT (SEQ ID NO:193)
 AAAAGCCTATGTTTGCCCACTTTGTGGAAG (SEQ ID NO:194)
 TGTCACCTTTCTCTTTCTGGGTTGTGCCAAT (SEQ ID NO:195)
 CATACTTTTCCATCTGTTTGTGTTTAAAA (SEQ ID NO:196)
 TGAGAGTGTCTGATGGATTTATTGGCAGCC (SEQ ID NO:197)
 GGGGTTATTTTCCATTTTACCGTCTATCTA (SEQ ID NO:198)
 TATCACGCCCATTTTCATTTCCGCATCTGT (SEQ ID NO:199)
 AACATTTTAATATAATTTCTAAATCTATTG (SEQ ID NO:200)
 TACAAAATTCCTTCAAACGCTATTTATTGA (SEQ ID NO:201)
 AGAGTTTGAAAATTATTTTTCAGTTTCTA (SEQ ID NO:202)
 TTCCTCATCTTTCTCCGCTTTTGCTAGCTT (SEQ ID NO:203)
 TTGAGCGTTCTAGTGTGTGGCTTGTAATGAA (SEQ ID NO:204)
 TGAAAGAAATACAATACAACGATAATGACC (SEQ ID NO:205)
 CTAGTTTTAAGAGATAGCTCTCTAAGTAGG (SEQ ID NO:206)
 AAATTCGACATAAGCACTACAGTTATATT (SEQ ID NO:207)
 CTATTTTCGAGAGAACGTCAGTCATTTTAA (SEQ ID NO:208)
 GTGCTAACTATATCAGTCGCATCAATAACA (SEQ ID NO:209)
 TTAGCGGTGATTGGAATAGAATAAGCGAAT (SEQ ID NO:210)
 CTTCTACAGCAGTTTAAGACACATTATCAT (SEQ ID NO:211)
 CGTATCGAAAACGGCGATAATCCAACAGT (SEQ ID NO:212)
 CAATACCTTTTTTTAATTCATCTTGATAAGT (SEQ ID NO:213)
 TTAAGAACAATATCATCAATACGACTTTCA (SEQ ID NO:214)
 CATCTATCAAATTCAAATTCGGATAAACTA (SEQ ID NO:215)
 TGAGAGTGTCTGATGGATTTATTGGTAACC (SEQ ID NO:216)
 ACCTCATACATGGGGAAAACCTTGTAAGTA (SEQ ID NO:217)
 TATTTACGAATTTCTACACTTTTCAACCT (SEQ ID NO:218)
 CTGAAACCTTGTTTTGAAGCGCTTGGAAGT (SEQ ID NO:219)
 GTCAATTGATACTGCAATCTCTTTAACATT (SEQ ID NO:220)
 ACTTCAATATGGTCAACATCTTGATCACCGA (SEQ ID NO:221)
 TAAACTCGACAAAAGCACTACATGAATATT (SEQ ID NO:222)
 ATTTTTTAAGGAAAGGAGGAAAATAATATA (SEQ ID NO:223)
 CGTTCAAAAACAGCGAAAACCTTAACCTAAC (SEQ ID NO:224)
 CATTAAGTCGCTTGAGGCAGACATTGAAGC (SEQ ID NO:225)
 CCAAACCTCAAATTGTCTATAATAATAACCG (SEQ ID NO:226)
 TATCTCTATTTTCAAGGTGGTTTAAACATT (SEQ ID NO:227)
 AAACGAAGATGGAAGCGTTGATGTTTATTC (SEQ ID NO:228)
 GATTGCATTTGCCAGTATTTCTTTTGATTA (SEQ ID NO:229)

TGAAGACAACGGAACAATCAACCTATTA (SEQ ID NO:230)
 ACTTCTTTTTTAATGTCATCTAAGACAATA (SEQ ID NO:231)
 GCCAATGATGTTCAATTCGTTAATGGAATT (SEQ ID NO:232)
 TCAACATGGGATATTTTCGTTGGTCAGGATG (SEQ ID NO:233)
 TATGGCTCTCTTGTGGAATAAAGATGATT (SEQ ID NO:234)
 ATAACATAGCAGTCTATTTCTTTGCTGATG (SEQ ID NO:235)
 GTTACCACGCGCCCTACTGTATTAGTGGAG (SEQ ID NO:236)
 TACATACCCAAGGTTGTAAGTCGTTAAATT (SEQ ID NO:237)
 TGTAAGTAGTCAATATTCACCTTCTGATAAC (SEQ ID NO:238)
 GATAGCAATAGCTTTCTTGACCTAAAAGAC (SEQ ID NO:239)
 GAGGTCTGTAATTTCAATCCCTCGTAATCT (SEQ ID NO:240)
 AAAGGTTTCTCTAAACACATGCGGAATAT (SEQ ID NO:241)
 GTCATAGTACCAAGCACAAATAACGTTAGT (SEQ ID NO:242)
 GTGTATTTAGTAATGGTGATTTTTTAAATT (SEQ ID NO:243)
 CATTCAATTTTTATATATCAATAAAACTTT (SEQ ID NO:244)
 GGGGATTCTTATTTCACTGTAGTTACGATG (SEQ ID NO:245)
 CAAAAATTGATGTCACAATTAATAAAGGTG (SEQ ID NO:246)
 CTATTTCTGACAATGGTTGAAATTGTGTTC (SEQ ID NO:247)
 CTTTTTTTAAATTAATTTATCGTAAGCAA (SEQ ID NO:248)
 AACAACTTATGAGAACGGTTGAACGGCTT (SEQ ID NO:249)
 AGCCCGCTTATTGCTTCAGTTGTTTATAT (SEQ ID NO:250)
 TGGAGCAACAAGAATGATTAACCTAATGC (SEQ ID NO:251)
 TTTGATGGATATCATTGATAAACTATACGA (SEQ ID NO:252)
 TAACGAAAGCAATACCAATCGTGCTAAAGC (SEQ ID NO:253)
 TATTCCTATGGTCGATATTCGAACAGTCAA (SEQ ID NO:254)
 CAGGGGACAAAGGACTTTGACCCAACAGAAG (SEQ ID NO:255)
 AGAAACACCTAATGGTCTCTTAGAACCCGA (SEQ ID NO:256)
 AAGAAGTTAAAGACAACCTTGTTAAAGACT (SEQ ID NO:257)
 GAAAAAGCATCCATGATAGTGCTTAGACCT (SEQ ID NO:258)
 CGGAATGGTATAAAGAATACAAAGAAAACG (SEQ ID NO:259)
 CCAAGTATCAGCAAAGAAATCAACGAGA (SEQ ID NO:260)
 TTGACCTGTTTATCCTTGTTAACTAGAATAG (SEQ ID NO:261)
 AGAGCACTAGCATACTGTTTAGTCCGAACG (SEQ ID NO:262)
 AGGCAAGGTATTTGATCCAACAGAAGCCAA (SEQ ID NO:263)
 CATGATTTACAACCACGCGCTAGACCAAG (SEQ ID NO:264)
 ACCTAGAAGCATTTGAGCGTATATTGATTG (SEQ ID NO:265)
 AATTTTGCCCCTTCTTTGCCCTTGACTAG (SEQ ID NO:266)
 TAATAGTTTACCAAATCGTCCTTGTTCCAA (SEQ ID NO:267)
 ACCATTAGCAATCATTTGTGCCCATTGAGT (SEQ ID NO:268)
 ACGTCTGTCTAACTGGAAAGTACCTGTTAAT (SEQ ID NO:269)
 TTTTATACTTTGGGTAATTACAAAATAG (SEQ ID NO:270)
 AAGAAAGAAATATTCTAGATATAGATATAA (SEQ ID NO:271)
 CAACGACCAACACAACAATAAAGTTACTG (SEQ ID NO:272)
 TGATTATGGGTGTTAAACAAGGAGCTTATG (SEQ ID NO:273)
 TGAGTGGTAAGTACAAATACGCAGGACTGA (SEQ ID NO:274)
 TTATTTCCCTCCTTTCCTTAAAAAATTAGA (SEQ ID NO:275)
 GGATGTATCTGTTGAAAGAGGTGTGTATAT (SEQ ID NO:276)
 AATAGGTGAAAAATATGCAAGTCACACAAA (SEQ ID NO:277)
 AAAATGGCATTAAAAATTAACATAGGAATA (SEQ ID NO:278)
 TATCAGCTCGTAAATGTTTCGATAGACTCTT (SEQ ID NO:279)
 ATTCCATTAAACGTATTTGACTTCACTAGCT (SEQ ID NO:280)
 CTGTTACCGATCCAAGAGCAGACATCATAC (SEQ ID NO:281)
 AAGAAGCGGTAAATGCTTCAACTGAATAG (SEQ ID NO:282)
 AATTGCTAAACATCTAAAAGACTTAACGGG (SEQ ID NO:283)
 GATGAAGATTTGACTGATGATAAAGAGAAA (SEQ ID NO:284)

GACATCAGAAAGCAGTTTATAAATATTTTA (SEQ ID NO:285)
 TTTGAATTTAACAACCTTGATTTTGATATC (SEQ ID NO:286)
 TGATACGGTCAAAGTTTTTCCACTAATAGCG (SEQ ID NO:287)
 ATGGTTTTTCATTTCTGAACCCCTAAGAGG (SEQ ID NO:288)
 AAGTTATTGAAAAACGCCAACATGATGAGT (SEQ ID NO:289)
 ATATAAGTCCTCCTATTAATATCCACAATA (SEQ ID NO:290)
 TTGCCTCAAGAGATCCTGCTTGTTGCCAAG (SEQ ID NO:291)
 TCCCATAGTTTTAATGAGTCGGTTAACTTA (SEQ ID NO:292)
 GTGTACTAAAAGTGTGCTAAGTTCATAAGG (SEQ ID NO:293)
 ATATAGTGATTGTATCCAGCTGCGGCGTAG (SEQ ID NO:294)
 AAAAGCAAATCGCGAGTATAAAGGATATA (SEQ ID NO:295)
 TTTTAATTGATCTAGACACCCTATGAAATA (SEQ ID NO:296)
 ACAGAGGAGAGAAACCATGGCTATTTTAGA (SEQ ID NO:297)
 TGGCAGCAGTGAATTCGATGCCGAGCAAT (SEQ ID NO:298)
 CCAAGGAATACCAGGTCCTAAAGGTGCCGA (SEQ ID NO:299)
 CTAAATGAACTACAACAACAGCTTGATGA (SEQ ID NO:300)
 TACCTTAACATTTTCGATATTTTCAAATT (SEQ ID NO:301)
 TTTGACTGCTTTTTTATCTGAATTGTAATT (SEQ ID NO:302)
 CAGTAACCTAAAGCTCTATCAAGCCTATTT (SEQ ID NO:303)
 CGTCAAGCTGACAGACCTTGACAACAAATC (SEQ ID NO:304)
 AGGCATAAATAACATTGATAACCCTAACA (SEQ ID NO:305)
 GCCAACGAGGTCAAATATGTCAACGGCATT (SEQ ID NO:306)
 GAAATAGGAACTCAAAGGTAATTTCTTTA (SEQ ID NO:307)
 ATTTAGAGCAAGGAAAGCAGTACATCATTA (SEQ ID NO:308)
 CTGTAATCATTTTTAAATCAGGATTATCAA (SEQ ID NO:309)
 TTAAATGTATCCTAGTATTTTTGTACTATA (SEQ ID NO:310)
 CCATCAGCCAACTGTATCGGCTACTTTCTA (SEQ ID NO:311)
 ATGCTCTTGGCGACTATCTCATGGAGCGTG (SEQ ID NO:312)
 AGGAAAAAACCCAAACAACCCAAAATGTTA (SEQ ID NO:313)
 TCTAATTCTGTCAACGACTATATCGCCA (SEQ ID NO:314)
 AATCTGTGTGGGAAGTAAAGATTGAAGATG (SEQ ID NO:315)
 ATAGTTTGTTAAGTCATACCCATTAAATTG (SEQ ID NO:316)
 TCCACATGATTACAAAGCCACGCAAGACCT (SEQ ID NO:317)
 GAAGACCAAAATTTGACAATGAGTCCTGC (SEQ ID NO:318)
 ATTATATTTAAGTTGTAATGTTGCTTTTC (SEQ ID NO:319)
 GCAGACATTGGCTCAACAAGTGATTATGAA (SEQ ID NO:320)
 TGTTCTCATAAATTGCCTTTCTTTTATG (SEQ ID NO:321)
 CTTATCAAACATCAAGGATTGTAGATGAGG (SEQ ID NO:322)
 ATTTCATAGTAGCTTGATAAATGTTTCTA (SEQ ID NO:323)
 GAAAATACTATACTTTAAAAGAAATTTTAA (SEQ ID NO:324)
 TCTCCTCCGACATAATCTTTGTCTTCCG (SEQ ID NO:325)
 ACAAAGCACTGCCACCTATAGAAGCATTT (SEQ ID NO:326)
 AAAAAGCTTTATGCTATCCGTGTCAGTATAT (SEQ ID NO:327)
 TTTTCAATGATTGAAAGCCCATAACTAACA (SEQ ID NO:328)
 CTTTCATAGTTGTTACGAAATGTTTGCCAT (SEQ ID NO:329)
 CGATTTGCAATATGATGATATTGATGAATT (SEQ ID NO:330)
 TTTAGATGCTAGTCCTAAGACTGTAGAGAC (SEQ ID NO:331)
 GTAATCAAGCGTATATAAGTCAGGACTATC (SEQ ID NO:332)
 ATAACAGAAGGAGTAGGGGACGTAGGCGCG (SEQ ID NO:333)
 TTATTTGATAGGAATGTCAGTAATTTTGA (SEQ ID NO:334)
 AACATTTACGCGCTTACTTATCAATCTAAT (SEQ ID NO:335)
 GTATTAGTAGGCATACGATTATGGAAGTA (SEQ ID NO:336)
 CATATATATATATATATTTATTTTAAATAT (SEQ ID NO:337)
 TTGTCATAATAATTAATCCAATAGGACTT (SEQ ID NO:338)
 GAAAATTTCTGTTGTGTTCTTAATATTAGC (SEQ ID NO:339)

GTACTTCAAAGGTTCTAACTACATAACACA (SEQ ID NO:340)
 TAAACCAGATGGTGGTTCTTCTGATACTA (SEQ ID NO:341)
 CATTTTCTTCAGTCAATTCGTTCTCAAGCG (SEQ ID NO:342)
 AAAGGACGGGGCAATGAACAAACGACAAC (SEQ ID NO:343)
 TAATATCATTGATAGCTTCATCAAAGGCT (SEQ ID NO:344)
 TAAATTGTTCCCTGACTCCGAAGTCCCT (SEQ ID NO:345)
 AAACAATCGTTTATCTATCCTCAAAGGATG (SEQ ID NO:346)
 ATAAAAAACGCCTCAAAAACCGAGACAAC (SEQ ID NO:347)
 TGGAAATCCCTTATATCGACAAATACGTTA (SEQ ID NO:348)
 TTCCCAGTCGTTGATTTTTATTGAATACCC (SEQ ID NO:349)
 GGACATCGAACAAGTCAATGCCGTAAGCTT (SEQ ID NO:350)
 AATCTTTAACCGGATTGTAGAACCGTTCGG (SEQ ID NO:351)
 TGCCTTTAAAAATACTAGATTTTACCATCA (SEQ ID NO:352)
 GAGCAAGCAAGCAAGCTTTACTATCCT (SEQ ID NO:353)
 CAGATTGGTTTATCGAACAAGGTCGCAAGT (SEQ ID NO:354)
 CAAAAGCTGTTGGTTAACGGTGCTTTGGGCA (SEQ ID NO:355)
 CTTGTTTTCTCTGGGGTCTCTGCGACTT (SEQ ID NO:356)
 GAAATAAACTGCCCCAACATTTTTATTTTC (SEQ ID NO:357)
 TGAGTAAGCGACAAGCTAGAAATCAAGTCA (SEQ ID NO:358)
 ATAGCTAAGATGGAAGAAGCATCAAGCACC (SEQ ID NO:359)
 CAGTATCTCAAACGCTGGATACAACAAGAT (SEQ ID NO:360)
 CCTACTCAGTGGACACCTGCAATTGAAGAC (SEQ ID NO:361)
 CGATTGGAACGGGTGCTTATGGCCTTAAC (SEQ ID NO:362)
 GCGAACAATTGAATTTGTTAGAAAATGTCG (SEQ ID NO:363)
 GAAGCATTTATTAATATAGATGGTCTGCT (SEQ ID NO:364)
 TGCTGACGTATCTGTCCACTGTGTGCCA (SEQ ID NO:365)
 TTTTATACTTTGGGTAAATTACAAAATAG (SEQ ID NO:366)
 TCAAGGTGTCGCCTTATGGAAAAGATGCTTG (SEQ ID NO:367)
 TGTA AAAAATTTCTAGACGTTTAGACACTTTA (SEQ ID NO:368)
 AAATGATGATTGAATGCTTGAGATAGCAGT (SEQ ID NO:369)
 AATAAGAAGTTCTTGACGACCAACCGACAT (SEQ ID NO:370)
 TCGTCAACGTCGATACAGAACAACGTGCTT (SEQ ID NO:371)
 TGATTAGCAAATTTAAACAGGATATTTGG (SEQ ID NO:372)
 AAAGACAAGCCCAAGGGATTGAACTAGCAA (SEQ ID NO:373)
 CGAACAGTTGGCGAGAAATCCGTCTGGCGT (SEQ ID NO:374)
 CTACATTATTGATCATGTTTTTCTCCTGT (SEQ ID NO:375)
 TAGAAGGCTCTGGAAATACAAAGCAATTCT (SEQ ID NO:376)
 TAGAAGGCTCTGGTAAATACAAAGCAATTCT (SEQ ID NO:377)
 TCTGATGGCTCTTGGTAGGGAAGTGGATAT (SEQ ID NO:378)
 TTTGATGGCTCTTGGTAGGGAAGTGGATAT (SEQ ID NO:379)
 TTTTGATGGCTCTTGGTAGGGAAGTGGATAT (SEQ ID NO:380)
 ACAGAACAAAATGGTAGAATATATCATCT (SEQ ID NO:381)
 CCCTGGACAAGCTATCAGCACATATCCTTG (SEQ ID NO:382)
 CGCTGTTGATGTAACCCGCTTTATATATAT (SEQ ID NO:383)
 GAATGAATGTATTAGAGCAAGCACTTGACC (SEQ ID NO:384)
 TAGACGAAAAGGAAGGAAAATAGCATGAGC (SEQ ID NO:385)
 ATAACCTGATTGCTAACTTAAGCAAGCAGT (SEQ ID NO:386)
 CTGCATGTGTAACCATGACTTCTTCGTCGT (SEQ ID NO:387)
 CTTGCTGGAACTTCGTAGTCATACATAC (SEQ ID NO:388)
 AAGACCGCTGTAAGTTGGTATTCGTACC (SEQ ID NO:389)
 CAACCAAGCGAACACAGCAGTAGCACCGCA (SEQ ID NO:390)
 ATGATGATGAAGTATCGTCATCTACTAAC (SEQ ID NO:391)
 CTTACCTCAAATCTTAGAGATGGACTAAA (SEQ ID NO:392)
 AAAAGGTGCGTATGAAACTCATCCAGCGG (SEQ ID NO:393)
 AAGGGTTTAAGTCCTTCATAGAGTGAAAA (SEQ ID NO:394)

CCTCAAAGCTTAAAATTGGGCTGAAGTAGA (SEQ ID NO:395)
 GCAATTTATTTCGCTTGATGTACTCACGTTT (SEQ ID NO:396)
 TATTTATTGCAAATGGTTACCATATTTTAA (SEQ ID NO:397)
 TATTTTAGCACTACGGTATCAGCGTATCTC (SEQ ID NO:398)
 TGCTACGTGCTCTGGACGGGCGCTATCAGC (SEQ ID NO:399)
 AAATGAACAGACAAGAAGCAACAGAAATTG (SEQ ID NO:400)
 AAGTTGATCGTATCTATTTAGAATATCGCA (SEQ ID NO:401)
 ATTCACCTTTGACAGATACTAATGCTACATC (SEQ ID NO:402)
 CAAGCAGTGTAAGGTGGTTTATATGTAA (SEQ ID NO:403)
 CATAGTATAGCCGCTCTTCTTTGATTGATTG (SEQ ID NO:404)
 CCATGGGTGCTAAAGGTGATGACTACCGCT (SEQ ID NO:405)
 TTTCTAGGAATGGGTAATTATAGCGAGCTAGAAAGC (SEQ ID NO:406)
 AGTTGGGAAGGTCTTGGAAAATCTATGGCAAAAACCT (SEQ ID NO:407)
 TATATGGTTCAAATGCGATTCAAAGACTATTCAA (SEQ ID NO:408)
 TAATTGCCAATGCTTACAATATCTTCGTCA (SEQ ID NO:409)
 ATGTTCTGAATTACCTTTCTCGACACTCCG (SEQ ID NO:410)
 ACCATCAAGGCTCTTATCTGCAGATTGTTA (SEQ ID NO:411)
 AAATGGTTGCCAATGACTTTCTAGAGTGAT (SEQ ID NO:412)
 ACAAATCTTTTGTTGCTCCTGGACGTATT (SEQ ID NO:413)
 ATGTAAGGTATTGTAAACTTCTTCTTGCG (SEQ ID NO:414)
 ACTGTTCTATAATTAATAAAGAGGTA (SEQ ID NO:415)
 TGTTCAGTAAAAGTAATTTAAAGCATT (SEQ ID NO:416)
 CGCTCGATTGATGCTATCAACTATATTGAA (SEQ ID NO:417)
 TTCTCAAGAGAAGTTGTAGAAGAGCTTCA (SEQ ID NO:418)
 AAGGTACTTTAGCTTGTCTTGTGGTGTT (SEQ ID NO:419)
 ACAGCTACTGTAAATTCTGCTTTTACGGTT (SEQ ID NO:420)
 TAGTGAGTTGTCAAGGAGATTGTGAGCGA (SEQ ID NO:421)
 TTAAACCTTTGAAAATGTGAAAGGCTCGTA (SEQ ID NO:422)
 GCGATGATGGTAAGTCATCATGGACAGCGT (SEQ ID NO:423)
 TTTTACACACGATGTCAGATATAATGTCAA (SEQ ID NO:424)
 AGTACTGCACTAGGAATTGTAGAGATCAA (SEQ ID NO:425)
 CGTACCATCTATCAATTTACCGCAAGCTGT (SEQ ID NO:426)
 TTAAAAGATTTAACTATCAAGCGTCAATT (SEQ ID NO:427)
 TTCTAAATGCTGGTGACTGCTTTGCATAAA (SEQ ID NO:428)
 TTGCTGCTAGACCCAAACAGTTTATTTTAG (SEQ ID NO:429)
 TCCTTTTATAGATAATGTGCGATCACGGAC (SEQ ID NO:430)
 TTTTACCAATGCTTCCATATCGCTTATAT (SEQ ID NO:431)
 TGGTTATACATTTACTAATCCATCAGCATT (SEQ ID NO:432)
 AAGCTAATTCTCATCTCACCGAGATGGATA (SEQ ID NO:433)
 AAAAACTCTTACCACTTACATACATGTATG (SEQ ID NO:434)
 GCTGGAGATTTTACAAGCAGTTTGAATTTT (SEQ ID NO:435)
 ATCACACCAAGTCGTTATGATGGATGACTAT (SEQ ID NO:436)
 TGTCACAGTACGTGAGACGAGTGTGTAGG (SEQ ID NO:437)
 TGAAGTTGATGGATATGTTGATTTAGAGCT (SEQ ID NO:438)
 TAATCATTTTATGAGAGATACCGCCTCAAG (SEQ ID NO:439)
 TTTAAAGAGATATCTGTTTCATCTTGCGGA (SEQ ID NO:440)
 AATCACTTCTGCATAAATATCTTTTACTTC (SEQ ID NO:441)
 AAACATCCGCAACGGGATAAATAAAGCTAG (SEQ ID NO:442)
 AGTTTCTTGTGGGTTAGCTTGTCCACCGTA (SEQ ID NO:443)
 GAACATGAAAGATTTTAAAAAAGAACATTT (SEQ ID NO:444)
 AGAGGGGAAAATATCAATGCCGAATGCTGA (SEQ ID NO:445)
 GATGGTACAAAATCATTTGTTGGTACTGAT (SEQ ID NO:446)
 AAAAGGAAACGCCATTAATTAATATGGTGA (SEQ ID NO:447)
 GATTGAACCAAGCTAGCGCAGTTAGTGCTCT (SEQ ID NO:448)
 CGCTAAAAGCTGTTGTGTCATCATAGTTAG (SEQ ID NO:449)

TAAATATTTTCAATTAGACAATAGACAAAC (SEQ ID NO:450)
 TGCCTATGTATTCGGACATGACTTGCCACA (SEQ ID NO:451)
 ATGTGAAAAGAAAGTAACTACTACATTTGA (SEQ ID NO:452)
 TGCCTGGTTGATTTCTTCTTGGCTTTTT (SEQ ID NO:453)
 TTATATGAACATAACTCAATTTGTAAAAAA (SEQ ID NO:454)
 AGGAATATCCGCAATAATTAATTGCGCTCT (SEQ ID NO:455)
 TAAATTTGTTTAGCAGGTAAACCGTGCTTT (SEQ ID NO:456)
 TTCAGCACACTGAGACTTGTTGAGTTCCAT (SEQ ID NO:457)
 CTGTGACATTGCGGGATGTAATCAAAGTAAAA (SEQ ID NO:458)
 AAAGCAAACCTAGCAGAAGCAGAAAATGACTT (SEQ ID NO:459)
 TGATGTAATTGGTGATTTTCGTGATATGCTTTTT (SEQ ID NO:460)
 CAACACATTCAACAGATTAATGAAGAATAC (SEQ ID NO:522)
 TCCACTCAGTACAAATAGTGAGTGTAATC (SEQ ID NO:523)
 GCCCTTCTAATTGGATTACCTTCCGAGGTG (SEQ ID NO:524)
 CTCAGTCGTTACTGGTGAACCAAGTTTCAAT (SEQ ID NO:525)
 ATTGTCTATTACGACAACATGGAAGATGAT (SEQ ID NO:526)
 GAGTTTCTTTGTCAGACTCTAACACAGCCGC (SEQ ID NO:527)
 TTACTAGAGCGTGTGCTTAACCACTTTAAA (SEQ ID NO:528)
 TTCGTTAAAGTCACCTCGTGCTAGCGTTGC (SEQ ID NO:529)
 ATAACGGTAGCAAATATAAACCTGTTACTG (SEQ ID NO:530)
 GAAGTAGCCATACAAGAAGATGGATCAGCA (SEQ ID NO:531)
 ATGTCACTGAGTGTCTAAGCATTGCGTAC (SEQ ID NO:532)
 TGAATAAGCAGTTCTTGACGACCAACCGAC (SEQ ID NO:533)
 TTACGTTTGAAAAGAATATCAAATCAATGA (SEQ ID NO:535)
 GCTCTACGACTTCTTCCACGAGTTCTGCC (SEQ ID NO:536)
 AACACAGCAAGACAAGAGGATGATGCTATG (SEQ ID NO: (SEQ ID NO:537))
 AAGTAGTTGATGACCTCTACAATGGTTTAT (SEQ ID NO:538)
 AATAATTTATGGTATAGCTTAATATCATTG (SEQ ID NO:539)
 AATCAATACGACAAGAGTTAAAATGGTCTT (SEQ ID NO:540)
 AATCGTTCAAATTCTGTTTTAGGTACATT (SEQ ID NO:541)
 AATGACGAGGAGCTATTGGCACAACCTACA (SEQ ID NO:542)
 AATTAAGGGCATAGAAAGGGAGACAACATG (SEQ ID NO:543)
 ACAATTCTTCATCCGGTAACTGCTCAAGTG (SEQ ID NO:544)
 ACACTTGGCAGGCTTATTACTCAACAGCGA (SEQ ID NO:545)
 ATAACTATGAAATTTTATAATTTTAAAGA (SEQ ID NO:546)
 ATAAGTGAAGGATAGGAGCTTGTAAGTCT (SEQ ID NO:547)
 ATAATGCCGTTGAATTACACGGCAAGTCA (SEQ ID NO:548)
 CAACCAACGGTAACAGCTACTTTTTACAGT (SEQ ID NO:549)
 CATAGAGTGGAAGTAACTAGAAACAGATTCAA (SEQ ID NO:550)
 CGACACAAGAACGTATGCAAGAGTTCAAG (SEQ ID NO:551)
 CGATATTTAAATCATTTTCATAACTTCAT (SEQ ID NO:552)
 CGATTTGACAATCTGCTGACCACTGTTATC (SEQ ID NO:553)
 CTGTTCCCTGTTCTTTTGTGTATCTTTTC (SEQ ID NO:554)
 GAGCGAGCTCGAAATAATCTTAATTACAAG (SEQ ID NO:555)
 GCAGTATCAGCAAGCAAGCTGTTAGTTACT (SEQ ID NO:556)
 GCTGGCGAGGAAACGAACAAGGCCTCAACA (SEQ ID NO:557)
 GCTTAGCTGTCCAATCCACGAACGTGGATG (SEQ ID NO:558)
 GGCGTCCCAATCCTGATTAATACTTACTCG (SEQ ID NO:559)
 GTTCGCTAGCGTCATGTGGTAACGTATTTA (SEQ ID NO:560)
 TCTATATCGAGGTCAACTAACAATTATGCT (SEQ ID NO:561)
 TGCATCGAGCACGTTTCGAGTTTACCGTTTC (SEQ ID NO:562)
 TGTTTGACAGCAAATCAAGATTCTGAATTGT (SEQ ID NO:563)
 TTCATTCTTCCGTTTTTGTGCGAATCCT (SEQ ID NO:564)
 TGACTTAGCGAATTTAATCGCTAAGATATC (SEQ ID NO:565)
 TTTATACTTTATCTTTTAAAGAATGTATT (SEQ ID NO:566)

CCTAAAATCATTTCACGAGTTGCGATAC (SEQ ID NO:567)
 AATAAATTGCTATGATACAGCGTACCGATA (SEQ ID NO:568)
 TGCTCTCTATGCGATTGGACGTCTGTCTAA (SEQ ID NO:569)
 AAGAAAGATAAGAAAAAAGTAACACTACTT (SEQ ID NO:570)
 TCTCTTTCCATCGGTACTGGTATATCTCAT (SEQ ID NO:571)
 ATTGGTAGCCAAGTAAATATCACCATTGAT (SEQ ID NO:572)
 TTCTTCAAATTCACCGACTGCAAAATTACA (SEQ ID NO:573)
 GCTTCCTAAGTGCATGAAAATCGCAAACGG (SEQ ID NO:574)
 TATACCTGTCTATGTAAGGGAATTTAACTC (SEQ ID NO:575)
 GGTGTAGGTGCTGTTGGTAAGTTGTTAAT (SEQ ID NO:576)
 GTGAAACAGGTTATCAAAAAACGTATATTG (SEQ ID NO:577)
 TTATTCTTGGAATTATTACAGACCCTACTA (SEQ ID NO:578)
 GCTTTTCATTACTACTTACTCATAAATCT (SEQ ID NO:579)
 TAATCACCCCTTTTCTAGCTCTTGATTGA (SEQ ID NO:580)
 CAAGCAGTGTAAGGTTGGTTTAAATGTAA (SEQ ID NO:581)
 AACCCGCGTGGTTATGGGCTTGAGGAGTGT (SEQ ID NO:582)
 ATATTAATAGCGATTCTATGCTACAACGTG (SEQ ID NO:583)
 TCATCTTCTAAGTAAATACCACTGTCAGGG (SEQ ID NO:584)
 TTTTCGCAAAGTAAGCGAAGCTCTACGTG (SEQ ID NO:585)
 TTCTGTAGCCACTCCGTGGATGCCTTCAGC (SEQ ID NO:586)
 TTCTTTAGTTTCGGACACCCCTCAACACCTAT (SEQ ID NO:587)
 GCTTTGATTGGACGGAAAATGGTATCCCTG (SEQ ID NO:588)
 TTCCTCATCTTTCTCCGCTTTTGCTAGACTT (SEQ ID NO:589)
 TTAGACCAGATGGACAGATATTCTTCATCG (SEQ ID NO:590)
 TCATCAGAGTCAACAATCACGGGAAAGACCT (SEQ ID NO:591)
 ACACTACTCTTATCCTGTAGTTCAAAACA (SEQ ID NO:592)
 CAGCACTAGCCGCAAGCCCTTGTATATTAA (SEQ ID NO:593)
 TAGAAATCAAGGAACTTGGATGAAAAGTAA (SEQ ID NO:594)
 ATATGAAAGGGAAATGATATGAAGAATGAA (SEQ ID NO:595)
 TTTTGGGATACAACACGCAGTCGTTGACTTG (SEQ ID NO:596)
 GTTTGAGATGCCAATGTTTTTCAATCCTTG (SEQ ID NO:597)
 GTATCAAAAGACGCATTTCATGAAGCGAGCT (SEQ ID NO:598)
 AAAACAATTGAAATTCATAATCAGCGCTT (SEQ ID NO:599)
 GCTTTTAACGTTTAAAGAGAATACCCTCT (SEQ ID NO:600)
 GTGACGCTGCAATGACTTGCCATAGTAATT (SEQ ID NO:601)
 ATACTGGTATATAGTAATTCATACTTCATC (SEQ ID NO:602)
 TTGGTTTCATATTTACTCCTTTGTGTTTTG (SEQ ID NO:603)
 CTGATTTGGTCTTGTCTTTTGTCCCTTTT (SEQ ID NO:604)
 GCAGCAGTTGAGAACTTTAGCGTCCAGTGG (SEQ ID NO:605)
 TGCTACTATGAAGGACGCTGTTGATACTTT (SEQ ID NO:606)
 TCTTCTTTAATCTTTTTTAACGTCAACGTT (SEQ ID NO:607)
 GTATCCATTAATATAGTAGCATTTCTATCA (SEQ ID NO:608)
 ATTCATTAATATCTGCAAGGATGTCTTGTT (SEQ ID NO:609)
 GAGAAAGTAGCCCATTCGGCCCATTCGGGG (SEQ ID NO:610)
 TACTTGAGTTAGCTCTGGAAGTCATTTATC (SEQ ID NO:611)
 CTGCATTTGTAACCATGACTTCTTCGTCGT (SEQ ID NO:612)
 AATTTGTCATCGACATCTACCAACGCCAG (SEQ ID NO:613)
 ATAAAATTATGCCACGTTTTGGCACTAGAT (SEQ ID NO:614)
 ATGTCTCTGAGGCTGTAGTAATTTACTTGT (SEQ ID NO:615)
 CTTTAAAGAGTTGATTAAGTGCGTTACTGT (SEQ ID NO:616)
 AAATGGGTTATGCTGTTCAATATGCGTCCC (SEQ ID NO:617)
 AAAGTGAACAAACACAGACAATTCAACAA (SEQ ID NO:618)
 GCCCAAAATGCTAGACGTTTGAATGACGGC (SEQ ID NO:619)
 ATGAAGAACGTGATTCACCTACGGTATGCT (SEQ ID NO:620)
 GCTTTTGCAGAATTGTCTCCAGTGCCGATTT (SEQ ID NO:621)

TGTACTCTATTGATTGCTTCATCTTTATTA (SEQ ID NO:622)
 CTTTCAAGATACTCATCAACCATGATGTCA (SEQ ID NO:623)
 CTATGTCTTTACTGTTCTTCCAAAACCAACC (SEQ ID NO:624)
 TGCTACGTGCTCTGTACGGGCGCTATCAGC (SEQ ID NO:625)
 CGTGGCAGCGTGGTCGGGTTTAATAGCCCG (SEQ ID NO:626)
 AAGCCCAAGTCAGAGCATCCGTCCAAGCC (SEQ ID NO:627)
 ATTGGGTTTCGGTAAGAACTAAACATACCA (SEQ ID NO:628)
 CACAAAATAATTCCGGTAGTTTTTACTAACT (SEQ ID NO:629)
 TTTGACCGTTTATTTAGACGTGCTAAAGT (SEQ ID NO:630)
 CTTACCTCAAATCTTAGAGCTGGACTAAA (SEQ ID NO:631)
 ATGTCTGAAAAATAACCGACCATCATTACT (SEQ ID NO:632)
 GAAGCTCATCATGTAAAGGCTAAAACCTAT (SEQ ID NO:633)
 TAGTCTAAATAGATTTCTTGCACCATTTGTA (SEQ ID NO:634)
 ATTCGTGAAAAAATATCGTGAAATAGGCAA (SEQ ID NO:635)
 TCTAGGCTCATCTAAAGATAAATCAGTAGC (SEQ ID NO:636)
 TAAAAACATGGGGCGGCGGTAATAGTGTAAG (SEQ ID NO:637)
 ACAACCAGCAAAGAGAGCGCCGACAACATT (SEQ ID NO:638)
 TATAACACAGGTTTAGAGGATGTTATACTT (SEQ ID NO:639)
 CTAGAAGCTCAAGCGGTAAAAGTTGATGGCG (SEQ ID NO:640)
 CTTTGAGGGCAAGCCCTCGCCGTTCCATTT (SEQ ID NO:641)
 AACTACCAAGCAAATCAGCAATCAATAAGT (SEQ ID NO:642)
 CTATAAGTGACAATCAGCGTAGGGAATACG (SEQ ID NO:643)
 ATCAGTGCGGTATATTTACCCTAGACGCTA (SEQ ID NO:644)
 AACAGTTACTATTAATCACGATTCCAACGG (SEQ ID NO:645)
 AATTAGGGCGTCTTCCTTTATTCCGTGGTT (SEQ ID NO:646)
 ATAGCTTCATTGCGCTTTTTAATTTGACCT (SEQ ID NO:647)
 AACAACAAAGCAAATACAACAGTAACAACC (SEQ ID NO:648)
 CTAAGTACGTTTGAAGGTCTCAACTCCGT (SEQ ID NO:649)
 GAGGTTGAATAGTGAGTGACCATGTTTGT (SEQ ID NO:650)
 AGTAGAGAGACCAGCACACTACTGTACTAC (SEQ ID NO:651)
 CTTCCGACGAAAGTTTATTAGACAACCTCGC (SEQ ID NO:652)
 TGATAGAGCTAGAATTGTCTTTTTTACCGA (SEQ ID NO:653)
 AGATACTCTTGCTCGCCTCTGAACAACCAG (SEQ ID NO:654)
 GGTGAAAAAGGTTCACTGTACGAGTACTTA (SEQ ID NO:655)
 TCAATGAGTGGTATCCAAGACGAAAACTTA (SEQ ID NO:656)
 CCTTGTCGTGGCTCTCCATACGCCCATATA (SEQ ID NO:657)
 TGTTTTGGGAAACCGCAGTAGCCATGATTAA (SEQ ID NO:658)
 ACAGAGTACAATATTGTCCTCATTGGAGACAC (SEQ ID NO:659)
 CTCATATTCGTTAGTTGCTTTTGTCAATAA (SEQ ID NO:660)
 AGAACTTTATCAAGATAAACTACTTTAAA (SEQ ID NO:661)
 ATAGTATTAATTTCAATTGAAAAATAATTGT (SEQ ID NO:662)
 GCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAA (SEQ ID NO:663)
 TCAAAATATGTTATTACCTTGATTTCATAATTCAATTAA (SEQ ID NO:664)
 CCACTTGCTGTGTACATCCTACCAGTCCGCCTATGATG (SEQ ID NO:665)

En modalidades particularmente preferidas, los espaciadores CRISPR se flanquean por dos repeticiones CRISPR (es decir, un espaciador CRISPR que tiene por lo menos una repetición CRISPR a cada lado). Aunque no se pretende que la presente invención se limite a cualquier mecanismo, teoría o hipótesis particular se contempla que cuanto más lejos esté un espaciador CRISPR dado del extremo 5' de la ubicación CRISPR que comprende los genes cas y/o la secuencia líder, menor será la resistencia conferida por ese espaciador CRISPR. Así, en algunas modalidades de la presente invención, uno o más de los primeros 100 espaciadores CRISPR del extremo 5' de la ubicación CRISPR se modifican, mientras que en otras modalidades, se modifican uno o más de los primeros 50 espaciadores CRISPR del extremo 5' de la ubicación CRISPR. En algunas modalidades adicionales, se modifican uno o más de los primeros 40 espaciadores CRISPR del extremo 5' de la ubicación CRISPR, mientras que en algunas modalidades adicionales, se modifican uno o más de los primeros 30 espaciadores CRISPR del extremo 5' de la ubicación CRISPR, y aún en modalidades adicionales, se modifica uno o más de los primeros 20 espaciadores CRISPR del extremo 5' de la ubicación CRISPR, y aún en más modalidades, se modifica uno o más de los primeros 15 espaciadores CRISPR del extremo 5' de la ubicación CRISPR. En algunas modalidades preferidas, se modifica uno o más de los primeros 10 espaciadores CRISPR del extremo 5' de la ubicación CRISPR. Como se indica en la presente, las diferentes bacterias tienen números diferentes de espaciadores CRISPR, así, en algunas modalidades, se modifican varios espaciadores.

Núcleo espaciador CRISPR

Para un tipo CRISPR específico dentro de una especie microbiana, el espaciador CRISPR se representa típicamente por una longitud predominante definida, aunque el tamaño puede variar. Se ha encontrado actualmente que los tipos CRISPR descritos actualmente contienen una longitud espaciadora predominante de entre aproximadamente 20 pb y aproximadamente 58 pb.

Como se usa en la presente, el término “centro espaciador CRISPR” se refiere a la longitud del espaciador más corto observado dentro de un tipo CRISPR. Así, por ejemplo, dentro del CRISPR tipo 1 *S. thermophilus* (CRISPR1), la longitud espaciadora dominante es 30 pb, con una minoría de espaciadores entre 28 pb y 32 pb en tamaño. Así, en el CRISPR Tipo 1 *S. thermophilus*, el centro espaciador CRISPR se define como un estiramiento continuo de 28 pb.

En algunas modalidades preferidas de la presente invención, el centro espaciador CRISPR es homólogo al ácido nucleico objetivo, un producto de la transcripción del mismo, o una secuencia identificada sobre la longitud de la secuencia del centro. Como se ha indicado anteriormente, aunque la homología también puede ser considerada por lo que se refiere a la similitud, en algunas modalidades preferidas de la presente invención, la homología se expresa en términos de identidad de la secuencia. Así, en algunas modalidades, una secuencia homóloga abarca un centro espaciador CRISPR, que puede ser por lo menos aproximadamente 90% idéntico, o por lo menos aproximadamente 91, aproximadamente 92, aproximadamente 93, aproximadamente 94, aproximadamente 95, aproximadamente 96, aproximadamente 97, aproximadamente 98 o aproximadamente 99% idéntico a la secuencia del ácido nucleico objetivo, un producto de la transcripción del mismo, o una secuencia identificada sobre la longitud de la secuencia central. En algunas modalidades particularmente preferidas, el centro espaciador CRISPR es aproximadamente 100% idéntico a la secuencia del ácido nucleico objetivo, un producto de la transcripción del mismo, o una secuencia identificada sobre la longitud de la secuencia del centro.

Durante el desarrollo de la presente invención, las secuencias CRISPR de varias cepas *S. thermophilus*, que incluyen cepas industriales estrechamente relacionadas y variantes resistentes al fago fueron analizadas. Se observaron diferencias en el número y tipo de espaciadores principalmente en la ubicación CRISPR1. Notablemente, la sensibilidad del fago parecía estar en correlación con el contenido del espaciador CRISPR1. Específicamente, el contenido del espaciador fue casi idéntico entre las cepas parentales y los derivados resistentes al fago, excepto para los espaciadores adicionales presentes en el último. Estos resultados sugirieron una relación potencial entre la presencia de espaciadores adicionales y las diferencias observadas en la sensibilidad del fago de una cepa dada. Esta observación impulsó la investigación del origen y función de espaciadores adicionales presentes en los mutantes resistentes al fago.

Espaciador Pseudo-CRISPR

Como se usa en la presente, el término “espaciador pseudo-CRISPR” se refiere a una secuencia del ácido nucleico presente en un organismo (por ejemplo, un organismo del donador, que incluye pero no se limita a un bacteriófago), que es preferentemente esencial para la función y/o supervivencia y/o replicación e/o infectividad, etc., y que comprende una secuencia espaciadora CRISPR. En algunas modalidades, los espaciadores pseudo-CRISPR encuentran uso en la producción de secuencias espaciadoras CRISPR que son complementarias al espaciador pseudo-CRISPR u homólogas a este. En algunas modalidades particularmente preferidas, estas secuencias encuentran uso en la resistencia a la modulación.

En algunas modalidades, por lo menos un espaciador pseudo CRISPR y espaciadores CRISPR que son complementarios u homólogos a por lo menos un espaciador pseudo-CRISPR se usa para diseñar una célula receptora. En algunas modalidades preferidas, por lo menos un espaciador pseudo CRISPR o espaciadores CRISPR que son complementarios u homólogos a por lo menos un espaciador pseudo-CRISPR se usan en combinación con uno o más genes o proteínas *cas* y/o uno o más repeticiones CRISPR (por ejemplo, una o más combinaciones funcionales de las mismas), para diseñar una célula receptora, tal que la resistencia de la célula receptora se module contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

En algunas modalidades, el espaciador pseudo CRISPR o espaciadores CRISPR que son complementarios u homólogos a uno o más espaciadores pseudo-CRISPR se insertan en el plásmido y/o ADN genómico de una célula receptora usando cualquier método conveniente conocido en el arte.

En algunas modalidades adicionales, los espaciadores pseudo-CRISPR se usan como una plantilla a partir de la cual se modifica (por ejemplo, muta) el plásmido y/o ADN genómico de una célula receptora, tal que los espaciadores CRISPR se creen en el plásmido y/o ADN genómico de la célula. En algunas modalidades adicionales, el espaciador pseudo-CRISPR o espaciadores CRISPR que son complementarios u homólogos a uno o más espaciadores pseudo-CRISPR se clonan en un constructo, plásmido y/o vector, etc., son introducidos en la célula hospedadora usando cualquier método conveniente conocido en el arte.

CAS y Genes *cas*

Como se usa en la presente, el término “gen *cas*” tiene el significado convencional como se ha usado en el arte y se refiere a uno o más genes *cas* que son generalmente acoplados, asociados o cerca de o en vecindad del flanco de las ubicaciones CRISPR. Una revisión exhaustiva de la familia de la proteína *Cas* se presenta por *Haft et al.* (*Haft et al.*, PLoS. Comput. Biol., 1(6): e60 [2005]) en el que se describen 41 familias del gen recientemente reconocido asociado a CRISPR (*cas*), además de las cuatro familias del gen previamente conocidas. Como se indica en esta, los sistemas CRISPR pertenecen a clases diferentes, con diferentes patrones de repetición, conjuntos de genes, e intervalos de la especie. Como se indica en la presente, el número de genes *cas* en una ubicación CRISPR dada puede variar entre las especies.

En algunas modalidades, la presente invención proporciona métodos y composiciones para el uso de uno o más genes o proteínas *cas*, solos o en cualquier combinación con uno o más espaciadores CRISPR para modular la resistencia en una célula (por ejemplo, una célula receptora) contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

En algunas modalidades, uno o más de los genes y/o proteínas *cas* se presentan naturalmente en una célula receptora y uno o más espaciadores heterólogos se integra/son integrados o insertados adyacentes a uno o más de los genes o proteínas *cas*. En algunas modalidades, uno o más de los genes y/o proteínas *cas* es/son heterólogas a la célula receptora y uno o más de los espaciadores es/son homólogos o heterólogos. En algunas modalidades preferidas, los espaciadores están integrados o insertados adyacentes a uno o más de los genes o proteínas *cas*.

En algunas modalidades adicionales, la presente invención proporciona métodos y composiciones para el uso de uno o más genes o proteínas *cas* y por lo menos dos repeticiones CRISPR para modular la resistencia en una célula (por ejemplo, una célula receptora) contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

En otras modalidades adicionales, la presente invención proporciona métodos y composiciones para el uso de uno o más genes o proteínas *cas*, por lo menos dos repeticiones CRISPR y por lo menos un espaciador CRISPR para modular la resistencia en una célula (por ejemplo, una célula receptora) contra un ácido nucleico objetivo o un producto de transcripción del mismo.

Las estructuras CRISPR se encuentran típicamente en la vecindad de cuatro genes nombrados *cas1* a *cas4*. El arreglo más común de estos genes es *cas3-cas4-cas1-cas2*. La proteína *cas3* parece ser una helicasa, considerando que *Cas4* se parece a la familia RecB de las exonucleasas y contiene una porción rica en cisteína, sugestiva del enlace del ADN. *Cas1* es generalmente muy básica y es la única proteína *Cas* encontrada de forma consistente en todas las especies que contienen las ubicaciones CRISPR. *Cas2* aún no se ha caracterizado, *cas1-4* son caracterizados típicamente por su proximidad cercana a las ubicaciones CRISPR y su distribución amplia a través de las especies bacterianas y arqueobacterianas. Aunque no todos los genes *cas1-4* se asocian con todos las ubicaciones CRISPR, todos se encuentran en los subtipos múltiples.

Además, existen otros grupos de tres genes asociados con las estructuras CRISPR en muchas especies bacterianas, referidas en la presente como *cas1 B*, *cas5* y *cas6* (*Ver, Bolotin et al.*, [2005], *supra*). Se nota que la nomenclatura de los genes *cas* está en el flujo. Así, el texto en la presente debe tomarse en el contexto. En algunas modalidades preferidas, el gen *cas* se selecciona del *cas1*, *cas2*, *cas3*, *cas4*, *cas 1B*, *cas5* y/o *cas6*. En algunas modalidades preferidas, el gen del *cas* es *cas1*. En aún otras modalidades, el gen *cas* se selecciona de los fragmentos *cas1*, *cas2*, *cas3*, *cas4*, *cas1B*, *cas5* y/o *cas6*, variantes, homólogos y/o derivados de los mismos. En algunas modalidades adicionales, una combinación de dos o más genes *cas* encuentra uso, que incluye cualquier combinación adecuada, incluyendo aquéllas proporcionadas en WO 07/025097. En algunas modalidades, se proporciona una pluralidad de genes *cas*. En algunas modalidades, existe una pluralidad de diferentes y/o mismos genes *cas*, o cualquier combinación de los mismos, proporcionados en la WO 07/025097.

En algunas modalidades, los genes *cas* comprenden ADN, mientras que en otras modalidades, los *cas* comprenden ARN. En algunas modalidades, el ácido nucleico es de origen genómico, mientras que en otras modalidades, es de origen sintético u origen recombinante. En algunas modalidades, los genes *cas* son de doble hebra o de una sola hebra que representan ya sea la hebra sentido o antisentido o combinaciones de la misma. En algunas modalidades, los genes *cas* son preparados por el uso de técnicas de ADN recombinantes (por ejemplo, ADN recombinante), como se describió en la presente.

Como se describió en la presente, en algunas modalidades, el gen *cas* comprende un fragmento de un gen *cas* (es decir, este fragmento del gen *cas* comprende una porción de una secuencia del tipo natural). En algunas modalidades, la secuencia comprende por lo menos aproximadamente 30%, por lo menos aproximadamente 40%, por lo menos aproximadamente 50%, por lo menos aproximadamente 60%, por lo menos aproximadamente 65%, por lo menos aproximadamente 70%, por lo menos aproximadamente 75%, por lo menos aproximadamente 80%, por lo menos aproximadamente 85%, por lo menos aproximadamente 90%, por lo menos aproximadamente 95%, por lo menos aproximadamente 96%, por lo menos aproximadamente 97%, por lo menos aproximadamente 98%, o por lo menos aproximadamente 99% de la secuencia del tipo natural.

En algunas modalidades se prefiere que el gen *cas* sea el gen *cas* que está más cercano a la secuencia líder o la primera repetición CRISPR en el extremo 5' de la ubicación CRISPR – tal como *cas4* o *cas6*.

En algunas modalidades, la proteína *Cas* se selecciona de *Cas1*, *Cas2*, *Cas3*, *Cas4*, *Cas1 B*, *Cas5* y/o *Cas6*, así como los fragmentos, variantes, homólogos y/o derivados de los mismos. En algunas modalidades adicionales, la proteína *Cas* se selecciona de *Cas1*, *Cas2*, *Cas3*, *Cas4*, *Cas1B*, *Cas5* y/o *Cas6*, y combinaciones de los mismos, como se describió en la WO 07/025097. En aún otras modalidades adicionales, la proteína *Cas* se selecciona de uno o más de *Cas1*, *Cas2*, *Cas3*, *Cas4*, *Cas1 B*, *Cas5* y/o *Cas6* o una pluralidad de mismas y/o diferentes proteínas *Cas*, en cualquier número conveniente y/o combinación.

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

En algunas modalidades, las proteínas *Cas* se codifican por genes *cas* que comprenden ADN, mientras que en otras modalidades, las *cas* comprenden ARN. En algunas modalidades, los ácidos nucleicos son de origen genómico, mientras que en otras modalidades, es de origen sintético u origen recombinante. En algunas modalidades, los genes *cas* que codifican a las proteínas *Cas* son de doble hebra o de una sola hebra ya sea que representen la hebra sentido o antisentido o combinaciones de las mismas. En algunas modalidades, los genes *cas* son preparados mediante el uso de técnicas de ADN recombinantes (por ejemplo, el ADN recombinante), como se describe en la presente.

Se proporcionan métodos para identificar un gen *cas* para el uso en la modulación de la resistencia de una célula contra un ácido nucleico objetivo o producto de transcripción del mismo que comprende las etapas de: preparar una célula que comprende por lo menos un espaciador CRISPR y por lo menos dos repeticiones CRISPR; diseñando la célula tal que comprenda por lo menos un gen *cas*; y determinando si la célula modula la resistencia contra el ácido nucleico objetivo o producto de la transcripción del mismo, en donde la modulación de la resistencia de la célula contra los ácidos nucleicos objetivo o producto de la transcripción del mismo es indicativa de que el gen *cas* es útil para modular la resistencia de la célula.

Se proporcionan métodos y uno o más de los genes *cas* útiles en el diseño de células (por ejemplo, células receptoras). En algunas modalidades preferidas, se usan uno o más genes *cas* para diseñar una célula (por ejemplo, una célula receptora), que en combinación con uno o más, preferentemente, dos o más repeticiones CRISPR y uno o más espaciadores CRISPR encuentran uso en la modulación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de transcripción del mismo. Por ejemplo, en algunas modalidades, los genes *cas* son insertados en el ADN de una célula (por ejemplo, plásmido y/o ADN genómico de una célula receptora) usando cualquier método conveniente conocido en el arte. En algunas modalidades adicionales, los genes *cas* se usan como una plantilla a partir de la cual se modifica (por ejemplo, mutar) el ADN de una célula (por ejemplo, plásmido y/o ADN genómico de una célula receptora), tal que los genes *cas* se crean o forman en el ADN de la célula. En algunas modalidades, los genes *cas* están presentes en por lo menos un constructo, por lo menos un plásmido, y/o por lo menos un vector que son introducidos en la célula, usando cualquier método conveniente conocido en el arte.

En algunas modalidades, los genes *cas* comprenden por lo menos un grupo *cas* seleccionado de cualquiera de una o más SEQ ID NOS: 461, 466, 473, 478, 488, 493, 498, 504, 509, y 517. En modalidades adicionales, los genes *cas* comprenden cualquiera de una o más de las SEQ ID NOS: 462-465, 467-472, 474-477, 479-487, 489-492, 494-497, 499-503, 505-508, 510-517, usadas exclusivamente o juntas en cualquier combinación conveniente. En algunas modalidades preferidas, los grupos son usados en combinación con uno o más, preferentemente, dos o más repeticiones CRISPR y opcionalmente uno o más espaciadores CRISPR. En algunas modalidades adicionales, uno o más genes o proteínas *cas* son usados en combinaciones convenientes.

Como se indica en la presente, un conjunto dado de genes o proteínas *cas* siempre está asociado con una secuencia repetida dada dentro de una ubicación CRISPR particular. Así, genes o proteínas *cas* parecen ser específicas para una repetición de ADN dado (es decir, genes o proteínas *cas* y la secuencia repetida forman un par funcional).

Por consiguiente, las combinaciones particulares de uno o más genes o proteínas *cas* y uno o más, preferentemente, dos o más repeticiones CRISPR se usan con objeto de que un espaciador CRISPR confiera resistencia contra un ácido nucleico objetivo o producto de la transcripción del mismo en una célula (por ejemplo, una célula receptora). Por consiguiente, se ha encontrado sorprendentemente que no es posible usar simplemente cualquier gen o proteína *cas* o cualquier repetición CRISPR. En cambio, es una característica de la presente invención que la combinación sea funcional.

En el contexto de la combinación de la proteína o gen *cas* y las repeticiones CRISPR descrita en la presente, el término “funcional” significa que la combinación es capaz de conferir resistencia a un ácido nucleico objetivo o un producto de la transcripción del mismo cuando se usó junto con un espaciador CRISPR que se alinea con o es homólogo a un ácido nucleico objetivo o producto de la transcripción del mismo. Como se usa en la presente, los términos “combinación *cas* y la repetición CRISPR funcional” y “combinación del gen *cas* y la repetición CRISPR funcional” incluye una combinación funcional en la que *cas* es un gen *cas* o una proteína *Cas*.

Adecuadamente, uno o más genes o proteínas *cas* y/o uno o más, preferentemente, dos o más repeticiones CRISPR se derivan de la misma célula (por ejemplo, la misma célula receptora). En algunas modalidades, el término “derivable” es sinónimo del término “obtenible,” como se usó en el contexto. En algunas modalidades preferidas, el término “derivable” también es sinónimo de “derivado”, como se usó en el contexto, debido a que no se pretende que la presente invención se limite específicamente a elementos que se “deriven”. En algunas modalidades, el término “derivado” es sinónimo del término “obtenido,” como se usó en el contexto.

En algunas modalidades, uno o más genes o proteínas *cas* y/o uno o más, preferentemente, dos o más repeticiones CRISPR se derivan de la misma ubicación CRISPR dentro de un genoma o plásmido, preferentemente un genoma o plásmido de la misma cepa, especie o género. En algunas modalidades adicionales, uno o más genes o proteínas *cas* y/o uno o más, preferentemente, dos o más repeticiones CRISPR se derivan de la misma ubicación CRISPR dentro de un solo genoma o plásmido, preferentemente, un solo genoma o plásmido de la misma cepa, especie o género. En algunas modalidades, uno o más genes o proteínas *cas* y uno o más, preferentemente, dos o más repeticiones CRISPR que se co-presentan naturalmente. En modalidades adicionales, uno o más genes o proteínas *cas* y uno o más, preferentemente, dos o más repeticiones CRISPR se co-presentan naturalmente en la misma célula (por ejemplo, la célula receptora). En modalidades adicionales, uno o más genes o proteínas *cas* y uno o más, preferentemente, dos o más repeticiones CRISPR se co-presentan naturalmente en el mismo genoma de una célula (por ejemplo, la célula receptora). Aún en modalidades adicionales, uno o más genes o proteínas *cas* y uno o más, preferentemente, dos o más repeticiones CRISPR se co-presentan naturalmente en el mismo genoma de una cepa, especies o géneros. En algunas modalidades preferidas adicionales, la presente invención proporciona cualquier combinación conveniente de ácidos nucleicos que consisten esencialmente en por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas*.

En algunas modalidades, el término “consiste esencialmente de” se refiere a una combinación de por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas* excluyendo por lo menos a un componente adicional de una ubicación CRISPR (por ejemplo, la ausencia de uno o más espaciadores CRISPR) y/o la ausencia de una o más secuencias líder comunes de una ubicación CRISPR. En algunas modalidades alternativas, el término “consiste esencialmente de” se refiere a una combinación de por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas* solamente, que excluye a todos los otros componentes de una ubicación CRISPR (por ejemplo, una ubicación CRISPR que se presenta naturalmente). En algunas modalidades adicionales, el término “consiste esencialmente de” se refiere a una combinación de por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas* solamente, que excluye a por lo menos un componente adicional de una ubicación CRISPR, que excluye preferentemente a por lo menos un componente adicional de una ubicación CRISPR que se presenta naturalmente. Incluso, en alguna modalidad adicional, el término “consiste esencialmente de” se refiere a una combinación de por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas*, con la condición de que por lo menos un componente adicional de la ubicación CRISPR natural esté ausente (por ejemplo, substancialmente ausente). Así, se pretende que el término encuentre un uso en el contexto. En algunas modalidades, la presente invención proporciona cualquier combinación conveniente de por lo menos dos repeticiones CRISPR y por lo menos un gen o proteína *cas*, con la condición que todos los otros componentes de la ubicación CRISPR estén ausentes (por ejemplo, substancialmente ausentes), preferentemente que todos los otros componentes de la ubicación CRISPR de la combinación natural de repeticiones CRISPR y genes *cas* estén ausentes. En algunas modalidades adicionales, se usan uno o más genes o proteínas *cas* en combinación o junto con por lo menos uno o más espaciadores CRISPR. En algunas modalidades adicionales, se usan uno o más genes o proteínas *cas* en combinación o junto con por lo menos uno o más espaciadores CRISPR y por lo menos, uno o más, preferentemente, dos o más repeticiones CRISPR. En algunas modalidades, los espaciadores CRISPR se o son derivados de un organismo (por ejemplo, un organismo donador) que es diferente a la célula (por ejemplo, la célula receptora) a partir de la cual uno o más genes o proteínas *cas* y/o uno o más, preferentemente, dos o más repeticiones CRISPR se derivan.

Se proporcionan varias disposiciones de repeticiones CRISPR y genes o proteínas *cas*, combinaciones repetición *cas*-CRISPR funcionales. En algunas modalidades, la combinación comprende, consiste o consiste esencialmente de por lo menos cualquiera de aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12, aproximadamente 13, aproximadamente 14, aproximadamente 15, aproximadamente 16, aproximadamente 17, aproximadamente 18, aproximadamente 19, o aproximadamente 20 repeticiones CRISPR en combinación con cualquiera de aproximadamente 1, aproximadamente 2, aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 11, aproximadamente 12, aproximadamente 13, aproximadamente 14, aproximadamente 15, aproximadamente 16, aproximadamente 17, aproximadamente 18, aproximadamente 19, o aproximadamente 20 genes o proteínas *cas* (por ejemplo, 16 repeticiones CRISPR y 12 genes o proteínas *cas* o 18 repeticiones CRISPR

y 20 genes o proteínas cas o cualquier otra combinación de las mismas). La presente invención proporciona repeticiones CRISPR y genes cas dispuestos de varias maneras, como se establece en la WO 07/025097. En algunas modalidades en las que la combinación de un gen cas y una repetición CRISPR comprende más de un gen cas, se entenderá que las repeticiones CRISPR se insertan en el extremo 3' de los genes cas, el extremo 5' de los genes cas, o entre los genes cas, con tal de que por lo menos uno de los genes cas permanezcan en funcionamiento.

En algunas modalidades, una primera combinación de proteína o gen cas y la repetición CRISPR (que comprende por lo menos un gen o proteína cas y por lo menos dos repeticiones CRISPR, en las que ambas se derivan de la misma ubicación CRISPR dentro de un genoma) se usa en combinación con una segunda combinación de proteína o gen cas y la repetición CRISPR (que comprende por lo menos un gen o una proteína cas y por lo menos dos repeticiones CRISPR, en donde ambos se derivan de la misma o diferente ubicación CRISPR dentro de un genoma). Por consiguiente, en estas modalidades de la invención, las primeras y segundas combinaciones se derivan de la misma o diferente ubicación CRISPR dentro de un genoma. Así, en algunas modalidades, la primera y segunda combinaciones de proteína o gen cas y la repetición CRISPR son de diferentes genomas (por ejemplo, de genomas diferentes dentro del mismo grupo), como se describió a detalle adicionalmente en la presente.

Aún en modalidades adicionales de la presente invención, una primera y/o una segunda combinación de proteína o gen cas y la repetición CRISPR (que comprende por lo menos un gen cas y por lo menos dos repeticiones CRISPR derivadas de la misma ubicación CRISPR dentro de un genoma) se usa en la combinación con aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9, o aproximadamente 10 o más combinaciones de la proteína o gen cas y la repetición CRISPR (que comprende cada una de por lo menos un gen o proteína cas y por lo menos dos repeticiones CRISPR derivadas del mismo o una ubicación CRISPR diferente dentro de un genoma). Por consiguiente, en estas modalidades de la invención, las combinaciones se derivan de los mismos o diferentes ubicaciones CRISPR dentro de un genoma. En algunas modalidades adicionales de la invención, las combinaciones son de genomas diferentes (por ejemplo, genomas diferentes dentro del mismo grupo), como se describió a detalle adicionalmente en la presente.

Así, en algunas modalidades, para que la combinación de la proteína o gen cas y la repetición CRISPR confiera resistencia, las proteína(s) o genes cas y repeticiones CRISPR se co-presentan naturalmente dentro de una ubicación CRISPR dada de un genoma. En algunas modalidades, la repeticiones CRISPR y proteína(s) o genes cas se co-presentan naturalmente dentro de la misma ubicación CRISPR de un genoma. En algunas modalidades, estas combinaciones funcionales tomadas juntas, confieren resistencia contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

En algunas modalidades adicionales, la presente invención emplea métodos para identificar una combinación funcional de una proteína o gen cas y una repetición CRISPR que comprende las etapas de: analizar las secuencias (por ejemplo, secuencias de los ácidos nucleicos o proteínas) del gen o proteína cas y la repetición CRISPR; identificando uno o más grupos de genes o proteínas cas; identificando uno o más grupos de repeticiones CRISPR; y combinando esos genes o proteínas cas y secuencias de repeticiones CRISPR que caen dentro del mismo grupo.

En algunas modalidades adicionales, la presente invención emplea los métodos para identificar una combinación funcional de un gen o proteína cas y una repetición CRISPR para usarse en la modulación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo, que comprende las etapas de: preparar una célula que comprende una combinación de uno o más genes o proteínas cas y una o más, preferentemente, dos o más repeticiones CRISPR; diseñar la célula tal que contenga uno o más espaciadores CRISPR; y determinar si la célula modula la resistencia contra un ácido nucleico objetivo, en donde la modulación de la resistencia de la célula contra los ácidos nucleicos objetivo o un producto de la transcripción de los mismos es indicativo de que la combinación puede usarse para modular la resistencia de la célula contra los ácidos nucleicos objetivo.

En algunas modalidades, las secuencias del gen y/o proteína cas y/o las repeticiones CRISPR son o derivan de las mismas o diferentes cepas, especies, géneros, y/u organismos. En algunas modalidades, la combinación comprende ADN y/o ARN de origen genómico, recombinante, y/o sintético. En algunas modalidades, las repeticiones CRISPR comprenden ADN y/o ARN de origen genómico, recombinante y/o sintético. En algunas modalidades, los genes cas comprenden ADN y/o ARN de origen genómico, recombinante y/o sintético. De hecho, se pretende que la presente invención abarque cualquier combinación de ADN y/o ARN para cada uno de los elementos (por ejemplo, el gen cas y/o las repeticiones CRISPR). En algunas modalidades, los elementos se analizan usando cualquier método conveniente conocido en el arte. En algunas modalidades preferidas, el análisis se conduce usando el análisis dotplot (grafica de puntos). En algunas modalidades, el gen cas y/o repeticiones CRISPR son de hebra doble, mientras que en otras modalidades, ambas son de una sola hebra, ya representen la hebra sentido o antisentido o combinaciones de las mismas.

En algunas modalidades, una o más de las combinaciones funcionales descritas en la presente se usan para diseñar una célula (por ejemplo, una célula receptora). En algunas modalidades preferidas, se usan una o más

combinaciones funcionales para diseñar una célula (por ejemplo, una célula receptora), que en combinación con uno o más espaciadores CRISPR encuentran un uso en la modulación de la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo. En algunas modalidades, las combinaciones funcionales se insertan en el ADN de una célula receptora (por ejemplo, como un plásmido o un ADN genómico de una célula) usando cualquier método conveniente conocido en el arte. En algunas modalidades adicionales, las combinaciones funcionales se usan como una plantilla después a partir de la cual se modifica (por ejemplo, mutar) el ADN de una célula receptora (por ejemplo, ADN del plásmido o ADN genómico), tal que se crean combinaciones funcionales en el ADN de la célula. Aún en modalidades adicionales, las combinaciones funcionales se clonan en un constructo, plásmido, o vector, etc., que se transforma después en la célula, usando métodos tales como aquéllos descritos en la presente y conocidos en el arte.

En algunas modalidades, la combinación funcional se obtiene o es obtenible mediante un método que comprende las etapas de: analizar las secuencias de un gen *cas* y una repetición CRISPR; identificar uno o más grupos de genes *cas*; identificar uno o más grupos de repeticiones CRISPR; y combinar esos genes *cas* y secuencias de repeticiones CRISPR que caen dentro del mismo grupo, en donde la combinación del gen *cas* y secuencias de repetición CRISPR dentro del mismo grupo es indicativo de que la combinación es una combinación funcional.

Como se indicó anteriormente, se encontró sorprendentemente que no es posible intercambiar simplemente combinaciones de repetición CRISPR-*cas* entre cualquier célula (por ejemplo, cualquier cepa, especies o géneros de células), ya que esto no resulta necesariamente en combinaciones de repetición CRISPR-*cas* funcional. De hecho, para que las combinaciones de repetición CRISPR-*cas* sean funcionales, necesitan ser compatibles. Así, se contempla que no es posible cambiar los genes *cas* o repeticiones CRISPR entre diferentes ubicaciones CRISPR a menos que sean del mismo grupo. Aun más sorprendente es que los grupos no siguen la filogenia del "organismo". Específicamente, dentro de un organismo, puede haber más de un CRISPR. Estos CRISPR pueden pertenecer a grupos diferentes, aunque estén presentes en el mismo organismo. Como resultado, se cree que una combinación de repetición CRISPR-*cas* funcional requiere que la combinación se cambie dentro de un grupo como opuesto dentro de un organismo.

Para evitar dudas, el término "grupo" como se usa en la presente no se refiere a un grupo de genes localizados en la misma ubicación (que forman típicamente un operon) sino al rendimiento del análisis de comparación de la secuencia (por ejemplo, el análisis de comparación de secuencia múltiple y/o alineaciones de la secuencia múltiples y/o análisis de gráfica de puntos). Por consiguiente, en algunas modalidades, el análisis del grupo de ubicaciones CRISPR se realiza usando varios métodos que son conocidos en el arte (por ejemplo, tal como el análisis de gráfica de puntos como el descrito en la presente) o la alineación múltiple seguida por el cálculo del dendrograma. En algunas modalidades, el grupo es una clase, una familia o un grupo de secuencias.

Ventajosamente, el uso de combinaciones de repetición CRISPR-*cas* que se co-presentan naturalmente proporciona para el intercambio de la combinación, ambos dentro y entre una especie dada, haciendo posible diseñar la resistencia de una cepa usando la combinación de una cepa diferente.

Bacteriófago

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

Un bacteriófago lítico es uno que sigue la trayectoria lítica a través de la realización del ciclo lítico, en lugar de ingresar a la trayectoria lisogénica. Un bacteriófago lítico experimenta una replicación vírica que conduce a la lisis de la membrana celular, destrucción de la célula, y liberación de partículas bacteriófagas de la progenie capaces de infectar otras células.

Un bacteriófago lisogénico es aquel con capacidad para ingresar en la trayectoria lisogénica, en la que el bacteriófago se convierte en una parte pasiva, inactiva, del genoma de la célula a través de una previa conclusión de su ciclo lítico.

Los bacteriófagos que encuentran uso en la presente invención incluyen, pero no se limita a bacteriófagos que pertenecen a cualquiera de las siguientes familias de virus: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, o Tectiviridae. En algunas modalidades, el bacteriófago que infecta bacterias que son patógenicas para las plantas y/o animales (incluso a humanos) encuentran un uso particular. En algunas modalidades particularmente preferidas, se modula la resistencia de una célula contra un bacteriófago.

En algunas modalidades particularmente preferidas, los bacteriófagos de la presente invención incluyen, pero no se limitan a, aquellos bacteriófagos capaces de infectar una bacteria que naturalmente comprende una o más ubicaciones CRISPR. Las ubicaciones CRISPR se han identificado en más de 40 procariotas (Ver por ejemplo, Jansen *et al.*, *Mol. Microbiol.*, 43: 1565-1575 [2002]; y Mojica *et al.*, [2005]) que incluyen, pero que no se limita a

5 *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema* y *Thermotoga*.

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

Aún en modalidades adicionales, los bacteriófagos incluyen, pero no se limitan a, aquellos bacteriófagos capaces de infectar (o transducir) bacterias de ácido láctico, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (por ejemplo, *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc*, y *Oenococcus*.

Aún en modalidades adicionales, los bacteriófagos incluyen, pero no se limitan a, aquellos bacteriófagos capaces de infectar *Lactococcus lacti* (por ejemplo, *L. lactis* subsp. *lactis* y *L. lactis* subsp. *cremoris*, y *L. lactis* subsp. *lactis* biovar *diacetylactis*), *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*, *Bifidobacterium infantis*, *Lactobacillus paracasei*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus reuteri*, *Lactobacillus gasseri*, *Lactobacillus johnsonii* o *Bifidobacterium longum*.

Aún en modalidades adicionales, los bacteriófagos incluyen, pero no se limitan a, aquellos bacteriófagos capaces de infectar cualquier bacteria fermentativa susceptible a una alteración mediante una infección del bacteriófago, que incluye pero que no se limita a procesos para la producción de antibióticos, aminoácidos, y solventes. Productos producidos por fermentación que se sabe que experimentan una infección bacteriófaga, y las bacterias de fermentación infectadas correspondientes, incluidas las de los tipos de queso Cheddar y el queso cottage (*Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris*), yogur (*Lactobacillus delbrueckii* subsp. *bulgaricus*, *Streptococcus thermophilus*), queso suizo (*S. thermophilus*, *Lactobacillus lactis*, *Lactobacillus helveticus*), queso azul (*Leuconostoc cremoris*), queso italiano (*L. bulgaricus*, *S. thermophilus*), vili (*Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *lactis* biovar *diacetylactis*, *Leuconostoc cremoris*), yakult (*Lactobacillus casei*), caseína (*Lactococcus 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*, *Microbacterium ammoniaphilum*), y acetona y butanol (*Clostridium acetobutylicum*, *Clostridium saccharoperbutylacetonicum*).

En algunas modalidades preferidas, las bacterias que encuentran uso en la presente invención incluyen, pero no se limitan a *S. thermophilus*, *L.*, *delbrueckii*. subsp. *bulgaricus* y/o *L. acidophilus*.

En algunas modalidades particularmente preferidas, los bacteriófagos incluyen, pero no se limitan a, aquellos bacteriófagos capaces de infectar bacterias que comprenden una o más ubicaciones CRISPR heterólogas. En algunas modalidades, las bacterias comprenden una o más ubicaciones heterólogas CRISPR, y/o uno o más genes cas heterólogos, y/o uno o más repeticiones CRISPR heterólogas, y/o uno o más espaciadores heterólogos CRISPR.

La infección de bacterias por fagos es el resultado de la inyección o transferencia del ADN del fago en las células. En algunas modalidades, la infección conduce a la expresión (es decir, la transcripción y traducción) de los ácidos nucleicos del bacteriófago dentro de la célula y la continuación del ciclo de vida del bacteriófago. En algunas modalidades que involucran al bacteriófago recombinante, también se expresan las secuencias recombinantes dentro del genoma del fago (por ejemplo, los ácidos nucleicos reporteros).

Se ha encontrado que las secuencias espaciadoras CRISPR en los procariotes tienen a menudo similitudes significativas para una variedad de moléculas de ADN, que incluyen tales elementos genéticos como los cromosomas, bacteriófagos, y plásmidos conjugativos. Se ha reportado que células que llevan estos espaciadores CRISPR son incapaces de ser infectados por moléculas de ADN que contienen secuencias homólogas a los espaciadores (Ver, Mojica *et al.*, [2005]).

En algunas modalidades de la presente invención, uno o más particular pseudoespaciadores particulares derivados del ADN bacteriófago o espaciadores CRISPR que son complementarios u homólogos a uno o más espaciadores pseudo-CRISPR se agregan dentro de una ubicación CRISPR de una célula (por ejemplo, una célula receptora) para

modular (por ejemplo, proporcionar) resistencia contra un bacteriófago particular, previniendo así, substancialmente el ataque del fago.

5 En algunas modalidades preferidas, las regiones particulares dentro del genoma del fago son dirigidas para preparar los pseudoespaciadores, que incluyen pero que no se limitan a genes que codifican a proteínas con especificidad hospedadora, que incluyen aquéllas que proporcionan el reconocimiento del hospedador de fago particular, tal como las helicasas, primasas, proteínas estructurales de cabeza o cola, proteínas con un dominio conservado (por ejemplo, huecas, lisina, y otras) o secuencias conservadas entre los genes del fago importantes.

10 Cualquiera de los ácidos nucleicos que se originan del genoma del fago puede conferir inmunidad contra el fago cuando se inserta, por ejemplo, entre dos repeticiones en una ubicación CRISPR activo. En algunas modalidades, la inmunidad es más “eficiente” cuando el espaciador CRISPR corresponde a una secuencia interna de un gen del fago. En algunas modalidades particularmente preferidas, la inmunidad se hace aún más “eficiente” cuando el gen codifica una proteína “esencial” (por ejemplo, el antirreceptor).

15 En algunas modalidades preferidas, la presente invención emplea métodos para conferir resistencia a una célula (por ejemplo, una célula bacteriana) contra un bacteriófago que comprende las etapas de: (a) proporcionar uno o más pseudoespaciadores CRISPR de por lo menos un bacteriófago; (b) identificar una o más combinaciones de repetición CRISPR-cas funcionales en por lo menos una célula que es substancialmente sensible al bacteriófago; y (c) diseñar una o más ubicaciones CRISPR en la célula substancialmente sensible tal que comprenda uno o más espaciadores pseudo CRISPR de un bacteriófago o uno o más espaciadores CRISPR que son complementarios u homólogos a uno o más pseudoespaciadores CRISPR para suministrar resistencia a la célula.

25 Aún en modalidades adicionales, la presente invención emplea métodos para conferir resistencia a una célula (por ejemplo, una célula bacteriana) contra un bacteriófago que comprende las etapas de: (a) proporcionar uno o más pseudoespaciadores CRISPR de por lo menos un bacteriófago; (b) identificar una o más combinaciones de repetición CRISPR-cas funcionales en por lo menos una célula que es substancialmente sensible al bacteriófago; y (c) insertar uno o más pseudoespaciadores CRISPR de los bacteriófagos o uno o más espaciadores CRISPR que son complementarios u homólogos a uno o más pseudoespaciadores CRISPR en la célula substancialmente sensible tal que la célula sea substancialmente resistente al bacteriófago.

30 Aún en modalidades adicionales, la presente invención emplea métodos para modular el lisotipo de una célula bacteriana que comprende las etapas de: (a) proporcionar uno o más pseudoespaciadores CRISPR de por lo menos un bacteriófago; (b) identificar una o más combinaciones de repetición CRISPR-cas funcionales en por lo menos una célula que es substancialmente sensible al bacteriófago; y (c) diseñar una o más ubicaciones CRISPR en la célula substancialmente sensible tal que comprendan uno o más pseudoespaciadores CRISPR de un bacteriófago o uno o más espaciadores CRISPR que son complementarios u homólogos a uno o más pseudoespaciadores CRISPR.

40 Aún en modalidades adicionales, la presente invención emplea métodos para modular el lisotipo de una célula bacteriana que comprende las etapas de: (a) proporcionar uno o más pseudoespaciadores CRISPR de por lo menos un bacteriófago; (b) identificar una o más combinaciones de repetición CRISPR-cas funcionales en por lo menos una célula que es substancialmente sensible al bacteriófago; y (c) insertar uno o más de un pseudoespaciadores CRISPR de los bacteriófago o uno o más espaciadores CRISPR que son complementarios u homólogos a uno o más los pseudoespaciadores CRISPR en la célula substancialmente sensible.

45 En algunas modalidades preferidas adicionales, la presente invención emplea métodos para conferir resistencia a una célula (por ejemplo, una célula bacteriana) contra un bacteriófago que comprende las etapas de: (i) identificar un pseudoespaciador CRISPR en un bacteriófago que comprende un ácido nucleico objetivo o un producto de la transcripción del mismo contra el cual la resistencia será modulada; y (ii) modificar la secuencia del espaciador CRISPR de la célula tal que el espaciador CRISPR de la célula tenga homología para el pseudoespaciador CRISPR del bacteriófago que comprende los ácidos nucleicos objetivo.

50 En algunas modalidades adicionales, la presente invención emplea métodos para conferir resistencia a una célula (por ejemplo, una célula bacteriana) contra un bacteriófago que comprende las etapas de: (i) identificar un pseudoespaciador CRISPR en un bacteriófago que comprende un ácido nucleico objetivo o un producto de la transcripción del mismo contra el cual la resistencia será modulada; y (ii) modificar la secuencia del espaciador CRISPR de la célula tal que el espaciador CRISPR de la célula tenga 100% de homología o identidad para el pseudoespaciador CRISPR del bacteriófago que comprende los ácidos nucleicos objetivo.

60 Aún en modalidades adicionales, la presente invención emplea métodos para modular el lisotipo de una célula bacteriana que comprende las etapas de: (i) identificar un pseudoespaciador CRISPR en un bacteriófago que comprende un ácido nucleico objetivo o un producto de la transcripción del mismo contra el cual la resistencia será modulada; y (ii) modificar la secuencia del espaciador CRISPR de la célula tal que el espaciador CRISPR de la célula tenga homología para el pseudoespaciador CRISPR del bacteriófago que comprende los ácidos nucleicos objetivo.

65

En algunas modalidades adicionales, la presente invención emplea métodos para modular el lisotipo de una célula bacteriana que comprende las etapas de: (i) identificar un pseudoespaciador CRISPR en un bacteriófago que comprende un ácido nucleico objetivo o un producto de la transcripción del mismo contra el cual la resistencia será modulada; y (ii) modificar la secuencia del espaciador CRISPR de la célula tal que el espaciador CRISPR de la célula tenga 100% de homología o identidad para el pseudoespaciador CRISPR del bacteriófago que comprende los ácidos nucleicos objetivo.

En algunas modalidades, el espaciador CRISPR de la célula bacteriana tiene 100% de homología o identidad para una secuencia (por ejemplo, como un pseudoespaciador CRISPR) en el bacteriófago que comprende los ácidos nucleicos objetivo.

En algunas modalidades alternativas, el espaciador CRISPR de la célula bacteriana forma una parte componente de una ubicación CRISPR que comprende una combinación de repetición cas CRISPR funcional como se describe en la presente.

En algunas modalidades particularmente preferidas, los ácidos nucleicos objetivo o un producto de la transcripción de los mismos en el bacteriófago es una secuencia de ácidos nucleicos altamente conservada. En algunas modalidades adicionales, los ácidos nucleicos objetivo o producto de la transcripción del mismo en el bacteriófago es un gen que codifica a una proteína de especificidad hospedadora. En algunas modalidades adicionales, los ácidos nucleicos objetivo o producto de la transcripción del mismo en el bacteriófago codifica una enzima que es esencial para la supervivencia, replicación y/o crecimiento del bacteriófago. Aún en modalidades adicionales, los ácidos nucleicos objetivo o producto de la transcripción de los mismos en el bacteriófago codifican una helicasa, primasa, una proteína estructural de cabeza o cola, o una proteína con un dominio conservado (por ejemplo, holina, lisina, etc.).

En algunas modalidades preferidas, las células bacterianas se preparan para tener "susceptibilidad reducida para una multiplicación o infección del bacteriófago". Como se usa en la presente, este término se refiere a bacterias que tienen baja o ninguna susceptibilidad a la multiplicación o infección del bacteriófago cuando se compara con las bacterias del tipo natural cuando se cultivan (por ejemplo, en un medio lácteo).

En algunas modalidades, algunas células bacterianas exhiben "susceptibilidad baja a la multiplicación del bacteriófago". Este término se refiere a que el nivel de multiplicación del bacteriófago en una bacteria está por debajo de un nivel, lo cual causaría un efecto nocivo a un cultivo en un período dado de tiempo. Tales efectos nocivos en un cultivo incluyen, pero no se limitan a, ninguna coagulación de leche durante la producción de productos de leche fermentados (por ejemplo, yogur o queso), reducción inadecuada o lenta del pH durante la producción de productos de leche fermentados (por ejemplo, yogur o queso), maduración lenta del queso, y/o deterioración de la textura de un alimento al punto en el que sea poco apetecible o impropio para el consumo.

Para un conjunto equivalente de condiciones del cultivo la susceptibilidad bacteriana hacia un bacteriófago de la presente invención se expresa generalmente en comparación con las bacterias del tipo natural. En algunas modalidades, las bacterias tienen aproximadamente 100 veces más baja (eficiencia de formación de placa [EOP] = 10^{-2}), preferentemente aproximadamente 1000 veces más baja (EOP = 10^{-3}), más preferentemente 10,000 veces más baja (EOP = 10^{-4}), y más preferentemente aproximadamente 100,000 veces más baja (EOP = 10^{-5}). En algunas modalidades preferidas, el nivel de multiplicación del bacteriófago en un cultivo se mide después de aproximadamente 14 horas de incubación del cultivo, más preferentemente después de aproximadamente 12 horas, más aun preferentemente después de aproximadamente 7 horas, aún más preferentemente después de aproximadamente 6 horas, aún más preferentemente después de aproximadamente 5 horas, y más preferentemente después de aproximadamente 4 horas.

En modalidades adicionales, la presente invención emplea métodos para conferir la sensibilidad a una célula (por ejemplo, una célula bacteriana) contra un bacteriófago que comprende las etapas de: (a) proporcionar un pseudoespaciador CRISPR de por lo menos un bacteriófago; (b) identificar una o más combinaciones de repetición CRISPR-cas funcionales en una célula que es substancialmente resistente al bacteriófago; y (c) diseñar una o más ubicaciones CRISPR en la célula substancialmente sensible tal que comprendan uno o más pseudoespaciadores CRISPR o uno o más espaciadores CRISPR que son complementarios u homólogos a uno o más pseudoespaciadores CRISPR que tienen un grado reducido de homología comparado con uno o más de las ubicaciones CRISPR en la célula substancialmente resistente.

En modalidades adicionales, la presente invención emplea métodos para modular (por ejemplo, reducir) el lisotipo de una célula (por ejemplo, una célula bacteriana), que comprende uno o más genes o proteínas cas y uno o más, preferentemente, dos o más repeticiones CRISPR que comprenden las etapas de: (i) identificar un pseudoespaciador CRISPR en un bacteriófago contra el cual la resistencia será modulada; y (ii) modificar la secuencia del espaciador CRISPR de la célula tal que el espaciador CRISPR de la célula tenga un grado reducido de homología al pseudoespaciador CRISPR del bacteriófago que comprende los ácidos nucleicos objetivo.

Aún en modalidades adicionales, la presente invención emplea métodos para modular (por ejemplo, reducir o disminuir) la resistencia de una célula bacteriana que comprende uno o más genes o proteínas cas y uno o más, preferentemente, dos o más repeticiones CRISPR contra un bacteriófago que comprende las etapas de: (i) identificar uno o más pseudoespaciadores CRISPR en un bacteriófago contra el cual la resistencia será modulada; (ii)

identificar un espaciador CRISPR en la célula bacteriana en la que la resistencia será modulada, que es homóloga a los pseudoespaciadores CRISPR; y (iii) modificar la secuencia del espaciador CRISPR en la célula bacteriana en la que la resistencia será modulada tal que el CRISPR espaciador tenga un grado más bajo de homología para los pseudoespaciadores CRISPR del bacteriófago contra el cual la resistencia será modulada.

En algunas modalidades, el espaciador CRISPR de la célula tiene un grado reducido de homología (por ejemplo, aproximadamente 1, aproximadamente 2, aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9, aproximadamente 10, aproximadamente 15, aproximadamente 20, aproximadamente 25, aproximadamente 30, aproximadamente 35, aproximadamente 40, aproximadamente 45, aproximadamente 50, aproximadamente 55, aproximadamente 60, aproximadamente 65, aproximadamente 70, aproximadamente 75, aproximadamente 80, aproximadamente 90, o aproximadamente 95% de reducción en homología) comparado con los pseudoespaciadores CRISPR del bacteriófago contra el cual la resistencia será modulada.

En algunas modalidades, las células bacterianas se preparan usando métodos de la presente invención, tal que las células tengan una “susceptibilidad aumentada para la multiplicación del bacteriófago”. Como se usa en la presente, este término se refiere a bacterias que tienen un aumento o susceptibilidad superior para la multiplicación del bacteriófago cuando se compararon con las bacterias del tipo natural cuando se cultivan (por ejemplo, en un medio lácteo).

En algunas modalidades, el término “susceptibilidad alta para la multiplicación del bacteriófago” se refiere al nivel de multiplicación del bacteriófago en una bacteria que está por arriba de un nivel que causaría un efecto nocivo a un cultivo en un período dado de tiempo. Tales efectos nocivos en un cultivo incluyen, pero no se limitan a, ninguna coagulación de leche durante la producción de productos de leche fermentados (por ejemplo, yogur o queso), reducción lenta o inadecuada del pH durante la producción de productos de leche fermentados (por ejemplo, yogur o queso), maduración lenta del queso, y/o deterioración de la textura de un alimento al punto en el que sea poco apetecible o impropio para el consumo.

Para un conjunto equivalente de condiciones del cultivo la susceptibilidad bacteriana hacia un bacteriófago de la presente invención se expresa generalmente en comparación con las bacterias del tipo natural. En algunas modalidades, las bacterias tienen aproximadamente 100 veces más baja (eficiencia de formación de placa [EOP] = 10^{-2}), preferentemente aproximadamente 1000 veces más baja (EOP = 10^{-3}), más preferentemente 10,000 veces más baja (EOP = 10^{-4}), y más preferentemente aproximadamente 100,000 veces más baja (EOP = 10^{-5}). En algunas modalidades preferidas, el nivel de multiplicación del bacteriófago en un cultivo se mide después de aproximadamente 14 horas de incubación del cultivo, más preferentemente después de aproximadamente 12 horas, aún más preferentemente después de aproximadamente 7 horas, aún más preferentemente después de aproximadamente 6 horas, aún más preferentemente después de aproximadamente 5 horas, y más preferentemente después de aproximadamente 4 horas.

En algunas modalidades preferidas, un espaciador CRISPR se flanquea por dos repeticiones CRISPR (es decir, un espaciador CRISPR tiene por lo menos una repetición CRISPR a cada lado).

En algunas modalidades de la presente invención, la bacteria precursora (por ejemplo, la “cepa bacteriana precursora”) se expone (por ejemplo, iterativamente, secuencialmente, simultáneamente o substancialmente simultáneamente) a más de un bacteriófago (por ejemplo, una mezcla de uno o más fagos). En algunas modalidades, la cepa bacteriana precursora es sensible a cada uno de los bacteriófagos que se exponen en la mezcla, mientras que en otras modalidades, la cepa bacteriana es sensible a alguno de los bacteriófagos, pero es resistente a otros.

Como se usa en la presente, el término “secuencias de marcado” se refiere a la porción de un fragmento de ADN adicional que se deriva del genoma de uno o más bacteriófagos (por ejemplo, la mejor hebra del genoma de uno o más bacteriófagos) en el que la bacteria precursora se expone de acuerdo con métodos de la presente invención y se usa como una etiqueta o una marca (por ejemplo, proporcionando una única etiqueta o una única marca).

Las secuencias de marcado son típicamente una secuencia que es una secuencia que se presenta naturalmente en el bacteriófago. Preferentemente, las secuencias de marcado tienen por lo menos aproximadamente 90%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98% o aproximadamente 99% de identidad para la secuencia que se presenta naturalmente en el bacteriófago (por ejemplo, el genoma del bacteriófago del cual se deriva). En algunas modalidades más preferidas, las secuencias de marcado tienen aproximadamente 100% de identidad para la secuencia que se presenta naturalmente en el bacteriófago (por ejemplo, el genoma del bacteriófago del cual se deriva).

En algunas modalidades, las secuencias de marcado tienen 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 para cualquiera de otros espaciadores CRISPR o centro espaciador CRISPR en una o más ubicaciones CRISPR de la bacteria etiquetada.

En algunas modalidades, las secuencias de marcado 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 para cualquier otra secuencia en una o más ubicaciones CRISPR de la bacteria etiquetada.

En algunas modalidades adicionales, las secuencias de marcado tienen una secuencia que es idéntica a una secuencia (por ejemplo, como un espaciador CRISPR) en la ubicación CRISPR de la bacteria. En algunas modalidades adicionales, las secuencias de marcado tienen una secuencia que es idéntica a una secuencia (por ejemplo, un espaciador CRISPR) en la ubicación CRISPR de la bacteria aparte de una o más polimorfismos del nucleótido simples (por ejemplo, uno o dos polimorfismos del nucleótido simples).

En algunas modalidades preferidas, las secuencias de marcado tienen por lo menos aproximadamente 20 nucleótidos en longitud, mientras que en algunas modalidades particularmente preferidas es de aproximadamente 20 hasta aproximadamente 58 nucleótidos en longitud.

En algunas modalidades particularmente preferidas, por lo menos una secuencia de marcado se integra en la bacteria precursora. En algunas modalidades adicionales, también se integra por lo menos una secuencia duplicada (por ejemplo, una secuencia de repetición duplicada CRISPR) derivada del genoma de la bacteria precursora o uno o más plásmidos de la bacteria precursora (por ejemplo, megaplásmidos). No se pretende que la presente invención se limite a cualquier mecanismo particular o teoría. Sin embargo, se cree que por lo menos una secuencia duplicada se copie o replique desde el genoma de la bacteria precursora. En particular, se cree típicamente que la secuencia de repetición CRISPR en una ubicación CRISPR se duplica y la secuencia de marcado se integra en el genoma de la bacteria inmediatamente después (es decir, en corriente abajo) de las nuevas repeticiones CRISPR duplicadas.

En algunas modalidades particularmente preferidas, por lo menos una secuencia duplicada es una secuencia de repetición CRISPR que tiene por lo menos aproximadamente 90%, aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98%, o aproximadamente 99% de identidad para las repeticiones CRISPR en una o más ubicaciones CRISPR de la bacteria precursora y/o bacteria etiquetada. Más preferentemente, por lo menos una secuencia duplicada es una secuencia de repetición CRISPR que tiene por lo menos aproximadamente 100% de identidad para las repeticiones CRISPR en una o más ubicaciones CRISPR de la bacteria precursora y/o bacteria etiquetada. En algunas modalidades preferidas, la secuencia duplicada es por lo menos de aproximadamente 24 nucleótidos en longitud, mientras que en algunas modalidades particularmente preferidas es de aproximadamente 24 hasta aproximadamente 40 nucleótidos en longitud.

En algunas modalidades preferidas, por lo menos una secuencia de marcado y por lo menos una secuencia duplicada se integran en la bacteria precursora. No se pretende que la presente invención se limite a cualquier mecanismo particular o teoría. Sin embargo, se cree que cada vez que, una secuencia de marcado se integra en el genoma de la bacteria precursora, esta se acompaña por la integración iterativa, secuencial, simultánea o substancialmente simultánea de por lo menos una secuencia duplicada. Por consiguiente, por lo menos un par de secuencias que comprenden las secuencias de marcado y la secuencia duplicada se integran en la bacteria precursora, resultando en una bacteria etiquetada.

En algunas modalidades preferidas, por lo menos una secuencia de marcado y por lo menos una secuencia duplicada se integran adyacentes entre sí. Más preferentemente, por lo menos una secuencia de marcado y por lo menos una secuencia duplicada se integran directamente adyacente entre sí, tal que no existan nucleótidos que intervengan entre las secuencias.

En algunas modalidades, la secuencia duplicada se adjunta, enlaza o fusiona a un extremo (por ejemplo, el extremo 5' o 3') de la secuencia de marcado. Preferentemente, la secuencia duplicada se adjunta, enlaza o fusiona al extremo 5' de la secuencia de marcado. Por consiguiente, siguiendo la integración de un solo par de secuencias, la secuencia duplicada es la primer secuencia en el extremo 5' de la ubicación CRISPR y secuencia de marcado será la segunda (por ejemplo, la próxima) secuencia en la ubicación CRISPR, en corriente abajo de la secuencia duplicada. En algunas modalidades preferidas, las secuencias se adjuntan directamente, se enlazan directamente o fusionan directamente tal que no existan nucleótidos que intervengan entre la secuencia duplicada y la secuencia de marcado.

Así, en algunas modalidades, una secuencia duplicada y un par de la secuencia de marcado se integran en el genoma de la bacteria precursora para dar lugar a una bacteria etiquetada. La secuencia duplicada se deriva, se puede derivar, se obtiene o se puede obtener del genoma de la bacteria precursora y la secuencia de marcado se

deriva, se puede derivar, se obtiene o se puede obtener del genoma del bacteriófago que se usa para infectar a la bacteria precursora.

Sorprendentemente, se ha encontrado incluso que en algunas modalidades, se integran pares múltiples de secuencias en el genoma de la bacteria precursora. De acuerdo a estas modalidades, los pares múltiples comprenden un primer par que comprende una secuencia duplicada y una secuencia de marcado, y un segundo par que comprende una segunda secuencia duplicada y una segunda secuencia de marcado. La segunda secuencia duplicada comprende típicamente la misma secuencia (por ejemplo, mayor que aproximadamente 95%, aproximadamente 96%, aproximadamente 97%, aproximadamente 98%, aproximadamente 99%, o aproximadamente 100% de identidad) como la primera secuencia duplicada. Las secuencias de marcado comprenden típicamente una secuencia diferente (por ejemplo, menos de aproximadamente 40%, aproximadamente 30%, aproximadamente 20%, aproximadamente 10%, aproximadamente 5%, aproximadamente 4%, aproximadamente 3%, aproximadamente 2%, aproximadamente 1%, o aproximadamente 0% identidad) a la primera secuencia de marcado. Este también es el caso en modalidades en las que se integran pares adicionales de secuencias.

Por consiguiente, la configuración de los pares múltiples comprende típicamente:

[secuencia de marcado - secuencia duplicada]_n

en donde n = 2, 3, 4, 5, 6, o más.

Preferentemente, la configuración de los pares múltiples comprende típicamente:

[secuencia de marcado - repetición CRISPR]_n,

en donde n = 2, 3, 4, 5, 6, o más.

En algunas modalidades, la configuración de los pares múltiples es:

5'-[secuencia de marcado - secuencia duplicada]_n-3'

en donde n = 2, 3, 4, 5, 6, o más.

Preferentemente, la configuración de los pares múltiples es:

5'-[secuencia de marcado - repetición CRISPR]_n-3',

en donde n = 2, 3, 4, 5, 6, o más.

Por consiguiente, en algunas modalidades, los pares múltiples de secuencias se integran en la bacteria precursora. En algunas modalidades, las secuencias de marcado se integran adyacente a: (i) una secuencia duplicada que es homóloga (por ejemplo, idéntica) a una secuencia que se presenta naturalmente en la bacteria precursora; (ii) una secuencia duplicada que es homóloga (por ejemplo, idéntica) a una secuencia que se presenta naturalmente en la ubicación CRISPR de la bacteria precursora; o (iii) más preferentemente, una secuencia duplicada que es homóloga (por ejemplo, idéntica) a una repetición CRISPR que se presenta naturalmente en la ubicación CRISPR de la bacteria precursora.

Siguiendo cada exposición de una bacteria precursora para un bacteriófago dado en experimentos independientes, la secuencia de marcado en cada una de las bacterias etiquetadas presenta una secuencia del nucleótido diferente, creando así, una secuencia que es única para cada bacteria. Así, sin limitarse por cualquier teoría particular, se cree que luego de la exposición de una bacteria precursora para un bacteriófago dado, la secuencia de marcado que se integra en una bacteria precursora se selecciona aparentemente al azar del genoma del bacteriófago. Sin embargo, no se pretende que la presente invención se limite a eventos de integración aleatorios.

Ventajosamente, este descubrimiento sorprendente se utiliza en el contexto de la presente invención en virtud del hecho de que la secuencia de marcado seleccionada aleatoriamente proporciona una única marca o etiqueta en la bacteria etiquetada. Sorprendentemente, también se ha encontrado que cuando la misma bacteria precursora se expone al mismo bacteriófago, la secuencia de marcado que se integra en los experimentos independientes/distintos es de una secuencia diferente, produciendo una única etiqueta en la bacteria etiquetada siguiendo cada exposición.

En algunas modalidades, las secuencias de marcado seleccionadas aleatoriamente se identifican en la bacteria etiquetada en virtud de una o más de las siguientes propiedades de la secuencia de marcado:

(1) La ubicación de la secuencia de marcado en una o más ubicaciones CRISPR del mutante insensible al bacteriófago. Como se describe en la presente, las secuencias de marcado se localizan típicamente en un y/o

ambos extremos (por ejemplo, los extremos 5' y/o 3', más preferentemente, el extremo 5') de la ubicación CRISPR de la bacteria etiquetada.

(2) La secuencia de marcado tiene un grado alto de homología o identidad (por ejemplo, 100% de identidad) a una secuencia en el genoma del bacteriófago en el que la bacteria precursora fue expuesta; y/o

(3) La secuencia de marcado se fusiona, enlaza o anexa a (por ejemplo, directamente fusionado, enlazado o anexado) por lo menos a una secuencia (por ejemplo, una repetición CRISPR) que se duplica desde el genoma de la bacteria precursora. Típicamente, como se describe en la presente, este par adicional de secuencias se localiza en uno y/o ambos extremos (por ejemplo, los extremos 5' y/o 3', preferentemente, el extremo 5') de la ubicación CRISPR de la bacteria etiquetada.

En algunas modalidades de marcado/etiquetación proporcionadas en la presente, una o más secuencias de marcado y/o una o más secuencias duplicadas (por ejemplo, la repetición CRISPR duplicada desde la bacteria precursora) se integra en la ubicación CRISPR de la bacteria precursora. En algunas modalidades preferidas, uno o más pares de secuencia de marcado-secuencia duplicada se describen en la presente para integrarse en la ubicación CRISPR de la bacteria precursora. En algunas modalidades, las secuencias marcadas y/o secuencias duplicadas se integran dentro de la ubicación CRISPR de la bacteria precursora. En algunas otras modalidades, las secuencias de marcado y/o secuencias duplicadas se integran en uno o en ambos extremos de la ubicación CRISPR de la bacteria precursora. Aún en modalidades adicionales, las secuencias de marcado y/o secuencias duplicadas se integran a ambos extremos de la ubicación CRISPR de la bacteria precursora, tal que las secuencias están en el extremo 5' y 3' de la ubicación CRISPR. Una de las secuencias duplicadas será típicamente la primera secuencia en el extremo 5' de la ubicación CRISPR y las secuencias de marcado estarán inmediatamente en corriente abajo de la secuencia duplicada. La otra secuencia duplicada estará en la última secuencia en el extremo 3' de la ubicación CRISPR y la secuencia de marcado estará inmediatamente en corriente arriba de la secuencia duplicada.

En algunas modalidades, las secuencias de marcado y/o secuencias duplicadas integran una o más ubicaciones CRISPR. Aún en modalidades adicionales, las secuencias de marcado y/o secuencias duplicadas se integran en un extremo de la ubicación CRISPR de la bacteria precursora, tal que las secuencias estén en el extremo 3' de la ubicación CRISPR. La secuencia duplicada estará en la última secuencia en el extremo 3' de la ubicación CRISPR y las secuencias de marcado estarán inmediatamente en corriente arriba de la secuencia duplicada. Preferentemente, las secuencias de marcado y/o secuencias duplicadas se integran en un extremo de la ubicación CRISPR de la bacteria precursora tal que las secuencias estén en el extremo 5' de la ubicación CRISPR. La secuencia duplicada es la primera secuencia en el extremo 5' de la ubicación CRISPR y la secuencia de marcado está inmediatamente en dirección descendente de la secuencia duplicada.

Como se describe en la presente, las secuencias de marcado son una marca específica de cepa en el sentido en el que las secuencias de marcado que se integran o se insertan desde el bacteriófago en la bacteria precursora es diferente en cada momento en el que la bacteria precursora (por ejemplo, la misma bacteria precursora) se expone al bacteriófago (por ejemplo, el mismo bacteriófago). Aquí, la secuencia de marcado encuntra un uso como una única etiqueta para una cepa bacteriana dada.

Las secuencias de marcado y/o secuencias duplicadas se integran en una o más ubicaciones CRISPR diferentes, mientras que en otras modalidades, dos o más secuencias de marcado diferentes y/o secuencias duplicadas se integran en una ubicación CRISPR, y aún en modalidades adicionales, dos o más secuencias de marcado diferentes y/o secuencias duplicadas se integran cada una en dos o más ubicaciones CRISPR diferentes. Cada una de las secuencias de marcado de cada uno de los bacteriófagos y/o cada una de las secuencias duplicadas (por ejemplo, las repeticiones CRISPR duplicadas) de la bacteria precursora puede integrarse en la misma ubicación CRISPR.

En algunas modalidades, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas se integran en uno o ambos extremos de la misma ubicación CRISPR. En algunas modalidades adicionales, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas se integran en el extremo 5' y/o 3' de la misma ubicación CRISPR. Preferentemente, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas se integran en el extremo 5' de la misma ubicación CRISPR. En algunas modalidades adicionales, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas de la bacteria precursora se integran iterativamente, simultáneamente o substancialmente de forma simultánea. En algunas modalidades, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas se integran secuencialmente, por medio de las cuales la primera secuencia de marcado y/o la primera secuencia duplicada se integra en la bacteria precursora. Una segunda secuencia de marcado desde un segundo bacteriófago y/o otra secuencia duplicada se integran en la bacteria precursora. Adecuadamente, las secuencias de marcado y/o la secuencia duplicada se integran en el ADN cromosómico de la bacteria precursora.

En algunas modalidades, cada una de las secuencias de marcado y/o cada una de las secuencias duplicadas se integran en un extremo (por ejemplo, el extremo 5') de la misma ubicación CRISPR adyacente (por ejemplo, contiguo) entre sí. Así, en algunas modalidades, cada una de las secuencias de marcado y/o secuencias duplicadas se integran secuencialmente, por medio de las cuales las primeras secuencias se integran en la bacteria precursora

en un extremo (por ejemplo, dentro o en el extremo 5' y/o 3') de la ubicación CRISPR. Una segunda secuencia de marcado y/o secuencia duplicada puede integrarse entonces en la bacteria precursora adyacente (por ejemplo, directamente adyacente) al primer par de secuencias. En algunas modalidades, las segundas secuencias se integran en la bacteria precursora adyacente (por ejemplo, directamente adyacente) al extremo 5' o 3' de las primeras secuencias. Preferentemente, las segundas secuencias se integran en la bacteria precursora adyacente (por ejemplo, directamente adyacente) al extremo 3' de las primeras secuencias y así sucesivamente. En algunas modalidades, cada una de las secuencias se integran adyacentes (por ejemplo, próximo a) entre sí dentro de o en el extremo 3' y/o en el extremo 5' de la misma ubicación CRISPR de la bacteria precursora. En algunas modalidades preferidas, cada una de las secuencias se integran adyacentes (por ejemplo, próximo a) entre sí en el extremo 5' de la misma ubicación CRISPR de la bacteria precursora. Más preferentemente, cada una de las secuencias se integran adyacentes (por ejemplo, próximo a) entre sí en corriente arriba del extremo 5' de la ubicación CRISPR de la bacteria precursora. Más preferentemente, cada una de las secuencias se integran adyacentes (por ejemplo, próximo a) entre sí en corriente arriba de la repetición 5' CRISPR de la ubicación CRISPR de la bacteria precursora. El más preferentemente, cada una de las secuencias se integra adyacente (por ejemplo, próximo a) entre sí río arriba del primero 5' repeticiones CRISPR de la ubicación CRISPR de la bacteria precursora.

Bacterias etiquetadas.

Como se usa en la presente, el término "bacterias etiquetadas" y "bacteria etiquetada" se refiere a una bacteria precursora, bacterias precursoras, o cepa bacteriana precursora en la cual una o más ubicaciones CRISPR o una porción de la misma ha sido modificada (por ejemplo, mutada) de tal manera que es insensible a uno o más a los bacteriófagos que se han expuesto. Como se describe adicionalmente a detalle en la presente, en algunas modalidades, la bacteria etiquetada se expone en más de un bacteriófago (por ejemplo, ya sea iterativamente, secuencialmente o simultáneamente), tal que acumula una o más modificaciones genómicas dentro de una o más ubicaciones CRISPR de tal manera que se vuelve insensible a cada uno de los bacteriófagos a los que han sido expuestos.

Para infectar las células, un bacteriófago inyecta o transfiere sus ácidos nucleicos en la célula con los ácidos nucleicos del fago que existen independientemente del genoma de la célula. En algunas modalidades, la infección resulta en la expresión (es decir, la transcripción y traducción) de ácidos nucleicos del bacteriófago dentro de la célula y continuación del ciclo de vida del bacteriófago.

En algunas modalidades de la presente invención, después de la exposición al bacteriófago, la bacteria etiquetada tiene una reducida o ninguna susceptibilidad a la infección del bacteriófago y/o multiplicación cuando se comparó con la bacteria precursora. Como se usa en la presente, el término "susceptibilidad reducida a la infección y/o multiplicación del bacteriófago" significa que el nivel de infección del bacteriófago y/o multiplicación en la bacteria etiquetada no causa un efecto nocivo a la bacteria etiquetada.

Así, en algunas modalidades de la presente invención, una bacteria precursora no se asesina después de la exposición al bacteriófago, debido a la mutación de la bacteria precursora de tal manera que se vuelve insensible al bacteriófago.

En algunas modalidades, la bacteria etiquetada es insensible o substancialmente insensible a la infección y/o multiplicación adicional mediante el bacteriófago. En modalidades adicionales, la bacteria etiquetada es insensible o substancialmente insensible a uno o más de los mecanismos que el bacteriófago usa para infectar y/o multiplicarse en una bacteria. Aún en modalidades adicionales, la bacteria etiquetada es insensible o substancialmente insensible a todos los mecanismos que el bacteriófago usa para infectar y/o multiplicarse en una bacteria. Aún en modalidades adicionales, la bacteria etiquetada desarrolla uno o más mecanismos que atenúan, inactivan o destruyen al bacteriófago durante el ciclo de infección. En algunas modalidades adicionales, la presente invención proporciona cepas etiquetadas seleccionadas mediante procedimientos de separación por exclusión estándar que son conocidos en el arte para aislar mutantes insensibles a bacteriófagos.

En algunas modalidades de la presente invención, una bacteria etiquetada que comprende una secuencia de marcado en la ubicación CRISPR que no está presente en la bacteria precursora se selecciona después de la comparación de la ubicación CRISPR o una porción de la misma de la bacteria precursora y la bacteria etiquetada.

En algunas modalidades preferidas, se selecciona una bacteria etiquetada que comprende un fragmento de ADN adicional dentro de o en el extremo 5' y/o 3' de la ubicación CRISPR que no está presente en la bacteria precursora. Más preferentemente, se selecciona una bacteria etiquetada que comprende una secuencia de marcado adyacente (por ejemplo, directamente adyacente) al extremo 3' de una secuencia recientemente duplicada en la ubicación CRISPR de la bacteria etiquetada que no está presente en la bacteria precursora. Más preferentemente, se selecciona una bacteria etiquetada que comprende una secuencia de marcado adyacente (por ejemplo, directamente adyacente) al extremo 3' de la primera repetición CRISPR de la ubicación CRISPR en la bacteria etiquetada que no está presente en la bacteria precursora.

En algunas modalidades, las secuencias de marcado (por ejemplo, una o más secuencias de marcado) se aíslan y/o clonan. En algunas modalidades adicionales, se forman secuencias de las secuencias de marcado (por ejemplo, una o más secuencias de marcado). Estas modalidades proporcionan ventajas, que no sólo proporcionan información acerca de la ubicación de las secuencias de marcado dentro de la ubicación CRISPR, sino también la secuencia específica de la misma. En algunas modalidades, esta información se guarda en una base de datos, que proporciona una etiqueta única para la bacteria dada y también un medio para rastrear subsecuentemente y/o para identificar la bacteria.

Una vez que se conoce la secuencia de las secuencias de marcado en la bacteria etiquetada, las secuencias de marcado solo encuentran uso en la identificación de las bacterias. Usando varios métodos que son conocidos en el arte y descritos en la presente, se determina la secuencia y/o ubicación de las secuencias de marcado. Esta secuencia se compara entonces con, por ejemplo, una base de datos de la secuencia bacteriana y/o una base de datos de la secuencia del bacteriófago y/o una base de datos o etiquetas/marcas para identificar la bacteria.

Organismo del donador

Como se usa en la presente en algunas modalidades, el término “organismo del donador” se refiere a un organismo o célula a partir del cual la repetición CRISPR y/o gen *cas* y/o combinaciones y/o espaciadores CRISPR se derivan. Estos pueden ser el mismo o diferente. En algunas modalidades, el término “organismo del donador” se refiere a un organismo o célula a partir del cual una o más, preferentemente, dos o más repeticiones CRISPR y/o uno o más genes *cas* y/o combinaciones de los mismos y/o espaciadores CRISPR se derivan. Estos pueden ser el mismo o diferente. En algunas modalidades, el espaciador CRISPR y/o pseudoespaciador CRISPR se deriva sintéticamente. Aún en modalidades adicionales, el organismo del donador o célula comprende uno o más espaciadores CRISPR, que confieren inmunidad específica contra un ácido nucleico objetivo o producto de la transcripción del mismo. En modalidades adicionales, el organismo del donador o célula a partir del cual el gen *cas* y/o repeticiones CRISPR y/o combinación de las mismas se deriva, es también la célula receptora/organismo para la ubicación CRISPR recombinante. Estos pueden ser el mismo o diferente. En otras modalidades, el organismo del donador o célula a partir del cual el espaciador CRISPR se deriva, es también la célula receptora/organismo para la ubicación CRISPR recombinante. Estos pueden ser el mismo o diferente. En modalidades en las que el organismo del donador es una célula bacteriana, el organismo del donador comprende típicamente un espaciador CRISPR que confiere inmunidad específica contra los ácidos nucleicos objetivo o productos de la transcripción de los mismos. En algunas modalidades, el organismo es una célula bacteriana mientras que en otras modalidades es un bacteriófago.

Célula Hospedadora

Como se usa en la presente, el término “célula hospedadora” se refiere a cualquier célula que comprende la combinación, constructo o vector y similares de acuerdo a la presente invención. En algunas modalidades, las células hospedadoras se transforman o transfectan con una secuencia del nucleótido contenida en un vector (por ejemplo, un vector de clonación). En algunas modalidades, una secuencia del nucleótido puede conducirse en un vector para la replicación y/o expresión de la secuencia del nucleótido. Las células se eligen para ser compatibles con el vector y en algunas modalidades, son células procarióticas (por ejemplo, bacterianas).

Células Receptoras

Como se usa en la presente, el término “célula receptora” se refiere a cualquier célula cuya resistencia contra un ácido nucleico objetivo o un producto de la transcripción del mismo se modula o va a ser modulado. En algunas modalidades, la célula receptora se refiere a cualquier célula que comprende ácidos nucleicos recombinantes de acuerdo a la presente invención. En algunas modalidades, la célula receptora comprende una o más, preferentemente, dos o más repeticiones CRISPR y uno o más genes o proteínas *cas*. Apropriadamente, las repeticiones CRISPR y los genes o proteínas *cas* forman una combinación funcional en la célula receptora, como se describe en la presente. En algunas modalidades adicionales, la célula receptora comprende una o más repeticiones CRISPR modificadas y/o uno o más genes o proteínas *cas* modificadas. Apropriadamente, las repeticiones CRISPR modificadas y/o los genes o proteínas *cas* modificadas forman una combinación funcional en la célula receptora, como se describe en la presente. En algunas modalidades, la célula receptora comprende una o más repeticiones CRISPR genéticamente diseñadas y/o uno o más genes o proteínas *cas* genéticamente diseñadas. Apropriadamente, las repeticiones CRISPR genéticamente diseñadas y/o los genes o proteínas *cas* genéticamente diseñados forman una combinación funcional en la célula receptora, como se describe en la presente. En algunas modalidades alternativas, la célula receptora comprende una o más repeticiones CRISPR recombinantes y/o uno o más genes o proteínas *cas* recombinantes. Apropriadamente, las repeticiones CRISPR recombinantes y/o los genes o proteínas *cas* recombinantes forman una combinación funcional en la célula receptora, como se describe en la presente. Aún en modalidades adicionales, la célula receptora comprende una o más repeticiones CRISPR y uno o más genes o proteínas *cas* que se presentan naturalmente. Apropriadamente, las repeticiones CRISPR y los genes o proteínas *cas* forman una combinación funcional.

En algunas modalidades, la célula receptora comprende combinaciones de una o más repeticiones CRISPR que se presentan naturalmente o recombinantemente, genéticamente diseñadas, modificadas, y uno o más genes o

proteínas cas que se presentan naturalmente o recombinantemente, genéticamente diseñadas o modificadas. Apropiadamente, uno o más espaciadores CRISPR que se presentan naturalmente o recombinantemente, genéticamente diseñados, modificados, o uno o más genes o proteínas cas que se presentan naturalmente o recombinantemente, genéticamente diseñadas, modificadas forman una combinación funcional.

En algunas modalidades, la célula receptora es una célula procariótica. En algunas modalidades preferidas, la célula receptora es una célula bacteriana. Se describen las células bacterianas adecuadas en la presente. En algunas modalidades, la célula bacteriana se selecciona de una especie de bacteria del ácido láctico, una especie de *Bifidobacterium*, una especie de *Brevibacterium*, una especie de *Propionibacterium*, una especie de *Lactococcus*, una especie de *Streptococcus*, una especie de *Lactobacillus* que incluye las especies *Enterococcus*, especies *Pediococcus*, una especie *Leuconostoc* y especies de *Oenococcus*. Las especies convenientes incluyen, pero no se limitan a *Lactococcus lactis*, que incluyen *Lactococcus lactis* subsp. y *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *cremoris*, *Leuconostoc* sp, *Lactococcus lactis* subsp. *lactis* biovar, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Lactobacillus helveticus*, *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei*.

En algunas modalidades en la que la resistencia de la célula será modulada, la célula bacteriana se usa para la fermentación de la carne (que incluye ternera, carne de cerdo, y pollos) que incluye, pero no se limita a, bacterias de ácido láctico, *Pediococcus cerevisiae*, *Lactobacillus plantarum*, *Lactobacillus brevis*, especies de *Micrococcus*, *Lactobacillus sakei*, *Lactobacillus curvatus*, *Pediococcus pentosaceus*, *Staphylococcus xylosus* y *Staphylococcus vitulinus* y mezclas de las mismas, como se conoce en el arte. En algunas modalidades alternativas, la célula bacteriana se usa para la fermentación de mezclas formadas de verduras (por ejemplo, zanahorias, pepinos, tomates, pimientos, y col) que incluye, pero no se limita a, *Lactobacillus plantarum*, *Lactobacillus brevis*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, y mezclas de las mismas, como se conoce en el arte. En algunas modalidades alternativas, la célula bacteriana se usa para la fermentación de masa formada de cereales (por ejemplo, trigo, centeno, arroz, avena, cebada, y maíz). En algunas modalidades adicionales, la célula bacteriana se usa para la producción de vino. Típicamente, esto se logra por la fermentación de jugo de fruta, típicamente el jugo de la uva. Aún en modalidades adicionales, la célula bacteriana se usa para la fermentación de la leche para producir el queso (por ejemplo, *Lactobacillus delbrueckii* subsp. *bulgaricus*, *Lactobacillus helveticus*, *Streptococcus thermophilus*, *Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis* subsp. *biovar diacetylactis*, *Lactococcus*, *Bifidobacterium*, y *Enterococcus*, etc., y mezclas de las mismas), como se conoce en el arte. Aún en modalidades adicionales, la célula bacteriana se usa para la fermentación del huevo (por ejemplo, *Pediococcus pentosaceus*, *Lactobacillus plantarum*, y mezclas de los mismos), como se conoce en el arte. En algunas modalidades adicionales, la célula bacteriana se usa en cosméticos o composiciones farmacéuticas.

En algunas modalidades, la célula en la cual la resistencia será modulada es una bacteria que naturalmente comprende una o más ubicaciones CRISPR. Las ubicaciones CRISPR se han identificado en más de 40 procariotas (Ver, Haft et al., 2005, supra) que incluye, pero no se limita a *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*.

Cepa Bacteriana Precursora

Como se usa en la presente los términos “bacteria precursora”, “bacterias precursoras” y “cepa precursora” se refieren a cualquier bacteria/bacterias/cepas que son expuestos a uno o más bacteriófagos. En algunas modalidades, los bacteriófagos son virulentos para la cepa bacteriana precursora, mientras que en otras modalidades, no son virulentos. En algunas modalidades particularmente preferidas, las bacterias precursoras son sensibles al fago virulento. En algunas modalidades preferidas, la cepa precursora se infecta por el bacteriófago. En algunas modalidades particularmente preferidas, la infección por el fago da como resultado las bacteria/bacterias/cepas precursoras o una subpoblación insensible a las mismas para una infección adicional por el bacteriófago. En algunas modalidades preferidas, la infección de una “bacteria precursora” por uno o más bacteriófagos resulta en la creación de una cepa etiquetada que puede seleccionarse en base a su insensibilidad al bacteriófago. En algunas modalidades preferidas, el “mutante resistente a bacteriófagos” es una bacteria que se marca o etiqueta de acuerdo a los métodos de la presente invención. En algunas modalidades, las bacterias precursoras son cepas bacterianas del tipo natural. En algunas modalidades preferidas, las bacterias precursoras son cepas del tipo natural de bacterias que no se han infectado previamente con cualquier bacteriófago. En algunas modalidades preferidas, las bacterias precursoras son cepas del tipo natural de bacterias que no se han marcado o etiquetado previamente, mientras que en algunas modalidades alternativas, las bacterias de patente son mutantes resistentes a bacteriófagos que se han marcado o etiquetado previamente.

En algunas modalidades particularmente preferidas, la bacteria precursora se selecciona de cualquier bacteria que naturalmente comprende una o más ubicaciones CRISPR. Como se indica anteriormente, las ubicaciones CRISPR

se han identificado en más de 40 procariotas (Ver, Haft et al., [2005], *supra*) que incluye, pero no se limita a *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacteria*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacteria*, *Mycobacteria*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacteria*, *Azarcus*, *Chromobacteria*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacteria*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema*, y *Thermotoga*.

En algunas modalidades, la bacteria precursora comprende uno o más espaciadores heterólogos CRISPR, una o más repeticiones CRISPR heterólogas, y/o uno o más genes cas heterólogos. En algunas modalidades alternativas, la bacteria precursora comprende una o más ubicaciones CRISPR heterólogas, preferentemente, una o más ubicaciones CRISPR completas. En algunas modalidades adicionales, la bacteria precursora comprende naturalmente una o más ubicaciones CRISPR y también comprende uno o más espaciadores CRISPR heterólogos, una o más repeticiones CRISPR heterólogas y/o uno o más genes cas heterólogos. En algunas modalidades adicionales, la bacteria precursora comprende naturalmente una o más ubicaciones CRISPR y también comprende una o más ubicaciones CRISPR heterólogas, preferentemente, una o más ubicaciones CRISPR completas.

La presente invención abarca cultivos mixtos de células y fagos. En algunas modalidades, el cultivo mixto es una mezcla de diferentes mutantes que corresponden a eventos de integración diferentes en el mismo y/o en ubicaciones CRISPR diferentes.

Aunque no se pretende que la presente invención se limite, los géneros bacterianos precursores preferidos son *Streptococcus* y *Lactobacillus*. De hecho, se pretende que cualquier especie bacteriana encuentre un uso en la presente invención, que incluye pero que no se limita a *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacteria*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Enterococcus*, *Pediococcus*, *Leuconostoc*, *Oenococcus*, y/o *Xanthomonas*. En algunas modalidades, las bacterias precursoras son o se derivan de las bacterias del ácido láctico, que incluye pero que no se limitan a *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Streptococcus*, *Lactobacillus* (por ejemplo, *L. acidophilus*), *Enterococcus*, *Pediococcus*, *Leuconostoc*, y/o *Oenococcus*. En modalidades adicionales, las bacterias precursoras son o se derivan de *Lactococcus lactis* (por ejemplo, *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 algunas modalidades de la presente invención, la bacteria precursora es una "bacteria de grado alimenticio" (es decir, una bacteria que se usa y generalmente se considera como segura para el uso en la preparación y/o producción de alimentos y/o alimentación). En algunas modalidades preferidas, la bacteria precursora es conveniente para uso como un cultivo de inicio, un cultivo probiótico, y/o un suplemento dietético. En modalidades adicionales, la bacteria precursora encuentra uso en la fermentación de la carne (por ejemplo, ternera, carne de cerdo, cordero, y pollería) que incluye, pero no se limita a, bacterias del ácido láctico, *Pediococcus cerevisiae*, *Lactobacillus plantarum*, *L. brevis*, *L. sakei*, *L. curvatus*, especies de *Micrococcus*, *Pediococcus pentosaceus*, *Staphylococcus xylosum*, *S. vitulinus* y mezclas de las mismas (Vea por ejemplo, Knorr (ed.), *Food Biotechnology*, en 538-39 [1987]; y Pederson, *Microbiology of Fermented Foods*, en 210-34, 2d ed., [1979]; y la Patente de los Estados Unidos No. 2,225,783). Aún en modalidades adicionales, la bacteria precursora encuentra uso en la fermentación de verduras (por ejemplo, zanahorias, pepinos, tomates, pimientos, y col) que incluye, pero no se limita a, *L. plantatum*, *L. brevis*, *Leuconostoc mesenteroides*, *Pediococcus pentosaceus*, y mezclas de las mismas (Ver por ejemplo, Knorr, *supra*; Pederson, *supra*; y la patente de los Estados Unidos Nos. 3,024,116, 3,403,032, 3,932,674, y 3,897,307). Aún en modalidades adicionales, la bacteria precursora encuentra uso en la fermentación de masa formada a partir de cereales (por ejemplo, trigo, centeno, arroz, avena, cebada, y maíz). Aún en modalidades adicionales, la bacteria precursora encuentra uso en la producción de vino a través de la fermentación del jugo de fruta (por ejemplo, el jugo de la uva). En algunas modalidades adicionales, la bacteria precursora encuentra uso en la fermentación de la leche (por ejemplo, *L. delbrueckii* subsp. *bulgaricus*, *L. acidophilus*, *S. thermophilus*, y mezclas de las mismas (Ver, Knorr, *supra*; y Pederson *supra*, páginas 105-35). En algunas modalidades preferidas, la bacteria precursora encuentra un uso en la producción de queso, que incluye pero que no se limitan a *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 los mismos (Ver por ejemplo, Knorr, *supra*, y Pederson, *supra*, en 135-51). Aún en modalidades adicionales, la bacteria precursora encuentra un uso en la fermentación de huevos, que incluye pero que no se limita a *Pediococcus pentosaceus*, *Lactobacillus plantarum*, y mezclas de los mismos (Ver, Knorr, *supra*). En algunas modalidades, la bacteria precursora encuentra un uso en la fermentación para producir varios productos, que incluyen pero que no se limitan a quesos y queso cotagge (por ejemplo, *L. lactis* subsp. *lactis*, *L. lactis* subsp. *cremoris*), yogur (*L. delbrueckii* subsp. *bulgaricus*, y *S. thermophilus*), queso suizo (por ejemplo, *S. thermophilus*, *L. lactis*, y *L. helveticus*), queso azul (*Leuconostoc cremoris*), queso italiano (*L. bulgaricus* 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 (*Brevibacteria lactofermentum* y *Microbacterium ammoniaphilum*), y acetona y butanol (*Clostridium acetobutyricum*, y *Clostridium saccharoperbutylacetonicum*). En algunas modalidades preferidas, las especies bacterianas precursoras son seleccionadas de *S. thermophilus*, *L. delbrueckii subsp. bulgaricus* y/o *L. acidophilus*.

Aún en modalidades adicionales, las bacterias precursoras encuentran un uso en métodos que incluyen pero no se limitan a la producción antibiótica, producción de aminoácidos, producción de solventes, y producción de otros materiales económicamente útiles. Aún en otras modalidades, las bacterias precursoras encuentran un uso en las composiciones cosméticas, terapéuticas, y/o farmacéuticas. En algunas modalidades, las composiciones tienen las actividades particulares, que incluyen pero que no se limitan a la regeneración de la piel, que incluyen pero que no se limitan a las propiedades anti arrugas, desvanecimiento de cicatrices antiguas, reparación de los tejidos dañados por quemaduras, promoción de la salud de la piel, eliminación de manchas por pigmentación, etc. En algunas modalidades adicionales, las composiciones promueven o inhiben el crecimiento de uñas, pelo o vello. En algunas modalidades, las composiciones comprenden por lo menos un cultivo microbiano y/o la bacteria etiquetada y/o un cultivo celular producido usando métodos y composiciones de la presente invención.

En modalidades adicionales, las bacterias precursoras son mutantes insensibles a bacteriófagos. Así, en algunas modalidades, las bacterias precursoras son insensibles a uno o más bacteriófagos. En algunas modalidades preferidas, la bacteria precursora no es un mutante insensible a bacteriófagos para el bacteriófago al que será expuesto durante un uso en la presente invención.

Cultivos de Inicio

Los cultivos de inicio se usan extensivamente en la industria de los alimentos, en la fabricación de productos fermentados que incluyen los productos derivados de la leche (por ejemplo, el yogur y el queso), así como los productos derivados de la carne, productos de panadería, vinos y productos de las verduras.

Los cultivos de inicio usados en la fabricación de la mayoría de la leche fermentada, queso y productos de mantequilla incluyen cultivos de bacterias, generalmente clasificados como bacterias del ácido láctico. Tales cultivos de inicio bacterianos imparten características específicas para varios productos lácteos realizando una variedad de funciones.

Los cultivos no concentrados comerciales de bacterias se refieren en la industria como "cultivos madre" y se propagan en el sitio de la producción, por ejemplo, en un lácteo, antes de que se agregue a un material de inicio comestible, como la leche, para la fermentación. El cultivo de inicio propagado en el sitio de la producción para la inoculación en un material de inicio comestible se denomina "de inicio a granel".

Los cultivos de inicio adecuados para uso en la presente invención incluyen cualquier organismo que es de uso en los alimentos, cosméticos o industria farmacéutica (es decir, "cultivos industrialmente útiles" o "cepas industrialmente útiles").

Los cultivos de inicio son preparados mediante técnicas bien conocidas en el arte (Ver por ejemplo, la patente de los Estados Unidos No. 4,621,058). En algunas modalidades, los cultivos de inicio se preparan mediante la introducción de inoculación, por ejemplo una bacteria, a un medio de crecimiento (por ejemplo, un producto o medio de fermentación) para producir un medio inoculado e incubando el medio inoculado para producir un cultivo de inicio.

Los cultivos de inicio secos se preparan mediante técnicas bien conocidas en el arte (Ver por ejemplo, Patente de los Estados Unidos. Nos. 4,423,079 y 4,140,800). Cualquier forma conveniente de cultivos de inicio secos encuentra un uso en la presente invención, que incluye las preparaciones sólidas (por ejemplo, comprimidos, peletizados, cápsulas, polvos, gránulos y productos en polvo) que son humedecibles, secados por rocío, secados por congelación, y liofilizados. En algunas modalidades, los cultivos de inicio secos para uso en la presente invención están en la forma de un peletizado congelado ahuecado o en forma de un polvo deshidratado por congelación. Los cultivos de inicio secos en un peletizado congelado ahuecado o polvo deshidratado por congelación se preparan de acuerdo con cualquier método conveniente conocido en el arte.

En algunas modalidades, los cultivos de inicio usados en la presente invención están en la forma de concentrados que comprenden una concentración substancialmente alta de una o más cepas bacterianas. En algunas modalidades, los concentrados se diluyen con agua o resuspenden en agua u otros diluyentes convenientes, (por ejemplo, un medio de crecimiento apropiado, aceite mineral, o aceite vegetal). Los cultivos de inicio secos de la presente invención están en la forma de concentrados que se preparan de acuerdo a los métodos conocidos en el arte (por ejemplo, centrifugación, filtración o una combinación de tales técnicas).

En algunas modalidades, el cultivo de inicio es conveniente para usarse en la industria láctea. Cuando se usa en la industria láctea, el cultivo de inicio es a menudo seleccionado de una especie de bacterias de ácido láctico, una especie de *Bifidobacterium*, una especie *Brevibacterium*, una especie de *Propionibacteria*. Los cultivos de inicio

adecuados del grupo de bacterias de ácido láctico incluyen comúnmente cepas usadas de especies *Lactococcus*, una especie *Streptococcus*, una especie *Lactobacillus* incluye *Lactobacillus acidophilus*, especies *Enterococcus*, especies *Pediococcus*, una especie *Leuconostoc* y especies *Oenococcus*.

5 Normalmente se usan cultivos de bacterias del ácido láctico en la fabricación de productos de leche fermentados (por ejemplo, el suero de leche, yogur o crema agria), y en la fabricación de mantequilla y queso (por ejemplo, brie o havarti). Las especies de *Lactococcus* incluyen *Lactococcus lactis* ampliamente usados, que incluye el *Lactococcus lactis* subsp. *lactis* y *Lactococcus lactis* subsp. *cremoris*.

10 Otras especies de bacteria de ácido láctico incluyen *Leuconostoc* sp., *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Lactobacillus helveticus*. Además, las cepas probióticas (por ejemplo, especies *Lactococcus*), incluyen el ampliamente usado *Lactococcus lactis*, que incluye el *Lactococcus lactis* subsp. *lactis*, y *Lactococcus lactis* subsp. *cremoris*.

15 Los cultivos mesofílicos de bacterias del ácido láctico usadas comúnmente en la fabricación de productos de leche fermentados tales como el suero de leche, yogur o crema agria, y en la fabricación de mantequilla y queso (por ejemplo, brie o havarti). Otras especies *Lactococcus* incluyen *Lactococcus lactis* subsp. *cremoris*, *Lactococcus lactis*, *Leuconostoc*, sp., *Lactococcus lactis* subsp. *lactis* biovar, *Streptococcus thermophilus*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Lactobacillus helveticus*. Además, en algunas modalidades, la cepa probiótica tal como *Bifidobacterium lactis*, *Lactobacillus acidophilus*, *Lactobacillus casei* se agregan durante la fabricación para mejorar el sabor o promover la salud.

20 Los cultivos de bacterias de ácido láctico normalmente usados en la fabricación de quesos cheddar y monterey jack incluyen *Streptococcus thermophilus*, *Lactococcus lactis* subsp. *lactis* y *Lactococcus lactis* subsp. *cremoris* o combinaciones de los mismos.

25 Los cultivos termofílicos de bacterias del ácido láctico normalmente usados en la fabricación de quesos italianos tales como de pasta filata o parmesanos, incluyen *Streptococcus thermophilus* y *Lactobacillus delbrueckii* subsp. *bulgaricus*. Otras especies de *Lactobacillus* (por ejemplo, *Lactobacillus helveticus*) se agregan durante la fabricación para obtener un sabor deseado.

30 En algunas modalidades preferidas, el organismo del cultivo de inicio comprende o consiste en una cepa genéticamente modificada preparada de acuerdo con los métodos proporcionados en la presente, de una de las cepas de bacterias del ácido láctico anteriores o cualquier otra cepa de cultivo de inicio.

35 La selección de organismos para el cultivo de inicio de la invención dependerá del tipo particular de productos a ser preparados y tratados. Así, por ejemplo, para la elaboración de queso y mantequilla, se usan los cultivos mesofílicos de especies *Lactococcus*, especies *Leuconostoc* y especies *Lactobacillus* ampliamente usadas, considerando que para el yogur y otros productos de leche fermentados, se usan típicamente las cepas termofílicas de especies *Streptococcus* y de especies *Lactobacillus*.

40 En algunas modalidades, el cultivo de inicio es un cultivo de inicio seco, un cultivo de inicio deshidratado, un cultivo de inicio congelado, o un cultivo de inicio concentrado. En algunas modalidades, el cultivo de inicio se usa en la inoculación directa del medio de fermentación o producto.

45 El cultivo de inicio es un cultivo mixto (es decir, que comprende por lo menos dos cepas bacterianas diferentes).

Bacterias del Ácido láctico

50 Los cultivos de inicio particularmente adecuados, en particular cultivos de inicio secos, para un uso en la presente invención comprenden las bacterias del ácido láctico.

55 Como se usa en la presente el término "bacterias del ácido láctico" se refiere a las bacterias anaeróbicas o microaerofílicas Gram positivas que fermentan el azúcar con la producción de ácidos que incluyen el ácido láctico como el ácido producido predominantemente, ácido acético, ácido fórmico y ácido propiónico. Las bacterias del ácido láctico más útiles industrialmente se encuentran entre las especies de *Lactococcus*, tales como *Lactococcus lactis*, especies *Lactobacillus*, especies *Bifidobacterium*, especies *Streptococcus*, especies *Leuconostoc*, especies *Pediococcus* y especies *Propionibacterium*.

60 Los cultivos de inicio de la presente invención pueden comprender una o más especies de bacterias del ácido láctico tales como, *Lactococcus lactis*, *Lactobacillus delbrueckii* subsp. *bulgaricus* y *Streptococcus thermophilus* o combinaciones de los mismos .

65 Los cultivos de inicio de bacterias del ácido láctico se usan comúnmente en la industria de alimentos como cultivos de cepa mixtos que comprenden una o más especies. Para una variedad de cultivos de cepa mixtas, tales como los cultivos de inicio para el yogur que comprenden cepas de *Lactobacillus delbrueckii* subsp. *bulgaricus* y

Streptococcus thermophilus, una relación simbiótica existe entre las especies, en donde la producción de ácido láctico es mayor comparado con los cultivos de bacterias de ácido láctico de una sola cepa (Ver por ejemplo, Rajagopal *et al.*, J. Dairy Sci., 73:894-899 [1990]).

5 Productos

Los productos adecuados para uso en la presente invención incluyen, pero no se limitan a, un comestible, productos cosméticos o productos farmacéuticos. Cualquier producto, que se prepara de o comprende, un cultivo se contempla de acuerdo con la presente invención. Estos incluyen, pero no se limitan a, frutas, legumbres, cosechas de forraje y verduras que incluyen productos derivados, granos y productos derivados de granos, alimentos lácteos y productos derivados de alimentos lácteos, carne, pollos, mariscos, cosméticos y productos farmacéuticos.

El término “alimentos” se usa en un sentido amplio e incluye alimentos, comestibles, ingredientes para alimentos, suplementos alimenticios, y alimentos funcionales.

Como se usa en la presente, el término “ingrediente para alimentos” incluye una formulación que es o puede agregarse a los alimentos e incluye formulaciones que pueden usarse a niveles bajos en una variedad amplia de productos que requieren, por ejemplo, acidificación o emulsión.

Como se usa en la presente, el término “alimentos funcionales” significa un alimento que es capaz de proporcionar no solo un efecto nutritivo y/o una satisfacción del sabor, sino también es capaz de suministrar un efecto beneficioso adicional al consumidor. Aunque no existe una definición formal de un alimento funcional, en la mayoría de las reuniones con interés en esta área se está de acuerdo que hay alimentos comercializados que tienen efectos saludables específicos.

El término “alimento” abarca a los alimentos para humanos así como los alimentos para animales (es decir, pienso). En un aspecto preferido, el alimento es para el consumo humano.

En algunas modalidades, las células descritas en la presente comprenden o se agregan a un ingrediente del alimento, un suplemento alimenticio, o un alimento funcional. En algunas modalidades, el alimento está en forma líquida (por ejemplo, una solución), gel, emulsión, o sólido, denominado por el modo de aplicación y/o administración.

Las células descritas en la presente encuentran uso en la preparación de productos alimenticios así como en uno o más de: productos de confitería, productos lácteos, productos de la carne, productos de pollería, productos de pescado y productos de panadería. En algunas modalidades, las bacterias encuentran un uso como ingredientes en bebidas ligeras, jugos de fruta, bebidas que incluyen suero de proteínas, té saludables, bebidas de cacao, bebidas de leche, bebidas con bacterias del ácido láctico, yogur, yogur bebible, y vinos, etc.

Se proporciona también un método para preparar alimentos, el método que comprende mezclar las células de acuerdo con la presente invención con un ingrediente alimenticio (tal como un material de inicio para alimentos). El método para preparar un alimento también es otro aspecto de la presente invención.

Apropiadamente, un alimento como se describe en la presente, es un producto lácteo. En algunas modalidades preferidas, el producto lácteo es el yogur, queso (por ejemplo, el queso de cuajada ácido, queso duro, queso semi-duro, queso cottage, etc.), el suero de leche, quark, crema agria, kefir, crème fraîche, bebida basada en suero fermentado, koumiss, bebida de leche, o bebida de yogur.

Como se usa en la presente, el término “alimento” tiene un sentido muy amplio, pretende abarcar alimentos para humanos así como alimentos para los animales no humanos (es decir, pienso). En algunas modalidades preferidas, los alimentos son para el consumo humano. El término “pienso” como se usa en la presente incluye material de plantas procesado y crudo y material de no plantas. El término abarca cualquier alimento adecuado para consumirse por un animal, que incluye, pero no se limita al ganado, pollería, pescado, crustáceos, y/o animales domésticos.

55 Desarrollo de Cepas Resistentes al Fago y Cultivos de inicio

Durante el desarrollo de la presente invención, la resistencia al fago que involucra los genes CRISPR-cas, así como su función en la resistencia a la entrada del ADN extranjero y en la función de los espaciadores insertados dentro del CRISPR en la especificidad de esta resistencia, ha sido elucidada. Cabe destacar, que la presente invención proporciona métodos y composiciones para el desarrollo de cepas resistentes al fago y cultivos de inicio. En algunas modalidades, una cepa precursora “A” se expone al fago “P” y a una variante resistente al fago seleccionada (Variante “A 1.0”). La variante A 1.0 se analiza (por ejemplo por PCR, y/o formación de secuencias de ADN) para confirmar la presencia de un espaciador insertado adicional dentro de una ubicación CRISPR. La secuencia del nucleótido del espaciador adicional (espaciador Sp1.0) se determina entonces. Típicamente, el espaciador Sp1.0 es un fragmento de aproximadamente 30 nucleótidos en tamaño desde el fago P, y proporciona resistencia al fago P y

fagos relacionados ("fagos relacionados" son aquéllos que contienen la secuencia del espaciador en sus genomas, y definen una familia de fagos).

Independientemente de la primera exposición al fago, la misma cepa precursora A se expone al mismo fago P y se selecciona una segunda variante resistente al fago (Variante A2.0). La variante A2.0 se selecciona para también tener un espaciador adicional insertado (espaciador Sp2.0) dentro de una ubicación CRISPR pero con la secuencia del espaciador Sp2.0 que es diferente del espaciador Sp1.0. Típicamente, el espaciador Sp2.0 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago P, y proporciona resistencia al fago P y fagos relacionados. De manera semejante, en algunas modalidades, se generan las variantes A3.0 a Ax.0 a través de la exposición de la misma cepa A al mismo fago P. Todas las variantes "A" son seleccionadas para también tener un espaciador adicional insertado (espaciador Sp3.0 a Spx.0) dentro de una ubicación CRISPR pero con la secuencia de todos los espaciadores "Sp" que son diferentes de cada uno de los otros. Típicamente, los espaciadores "Sp" son fragmentos de aproximadamente 30 nucleótidos en tamaño del fago P, y todos proporcionan resistencia al fago P y fagos relacionados.

Aunque estas variantes son útiles, están limitadas por lo que se refiere al alcance de su resistencia. Así, en algunas modalidades, es ventajoso desarrollar cepas resistentes al fago de nivel secundario. De hecho, es ventajoso desarrollar además estas variantes resistentes al fago aumentando y extendiendo su resistencia a los fagos. Típicamente, puede estimarse que el nivel de resistencia será aproximadamente de una sola mutación que ocurre dentro del genoma del fago dentro de la secuencia que corresponde al espaciador (es decir, aproximadamente 10^{-4} a 10^{-6}). Por consiguiente, las cepas resistentes al fago que acumulan espaciadores diferentes dentro de la ubicación CRISPR tienen un nivel incrementado de resistencia al fago que contiene la secuencia de éstos espaciadores dentro de su genoma (es decir, puesto que múltiples mutaciones simples necesitan ocurrir dentro del genoma del fago).

En algunas modalidades, las variantes de nivel secundario se producen aislando un fago mutado a través de la exposición de la variante A1.0 al fago P. Típicamente, este fago mutado (fago P1.0) tiene una mutación (eliminación, punto de mutación, etc.) en su genoma dentro de la región que contiene la secuencia del espaciador Sp1.0. La variante A1.0 es sensible al fago P1.0. Entonces, la variante A1.0 se expone al fago P1.0 y a una variante resistente al fago (Variante A1.1) seleccionada (Ver, Figura 15). La variante A1.1 también se selecciona tal que tiene un espaciador adicional insertado (espaciador Sp1.1) dentro de una ubicación CRISPR pero con la secuencia del espaciador Sp1.1 que es diferente de los espaciadores Sp1.0, Sp2.0 a Spx.0. Típicamente, el espaciador Sp1.1 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago P1.0, y proporcionará resistencia al fago P1.0 y fagos relacionados. La variante A1.1 es resistente al fago P1.0 y preferentemente, tiene una resistencia incrementada al fago P debido a la acumulación del espaciador Sp1.0 y Sp1.1.

En modalidades adicionales, un fago recientemente mutado (fago P1.1) se genera a través de la exposición de la variante A1.1 al fago P1.0. Entonces, después de la exposición de la variante A1.1 al fago P1.1, una nueva variante A1.2 se obtiene, que contiene un nuevo espaciador adicional (Sp1.2). Este espaciador proporciona resistencia al fago P1.1 y preferentemente incrementa la resistencia al fago P1.0 y P (es decir, debido a la acumulación de espaciadores Sp1.0, Sp1.1, Sp1.2). Aún en modalidades adicionales, los espaciadores diferentes (por ejemplo, 2, 3 o 4) se acumulan iterativamente dentro de la cepa A a través de la variante A1, después de la variante A1.1, después de la variante A1.2, etc. para obtener una variante altamente resistente a los fagos (variante A1.n). Aún en modalidades adicionales, los espaciadores diferentes adicionales pueden acumularse en la misma cepa a través de la variante A2, después de la variante A2.1, después de la variante A2.2, etc para generar otra variante de la cepa A altamente resistente a los fagos (variante A2.n) en paralelo. La misma estrategia encuentra uso con las variantes A3.0 a Ax.0.

En algunas modalidades, se proporcionan cepas que son resistentes a más de una familia de fagos. Como una cepa dada puede ser sensible a más de una familia de fagos, en algunas modalidades, se deseaba agrandar la resistencia de la cepa para familias del fago múltiples introduciendo los espaciadores adicionales dentro de una ubicación CRISPR que se origina de otras familias de fagos (Ver, la Figura 16). Por ejemplo, los fagos P, Q, y R son fagos representativos de tres familias de fagos capaces de infectar la cepa A. Usando el método descrito arriba y en la presente, se producen variantes resistentes a todas las tres familias del fago. En algunas modalidades, el fago P se usa para generar la variante A1^P (que contienen el espaciador Sp1) que es resistente al fago P. Entonces, se selecciona la variante A1^P que se expone al fago Q y un fago resistente a la variante (Variante A1^{PQ}). La variante A1^{PQ} tiene un espaciador adicional (Sq1) insertado dentro de una ubicación CRISPR. Típicamente, el espaciador Sq1 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago Q, y proporciona resistencia al fago Q y los fagos relacionados. La variante A1^{PQ} es resistente a ambos fagos P y Q. Luego, la variante A1^{PQ} se expone al fago R y se selecciona una variante resistente al fago (Variante A1^{PQR}). La variante A1^{PQR} tiene un tercer espaciador adicional (Sr1) insertado dentro de una ubicación CRISPR. Típicamente, Sr1 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago R, y también proporciona resistencia al fago R y fagos relacionados. La variante A1^{PQR} es resistente a todos los tres fagos. En algunas modalidades particularmente preferidas, la variante también es resistente a los fagos relacionados.

En modalidades adicionales, los métodos anteriores se usan en combinación para producir resistencia incrementada y expandida en los fagos. En algunas modalidades particularmente preferidas, estas variantes tienen resistencias

altas para familias del fago múltiples. Aún en modalidades adicionales, se producen cepas que son resistentes a fagos particulares o familias de fagos que son problemáticos en fábricas particulares y/o fermentadores.

Inmunidad mediada por CRISPR y Aplicaciones para las Cepas Resistentes al fago.

En contraste con las enseñanzas del arte previo que supone que CRISPR o los espaciadores CRISPR podrían estar involucrados para conferir inmunidad específica, la presente invención se basa, en parte, al descubrimiento sorprendente que se requieren genes o proteínas *cas* para la inmunidad contra un ácido nucleico objetivo o un producto de la transcripción del mismo. Sin embargo, no se pretende que la presente invención se limite a cualquier mecanismo particular, función, o medios de acción.

Aún más sorprendentemente, durante el desarrollo de la presente invención, se encontró que uno o más genes o proteínas *cas* están asociados con dos o más repeticiones CRISPR dentro de las ubicaciones CRISPR. En otros términos, los genes o proteínas *cas* parecen ser específicas para una repetición CRISPR de ADN dado, lo que significa que los genes o proteínas *cas* y la forma de la secuencia repetida forman un par funcional. Por consiguiente, uno o más espaciadores CRISPR encuentran uso junto con uno o más de estos pares funcionales (es decir, repeticiones CRISPR y genes *cas*) para modular la resistencia de una célula contra un ácido nucleico objetivo o un producto de la transcripción del mismo.

En una modalidad, para una o más espaciadores CRISPR que confieren inmunidad a la célula, las repeticiones CRISPR y los genes *cas* o las proteínas forman una combinación funcional (es decir, repeticiones CRISPR) y los genes *cas* o las proteínas son compatibles).

En modalidades preferidas adicionales, la presente invención proporciona genes/proteínas *cas* que influyen en la resistencia de las bacterias en los fagos. En modalidades preferidas adicionales, la presente invención proporciona por lo menos dos repeticiones CRISPR y por lo menos un gen/proteína *cas* útil en la predicción, determinación y/o modificación de la resistencia bacteriana en fagos. De hecho, la presente invención proporciona métodos para modificar el lisotipo (es decir, la resistencia/sensibilidad a varios fagos) de bacterias. Por consiguiente, la identificación y detección de las ubicaciones CRISPR en las células y fagos proporciona medios para determinar, predecir y modificar el perfil de resistencia de células, así como interacciones del fago-hospedador.

Ventajosamente, la aplicación de una o más ubicaciones CRISPR, dos o más repeticiones CRISPR, uno o más genes o proteínas *cas* y/o uno o más espaciadores CRISPR en la ingeniería genética proporciona medios para producir variantes resistentes o sensibles de células para su uso dentro de una variedad amplia de aplicaciones en la industria de la biotecnología.

Como se discutió con mayor detalle debajo, los fagos son parásitos naturales de bacterias que pueden desarrollarse durante la fermentación. Luego de una infección por fagos, se matan las bacterias que dañan el proceso de la fermentación. En la fermentación de lácteos, estas infecciones por fagos tienen a menudo impactos económicos mayores, que van desde una calidad reducida del producto fermentado hasta la pérdida completa del producto.

Para superar los problemas por fagos, las compañías de cultivos de inicio han desarrollado varias estrategias. Los programas de cultivo de inicio tradicionales han dependido de estrategias de rotación de defensa contra fagos (PDRS) para minimizar los fracasos debido al ataque de fagos (Ver por ejemplo, Klaenhammer, Adv. Appl. Microbiol., 30: 1 [1984]; Lawrence et al., J. Dairy res. 43: 141 [1976]; y Whitehead and Hunter, J., Dairy Res., 15: 1 12 [1947]). Estas estrategias confían en múltiples cepas genéticamente no relacionadas que tienen probabilidad de presentar espectros diferentes de sensibilidad a los fagos (es decir, lisotipos diferentes). Cuando un fago aparece durante un proceso de fermentación usando una cepa definida, una cepa que es idealmente de un lisotipo diferente (es decir, con un patrón de sensibilidad diferente a los fagos) se usa como reemplazo para la fermentación. La historia ha demostrado, sin embargo, que es difícil identificar suficientes números de lisotipos distintos para utilizar estas estrategias con éxito. De hecho, muchas cepas de interés industrial presentan características funcionales raras (por ejemplo, *S. thermophilus* con textura de acidificación rápida). Además, no todas las cepas presentan características apropiadas para producirse como cultivos de inicio. Además, debido a su rareza y al aumento en tamaño de las fábricas de lácteos, estas cepas se usan intensivamente.

Existen problemas adicionales con las estrategias de rotación de cultivos de inicio tradicionales. Aunque algunas cepas no son atacadas por fagos existentes cuando se introducen, los fagos aparecen a menudo eventualmente debido a la mutación del fago, modificación, y acumulación progresiva de ataques a la cepa recientemente introducida (Ver por ejemplo, Heap and Lawrence, N.Z.J. Dairy Sci. Technol., 11 :16 [1976]; Limsowtin y Terzaghi, N.Z.J. Dairy Sci. Technol., 11 :251 [1976]; Pearce, N.Z.J. Dairy Sci. Technol., 13: 166 [1978]; y Sanders y Klaenhammer, Appl. Environ. Microbiol., 40:500 [1980]). Por otra parte, en muchos casos, la longevidad y actividad de inicio de rotaciones de cepas complejas es imprevisible y a menudo conduce a un fracaso inicial (Ver por ejemplo, Limsowtin et al., N.Z.J. Dairy Sci. Technol., 13: 1 [1977]; y Thunell et al., J. Dairy Sci., 64, 2270 [1981]). Además, rotaciones prolongadas que involucran numerosas cepas incrementan el nivel y diversidad del fago que contamina a la planta (Ver por ejemplo, Heap and Lawrence, N.Z.J. Dairy Sci. Technol., 12:213 [1981]; Lawrence et.al, J. Dairy Sci., 61 :1 181 [1978]; y Thunell et al, J., Dairy Sci. 64, 2270 [1981]).

Para luchar contra la proliferación del fago, los programas de cultivo de inicio tradicionales han dependido del uso de cepas que presentan las mismas o similares funcionalidades tecnológicas pero sensibilidades diferentes a los fagos. Las cepas se usan en la rotación para realizar una fermentación sucesiva. Estos programas confían tradicionalmente en múltiples cepas genéticamente no relacionadas que por consiguiente presentan espectros diferentes de sensibilidad a los fagos (lisotipo). Las metodologías alternativas (Ver por ejemplo, la Patente de los Estados Unidos 5,593,885) utiliza programas de cultivo de inicio basados en el uso de conjuntos de cepas isogénicas que presentan diferente sensibilidad a los fagos, en lugar de cepas genéticamente no relacionadas que presentan lisotipos diferentes. El término “conjunto de cepas isogénicas” como se usa en la presente, define cepas que son idénticas a las del punto de vista cromosómico pero que cada una difiere por la presencia de uno o más mecanismos de resistencia del fago que se sostienen por medio del plásmido. En tal programa de rotación del cultivo de inicio, cuando un fago aparece durante un proceso de fermentación usando una cepa definida, una cepa que es idealmente de un lisotipo diferente (es decir, con un espectro diferente de sensibilidad a los fagos) se usa en el reemplazo para la fermentación. Debido a este lisotipo diferente, la segunda cepa no es afectada por los fagos que permanecen inactivos en el ambiente. La mayoría de la población de fagos inactivos se eliminan entonces mediante una fermentación sucesiva e higienización, y son erradicados en el momento en el que la primera cepa se usa nuevamente para la fermentación, si es que el sistema trabaja como se espera.

La presente invención proporciona métodos mejorados y composiciones adecuadas para adordar estos problemas en la industria de la fermentación. De hecho, la presente invención proporciona métodos y composiciones para la industria de la fermentación, en particular, para la industria láctea con una selección de cepas adecuadas que cumplan con las necesidades de las estrategias de rotación de defensa contra fagos. Además, la presente invención proporciona métodos y composiciones adecuadas para personalizar cepas que tienen lisotipos que se adaptan a un ambiente del fago particular. En particular, la presente invención proporciona métodos y composiciones adecuadas para dirigir la evolución de una cepa dada en varios lisotipos, para producir cepas que sólo difieren de otras por su espectro de sensibilidad a los fagos (lisotipo). Esta diferencia de lisotipos es una función del sistema CRISPR-cas, como se describe en la presente. En algunas modalidades preferidas, se obtienen lisotipos diferentes a través de la “modulación” de la resistencia del fago. En algunas modalidades particularmente preferidas, aunque los lisotipos son diferentes, las cepas de este tipo tienen un metabolismo idéntico (por ejemplo, de carbono, nitrógeno, etc.) y por lo tanto, funcionalidades idénticas (por ejemplo, acidificación, sabor, textura, etc.). Esto proporciona medios para amplificar la construcción de una rotación inicial. Además, la procesabilidad industrial de la cepa resistente a fagos es idéntica (por ejemplo, necesidades de nutrición, resistencia a la operación de procesamiento, etc.), reduciendo así, la necesidad de desarrollar procesos de producción específicos. De hecho, la presente invención proporciona métodos y composiciones adecuadas para minimizar las fallas en la fermentación debido al ataque de fagos. En algunas modalidades, se proporcionan métodos y composiciones para la producción de cultivos iniciales altamente resistentes a los fagos, mediante la asociación de cepas resistentes a fagos múltiples que difieren por su lisotipo. En algunas modalidades alternativas, se proporcionan métodos y composiciones para producir cultivos iniciales con funcionalidades industriales estrictamente idénticas para usarse en la fermentación de lácteos por rotación. En modalidades adicionales, se proporcionan métodos y composiciones que son adecuadas para reemplazar a los iniciales existentes previniendo ataques frecuentes a fagos en las plantas de lácteos, introduciendo una nueva cepa bacteriana que es resistente a los fagos involucrada en estos ataques hacia los fagos. En algunas modalidades, estos métodos y composiciones se usan iterativamente para combatir los ataques de fagos secuenciales.

El cultivo inicial es un cultivo bacteriano mixto. En algunas modalidades particularmente preferidas, los cultivos iniciales incluyen cantidades iguales de variantes resistentes al fago múltiples (es decir, por lo menos 2) que sólo difieren en sus CRISPR y su sensibilidad a los fagos. En algunas modalidades, estas variantes son del primer nivel de variantes resistentes a los fagos, (por ejemplo, las variantes A1.0 más A2.0, como se describe anteriormente). En algunas modalidades preferidas, las variantes se seleccionan de aquellas que están en el segundo nivel de variantes resistentes a los fagos, (por ejemplo, las variantes A 1.4 más A2.4, como se describe anteriormente). En algunas modalidades particularmente preferidas, las variantes se seleccionan de entre el tercer nivel de variantes resistentes a los fagos. En tales cultivos bacterianos mixtos, cuando una de las variantes es atacada por un fago dado, las otras variantes no son atacadas por el fago, debido a su distinta sensibilidad a los fagos y la fermentación no se ve afectada de forma adversa.

En algunas modalidades adicionales, se usa un iniciador principal y un respaldo del iniciador. El iniciador principal está compuesto de una sola cepa. En algunas modalidades, esta cepa es del primer nivel de las variantes resistentes a los fagos, mientras que en otras modalidades preferidas, la cepa es del segundo nivel, y aún en otras modalidades más preferidas, la cepa es del tercer nivel. En algunas modalidades preferidas, el respaldo del iniciador se basa en una variante resistente a los fagos obtenida independientemente de la misma cepa precursora. Esta segunda variante resistente al fago difiere de la otra variante por sus CRISPR y es del primer nivel de variantes resistentes a los fagos, mientras que en otras modalidades preferidas, la cepa es del segundo nivel, y aún en otras modalidades más preferidas, la cepa es del tercer nivel. Por ejemplo, en algunas modalidades, el iniciador principal está hecho de una variante A1.4 y un respaldo del iniciador está hecho de la cepa A2.4. Luego de una primera aparición de un fago durante la fermentación con el iniciador principal, este iniciador se descarta y reemplaza por el respaldo del iniciador. En otras modalidades preferidas más, se prepara también un tercer iniciador como respaldo

del iniciador, que actuará de respaldo para el respaldo. En modalidades preferidas, los iniciadores están hechos de variantes resistentes a múltiples fagos.

Aún en modalidades adicionales, la presente invención proporciona métodos y composiciones adecuadas en las estrategias de rotación. En algunas modalidades, en lugar de descartar al iniciador, frecuentemente atacado por los fagos, los iniciadores se usan de una manera cíclica aún cuando se observa un ataque por fagos. Esta estrategia limita el número de iniciadores a ser desarrollados. En algunas modalidades particularmente preferidas, los iniciadores son cada uno hechos de cepas resistentes a fagos múltiples en lugar de una sola. Esto proporciona robustez incrementada a los fagos emergentes. Aún en modalidades adicionales, se proporcionan iniciadores personalizados. En algunas modalidades preferidas, las variantes resistentes a los fagos se producen para combatir específicamente fagos que están presentes en una planta o instalación de fermentación dada.

TIPIFICACIÓN

Se proporciona un método para identificar (por ejemplo, tipificación) a una bacteria etiquetada.

La etapa de identificación puede realizarse amplificando (por ejemplo, amplificación por PCR) la ubicación CRISPR o una porción de la misma.

Un primer cebador puede diseñarse para hibridizar a una secuencia que se localiza corriente arriba de la primera repetición CRISPR de una ubicación CRISPR. A modo de ejemplo, el primer cebador puede hibridizarse a una parte de la secuencia líder común de la ubicación CRISPR. A modo de ejemplo adicional, el primer cebador puede hibridizar a un gen contiguo que se localiza en corriente arriba de la ubicación CRISPR.

El segundo cebador puede hibridizarse en corriente abajo desde por lo menos el primer espaciador CRISPR o por lo menos el primer centro espaciador CRISPR. El segundo cebador puede hibridizarse en una posición tan alejada como el remolque o incluso en un gen contiguo en corriente abajo. Preferentemente, el segundo cebador hibridiza dentro de la ubicación CRISPR. Preferentemente, el segundo cebador hibridiza por lo menos parcialmente a un espaciador CRISPR en corriente abajo o centro espaciador CRISPR.

Después de la amplificación, las secuencias de marcado pueden identificarse usando varios métodos que se conocen en el arte.

A modo de ejemplo, las secuencias de marcado pueden identificarse determinando el patrón de restricción del producto de amplificación. Por consiguiente, una vez que el ADN que comprende la ubicación CRISPR o una porción de la misma se ha amplificado, puede digerirse (por ejemplo, cortarse) con una o más enzimas de restricción.

Como se usa en la presente, el término “enzimas de restricción” se refiere a las enzimas (por ejemplo, enzimas bacterianas) cada una de las cuales corta el ADN de hebra doble en o cerca de una secuencia del nucleótido específica. Las enzimas de restricción son bien conocidas en el arte y pueden obtenerse fácilmente, por ejemplo, de una variedad de fuentes comerciales (por ejemplo, New England Biolabs, Inc., Beverly, Massachusetts). De manera semejante, los métodos para usar las enzimas de restricción también son generalmente bien conocidos y de rutina en el arte. Pueden usarse las enzimas de restricción que producen entre 10 y 24 fragmentos de ADN cuando cortan la ubicación CRISPR o una porción de la misma. Ejemplos de tales enzimas incluyen, pero no se limitan a, *Alu1*, *Mse1*, y *Tsp5091*. Pueden detectarse fragmentos de ADN obtenidos usando las enzimas de restricción, por ejemplo, como bandas mediante electroforesis por gel. Las enzimas de restricción pueden usarse para crear Polimorfismos de Longitud del Fragmento de Restricción (RFLP).

Los RFLP son generados cortando (“restringiendo”) una molécula de ADN con una endonucleasa de restricción. Se han aislado muchos centenares de tales enzimas, elaboradas naturalmente por bacterias. En esencia, las bacterias usan tales enzimas como un sistema defensivo, para reconocer y después desdoblar (restringir) cualquier molécula de ADN extranjera que podría entrar en la célula bacteriana (por ejemplo, una infección vírica). Cada uno de los muchos centenares de diferentes enzimas de restricción se han encontrado que cortan (es decir, “desdoblan” o “restringen”) ADN en una secuencia diferente de los 4 nucleótidos básicos (A, T, G, C) que constituyen a todas las moléculas de ADN, por ejemplo, una enzima podría específicamente y solamente reconocer la secuencia A-A-T-G-A-C, mientras que otra podría específicamente y sólo reconocer la secuencia G-T-A-C-T-A, etc. Dependiendo de la única enzima involucrada, tales secuencias de reconocimiento pueden variar en longitud, desde tan solo 4 nucleótidos hasta 21 nucleótidos. Cuanto mayor sea la secuencia de reconocimiento, menos fragmentos de restricción resultarán, ya que cuanto mayor sea el sitio de reconocimiento menor será la probabilidad de que se encuentre repetidamente a lo largo del ADN.

A modo de ejemplo adicional, las secuencias de marcado pueden identificarse mediante determinado o también determinando la diferencia en el tamaño del producto de amplificación.

La separación puede lograrse mediante cualquier método adecuado para separar ADN, que incluye, pero no se limita a, electroforesis por gel, cromatografía líquida de alta eficacia (HPLC), espectroscopia de masa, y uso de un

dispositivo microfluidico. En una modalidad, los productos de amplificación o fragmentos de ADN están separados por electroforesis de gel de agarosa. La electroforesis de gel separa las moléculas cargadas de tamaño diferente por su proporción de movimiento a través de un gel estacionario bajo la influencia de una corriente eléctrica. Estos productos de amplificación separados o fragmentos de ADN pueden visualizarse fácilmente, por ejemplo, mediante la tinción con bromuro de etidio y observando el gel bajo iluminación de UV. El patrón de bandeo refleja el tamaño del ADN de restricción digerido o los productos de amplificación.

A modo de un ejemplo adicional, las secuencias de marcado pueden identificarse mediante la formación de secuencias de los productos de amplificación.

La secuencia de los productos amplificados puede obtenerse por cualquier método conocido en el arte, que incluye los métodos automáticos y los métodos de formación de secuencias manuales. Ver, por ejemplo, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, Nueva York; Roe *et al.* (1996) *DNA Isolation and Sequencing* (Essential Techniques Series, John Wiley & Sons).

Los métodos de Hibridización también están dentro del alcance de la presente invención, usando cualquier molécula de ácidos nucleicos tal como una sonda, o una molécula de ácidos nucleicos capaces de hibridizar a una secuencia del nucleótido particular. Ver, por ejemplo, Sambrook *et al.* (1989) *Molecular Cloning: Laboratory Manual* (2d ed., Cold Spring Harbor laboratory Press Plainview, Nueva York).

En las técnicas de hibridización, las sondas de hibridización pueden ser fragmentos de ADN genómicos, productos amplificados por PCR, u otros oligonucleótidos, y pueden comprender toda o parte de una secuencia del nucleótido conocida. Además, puede ser etiquetada con un grupo perceptible tal como ^{32}P , o cualquier otro marcador perceptible, tal como otros radioisótopos, un compuesto fluorescente, una enzima, o un co-factor de la enzima. El término "etiquetado," con respecto a la sonda, pretende abarcar la etiquetación directa de la sonda acoplando (es decir, enlazando físicamente) una sustancia perceptible por la sonda, así como la etiquetación indirecta de la sonda mediante la reactividad con otro reactivo que se etiqueta directamente. Ejemplos de etiquetación indirecta incluyen etiquetación en extremos de una sonda de ADN con biotina, tal que puede detectarse con estreptavidina etiquetada fluorescentemente.

Se abarcan también los métodos que abarcan las técnicas hibridización para detectar o diferenciar cepas bacterianas. Estos incluyen, pero no se limitan a, transferencia Southern (ver, por ejemplo, Van Embden *et al.* (1993) *J. Clin. Microbiol.* 31 :406-409), ensayos de movilidad de cambio (ver, por ejemplo, la Solicitud Publicada de los Estados Unidos No. 20030219778), ensayos de formación de secuencias usando arreglos de oligonucleótidos (ver, por ejemplo, Pease *et al.* (1994) *Proc. Natl. Acad. Sci. USA.* 91 :5022-5026), espoligotipificación (ver, por ejemplo, Kamerbeek *et al.* (1997) *J. Clin. Microbiol.* 35:907-914), Hibridación fluorescente in situ (FISH) (ver, por ejemplo, Amann *et al.* (1990) *J. Bacteriol.* 172:762-770) y ensayos de rastreo heteroduplo o análisis de movilidad heteroduplo (ver, por ejemplo, White *et al.*, (2000) *J. Clin. Micro.* 38:477-482).

La secuencia de marcado que se identifica puede compararse con una base de datos de la secuencia del fago y/o una base de datos de la secuencia bacteriana. Típicamente, las secuencias de marcado coinciden con una o más secuencias en la base de datos de la secuencia del fago pero no con la base de datos de la secuencia bacteriana.

Debido a que las nuevas bacterias etiquetadas se preparan usando los métodos descritos en la presente, puede crearse una base de datos de etiquetas que permita la identificación específica de bacterias que se han etiquetado.

Se proporciona el uso de una secuencia obtenida u obtenible de un bacteriófago (por ejemplo, en la fabricación de una bacteria etiquetada) para etiquetar y/o identificar una bacteria, en donde dicha secuencia se integra a un extremo de la ubicación CRISPR de la bacteria precursora.

Se proporciona el uso de una secuencia obtenida u obtenible de un bacteriófago (por ejemplo, en la fabricación de una bacteria etiquetada) para etiquetar e/o identificar una bacteria, en donde dicha secuencia comprende: (i) por lo menos una secuencia que es homóloga (por ejemplo, es idéntica) a una repetición CRISPR en la ubicación CRISPR de dicha bacteria; y (ii) una secuencia de marcado.

Se proporciona el uso de una secuencia para etiquetar y/o identificar una bacteria (por ejemplo, en la fabricación de una bacteria etiquetada), en donde dicha secuencia se obtiene o es obtenible por: (a) exponer una bacteria precursora a un bacteriófago; (b) seleccionar un mutante insensible a bacteriófagos; (c) comparar la ubicación CRISPR o una porción de la misma de la bacteria precursora y el mutante insensible a bacteriófagos; y (d) seleccionar una secuencia en la ubicación CRISPR o una porción de la misma del mutante insensible a bacteriófagos que no están presentes en la bacteria precursora.

CRISPR y Generación de Mutantes Resistentes a los fagos

Durante el desarrollo de la presente invención, los experimentos se dirigieron para determinar si las ubicaciones CRISPR se alteraron durante la generación natural de mutantes resistentes a los fagos. Un sistema de modelo de

fago hospedador se seleccionó, que consiste de una cepa *S. thermophilus* del tipo natural sensible a los fagos usada ampliamente en la industria láctea, DGCC7710 (WT) y dos bacteriófagos virulentos distintos, pero estrechamente relacionados aislados de las muestras de yogur industriales, a saber, el fago 858 y el fago 2972 (Levesque et al., Appl. Environ. Microbiol., 71 :4057 [2005]). Nueve mutantes resistentes al fago se generaron independientemente mediante una prueba inmunogénica de la cepa WT con el fago 858, fago 2972 o simultáneamente con ambos, y sus ubicaciones CRISPR fueron analizadas. Se observaron las diferencias de forma consistente con la ubicación CRISPR1, donde se insertaron 1 a 4 espaciadores adicionales próximos a los 32 espaciadores presentes en la cepa WT (Ver, Figura 9). La adición de nuevos espaciadores en respuesta a la infección del fago pareció ser polarizada hacia un extremo de la ubicación CRISPR1. Esto es consistente con las observaciones previas de hipervariabilidad del espaciador en el extremo líder de la ubicación CRISPR en varias cepas (Ver por ejemplo, Ver, Pourcel et al., Microbiol., 151 :653 [2005]; y Lillestol et al., Archaea 2:59 [2006]). El análisis de secuencia de espaciadores adicionales insertados en la ubicación CRISPR1 de varios mutantes resistentes al fago reveló una similitud con las secuencias encontradas dentro de los genomas de los fagos usados en la prueba inmunogénica. Se observaron similitudes a lo largo de los genomas del fago, en la mayoría de los módulos funcionales, tanto en las hebras codificantes como las no codificantes. Ninguna secuencia particular, gen o grupo funcional pareció ser específicamente el objetivo. Estos resultados indican que luego de volverse resistentes a los bacteriófagos, la ubicación CRISPR1 se modificó por la integración de nuevos espaciadores derivados aparentemente desde el ADN del fago. Sin embargo, no se pretende que la presente invención se limite a cualquier mecanismo específico.

Sorprendentemente, se observó que algunas cepas fueron resistentes a ambos fagos mientras que otros sólo fueron resistentes al fago usado en la prueba inmunogénica (Ver, Figura 9). El perfil de resistencia al fago pareció estar correlacionado con el contenido del espaciador, por medio del cual, las cepas con espaciadores que muestran 100% de identidad para las secuencias conservadas en ambos fagos fueron resistentes a ambos fagos, tales como los espaciadores S3, S6 y S7. En contraste, cuando los polimorfismos del nucleótido se observaron entre el espaciador y la secuencia del fago (de 1 a 15 SNPs en más de 29 o 30 nucleótidos), el espaciador no pareció proporcionar resistencia, tal como los espaciadores S1, S2, S4, S5 y S8 (Ver, Figura 9).

Adicionalmente, cuando se insertaron varios espaciadores (S9-S14), los niveles de resistencia al fago fueron superiores. Estos resultados indican que la ubicación CRISPR1 está sujeta a cambios evolutivos dinámicos y rápidos manejados por la exposición al fago. Estos resultados indican que las ubicaciones CRISPR pueden alterarse de hecho durante la generación de mutantes resistentes al fago y establecerse un enlace entre el contenido del CRISPR y la sensibilidad al fago. Así, se contempla que la presencia de un espaciador CRISPR idéntico a una secuencia del fago proporciona resistencia contra fagos que contienen esta secuencia particular.

Para determinar si el contenido del espaciador CRISPR define la resistencia del fago, la ubicación CRISPR1 se alteró agregando y anulando espaciadores, y se probó la sensibilidad de la cepa a los fagos. Todos los constructos se generaron e integraron en el cromosoma *S. thermophilus* usando métodos conocidos en el arte (Ver por ejemplo, Russell y Klaenhammer, Appl. Environ. Microbiol., 67:4361 [2001]). Los espaciadores y repeticiones en la ubicación CRISPR1 de la cepa WT_{Φ858}^{+S1S2} se removieron y se reemplazaron con una repetición simple sin cualquier espaciador. La cepa resultante WT_{Φ858}^{+S1S2} ΔCRISPR1 fue sensible al fago 858, indicando que la resistencia a los fagos del mutante resistente al fago original (WT_{Φ858}^{+S1S2}) probablemente se enlazó para la presencia de S1 y S2 (Ver, Figura 10).

Además, para determinar si la adición de los espaciadores proporcionan nueva resistencia al fago, la ubicación CRISPR1 de la cepa WT_{Φ2972}^{+S4} se reemplazó con una versión que sólo contiene los espaciadores S1 y S2. La sensibilidad del fago al constructo resultante fue entonces probada. La cepa resultante WT_{Φ2972}^{+S4} : : pS1S2 ganó resistencia al fago 858, sugiriendo que estos dos espaciadores tienen la capacidad de proporcionar resistencia al fago *de novo* (Ver, Figura 10). Estas modificaciones observadas establecen un enlace entre el contenido del espaciador CRISPR y la resistencia del fago.

En el proceso de generación de la cepa WT_{Φ858}^{+S1S2} ΔCRISPR1, WT_{Φ858}^{+S1S2} : : pR, se creó una variante que contiene el vector de integración con una sola repetición insertada entre los genes *cas* y la ubicación CRISPR1 nativo (Ver, Figura 10). Inesperadamente, la cepa WT_{Φ858}^{+S1S2} : : pR, fue sensible al fago 858, aunque los espaciadores S1 y S2 permanecieron presentes en el cromosoma (Ver, Figura 10). Semejantemente, el constructo WT_{Φ2972}^{+S4} : : pS1S2 perdió resistencia al fago 2972, aunque el espaciador S4 está presente en el cromosoma (Ver, Figura 10). Estos resultados indican que los espaciadores no proporcionaron resistencia, y quizás tengan que tener un contexto genético particular para ser eficientes.

Aunque los experimentos recientes sugirieron involucrarse en la reparación del ADN (Makarova et al., Nucl. Acids Res., 30:482 [2002]), la hipótesis actual es que los genes *cas* (Jansen et al., Mol. Microbiol., 43: 1565 [2002]; y Haft et al., PLoS Comput. Biol., 1: e60 [2005]) están involucrados en la inmunidad mediada por CRISPR (Makarova et al., Biol. Direct. 1 :7 [2006]). En experimentos adicionales, dos genes *cas* en la cepa WT_{Φ858}^{+S1S2} se volvieron inactivos, principalmente *cas5* (COG3513) y *cas7* que son equivalentes al *str0657Istu0657* y *str0660Istu066*, respectivamente (Ver, Bolotin et al., Nat. Biotechnol., 22: 1554 [2004]; y Bolotin et al., Microbiol., 151 :2551 [2005]). La inactivación de *cas5* produjo una pérdida de la resistencia del fago (Ver, Figura 10). Además, es posible que

Cas5 actúe como una nucleasa, puesto que contiene una porción de nucleasa del tipo HNH. En contraste, al inactivar *cas7* no se alteró la resistencia al fago 858 (Ver, Figura 10). Sin embargo, no se pretende que la presente invención se limite a cualquier mecanismo particular. Además, no funcionaron los experimentos para generar los mutantes resistentes al fago CRISPR1 a partir de la supresión de *cas 7*. Aunque no se pretende que la presente invención se limite a cualquier mecanismo particular, esto puede ser debido a que *Cas7* está involucrado en la síntesis y/o inserción de nuevos espaciadores y repeticiones adicionales.

Al probar la sensibilidad de los mutantes resistentes al fago, se encontró que la formación de la placa se redujo dramáticamente, pero que una población relativamente pequeña de bacteriófagos retuvo la capacidad para infectar los mutantes. Se analizaron adicionalmente variantes del fago derivadas del fago 858 que retuvieron la capacidad para infectar WT_{Φ858}^{ΔS1S2}. En particular, se investigaron las secuencias de la región del genoma que corresponde a espaciadores adicionales S1 y S2 en dos variantes del fago virulentas. En ambos casos, la secuencia del genoma de la variante del fago se mutó y se identificaron dos distintos polimorfismos del nucleótido simple en la secuencia que corresponde al S1 espaciador (Ver, Figura 13).

En conjunto, los procariotes parecen haber desarrollado un sistema de “inmunidad” basado en ácidos nucleicos por medio del cual la especificidad se dicta por el contenido del espaciador CRISPR, mientras que la resistencia se proporciona por la maquinaria enzimática *Cas*. Adicionalmente, se especuló que algunos de los genes *cas* que no proporcionan resistencia directamente están realmente involucrados en la inserción de espaciadores y repeticiones CRISPR adicionales, como parte de una respuesta adaptable “inmune”. Este sistema basado en ácidos nucleicos contrasta con su contraparte basada en aminoácidos en eucariotas por medio de los cuales la inmunidad adaptativa no es heredable. La naturaleza heredable de espaciadores CRISPR soporta el uso de las ubicaciones CRISPR como objetivos para los estudios evolutivos, de tipificación y genómico comparativos (Ver, Pourcel et al, *supra*; Groenen et al., *Mol. Microbiol.*, 10: 1057 [1993]; Mongodin et al, J., *Bacteriol.*, 187:4935 [2005]; y DeBoy et al, J., *Bacteriol.*, 188:2364 [2006]). Debido a que este sistema es reactivo al ambiente del fago, desempeña probablemente una función significativa en la evolución procariótica y ecológica y proporciona una perspectiva histórica de exposición del fago, así como una herramienta predictiva para la sensibilidad del fago. Sin embargo, no se pretende que la presente invención se limite a cualquier mecanismo particular. No obstante, la presente invención proporciona métodos y composiciones para utilizar el sistema CRISPR/*cas* como un medio de defensa del virus, y también potencialmente para reducir la diseminación de elementos genéticos móviles y la adquisición de características indeseables tales como los genes de resistencia antimicrobiana y marcadores de virulencia. En algunas modalidades, se contempla además desde una perspectiva de evolución del fago, secuencias del fago integradas dentro de las ubicaciones CRISPR también proporcionan puntos de ancla adicionales para facilitar la recombinación durante las infecciones del fago subsecuentes, incrementando así, la concentración del gen para la cual los fagos tienen acceso (Ver, Hendrix et al, *Proc. Natl. Acad. Sci. USA.* 96:2192 [1999]). Puesto que el ubicaciones CRISPR se encuentra en la mayoría de los géneros bacterianos, y está omnipresente en Archaea (Ver, Jansen et al, *supra*; Lillestol et al, *supra*; y Goode y Bickerton J. *Mol. Evol.*, 62:718 [2006]), proporcionan nuevas visiones en la relación y evolución co-dirigida entre los procariotes y sus depredadores.

Fagos de Biocontrol

Se proporcionan métodos y composiciones para el desarrollo de fagos como agentes de biocontrol. Como se indica en la presente, las bacterias pueden volverse resistentes al ataque de fagos incorporando secuencias derivadas del fago (espaciadores) en ubicaciones CRISPR activas. El Fago puede escapar de esta resistencia mediante una mutación dentro de la secuencia del genoma que corresponde al espaciador o la secuencia de reconocimiento de la porción CRISPR, que corresponde a un sistema *Cas*-CRISPR dado. A través de rondas iterativas de pruebas inmunogénicas para el fago, con el fin de crear derivados resistentes al fago mediados por CRISPR de una cepa hospedadora y el aislamiento de los mutantes de escape a fagos, la presente invención proporciona fagos que han sido alterados dentro de secuencias objetivo CRISPR y/o sitios de reconocimiento CRISPR putativos que dirigen la inserción del espaciador directa. Además, la presente invención proporciona fagos que se ha diseñado sintéticamente tal que la secuencia de la porción CRISPR para un sistema CRISPR *Cas* dado ha sido eliminado. Estos fagos “alterados”, aplicados como un coctel o en un “esquema de la rotación secuencial” reduce la capacidad de las bacterias objetivo para adaptar la resistencia vía el sistema CRISPR. De hecho, la presente invención proporciona un conjunto diverso de fagos virulentos para usarse como agentes de biocontrol. En modalidades particularmente preferidas, esta diversidad se direcciona al mecanismo CRISPR dirigido de resistencia al fago, tal que la capacidad del organismo hospedador para desarrollarse rápidamente contra el ataque al fago (por medio de CRISPR) se reduce o elimina severamente. La administración del fago diverso, ya sea como un coctel o en una rotación secuencial reduce adicionalmente la posibilidad de que el organismo hospedador se adapte o desarrolle resistencia al fago CRISPR-dirigido.

Los fagos son agentes antimicrobianos naturales que se han estudiado extensivamente como un agente terapéutico alternativo a los antibióticos. Este interés se ha renovado recientemente, debido a la proliferación de patógenos resistentes a los múltiples antibióticos. Como con los antibióticos, las bacterias han desarrollado mecanismos múltiples para superar el ataque de fagos. La presente invención proporciona métodos y composiciones que involucran el uso de *Cas*-CRISPR mediando la resistencia del fago para generar una población del fago diversa,

crear fagos sintéticos desprovistos de secuencias con porción CRISPR, así como métodos para administrar tal fago que reducirá la capacidad de un organismo objetivo de desarrollar resistencia contra el fago.

Como se detalla en la presente, se han descrito sistemas Cas-CRISPR en una gama amplia de organismos que incluyen ejemplos de géneros patógenos. Luego de la infección del fago, pueden encontrarse bacterias que escapan a la lisis que contienen nuevas secuencias espaciadoras dentro de una ubicación CRISPR. El nuevo espaciador es típicamente de una longitud definida que es característica para una ubicación CRISPR dada y se deriva del genoma del fago atacante al que confiere la resistencia. Como el nivel de resistencia conferido por un solo espaciador no está a menudo completo, los fagos pueden escapar al mecanismo. El análisis de "fagos de escape" indicó que los genomas mutaron en o próximos a la secuencia del espaciador correspondiente encontrada en la variante hospedadora resistente. Además, los "fagos de escape" son totalmente virulentos a la variante hospedadora mediada por CRISPR a partir de la cual fueron derivados.

Un único aspecto de fago terapéutico, que lo distingue de los antibióticos tradicionales, es la capacidad para propagarse exponencialmente junto con las bacterias infectadas. Mientras que esto puede ser ventajoso desde una perspectiva farmacológica, también proporciona oportunidades únicas para que el fago desarrolle una respuesta de adaptabilidad de las bacterias objetivo hacia el ataque del fago.

Las bacterias han desarrollado varios mecanismos de defensa contra el fago virulento. Como se indica en la presente, las ubicaciones Cas-CRISPR desempeñan una función para conferir resistencia al fago bacteriano. Después de la infección del fago, el análisis de bacterias supervivientes encontró que algunos aislados habían insertado un nuevo elemento espaciador dentro de su ubicación CRISPR residente, cuya secuencia fue idéntica a la encontrada en el genoma del fago correspondiente. Cuando se realizó una prueba inmunogénica con el fago, estas variantes resistentes al fago de primera generación mediadas por CRISPR dieron surgimiento a placas; el fago que se encontró resultó ser totalmente infectivo en ambos, precursor y derivado. El análisis de estos fagos "CRISPR-escape" indicó que sus genomas mutaron en la secuencia que corresponde al espaciador CRISPR albergada por la variante resistente al fago o en una secuencia próxima que se cree dirige la inserción del espaciador y se identifica como la porción CRISPR específica para un sistema Cas-CRISPR dado. Por consiguiente, el fago "CRISPR-escape" es potencialmente más virulento que las variantes precursoras y de primera generación, puesto que este fago es capaz de infectar la cepa precursora y la variante CRISPR de primera generación.

Como se indicó anteriormente, se han identificado las ubicaciones CRISPR en varios géneros/especies de bacterias que incluyen ejemplos de patógenos conocidos y microorganismos de desechos. También como se describe en la presente, la presente invención proporciona métodos y composiciones para la utilización de las ubicaciones CRISPR en combinación con las proteínas Cas para conferir "inmunidad" para invadir el ADN extranjero, en particular, los bacteriófagos. También como se describe en la presente, las cepas bacterianas que albergan a ubicaciones CRISPR-cas "activas", que contienen un espaciador que es idéntico a una secuencia correspondiente dentro de un genoma del fago (es decir, un "protoespaciador"), confiere a la cepa bacteriana, resistencia al fago. En algunas modalidades preferidas, son conocidas las secuencias del genoma del fago de biocontrol. En algunos métodos particularmente preferidos, los microorganismos objetivo aislados se examinan para la presencia de las ubicaciones CRISPR. En algunas modalidades preferidas, se usa PCR utilizando cebadores específicos para secuencias conservadas que flanquean las ubicaciones CRISPR del microorganismo objetivo. En algunas modalidades preferidas, los productos de amplificación se secuencian en comparación con la secuencia del genoma del fago de biocontrol. En algunas modalidades preferidas, la generación de variantes resistentes al fago CRISPR y análisis del espaciador/protoespaciador proporciona medios para identificar la porción CRISPR específica. Una vez identificada, la información de la secuencia se usa para diseñar y sintetizar un fago desprovisto de la porción CRISPR. Así, el fago resultante es insensible a la resistencia mediada por CRISPR-cas. En estas valoraciones, la ausencia de espaciadores con similitud al genoma del fago indica la susceptibilidad del microorganismo objetivo al fago de biocontrol. Así, el fago de biocontrol tiene un grado mayor de virulencia y eficiencia como agente de biocontrol.

La presente invención proporciona métodos y composiciones adecuadas para uso en alimentos, pienso, industria médica y veterinaria para generar el fago con un intervalo hospedador más amplio y un método de aplicación para un biocontrol más eficiente de las bacterias. La presente invención proporciona medios para producir un número suficiente de fagos alterados (en respuesta a CRISPR) para reducir significativamente la capacidad de las bacterias nativas de desarrollar una resistencia mediada por CRISPR eficiente. La presente invención también proporciona métodos de aplicación/administración diseñados tal que la proporción de evolución de las bacterias nativas esté significativamente reducida.

EXPERIMENTAL

Los siguientes ejemplos se proporcionan con el fin de demostrar e ilustrar adicionalmente ciertas modalidades preferidas y aspectos de la presente invención y no debe interpretarse como una limitación del ámbito de aplicación del mismo.

En la experiencia de divulgación que se indica, se aplican las siguientes abreviaturas: °C (grados centígrados); rpm (revoluciones por minuto), H₂O (agua), HCl (ácido clorhídrico); aa (aminoácidos), pb (par de bases); kb (par de

kilobases); kD (kilodaltons), g (gramos); µg y ug (microgramos); mg (miligramos); ng (nanogramos); µl y ul (microlitros); ml (mililitros); mm (milímetros); nm (nanómetros); µm y um (micrómetro), M (molar); mM (millimolar); µM y uM (micromolar), U (unidades); V (voltios), MW (peso molecular), s (segundos), min(s) (minuto/minutos); hora (s) (hora / horas), MOI (multiplicidad de la infección); EOP (eficiencia de placa); UFP (unidades formadoras de placa); MgCl₂ (cloruro de magnesio); de NaCl (cloruro de sodio); OD₄₂₀ (densidad óptica a 420 nm); PAGE (electroforesis en gel de poliacrilamida); EtOH (etanol), PBS (buffer salino de fosfatos [150 mM NaCl, 10 mM de buffer de fosfatos salino, pH 7.2]); SDS (dodecil sulfato de sodio) ; Tris (tris (hidroximetil) aminometano); w/v (peso a volumen); v/v (volumen a volumen); Amicon (Amicon, Inc., Beverly, MA); ATCC (American Type Culture Collection, Manassas, VA); Amersham (Amersham Biosciences, Inc., Piscataway, NJ); NEB (New England Biolabs, Beverly, MA), Becton Dickinson (Becton Dickinson Labware, Lincoln Park, NJ); BioRad (BioRad, Richmond, CA); Clontech (CLONTECH Laboratories, Palo Alto, CA); Difco (Difco Laboratories, Detroit, MI); Gibco BRL o Gibco BRL (Life Technologies, Inc., Gaithersburg, MD), Sigma (Sigma Chemical Co., St. Louis, MO), y Sorvall (Sorvall Instruments, una subsidiaria de DuPont Co., Biotechnology Systems, Wilmington, DE).

La presente invención utiliza, salvo se indique lo contrario, las técnicas convencionales de la química, la biología molecular, microbiología, inmunología y la recombinación del ADN, que están dentro de las capacidades de una persona con competencias normales en la materia. Esas técnicas son bien conocidas por los expertos en el arte. Como se usa aquí, DGCC7710 también se conoce como "WT"; DGCC7710RH1 también se conoce como "DGCC7710-RH1" y "RHL"; DGCC7710RH2 también se conoce como "DGCC7710-RH2" y "RH-2"; DGCC7778cas1 también se conoce como "DGCC7778cas1 KO", "CAS1 KO" y "cas1 KO"; DGCC7778cas4 también se conoce como "DGCC7778cas4KO"; DGCC7778 también se conoce como "WTφ858 + S1 S2"; DGCC7778RT también se conoce como "WTφ858 S1 + S2:: PR"; DGCC7778RT es también conocida como "WTφ858 + S1 S2ΔCRISPR1 "; DGCC7710-R2 también se conoce como " WTφ2972 + S4 ", y DGCC7710-R2S1 S2 es también conocida como " WTφ2972 + S4:: PS1S2 ".

EJEMPLO 1

Manipulación de Espaciadores Fago-Específicos

En este ejemplo, se describen los experimentos realizados para manipular espaciadores fago-específicos. En algunos experimentos, se describe que se insertó un espaciador específico de fagos en un CRISPR existente y funcional, para proporcionar resistencia a los fagos correspondientes. La cepa bacteriana utilizada fue *Streptococcus thermophilus* ST0089 y el fago, fue el 2972. *S. thermophilus* ST0089 es una cepa de importancia industrial utilizada en la fabricación de yogur. Esto es genéticamente susceptible a la manipulación, y sensible a la conocida virulencia del fago 2972.

Las ubicaciones CRISPR fueron determinadas en la cepa ST0089. Esto se determinó preferencialmente, secuenciando el genoma entero de ST0089. Alternativamente, las ubicaciones CRISPR se identifican a través de PCR usando conjuntos de cebadores con secuencias idénticas a los elementos CRISPR de *S. thermophilus* CRISPR, previamente identificados.

Una vez identificados, las secuencias de las ubicaciones CRISPR fueron determinadas así como las regiones proximales que contenían a los genes relevantes cas.

Al menos una ubicación en particular CRISPR-cas, fue seleccionada para una mayor manipulación. La funcionalidad de la ubicación se determinó mediante análisis *in silico* de las regiones espaciadoras y sus homologías con secuencias de ADN de fagos (es decir, la ausencia y/o la presencia de secuencias espaciadoras y la correlación a la infección de fagos con la cepa ST0089). En ausencia de esta correlación, la funcionalidad se suponía, basándose en la presencia de todos los elementos documentados (es decir, repeticiones, espaciadoras, secuencias líderes, y los genes cas que supuestamente codifican proteínas de tamaño completo).

Las secuencias espaciadoras adecuadas, fueron elegidas entre el genoma de los fagos 2972. Los criterios utilizados para seleccionar el espaciador generalmente se basan en la longitud de los espaciadores en la ubicación CRISPR seleccionado y la identidad (de preferencia de aproximadamente el 100%) a la secuencia de los fagos. De hecho, cualquier secuencia adecuada de fagos, encuentra uso en diversas modalidades de la presente invención.

En algunas modalidades, una unidad CRISPR constituida por una secuencia espaciadora de un fago 2972, flanqueada por dos elementos de repetición (idénticos a la ubicación CRISPR seleccionada) fue sintetizada químicamente. Por definición, esta "unidad CRISPR" sintética es de aproximadamente 100 pb de longitud y es demasiado corta para garantizar la integración en la ubicación CRISPR.

Por lo tanto, fue construido ADN de flanqueo adicional junto con la unidad CRISPR. Un mínimo de 500 pb de ADN homólogos, idénticos a la ubicación CRISPR objetivo, que flanquea la unidad sintética CRISPR, se elaboró con el fin de facilitar la integración.

En otras modalidades, hay varios enfoques. En una modalidad, un constructo emula la adición de un nuevo espaciador de sobre el CRISPR existente. En algunas modalidades alternativas, toda la ubicación CRISPR se sustituye por la unidad sintética CRISPR.

- 5 El CRISPR integrante resultante, se verificó mediante la secuenciación del ADN de la ubicación CRISPR antes de las pruebas biológicas. Además, los patrones de sensibilidad a fagos del CRISPR integrante contra el fago 2972, fueron analizados y comparados con la cepa precursora.

10 La construcción CRISPR integrante, demostró exitosamente la correlación directa entre la presencia de un espaciador de específico en el contexto adecuado de CRISPR-cas.

En otros experimentos, se realizó la inserción de un espaciador homólogo al ADN de un fago en una célula receptora. En estos experimentos, se diseñó un nuevo espaciador de CRISPR a partir de ADN de fagos (con un 100% de identidad al ADN de fagos) dentro del gen antirreceptor y se insertó en la célula en una ubicación CRISPR.

15 El gen antirreceptor fue utilizado como objetivo porque se ha encontrado que los espaciadores de CRISPR de otras cepas muestran similitud con genes antirreceptor de fagos. Cuatro cepas con espaciadores que mostraron identidad con genes antirreceptores de fagos fueron resistentes al fago en particular. El mutante fue expuesto a fagos, y se encontró que era resistente.

20 En otros experimentos, un espaciador se insertó en hospedador original, pero no en una ubicación CRISPR. El mutante resultante conserva su sensibilidad a los fagos. Así pues, estos experimentos mostraron que el espaciador debe de estar en un entorno particular dentro del contexto de un CRISPR y genes cas.

25 En otros experimentos, un espaciador de CRISPR en particular, fue eliminado de una ubicación CRISPR que se presentaba naturalmente. Esta eliminación suprimió la inmunidad contra de un determinado fago y el hospedador se hizo sensible (es decir, perdió la resistencia) a los fagos a los que el espaciador era homólogo. Los resultados de estos experimentos se presentan en la Figura 10.

30 Aún en experimentos adicionales, todas las combinaciones de repeticiones CRISPR-cas se insertaron en una célula receptora a fin de proporcionar inmunidad contra la entrada de ácidos nucleicos.

En otros experimentos, se preparó un plásmido compuesto por un espaciador CRISPR utilizando los métodos establecidos en el presente. Los intentos de transferir este plásmido en las células que contienen el mismo espaciador, no fueron exitosos. Sin embargo, los plásmidos que no contienen el espaciador, pueden ser transformados en células. Las Figuras 11 y 12 ilustran estos resultados.

35

En otros experimentos, las combinaciones CRISPR-cas presentes en dos cepas diferentes se intercambiaron. Este intercambio de espaciadores se mostró para modificar sus fenotipos (sensibilidad / resistencia a fagos). Como se indica en este documento cuando S1S2 es introducido en una cepa con S4, se conmuta la sensibilidad del fago (Ver Figura 10).

40

En otros experimentos, se preparan las diferentes combinaciones de repeticiones de cas-CRISPR. No sólo se requieren los genes o las proteínas cas, sino los pares específicos de repeticiones de cas-CRISPR son necesarios para la funcionalidad. Cuando los genes o las proteínas cas provienen de otra ubicación CRISPR, la cepa se mantiene sensible a los fagos.

45

Todavía en experimentos adicionales, uno o más genes cas (de una unidad funcional CRISPR-cas) fueron suprimidos. Los genes cas son necesarios para proporcionar inmunidad. Las mutantes cas siguen siendo sensibles a los fagos, a pesar de la presencia del espaciador idéntico al ADN del fago. En estos experimentos, cas 5 (anteriormente conocido como cas 1) y cas 7 (anteriormente conocido como cas 4) fueron suprimidos. Cas 5 ha demostrado ser necesario para la resistencia. Además, cas 7 ha demostrado ser necesario para la integración de nuevos espaciadores.

50

En experimentos adicionales, los genes cas son proporcionados en trans al hospedador. Donde el gen cas es eliminado, la inmunidad se restaura.

55

EJEMPLO 2

Integración de los Espaciadores de CRISPR

60

En estos experimentos, la integración de un espaciador de CRISPR en una ubicación CRISPR, mostró proveer resistencia contra un bacteriófago con el que el espaciador de CRISPR mostró identidad. En estos experimentos se produjo la cepa de *S. thermophilus* DGCC7710RH1.

65 La cepa de *Streptococcus thermophilus* DGCC7710 (depositada en la "Colección Nacional de los Cultivos de Microorganismos" francesa con el número CNCM 1-2423) posee por lo menos 3 ubicaciones CRISPR: CRISPR1,

CRISPR2, y CRISPR3. En las cepas de *S. thermophilus* CNRZI066 y LMGI8311, para las cuales la secuencia completa del genoma es conocida (Bolotin et al. Microbiol., 151 :2551-1561 [2005]). CRISPR1 está situado en la misma ubicación cromosómica: entre str0660 (o stu0660) y str0661 (o stu0661).

5 En la cepa DGCC7710, CRISPR1 también se encuentra en la misma ubicación cromosómica, entre genes muy similares. CRISPR1 de la cepa DGCC7710, contiene 33 repeticiones (incluyendo la repetición terminal), y, por tanto, 32 espaciadores.

10 Todos estos espaciadores son diferentes unos de otros. La mayoría de estos espaciadores son nuevos (es decir, no se han descrito previamente dentro de las ubicaciones CRISPR), pero cuatro espaciadores cerca del remolque CRISPR1, son idénticos a espaciadores de CRISPR1 ya conocidos:

- El espaciador 28^o de DGCC7710 es 100% idéntico al espaciador 31^{ro} de CRISPR1 de la cepa CNRZ1575 (número de acceso al GenBank DQ072991);

15 - El espaciador 30^o de DGCC7710 es 100% idéntico al espaciador 27^o de CRISPR1 de la cepa CNRZ703 (número de acceso al GenBank DQ072990);

20 - El espaciador 31^{ro} de DGCC7710 es 100% idéntico al espaciador 28^o de CRISPR1 de la cepa CNRZ703 (número de acceso al GenBank DQ072990);

- El espaciador 32^{ndo} de DGCC7710 es 100% idéntico al espaciador 30^o de CRISPR1 de la cepa CNRZ703 (número de acceso al GenBank DQ072990).

25 Durante el desarrollo de la presente invención, la cepa *S. thermophilus* DGCC7710RH1 fue aislada como un fago natural resistente a mutantes, utilizando a DGCC7710 como la cepa precursora, y al fago D858 como el fago virulento. Se utilizó a D858, un bacteriófago perteneciente a la familia de los virus *Siphoviridae*.

30 CRISPR1 de la cepa DGCC7710-RH1, contiene 34 repeticiones (incluyendo la repetición terminal), y, por tanto, 33 espaciadores. En comparación con la secuencia CRISPR1 de la cepa de *S. thermophilus* DGCC7710, la secuencia CRISPR1 de la cepa de *S. thermophilus* DGCC7710-RH1 posee un espaciador adicional nuevo (y una repetición adicional nueva, la cual flanquea al nuevo espaciador) en un extremo terminal de la ubicación CRISPR (es decir, cerca del líder, en el extremo 5' de la ubicación CRISPR). Todos los demás espaciadores de la ubicación de CRISPR1 siguen siendo los mismos. La secuencia de CRISPR1 (5'-3') de la cepa DGCC7710-RH1 se presenta a continuación:

35

caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataacccactt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgttgcaatcct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaacagattcaa
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatcttaattacaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggtaacgtattta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatcctgattaatacttactcg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggtaactgctcaagtg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaaggagacaacatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattttaaattcattttcataacttcat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggtatagcttaatatcattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcaggttcgagtttaccgtttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtcaactaacaattatgct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaattctgttttaggtacattt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatggctct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgctacatctcaaaggatgatcccaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
 Ttgattcaacataaaaaagccagttcaattgaacttggcttt (SEQ ID NO: 682)

En la secuencia anterior, el líder tiene la secuencia:

5 5'caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag 3' (SEQ ID NO: 688)

La secuencia integrada (GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataacccactt; SEQ ID NO: 689) que comprende una repetición CRISPR se muestra en mayúsculas y un espaciador de CRISPR (es decir, la secuencia marcada) en minúsculas, ambas se muestran en gris arriba. La repetición terminal y las secuencias de remolque de la repetición CRISPR se muestran a continuación:

Repetidor Terminal: 5 'gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO: 3)

Secuencia de remolque: 5'ttgattcaacataaaaaagccagttcaattgaacttggcttt 3' (SEQ ID NO: 691)

La secuencia del nuevo espaciador 5-TCAACAATTGCAACATCTTATAACCCACTT (SEQ ID NO: 534) existe en el genoma de los fagos D858.

La secuencia del espaciador se encuentra entre las posiciones 31921 y 31950 pb (es decir, en la hebra de arriba) del genoma de D858 (y tiene una identidad del 100% con la secuencia genómica de D858, de más de 30 nucleótidos):

espaciador 1 tcaacaattgcaacatcttataaaccactt 30 (SEQ ID NO:534)
 |||
 D858 31921 tcaacaattgcaacatcttataaaccactt 31950 (SEQ ID NO:534)

El nuevo espaciador que está integrado en la ubicación CRISPR1 de la cepa de *S. thermophilus* DGCC7710-RH1, confiere resistencia a los fagos D858 para esta cepa, como se muestra en la Figura 1 y Tabla 2-1.

5 Tabla 2-1.

Fago 2972			Fago 858		
Cepas	BIM en ¹	Sensibilidad ² al Fago	Homología ³ del espaciador-fago	Sensibilidad ² al fago	Homología ³ del espaciador-fago
DGCC7710	-	S	Ctrl	S	Ctrl
DGCC7778	858	S	> 10 SNPs	R	100% (2 espaciadores)
DGCC7710-RH1	858	R	100%	R	100%
DGCC7710-RH2	858	R	100%	R	100%
DGCC7778RT	858	S	> 10 SNPs	S	100% pero no próximo a cas
DGCC7778RT'	858	S	> 10 SNPs	S	Sin espaciadores izquierdos
DGCC7778cas1	858	S	> 10 SNPs	S	100% (2 espaciadores) pero cas1 Ko
DGCC7778cas4	858	S	> 10 SNPs	R	100% (2 espaciadores) pero cas4 Ko
DGCC7710-R2	2972	R	100% (1 espaciador)	S	5 SNPs
DGCC7710-R2S1S2	2972	S	100% pero no próximo a cas	R	S1S2 son 100% idénticos al fago 858

1. Fago usado para generar mutantes insensibles al Bacteriófago (BIMs)

10 2 Sensibilidad del Fago de la cepa, S = sensible R= resistente como se determinó por los ensayos de placa y de mancha.

3. Homología entre el nuevo espaciador del mutante y la secuencia de ADN del fago usado para generar el mutante.

15 Fagos retenidos por su capacidad para adsorber a los mutantes.

Además, durante el desarrollo de la presente invención, la cepa *S. thermophilus* DGCC7710-RH2 se aisló como un fago natural resistente a mutantes usando la cepa DGCC7710 *S. thermophilus* como cepa precursora, y el fago D858 como fago virulento.

20 CRISPR1 de la cepa DGCC7710-RH2 *S. thermophilus* contiene 34 repeticiones (que incluyen la repetición terminal), y por lo tanto 33 espaciadores. Cuando se comparó con la secuencia CRISPR1 de la cepa DGCC7710 *S. thermophilus*, la secuencia CRISPR1 de la cepa DGCC7710-RH2 *S. thermophilus* posee un nuevo espaciador adicional (y una repetición adicional que se flanquea con el nuevo espaciador) en un extremo de la ubicación CRISPR (es decir, cerca del líder, en el extremo 5' de la ubicación CRISPR). Todos los otros espaciadores de la
 25 ubicación CRISPR1 permanecieron inalterados.

La secuencia CRISPR1 (5'-3') de la cepa DGCC7710-RH2 se muestra debajo:

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaataaatga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtcattcttccgtttttgtttgcaaatcct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggtaacgtattta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcggtcccaatcctgattaatacttactcg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggttaactgctcaagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaaggagacaacatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaaatcattttcataacttcat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaaactatgaaattttataatttttaaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataattttatggtatagcttaatatcattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcaggttcgagtttaccggttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtcaactaacaattatgct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacagacaagagttaaaatggctctt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatcccaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
 ttgattcaacataaaaaagccagttcaattgaacttggcttt

En la secuencia de arriba, la secuencia líder es:

5

5'-caaggacagttattgattttataatcactagtggtataaaaacgtcaaaatttcatttgag-3' (SEQ ID NO:688)

La secuencia integrada que comprende una repetición CRISPR se muestra en mayúsculas y un espaciador CRISPR (es decir, las secuencias de marcado) en minúsculas; ambas se muestran arriba en gris (GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaataaatga; SEQ ID NO:694). Las secuencias de repetición terminal y de remolque de la CRISPR se muestran debajo:

10

Repetición terminal: 5'-gttttgtactctcaagatttaagtaactgtacagt-3' (SEQ ID NO:3)

15

Secuencia de remolque: 5'-ttgattcaacataaaaaagccagttcaattgaacttggcttt-3' (SEQ ID NO:691)

Se ha mostrado que la secuencia del nuevo espaciador existe dentro del genoma de fago D858.

20

La secuencia del espaciador (SEQ ID NO: 535) se encuentra entre las posiciones 17215 y 17244 pb (es decir, en la hebra adicional) del genoma de D858 (y tiene 100% de identidad a la secuencia genómica D858 sobre 30 nucleótidos):

Espaciador	1	ttacgtttgaaaagaatatcaaataaatga	30	(SEQ ID NO:535)
D858	17215	ttacgtttgaaaagaatatcaaataaatga	17244	(SEQ ID NO:690)

El nuevo espaciador se integra en la ubicación CRISPR1 de la cepa DGCC7710- RH2 *S. thermophilus* confiere resistencia contra el fago D858 para la cepa DGCC7710-RH2 *S. thermophilus*, como se muestra en la Figura 2 y la Tabla 2-1 (Ver también, la Figura 10).

5 EJEMPLO 3

Integración y Extracción del Constructo

En este Ejemplo, se describen métodos para la integración y extracción del constructo.

- 10 Las cepas usadas en estos experimentos fueron:
- Cepa precursora *S.thermophilus* DGCC7710, sensible a los fagos 858 y 2972
- 15 Mutante CRISPR *S.thermophilus* DGCC7778 resistente a 858
- DGCC7778cas 1 KO *S.thermophilus*
- DGCC7778cas4KO *S.thermophilus*
- 20 DGCC7778RT *S.thermophilus*
- DGCC7778RT *S.thermophilus*
- 25 Mutante CRISPR *S.thermophilus* DGCC7710R2 resistente a 2972
- DGCC7710R2S1S2 *S.thermophilus*
- 30 La *E. coli* EC1,000 proporciona pORI28 (Ver, Russell y Klaenhammer, Appl. Environ. Microbiol., 67:43691 - 4364 [2001])
- La *Escherichia coli* pCR2. 1TOPO proporciona pTOPO (Ver, Invitrogen catálogo #K4500-01)
- Los siguientes plásmidos se usaron en estos experimentos:
- 35 pTOPO, un plásmido usado para sub-clonar varios constructos
- pTOPOcas/ko contiene un fragmento íntegro de *cas1*
- 40 pTOPOcas4ko contiene un fragmento íntegro de *cas4*
- pTOPOS1S2 contiene el constructo del espaciador S1S2
- pTOPO RT contiene el constructo de repetición RT terminal
- 45 pORI28 es un plásmido usado para la integración de varios constructos en el cromosoma de cepas de *S.thermophilus*.
- pORlcas1ko contiene un fragmento íntegro de *cas1*
- 50 pORIScas4ko contiene un fragmento íntegro de *cas4*
- pORIS1S2 contiene el constructo espaciador S1S2 purista contiene el constructo de repeticiones terminales RT
- 55 Los siguientes cebadores se usaron en estos experimentos:
- Cas1
- 5'-caaatggatagagaacgc-3' (SEQ ID NO:670) y 5'-ctgataaggtgttcgtgtcc-3' (SEQ ID NO:671)
- 60 Cas4
- 5'-ggagcagatggaatacaagaaagg-3' (SEQ ID NO:672) y 5'-gagagactaggtgtctcagca-3' (SEQ ID NO:673)
- 65 S1S2 y RT

P1 5'-acaacaacagagaagtatctcattg-3' (SEQ ID NO:666)

P2 5'-aacgagtacactcactattgtacg-3' (SEQ ID NO:667)

5 P3 5'-tccactcacgtacaaatagtgagtgtactcgttttgtattctcaagatttaagtaactgtacagtttgattcaacataaaaag-3 (SEQ ID NO:668)

P4 5'-ctttcctcatcctcgctttgggt-3' (SEQ ID NO:669)

10 Se obtuvieron cepas y fagos de la Colección del Cultivo Danisco, o del material referenciado (Russell and Klaenhammer, Appl. Environ. Microbiol., 67:43691-4364 [2001]; y Levesque et al., Appl. Environ. Microbiol., 71:4057-4068 [2005]).

15 La preparación de Fago, purificación y pruebas se llevaron a cabo usando métodos conocidos en el arte (Ver por ejemplo, Duplessis et al., Virol., 340: 192-208 [2005]; y Levesque et al., Appl. Environ. Microbiol., 71 :4057-4068 [2005]).

20 Las cepas de *S. thermophilus* se desarrollaron a 37°C o 42°C en M17 (Difco) suplementado con 0.5% de lactosa o sucrosa. Para la infección del fago, se agregaron 10mM de CaCl₂ al medio previo a la infección del fago, como se conoce en el arte (Ver por ejemplo, Duplessis et al., supra; y Levesque et al., supra).

25 Las enzimas usadas para llevar a cabo la digestión por restricción y PCR se adquirieron de Invitrogen y se usaron de acuerdo con las instrucciones del fabricante. Las PCR se llevaron a cabo en un termociclador de gradiente Mastercycler Eppendorf como se conoce en el arte (Ver por ejemplo, Barrangov et al., Appl. Environ. Microbiol., 68:2877-2884 [2002]).

30 La inactivación del gen y la inserción del plásmido en el sitio específico por medio de la recombinación homóloga en el cromosoma *S. thermophilus* se llevó a cabo por la subclonación en el sistema pCR2.1TOPO de Invitrogen, subsecuente a la clonación en el sistema pORI usando la *E. coli* como un hospedador y los constructos se purificaron y transformaron finalmente en *S.thermophilus* como se describió previamente (Russell y Klaenhammer, supra).

Integración del Constructo RT

35 Usando el constructo RT diseñado como se muestra en la Figura 4, el constructo se insertó justo después del *cas4*, como se muestra en la Figura 5. El DGCC7778 precursor es resistente al fago 858. El precursor tiene dos espaciadores (S1 y D2) que son idénticos al ADN del fago 858. La cepa resultante (RT) perdió resistencia al fago 858. Este resultado indica que los genes *cas* necesitan estar en vecindad inmediata de los espaciadores para conferir resistencia. Como se muestra en la Figura 3, el DGCC7778 precursor se diseñó de tal manera que el gen *cas1* se interrumpió, produciendo una pérdida de resistencia, lo que denota que *cas1* se necesita para conferir la

40 resistencia. Como se muestra en la Figura 3, el DGCC7778 precursor se diseñó de tal manera que el gen *cas4* se interrumpió. Además, el constructo S1S2 se integró en el DGCC7710 precursor, como se muestra en las Figuras 6-8.

EJEMPLO 4

45 Aislamiento de Mutantes Resistentes al fago y Confirmación de Secuencias CRISPR

En este Ejemplo, se describen los métodos usados en el aislamiento de mutantes resistentes al fago y confirmación de secuencias CRISPR. Los mutantes resistentes al fago *S. thermophilus* fueron obtenidos mediante la prueba inmógena de la cepa DGCC7710 hospedadora del tipo natural (también llamada "RD534") con el fago 2972 y/o el fago 858 (Levesque et al., Appl. Environ. Microbiol., 71:4057 [2005]). La cepa hospedadora se desarrolló a 42°C en 10 ml de un caldo M17 suplementado con 0.5% de lactosa (LM 17). Cuando la densidad óptica (600 nm) alcanzó 0.3, fagos y se agregó el cloruro de calcio 10 mM a una concentración final de 10⁷ pfu/ml y 50 mM, respectivamente. El cultivo que contiene el fago se incubó a 42°C durante 24 horas y se monitoreó para la lisis. Después, 100 µl del lisado se inoculaban en 10 ml del LM17 fresco. El lisado restante se centrifugó y los pelletizados se inoculaban en otro tubo que contenía 10 ml del medio LM17 fresco. Estos dos cultivos se incubaron a 42°C durante 16 horas. Finalmente, estos cultivos se diluyeron y colocaron en LM17. Las colonias aisladas se probaron para la sensibilidad del fago como se conoce en el arte (Ver, Moineau et al., Can J. Microbiol., 38:875 [1992]). La ubicación CRISPR de los aislados resistentes se verificaron mediante la secuenciación de productos PCR, usando la información del

60 genoma del fago pertinente conocida en el arte (Ver, Levesque et al., Appl. Environ. Microbiol., 71:4057 [2005]).

EJEMPLO 5

Diseño del espaciador CRISPR

65

En este Ejemplo, se describen métodos usados en algunas modalidades para diseñar el espaciador CRISPR. Las enzimas usadas para llevar a cabo la digestión por restricción y PCR se adquirieron de Invitrogen y se usaron de acuerdo a las instrucciones del fabricante. Las PCR se llevaron a cabo en un termociclador de Gradiente Mastercycler Eppendorf, usando métodos conocidos en el arte.

La inactivación del gen y la inserción del plásmido en el sitio específico por medio de la recombinación homóloga en el cromosoma *S. thermophilus* se llevó a cabo mediante la subclonación en el sistema pCR2.1 -TOPO (Invitrogen), mediante la clonación subsecuente en el sistema pORI usando *E. coli* como un hospedador, y los constructos se purificaron y transformaron finalmente en *S. thermophilus* como se conoce en el arte (Ver, Russell y Klaenhammer, Appl. Environ. Microbiol., 67:4361 [2001]).

El ADN del mutante WT_{Φ858}^{+S1S2} se usó como una plantilla para amplificar dos fragmentos PCR distintos usando P1 (5'-acaacaacagagaagtgatctcattg-3'; SEQ ID NO:666) y P2 (5'-aacgagtacactcactattgtacg-3'; SEQ ID NO:667) en una reacción, y P3

(5'-tccactcacgtacaatagtgagtgactcgtttgtattctcaagatttaagtaactgtacagtttgattcaacataaaaag-3'; SEQ ID NO:668) y P4 (5' - crtccctcatcctcgcttggtt-3'; SEQ ID NO:669) en otra reacción. Se usaron ambos productos del PCR subsecuentemente como plantillas en otra reacción de PCR usando los cebadores P1 y P4 para generar el constructo SIS2 de acuerdo a la Figura 11.

El constructo S1S2 se sub-clonó en el sistema Invitrogen pCR2.1-TOPO. Este constructo se digirió con *NotI* y *HindIII* y se clonó subsecuentemente en el pORI en los sitios *NotI* y *HindIII*, que proporcionan el constructo pS1S2. La integración de pS1S2 en la ubicación CRISPR1 de WT_{Φ2972}^{+S4} ocurrió por medio de la recombinación homóloga en el extremo 3' del cas7, para generar WT_{Φ2972}^{+S4}::pS1S2.

El constructo pR se generó usando el constructo pS1S2 como una plantilla. Específicamente, el constructo S1S2 subclonado en el pCR2.1-TOPO se digirió usando BsrGI, que se corta dentro de la repetición CRISPR. Después, la digestión se ligó nuevamente y se usó un plásmido que contenía una sola repetición y ningún espaciador subsecuentemente para clonarse en pORI usando *NotI* y *HindIII*, generando pR. La integración de pR en el cromosoma de WT_{Φ858}^{+S1S2} en el extremo 3' de cas7 vía la recombinación homóloga generó WT_{Φ858}^{+S1S2}::pR, un mutante donde la ubicación CRISPR1 se desplaza y una única repetición se inserta en su lugar.

El mutante WT_{Φ858}^{+S1S2}::pR se desarrolló subsecuentemente en ausencia de eritromicina, y se analizaron variantes sensibles a los antibióticos para encontrar un mutante que tenía una eliminación completa de la ubicación CRISPR1. La eliminación se derivó de una recombinación homóloga que se presenta en el extremo 3' de ORF (opuesto a un evento de recombinación que ocurre en el extremo 3' de cas7, que habría resultado en una restauración de la cepa WT_{Φ858}^{+S1S2}), generando WT_{Φ858}^{+S1S2} ΔCRISPR1 (Ver, Figura 12), un mutante donde la ubicación CRISPR1 se elimina (Ver también, Figura 10).

EJEMPLO 6

Inactivation de Genes cas

Para la inactivación cas5, una pieza interna de 801-pb cas5 se amplificó por PCR usando los cebadores 5'-caaatggatagagaaacgc-3' (SEQ ID NO:670) y 5'-ctgataaggtgttcgtgtcc-3' (SEQ ID NO:671) y se subclonó en la *E. coli* pCR2.1-TOPO (Invitrogen). Este constructo se digirió con EcoRV y HindIII y se clonó subsecuentemente en pORI en los sitios EcoRV y HindIII. La integración de este constructo en el gen cas5 de WT_{Φ858}^{+S1S2} ocurrió por medio de la recombinación homóloga de la pieza interna del gen, resultando en WT_{Φ858}^{+S1S2}::pcas5 -.

De manera semejante, una pieza interna 672-pb de cas7 se amplificó por PCR usando los cebadores 5'-ggagcagatggaatacaagaagg-3' (SEQ ID NO:672) y 5'-gagagactaggtgtctcagca-3' (SEQ ID NO:673) y se subclonó en la *E. coli* pCR2.1-TOPO (Invitrogen). Este constructo se digirió con EcoRV y HindIII y se clonó subsecuentemente en pORI en los sitios EcoRV y HindIII. La integración de este constructo en el gen cas7 de WT_{Φ858}^{+S1S2} ocurrió por medio de la recombinación homóloga de la pieza interna del gen, resultando en WT_{Φ858}^{+S1S2}::pcas7 - (Ver, Figuras 10-12).

EJEMPLO 7

Métodos Naturales para Insertar una Secuencia Adicional en una Ubicación CRISPR

En este Ejemplo, se describen los métodos usados para provocar naturalmente la inserción de una secuencia adicional dentro de una ubicación CRISPR de una cepa bacteriana. La "secuencia adicional" como se usa en la presente se define como una secuencia espaciadora asociada con la secuencia de repetición CRISPR. Más particularmente, la "secuencia adicional" se origina en parte de un fago donador capaz de infectar la bacteria dirigida y en parte de la duplicación de la secuencia de repetición CRISPR. La introducción del ADN del fago donador en la célula bacteriana resulta de la infección de la célula por el fago donador. La selección de células que contienen la

secuencia adicional se hace a través de la presión de selección con el fago donador de tal modo que las células modificadas seleccionadas sean resistentes al fago.

En estos experimentos, una cepa precursora se expuso a un fago donador y una variante resistente al fago de la cepa precursora (es decir, una cepa variante) seleccionada. La cepa variante se analizó (por ejemplo, por PCR y/o formación de secuencias de ADN) para confirmar la presencia de una secuencia adicional dentro de una ubicación CRISPR. La secuencia del nucleótido de la secuencia adicional se determinó. Típicamente, la secuencia adicional es un fragmento de aproximadamente 30 nucleótidos en tamaño, del fago donador asociado (fusionado) a una secuencia de repetición CRISPR, y confiere resistencia al fago donador.

En algunos experimentos, la cepa precursora se pre-cultivó toda la noche en un medio basado en leche a 42°C. Un medio basado en leche se inoculó entonces a 0.1% (v/v) con el pre-cultivo de la cepa precursora y con una suspensión del fago donador a un MOI de 10. Después de 6 horas de incubación a 42°C, se colocaron las diluciones del cultivo en un medio nutritivo para obtener las colonias aisladas. Los aislados se probaron entonces para su resistencia al fago donador (cualquier método adecuado en el arte encuentra uso en estos experimentos). Se analizaron después cepas variantes para la presencia de una secuencia adicional dentro de una de sus ubicaciones CRISPR.

Las ubicaciones CRISPR se amplificaron por PCR y las secuencias del nucleótido de los productos de PCR resultantes se determinaron mediante una formación de secuencias de ADN usando PCR estándar y métodos de secuenciación conocidos en el arte. Estas secuencias se compararon después con las de la cepa precursora usando métodos estándares conocidos en el arte.

En algunos experimentos, DGCC7710 se usó como cepa precursora y D2972 se usó como un fago donador. La cepa DGCC7710 *S. thermophilus* se expuso al fago donador D2972 como se describió anteriormente. Se obtuvo una cepa variante nombrada WT_{phi2972}^{+S6} (Ver, tabla 7-1). La Tabla 7-1 también incluye los resultados para cepas variantes descritas en otros Ejemplos. En la Tabla 7-1, el EOP se expresa relativamente como el fago D2972. El posicionamiento de la secuencia adicional en el genoma del fago se da relativamente al fago D2972, a menos que se especifique lo contrario.

Tabla 7.1 Descripción de cepas variantes modificadas con CRISPR de DGCC7710

Cepa variante	Cepa precursora	Fago donador	Secuencia adicional	Secuencia	EOP
—	DGCC7710	—	—		1
WT _{phi2972} ^{+S6}	DGCC7710	D2972	34521-34492	GCCCTTCTAATTGGATTACCTTCC GAGGTG (SEQ ID NO:524)	10 ⁻⁴
WT _{phi858} ^{+S1S2} ; pcas5 ^{+S19} phi858	WT _{phi858} ^{+S1S2} 2 ::pcas5	D858	33824-33853	TGTAGATAGCGTGGTGCAGAGA TGCACGG (SEQ ID NO:702)	10 ⁻⁵
WT _{phi2972} ^{+S4}	DGCC7710	D2972	31582-31611	CTCAGTCGTTACTGGTGAACCAG TTTCAAT (SEQ ID NO:525)	10 ⁻⁵
WT _{phi2972} ^{+S20}	DGCC7710	D2972	25693-25722	TCTAATCCCACTAGGAATAGTGG GTAGTAA (SEQ ID NO:703)	10 ⁻⁵

Tabla 7.1 Descripción de cepas variantes modificadas con CRISPR de DGCC7710					
Cepa variante	Cepa precursora	Fago donador	Secuencia adicional	Secuencia	EOP
WT _{phi2972} ^{+S21}	DGCC7710	D2972	27560-27589	TCGATAAATCAGCCAAAGTATTA AGTGGTT (SEQ ID NO:704)	10 ⁻⁵
WT _{phi2972} ^{+S22}	DGCC7710	D2972	24624-24653	CAGCTTGAAATGTTTATTGAAGC AGCAGTG (SEQ ID NO:705)	10 ⁻⁴
WT _{phi2972} ^{+S6} 724 _{+S15}	WT _{phi2972} ^{+S6}	D4724	1113-1142	TTATATCGAAGAACGACTGAAAG AGCTTGA (SEQ ID NO:706)	<10 ⁻⁸
WT _{phi2972} ^{+S6} 724 _{+S17}	WT _{phi2972} ^{+S6}	D4724	33968-33997	TCCAAGTTATTTGAGGAGTTATTA AGACAT (SEQ ID NO:707)	<10 ⁻⁸
WT _{phi2972} ^{+S6} 724 _{+S24}	WT _{phi2972} ^{+S6}	D4724	30803-30832	TACCGAAACGACTGGTTTGAAAA ATTCAAG (SEQ ID NO:708)	<10 ⁻⁸
WT _{phi2972} ^{+S6} 724 _{+S15} phi4733 _{+S16}	WT _{phi2972} ^{+S6} phi4724 _{+S15}	D4733	29923-29894	AGTTGATTGCGTAATCAACCATC TCCATAA (SEQ ID NO:709)	<10 ⁻⁸
WT _{phi2972} ^{+S4} 720 _{+S17}	WT _{phi2972} ^{+S4}	D4720	33968-33997	TCCAAGTTATTTGAGGAGTTATTA AGACAT (SEQ ID NO:707)	10 ⁻⁶
WT _{phi2972} ^{+S4} 58 _{+S18}	WT _{phi2972} ^{+S4}	D858	30338-30367 (relativamente para D858)	CTTCAAATGTACTGCAAGGCTGC AAAAGTA (SEQ ID NO:710)	<10 ⁻⁸
WT _{phi2972} ^{+S4} 58 _{+S25}	WT _{phi2972} ^{+S4}	D858	33886-33915 (relativamente para D858)	ATTGTCTATTACGACAACATGGA AGATGAT (SEQ ID NO:526)	<10 ⁻⁷
WT _{phi858phi2972} ⁺ S9S10S11S12	DGCC7710	D2972 + D858	7874-7903 20650-20621 8360-8389 18998-19027	TTACTAGAGCGTGTCTGTTAACCA CTTTAAA (SEQ ID NO:528) TTCGTTAAAGTCACCTCGTGCTAG CGTTGC (SEQ ID NO:529) ATAACGGTAGCAAATATAAACCT GTTACTG (SEQ ID NO:530) GAAGTAGCCATACAAGAAGATGG ATCAGCA (SEQ ID NO:531)	<10 ⁻⁷
WT _{phi858phi2972} ⁺ S13S14	DGCC7710	D2972 + D858	33602-33631 4830-4801	GATGTCACTGAGTGTCTAAGCAT TGC GTAC (SEQ ID NO:776) TGAATAAGCAGTTCTTGACGACC AACCGAC (SEQ ID NO:533)	10 ⁻⁸
WT _{phi2972} ^{+S26S27}	DGCC7710	D2972	29647-29-618 16681-16652	TTTCCGTCTTCTTTT TAGCAA GATACG (SEQ ID NO:711) TGTTTCAAGGTTTCGGGTCCAAGT ATCATT (SEQ ID NO:712)	<10 ⁻⁸

Tabla 7.1 Descripción de cepas variantes modificadas con CRISPR de DGCC7710					
Cepa variante	Cepa precursora	Fago donador	Secuencia adicional	Secuencia	EOP
DGCC9705	DGCC7710	D2972	25582-25611 18079-18108	AAATCAGTTTTTTTGTTCAGAAACT TGTTCT (SEQ ID NO:713) TAGAGGTAATGACGGCTTACCGG GTAAAGA (SEQ ID NO:714)	<10 ⁻⁸
DGCC9726	DGCC9705	D3821	26530-26559	GAAGTATTAGGTCTCTCAAAAGA TGATATT (SEQ ID NO:715)	<10 ⁻⁸
DGCC9733	DGCC9726	D3288	5339-5363 (relativamente para 7201)	TAAACTATGAAATTTTATAATTTT TGAACA (SEQ ID NO:716)	<10 ⁻⁸
DGCC9836	DGCC9733	D4753	15562-15590 (relativamente para D858) 362-333 24083-24112 (relativamente para D858) 30059-30088 (relativamente para D858)	AACCCAATAATTACAGTGAAGCA CAATAG (SEQ ID NO:717) AATTGTGTCGGTCTTTTTTATTGT TTTTACC (SEQ ID NO:718) CTCGTTAATTGCAAGTTTGGTCGG CACGTT (SEQ ID NO:719) TTGTGGCACAACAAAATGAATT AAAGATT (SEQ ID NO:720)	<10 ⁻⁸

Esta resistencia mostrada de la variante para D2972, como la eficiencia de la formación de placa (EOP) de D2972 en WT_{phi2972}^{+S6} se redujo por 4 logs. El ADN se extrajo de WT_{phi2972}^{+S6} y la ubicación CRISPR1 se analizó por PCR como se conoce en el arte (Ver por ejemplo, Bolotin et al. [2005], supra) usando la combinación de un cebador delantero (ya sea YC70 y/o SPIDR-arriba (5'-gTCTTTAgAACTgTgACACC-3'; SEQ ID NO:674) y un cebador inverso (ya sea YC31 y/o SPIDR-abajo (5'-TAAACAgAgCCTCCCTATCC; SEQ ID NO:675). La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR1 de DGCC7710. Comparado con DGCC7710, WT_{phi2972}^{+S6} se encontró que difiere por la adición de una secuencia espaciadora sencilla de 30 pb en el extremo 5' de su región CRISPR1 y por la duplicación de la secuencia repetida, como se muestra en la Figura 14. La comparación de la secuencia adicional con la secuencia del genoma D2972 muestra que la nueva secuencia espaciadora es 100% idéntica a la del genoma D2972 del nucleótido 34521 hasta el nucleótido 34492.

En algunos experimentos adicionales, WT_{phi858}^{+S1S2}::pcas5 se usó como la cepa precursora y D858 se usó como un fago donador. La cepa variante resultante nombrada WT_{phi858}^{+S1S2}::pcas5_{phi858}^{+S19} (Ver, Tabla 7-1) fue resistente a D858, con un EOP reducido por 5 logs. El ADN se extrajo de WT_{phi858}^{+S1S2}::pcas5_{phi858}^{+S19} y su ubicación CRISPR3 se analizó por PCR usando un cebador delantero (CR3_líderF1, 5'-CTGAGATTAATAGTGCGATTACG; SEQ ID NO:676) y un cebador inverso (CR3_pistaR2, 5'-GCTGGATATTCGTATAACATGTC; SEQ ID NO:677). La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR3 de WT_{phi858}^{+S1S2}::pcas5. Comparado con WT_{phi858}^{+S1S2}::pcas5, WT_{phi858}^{+S1S2}::pcas5_{phi858}^{+S19} difiere por la adición de una secuencia espaciadora sencilla de 30 pb en el extremo 5' de su región CRISPR3 y por la duplicación de la secuencia repetida. La comparación de la secuencia adicional con la secuencia del genoma D858 muestra que la nueva secuencia espaciadora es 100% idéntica a la del genoma D858 del nucleótido 33824 hasta el nucleótido 33853.

En otros experimentos adicionales, DGCC7809 se usó como la cepa precursora y D3743 se usó como el fago donador. La cepa variante resultante nombrada DGCC7809_{phiD3743}^{+S28} (Ver, Tabla 7-2) fue resistente a D3743 con un EOP reducido por 8 logs. El ADN se extrajo de DGCC7809_{phiD3743}^{+S28} y su ubicación CRISPR3 se analizó por PCR usando un cebador delantero (CR3_líderF1, 5'-CTGAGATTAATAGTGCGATTACG; SEQ ID NO:676) y un cebador inverso (CR3_pistaR2, 5'-GCTGGATATTCGTATAACATGTC; SEQ ID NO:677). La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR3 de ST0189. Comparado con DGCC7809, DGCC7809_{phiD3743}^{+S28} difiere por la adición de una secuencia espaciadora sencilla de 29 pb en el extremo 5' de su región CRISPR3 y por la duplicación de la secuencia repetida. La secuencia del fago D3743 se desconoce; sin embargo la comparación de la secuencia adicional con la secuencia de otro genoma de fago de estreptococos muestra que la nueva secuencia espaciadora es 100% idéntica a la del genoma DT1 de fago del nucleótido 6967 hasta el nucleótido 6996.

En otros experimentos adicionales, DGCC3198 se usó como la cepa precursora y D4241 se usó como el fago donador. La cepa variante resultante nombrada DGCC3198^{phi4241}^{+S29} (Ver, Tabla 7-2) fue resistente a D4241 con un EOP reducido por 8 logs. El ADN se extrajo de DGCC3198^{phi4241}^{+S1} y su ubicación CRISPR1 se analizó por PCR usando un cebador delantero (ya sea YC70 y/o SPIDR-arriba (5'-gTCTTTAgAAACTgTgACACC-3'; SEQ ID NO:674) y de un cebador inverso (ya sea YC31 y/o SPIDR-abajo (5'-TAAACAgAgCCTCCCTATCC; SEQ ID NO:675). La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR1 de DGCC3198. Comparado con DGCC3198, DGCC3198^{phi4241}^{+S29} difiere por la adición de una secuencia espaciadora sencilla de 30 pb en el extremo 5' de su región CRISPR1 y por la duplicación de la secuencia repetida. La secuencia del fago D4241 se desconoce; sin embargo la comparación de la secuencia adicional con la secuencia de otro genoma de fago de estreptococos muestra que la nueva secuencia espaciadora es 100% idéntica a la del genoma de fago DT1 del nucleótido 3484 hasta el nucleótido 3455.

La Tabla 7-2 a continuación proporciona una descripción de cepas variantes modificadas con CRISPR de DGCC7809 y de DGCC3198. En esta Tabla, EOP se expresa relativamente a los fagos donadores. La posición de la secuencia adicional en el genoma de fago se da relativamente al fago DT1.

Tabla 7-2. Descripción de cepas variantes modificadas con CRISPR de DGCC7809 y DGCC3198					
Cepa variante	Cepa precursora	Fago donador	Secuencia adicional	Secuencia	EOP
-	DGCC7809	-	-	-	1
DGCC7809 ^{phi3743} ^{S28}	DGCC7809	D3743	6967-6996 (relativamente para DT1)	ACAAGCAAAGATTACAACCG CTGGTGCTA (SEQ ID NO:721)	<10 ⁻⁸
-	DGCC3198	-	-	-	1
DGCC3198 ^{phi4241} ^{S29}	DGCC3198	D4241	3484-3455 (relativamente para DT1)	ACCAAGTAGCATTTGAGCAAA GATAGATTG (SEQ ID NO:722)	<10 ⁻⁸

EJEMPLO 8

Selección de un Conjunto de Cepas Variantes Modificadas con CRISPR de la Misma Cepa Precursora

En este Ejemplo, se describen métodos usados para la selección de un conjunto de cepas variantes de la misma cepa precursora que difiere por su secuencia adicional originada del mismo fago. Como varias porciones de un fago donador encuentran uso como fuentes para secuencias adicionales, múltiples cepas variantes diferentes se pueden generar del fago donador dado. Además, cada cepa variante tiene una secuencia adicional diferente. En consecuencia, las cepas múltiples se desarrollaron de la misma cepa precursora además de la cepa variante descrita en el Ejemplo 7. En algunos experimentos, estas cepas adicionales se generaron al exponer la cepa receptora al mismo fago donador. Las diversas cepas variantes resultantes se contemplaron para presentar diferente espectro de la sensibilidad de fago.

En cultivos independientes, la cepa precursora se sometió al mismo fago donador. Para cada cultivo, una variante resistente de fago sencillo se aisló como se describe en el Ejemplo 7 y luego se analizó. Las secuencias adicionales en cada una de las cepas variantes se comparó una con la otra. El espectro de sensibilidad de las cepas variantes para el fago donador y otros fagos se determinó usando métodos microbiológicos clásicos conocidos en el arte. Los espectros de sensibilidad de las diversas cepas luego se compararon. Las cepas variantes seleccionadas fueron aquellas que presentan secuencias adicionales diferentes y espectros diferentes de la sensibilidad de fago.

En algunos experimentos, se condujo la selección de diversas cepas variantes de DGCC7710 usando D2972 como un fago donador sencillo. La cepa precursora DGCC7710 se expuso al fago donador D2972 en cuatro cultivos independientes como se describe en el Ejemplo 7. De cada uno de los cultivos, una cepa variante se aisló y fue respectivamente nombrada WT^{phi2972}^{+S4}, WT^{phi2972}^{+S20}, WT^{phi2972}^{+S21} y WT^{phi2972}^{+S22} (Ver, Tabla 7-1).

Estas cepas variantes muestran resistencia a D2972, como la eficiencia de la formación de placa (EOP) de D2972 en las cuatro variantes resistentes al fago se redujo por 3 hasta 5 logs. El ADN se extrajo de WT^{phi2972}^{+S4}, WT^{phi2972}^{+S20}, WT^{phi2972}^{+S21} y WT^{phi2972}^{+S22} y se analizó por PCR usando métodos conocidos en el arte (Ver por ejemplo, Bolotin et al. [2005], supra), usando la combinación de un cebador delantero (ya sea YC70 y/o SPIDR-arriba (5'-gTCTTTAgAAACTgTgACACC; SEQ ID NO:674) y de un cebador inverso (ya sea YC31 y/o SPIDR-abajo (5'-TAAACAgAgCCTCCCTATCC; SEQ ID NO:675). La secuencia de los productos PCR se determinó y comparó con la de la ubicación CRISPR1 de DGCC7710. Comparado con DGCC7710, WT^{phi2972}^{+S4}, WT^{phi2972}^{+S20}, WT^{phi2972}^{+S21} y WT^{phi2972}^{+S22} difiere por la adición de una secuencia espaciadora de 30 pb en el extremo 5' de su

región CRISPR1 y por la duplicación de la secuencia repetida, como se muestra en la Figura 17). La comparación de estas nuevas secuencias espaciadoras con la secuencia del genoma D2972 muestra que las nuevas secuencias espaciadoras son 100% idénticas a la del genoma D2972 del nucleótido 31582 hasta el nucleótido 31611, del nucleótido 25693 hasta el nucleótido 25722, del nucleótido 27560 hasta el nucleótido 27589, y del nucleótido 24624 hasta el nucleótido 24653, respectivamente. Los cuatro espaciadores adicionales se encontraron para ser diferentes uno del otro y diferentes de los espaciadores descritos en el Ejemplo 7.

EJEMPLO 9

Métodos Naturales Usados para la Inserción de una Segunda Secuencia Adicional en una Ubicación CRISPR

En este Ejemplo, se describen métodos naturales usados para provocar la inserción de una segunda secuencia adicional en la ubicación CRISPR. Durante la inserción de una secuencia adicional de un fago donador dado en una ubicación CRISPR bacteriana, la cepa variante se vuelve resistente o al menos, menos sensible a este fago. Por lo tanto, el método descrito en el Ejemplo 7 no es más eficiente para la inserción de secuencias adicionales en la ubicación CRISPR de esta cepa variante. Por ejemplo, el método no puede aplicarse a la cepa variante $WT_{\text{phi2972}}^{+S6}$ (como una cepa precursora) usando D2972 como un fago donador, debido a que $WT_{\text{phi2972}}^{+S6}$ tiene sensibilidad significativamente disminuida para D2972 (Ver, Ejemplo 7).

En algunos experimentos, este problema se venció por el uso de un fago donador mutado derivado de D2972 que incluye al menos una modificación específica dentro de su genoma (esto es, un "fago mutado"). Este fago mutado se seleccionó al exponer el fago donador a la cepa variante, de manera que la modificación (esto es, mutación) del fago parenteral se vuelve virulenta para la cepa variante.

En algunos experimentos, el fago mutado tuvo una mutación en su genoma dentro de la región que contiene la secuencia del espaciador adicional que es parte de la secuencia adicional en la cepa variante. La cepa variante fue sensible a este fago mutado. La cepa variante se expuso al fago mutado y una nueva variante resistente al fago (variante de 2^a generación) de la cepa variante se seleccionó. La variante de 2^a generación se analizó usando métodos adecuados conocidos en el arte (por ejemplo, PCR y procesado por secuencia), para confirmar la presencia de una secuencia adicional dentro de una ubicación CRISPR. La secuencia de nucleótidos de la secuencia adicional se determinó. En algunos experimentos, la secuencia adicional se encontró que contiene un fragmento de aproximadamente 30 nucleótidos en tamaño del fago mutado que da resistencia al fago mutado.

En algunos experimentos, la cepa variante se pre-cultivó durante la noche en un medio con base en leche apropiado a 42°C. Un medio con base en leche adecuado luego se inoculó con el pre-cultivo de la cepa variante en una concentración de alrededor de 10^6 cfu/ml y con una suspensión del fago donador en un MOI mayor que 100. El cultivo se incubó durante la noche a 42°C, y luego se centrifugó. El sobrenadante se cosechó y filtró usando un filtro de 0.45 µm. Las diluciones del sobrenadante filtrado se usaron para inocular un medio de agar nutritivo sembrado con la cepa variante con objeto de obtener placas de fago aisladas, usando cualquier método adecuado conocido en el arte. Las placas aisladas se cultivaron en la cepa variante en medio nutritivo líquido, usando cualquier método adecuado conocido en el arte. Una suspensión del fago mutado se obtuvo al filtrar el cultivo a través de un filtro 0.45 µm. El fago mutado luego se usó como se describe arriba (Ver, Ejemplo 7) para provocar la inserción de una segunda secuencia espaciadora adicional en la ubicación CRISPR de la cepa variante.

En algunos experimentos, $WT_{\text{phi2972}}^{+S6}$ (Ver, Ejemplo 7, y Tabla 7-1) se usó como la cepa precursora y D4724 se usó como el fago donador. La cepa variante $WT_{\text{phi2972}}^{+S6}$ se cultivó en la presencia de alta concentración del fago D2972. Un fago mutado nombrado D4724 se aisló al formar en placa el sobrenadante de este cultivo en la cepa $WT_{\text{phi2972}}^{+S6}$ usando los métodos descritos arriba. La virulencia del fago mutado D4724 en $WT_{\text{phi2972}}^{+S6}$ se verificó. La cepa variante $WT_{\text{phi2972}}^{+S6}$ se expuso al fago mutado D4724 en un cultivo como se describe en el Ejemplo 7. Se obtuvo una cepa variante resistente al fago nombrada $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$ (Ver, Tabla 7-1).

Comparado con $WT_{\text{phi2972}}^{+S6}$, esta cepa variante muestra una resistencia incrementada para D2972, como la eficiencia de la formación de placa (EOP) de D2972 en $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$ se redujo por más de 8 logs (en lugar de 4 logs); además, su resistencia también se extendió comparada con $WT_{\text{phi2972}}^{+S6}$ ya que muestra alguna resistencia para D4724 (Ver, Tabla 9-1). El ADN se extrajo de $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$ y su ubicación CRISPR1 se analizó por PCR como se describe arriba usando las mismas combinaciones de cebadores como se describen arriba. La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR1 de $WT_{\text{phi2972}}^{+S6}$. Comparado con $WT_{\text{phi2972}}^{+S6}$, $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$ difiere por la adición de una secuencia espaciadora de 30 pb en el extremo 5' de su región CRISPR1 y por la duplicación de la secuencia repetida, como se muestra en la Figura 17. La comparación de esta secuencia espaciadora adicional con la secuencia del genoma D2972 muestra que la segunda secuencia espaciadora adicional es 100% idéntica a la del genoma D2972 del nucleótido 1113 hasta el nucleótido 1142.

De los cultivos independientes usando condiciones experimentales idénticas, las cepas variantes $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S17}$ y $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S24}$ se aislaron y analizaron (Ver, Tabla 7-1). Como se compara con $WT_{\text{phi2972}}^{+S6}$, estas cepas variantes muestran una resistencia incrementada para D2972, como la eficiencia de la

formación de placa (EOP) de D2972 en $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S17}$ y $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S24}$ se redujo por más de 8 logs para ambas cepas variantes; y su resistencia también se extendió, como se compara con $WT_{\text{phi2972}}^{+S6}$ ya que muestran resistencia para D4724 (Ver Tabla 9-1). Además, estas cepas variantes muestran secuencias espaciadoras adicionales en CRISPR1 que son 100% idénticas a la del genoma de D2972 del nucleótido 33968 hasta el nucleótido 33997 y el nucleótido 30803 hasta el nucleótido 30832, respectivamente.

En experimentos adicionales, $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$ se usó como la cepa precursora y D4733 se usó como el fago donador. Los métodos descritos arriba se usaron para generar el fago mutado D4733 del fago D4724. Luego, el fago D4733 se usó para obtener una cepa variante resistente al fago de $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$. La cepa variante resultante se nombró $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15} \text{ phi4733}^{+S16}$ (Ver, Tabla 7-1). Esta cepa variante contiene una secuencia adicional incluyendo una secuencia espaciadora que es 100% idéntica a una secuencia del genoma D2972, nucleótido 29923 hasta el nucleótido 29894. Mostró una resistencia incrementada para D2972, como la eficiencia de la formación de placa (EOP) de D2972 en $WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15} \text{ phi4733}^{+S16}$ se redujo por más de 8 logs y la resistencia se expandió al fago D4733 (Ver Tabla 9-1). La Tabla 9-1 proporciona una descripción de la resistencia de fago de algunas cepas variantes modificadas con CRISPR de DGCC7710. En esta Tabla, "nd" indica que los resultados no se determinaron.

Tabla 9-1: Descripción de la Resistencia de Fago de Algunas Cepas Variantes Modificadas con CRISPR de DGCC7710

Cepa variante	Cepa precursora	Fago donador	EOP D2972	EOP D4724	EOP D4733	EOP D4720	EOP D858
–	DGCC7710	–	1	1	1	1	1
$WT_{\text{phi2972}}^{+S6}$	DGCC7710	D2972	10^{-4}	1	1	nd	nd
$WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$	$WT_{\text{phi2972}}^{+S6}$	D4724	$<10^{-8}$	$<10^{-7}$	1	nd	nd
$WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S17}$	$WT_{\text{phi2972}}^{+S6}$	D4724	$<10^{-8}$	$<10^{-7}$	nd	nd	nd
$WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S24}$	$WT_{\text{phi2972}}^{+S6}$	D4724	$<10^{-8}$	$<10^{-7}$	nd	nd	nd
$WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15} \text{ phi4733}^{+S16}$	$WT_{\text{phi2972}}^{+S6} \text{ phi4724}^{+S15}$	D4733	$<10^{-8}$	$<10^{-7}$	$<10^{-6}$	nd	nd
$WT_{\text{phi2972}}^{+S4}$	DGCC7710	D2972	10^{-5}	nd	nd	1	nd
$WT_{\text{phi2972}}^{+S4} \text{ phi4720}^{+S17}$	$WT_{\text{phi2972}}^{+S4}$	D4720	10^{-6}	nd	nd	10^{-5}	nd
$WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S18}$	$WT_{\text{phi2972}}^{+S4}$	D858	$<10^{-8}$	nd	nd	nd	10^{-5}
$WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S25}$	$WT_{\text{phi2972}}^{+S4}$	D858	$<10^{-7}$	nd	nd	nd	10^{-3}

Todavía en experimentos adicionales, $WT_{\text{phi2972}}^{+S4}$ se usó como la cepa precursora y D4720 se usó como el fago donador. Usando los mismos métodos como se describen arriba, el fago mutado D4720 se generó del fago D2972. El fago D4720 se usó para obtener una variante resistente al fago de $WT_{\text{phi2972}}^{+S4}$. La cepa variante resultante se nombró $WT_{\text{phi2972}}^{+S4} \text{ phi4720}^{+S17}$ (Ver Tabla 7-1). Esta cepa variante contiene una secuencia adicional incluyendo una secuencia espaciadora que es 100% idéntica a una secuencia del genoma D2972 del nucleótido 33968 hasta 33997. Muestra una resistencia incrementada para D2972, como la eficiencia de la formación de placa (EOP) de D2972 en $WT_{\text{phi2972}}^{+S4} \text{ phi4720}^{+S17}$ se redujo por 6 logs (comparado con 5 logs); y su resistencia se extendió al fago D4720 (Ver, Tabla 9-1).

EJEMPLO 10

Métodos Naturales Alternativos para Insertar Segundas Secuencias Adicionales en la Ubicación CRISPR

En este Ejemplo, se describen métodos naturales alternativos útiles para insertar una segunda secuencia adicional en una ubicación CRISPR. Se conoce que una cepa precursora dada puede ser sensible para más de una familia de fagos. Esta diversidad de la sensibilidad se usó ventajosamente para insertar secuencias adicionales en una ubicación CRISPR de una cepa variante, como se describen en la presente. En estos experimentos, el segundo fago donador se seleccionó al probar la virulencia de una selección de fagos en la cepa precursora y en las cepas variantes. Los segundos fagos donadores de interés fueron aquellos que estuvieron virulentos en ambas cepas. Se contempló que estos fagos deben ser probables para representar una familia diferente de fagos que aquellos representados por el fago donador inicial. Después de su selección, el segundo fago donador se usó para infectar la

cepa variante. Como se describe en los métodos de arriba, una cepa variante resistente al fago de segunda generación se aisló y probó para una secuencia adicional dentro de la ubicación CRISPR.

En estos experimentos, una colección de fagos (o muestras que contienen fago) se probó contra la cepa precursora usando métodos microbiológicos clásicos conocidos en el arte. Los fagos (o muestras) que son virulentos a la cepa precursora se probaron a continuación contra la cepa variante usando los mismos métodos. Un fago (o muestra) que fue virulento a la cepa variante se seleccionó como un segundo fago donador. En el caso de muestras que contienen fago, un fago virulento se purificó para homogeneidad en la cepa variante usando métodos microbiológicos clásicos conocidos en el arte. En algunos experimentos, la secuencia del segundo fago donador se determinó. En algunos experimentos, el segundo fago donador luego se usó como se describe arriba (Ver, Ejemplo 7) para provocar la inserción de una segunda secuencia adicional en la ubicación CRISPR de la cepa variante.

En algunos experimentos, $WT_{\text{phi2972}}^{+S4}$ (Ver, Ejemplo 8 y Tabla 7-1) se usó como la cepa precursora y D858 se usó como el fago donador. Durante la prueba de varios fagos, la cepa DGCC7710 se encontró que era sensible tanto al fago D2972 como al fago D858. Además, D858 se encontró que era virulento contra la cepa variante $WT_{\text{phi2972}}^{+S4}$. El fago D858 por lo tanto se eligió como un segundo fago donador en algunos experimentos.

La cepa variante $WT_{\text{phi2972}}^{+S4}$ se expuso al segundo fago donador D858, como se describe en el Ejemplo 7. Una cepa variante resistente al fago nombrada $WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S18}$ (Ver, Tabla 7-1) se obtuvo que es resistente a D858 (Ver Tabla 9-1). Esta cepa muestra una resistencia incrementada para D2972, como la eficiencia de la formación de placa de $WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S18}$ se redujo por más de 8 logs (comparado con 5 logs para $WT_{\text{phi2972}}^{+S4}$, Ver Tabla 9-1). El ADN se extrajo de $WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S18}$ y su ubicación CRISPR1 se analizó por PCR usando los mismos métodos y los cebadores como se describen arriba. La secuencia del producto PCR se determinó y comparó con la de la ubicación CRISPR de $WT_{\text{phi2972}}^{+S4}$. Comparado con $WT_{\text{phi2972}}^{+S4}$, $WT_{\text{phi858}}^{+S18}$ difiere por la adición de una secuencia espaciadora de 30 pb en el extremo 5' de su región CRISPR1 y por la duplicación de la secuencia repetida, como se muestra en la Figura 17. La comparación de esta secuencia espaciadora adicional con la secuencia del genoma D858 muestra que la segunda secuencia espaciadora adicional es 100% idéntica a la del genoma D858 del nucleótido 30338 hasta el nucleótido 30367.

Otra cepa variante nombrada $WT_{\text{phi2972}}^{+S4} \text{ phi4720}^{+S25}$ (Ver, Tabla 7-1) también se obtuvo usando este método en trabajo experimental independiente. Esta cepa variante contiene una secuencia adicional incluyendo una secuencia espaciadora que es 100% idéntica a una secuencia del genoma D858 del nucleótido 33886 hasta 33915. Muestra una resistencia incrementada para D2972, como la eficiencia de la formación de placa de D2972 en $WT_{\text{phi2972}}^{+S4} \text{ phi4720}^{+S25}$ se redujo por más de 7 logs (Ver, Tabla 9-1).

EJEMPLO 11

Génesis de una Cepa Variante Modificada con CRISPR Resistente a Múltiples Fagos por Múltiples Inserciones de Secuencias Adicionales en la Ubicación CRISPR

En este Ejemplo, el desarrollo de una cepa resistente a fagos múltiples se describe a través de la adición iterativa de las secuencias de fago en la ubicación CRISPR, como la adición de 2 secuencias de fago en la ubicación CRISPR no es suficiente para conferir resistencia a todos los fagos para una cepa dada. Por ejemplo, la cepa $WT_{\text{phi2972}}^{+S4} \text{ phi858}^{+S18}$ (descrita en el Ejemplo 10) se encontró que era sensible a otros múltiples fagos. En el proceso de desarrollar una cepa resistente a fagos múltiples, la cepa precursora se sometió a un primer fago para seleccionar una cepa variante, luego la cepa variante se sometió a un segundo fago para seleccionar una cepa variante de segunda generación que es resistente a ambos fagos. Luego, la última cepa variante se sometió iterativamente a los fagos a los cuales todavía fue sensible, hasta que una cepa variante final se obtuvo que fue resistente a todos los fagos disponibles.

Usando métodos conocidos en el arte, un conjunto de 10 fagos de referencia se identificaron que son representativos de la diversidad de fagos que son capaces de desarrollar una cepa DGCC7710, particularmente fagos D858, D1126, D2766, D2972, D3288, D3821, D4083, D4752, D4753, y N1495. Como se describe en el Ejemplo 7, DGCC7710 se expuso al fago D2972 para generar la cepa variante DGCC9705. DGCC9705 se encontró para ser resistente al fago D2766 y D4752 además al fago D2972, pero todavía fue sensible a otros fagos como se muestra en la Tabla 11-1. DGCC9705 se describe en la Tabla 11-1 y en la Figura 17. DGCC9705 presenta 1 secuencia adicional en CRISPR1 y 1 secuencia adicional en CRISPR3. El análisis de la secuencia de la ubicación CRISPR1 y de la ubicación CRISPR3 se hizo en consecuencia con los métodos descritos en el Ejemplo 7. La secuencia de los productos PCR se determinó y comparó con la de la ubicación CRISPR1 y 3 de DGCC7710. DGCC9705 presenta 1 espaciador adicional en su ubicación CRISPR1 y un espaciador adicional en su ubicación CRISPR3. Las secuencias espaciadoras son idénticas a las secuencias del fago D2972. Usando los mismos métodos, DGCC9705 luego se expuso al fago D3821 y la cepa variante DGCC9726 luego se aisló. Además de ser resistente a D2972, DGCC9726 tiene resistencia a los fagos D858, D3821, D4083 y N1495 (Ver, Tabla 11-1). DGCC9726 tiene 1 secuencia espaciadora adicional en su ubicación CRISPR1 como se compara con DGCC9705 (Ver, Tabla 7-1 y Figura 17). La secuencia espaciadora adicional es idéntica a una secuencia del D2972. A través de la exposición de la cepa DGCC9726 al fago D3288, DGCC9733 se aisló. La cepa DGCC9733 es adicionalmente

resistente al fago D3288 y D1126 (Ver, Tabla 11-1). DGCC9733 tiene 1 secuencia espaciadora adicional en su ubicación CRISPR1 en comparación con DGCC9726 (Ver, Tabla 7-1 y Figura 17). Esta secuencia espaciadora tiene alguna identidad (25/30 pares base de identidad) con una secuencia de fago de estreptococos 7201. Finalmente, por una última exposición iterativa al fago D4753, DGCC9836 se aisló que es resistente a todos los fagos (Ver, Tabla 11-1). DGCC9836 tiene 2 secuencias espaciadoras adicionales en su ubicación CRISPR1 y 2 secuencias espaciadoras adicionales en su ubicación CRISPR3 (Ver, Tabla 7-1 y Figura 17). Una secuencia espaciadora es idéntica a una secuencia en el fago D2972 y las otras 3 secuencias espaciadoras son idénticas a las secuencias en el fago D858.

La Tabla 11-1 proporciona datos con respecto a la sensibilidad de fago de la cepa variante DGCC9836 modificada con CRISPR y cepas variantes modificadas con CRISPR intermedias. En esta Tabla, "S" indica sensibilidad y "R" indica resistencia.

Tabla 11-1. Sensibilidad de fago de la Cepa Variante DGCC9836 modificada con CRISPR y Cepas Variantes Modificadas con CRISP Intermedias												
Cepa	Cepa precursora	Fago donador	Fago									
			D2972	D2766	D4752	D3821	D858	D4083	N1495	D3288	D1126	D4753
DGCC7710	—	—	S	S	S	S	S	S	S	S	S	S
DGCC9705	DGCC7710	D2972	R	R	R	S	S	S	S	S	S	S
DGCC9726	DGCC9705	D3821	R	R	R	R	R	R	R	S	S	S
DGCC9733	DGCC7710	D3288	R	R	R	R	R	R	R	R	R	S
DGCC9836	DGCC7710	D4753	R	R	R	R	R	R	R	R	R	R

EJEMPLO 12

Método Natural para Insertar Secuencias Adicionales Múltiples en una Ubicación CRISPR

En este Ejemplo, los métodos se describen para insertar secuencias adicionales múltiples en la ubicación CRISPR. En estos métodos, en vez de usar los varios fagos iterativamente, la cepa precursora se expuso a una mezcla que contiene fagos múltiples. Una colección de fagos se probó contra cepas múltiples usando métodos microbiológicos clásicos, con objeto de determinar su espectro hospedador. Los fagos que fueron virulentos a la cepa precursora pero que tienen diferente espectro hospedador se seleccionaron. Los fagos seleccionados se mezclaron y usaron en los métodos proporcionados arriba (Ver, Ejemplo 7) para provocar la inserción de secuencias adicionales en la ubicación CRISPR de la cepa variante.

En algunos experimentos, DGCC7710 se usó como la cepa precursora y D858 y D2972 se usaron como los fagos donadores. Durante la prueba de varios fagos, la cepa DGCC7710 se encontró para ser sensible tanto al fago D2972 como al fago D858. Sin embargo, D2972 y D858 presentan diferentes espectros hospedadores cuando se prueban en la cepa DGCC7778, sugiriendo que los dos fagos fueron diferentes.

La cepa precursora DGCC7710 se expuso a una mezcla del fago D858 y D2972 como se describe en el Ejemplo 7. Una cepa variante resistente al fago nombrada WT_{phi858phi2972}^{+S9S10S11S12} (Ver, Tabla 7-1) se obtuvo. Exhibe resistencia para D858, como la eficiencia de la formación de placa de D858 en WT_{phi858phi2972}^{+S9S10S11S12} se redujo por más de 7 logs, así como la resistencia para D2972, como la eficiencia de la formación de placa de D2972 en WT_{phi858phi2972}^{+S9S10S11S12} se redujo por más de 7 logs. El ADN se extrajo de WT_{phi858phi2972}^{+S9S10S11S12} y su ubicación CRISPR se analizó por PCR usando los mismos métodos y cebadores como se describen arriba. La secuencia de los productos PCR se determinó y comparó con la de la ubicación CRISPR1 y de la ubicación CRIPR3 de DGCC7710. Comparado con DGCC7710, WT_{phi858phi2972}^{+S9S10S11S12} difiere por la adición de 4 secuencias espaciadoras de 30 pb en el extremo 5' de su región CRISPR1 y por la duplicación de las secuencias repetidas, como se muestra en la Figura 17. La comparación de las secuencias espaciadoras adicionales con la secuencia del genoma D2972 muestra que las secuencias espaciadoras adicionales son 100% idénticas a las del D2972 del nucleótido 7874 hasta el nucleótido 7903, del nucleótido 20650 hasta el nucleótido 20621, del nucleótido 8360 hasta el nucleótido 8389, y del nucleótido 18998 hasta el nucleótido 19027.

En experimentos adicionales, la cepa WT_{phi858phi2972}^{+S13S14} (Ver, Tabla 7-1) también se obtuvo siguiendo estos métodos. Muestra resistencia para D858, como la eficiencia de la formación de placa de D858 en WT_{phi858phi2972}^{+S13S14} se redujo por 7 logs, y resistencia para D2972, como la eficiencia de la formación de placa de D2972 en WT_{phi858phi2972}^{+S13S14} se redujo por 8 logs. La comparación de las secuencias espaciadoras adicionales con

la secuencia del genoma D2972 muestra que las secuencias espaciadoras adicionales son 100% idénticas a la del D2972 del nucleótido 33602 hasta el nucleótido 33631, y del nucleótido 4830 hasta el nucleótido 4801.

EJEMPLO 13

Fagos Combatientes en la Fermentación Usando una Cepa Variante Modificada con CRISPR

En este Ejemplo se describen los métodos de fagos combatientes en la fermentación por el uso de una cepa variante en vez de una cepa precursora (esto es, tipo natural, receptora). De esta manera, este Ejemplo proporciona aún otra descripción de los beneficios proporcionados por las cepas variantes.

En algunos experimentos, la comparación de la cepa DGCC7710 con la cepa $WT_{\text{phi2972}}^{+S20}$ y la cepa $WT_{\text{phi2972}}^{+S26S27}$ en la fermentación de leche en la presencia del fago D2972 se realizó. DGCC7710 es una cepa industrial usada en la fermentación de la leche. La cepa $WT_{\text{phi2972}}^{+S20}$ se describe en la Tabla 7-1 y en el Ejemplo 8 y muestra en su ubicación CRISPR1 un espaciador adicional, como se compara con la cepa DGCC7710. La cepa $WT_{\text{phi2972}}^{+S20}$ muestra resistencia mejorada para D2972, como se compara con DGCC7710. $WT_{\text{phi2972}}^{+S26S27}$ es otra variante que muestra alguna resistencia para D2972 (descrita en la Tabla 7-1) y muestra 2 espaciadores adicionales en su ubicación CRISPR1.

Las diversas fermentaciones se realizaron con cada cepa. Primero, medio en polvo de leche al 10% (p/v) se sembró con 1% (v/v) de un pre-cultivo de la cepa probada y con 10^4 pfu/ml del fago D2972. El cultivo se incubó a 42°C durante 6 h. Después de la primera fermentación, una segunda fermentación se estableció. Las mismas condiciones de fermentación exactas se usaron excepto que 0.1% de volumen del fermentado de la fermentación precedente se agregó (antes de la adición, el fermentado se filtró usando un filtro de 0.45 μm). Luego, las fermentaciones sucesivas se realizaron con las mismas condiciones experimentales que aquellas usadas para la segunda fermentación. Todas las fermentaciones se registraron por impedimetría. Al final de cada una de las fermentaciones, la coagulación de la leche se probó y la titulación de fagos se realizó usando métodos conocidos en el arte.

En el caso de la fermentación de leche en la ausencia de fago, la variación de la impedancia con DGCC7710 fue arriba de 2500 μS dentro de 6 horas. En la presencia de D2972, (DGCC7710 es muy sensible a los fagos), los fagos D2972 alcanzan un nivel alto de población durante el primer cultivo y la fermentación falla para coagular la leche. La variación de la impedancia en 6 horas fue siempre menos de 500 μS . Al contrario, la fermentación de leche con $WT_{\text{phi2972}}^{+S20}$ en la presencia de D2972 permitió la coagulación de la leche al menos hasta el 3er sub-cultivo y la evolución lenta del nivel de fago se notó. La variación de impedancia incrementa hasta más de 2500 μS también hasta el 3er sub-cultivo. Esto demuestra que la cepa variante $WT_{\text{phi2972}}^{+S20}$ es más apropiada que la cepa precursora DGCC7710 para la acidificación de leche en la presencia de fagos. Además, la fermentación de leche con $WT_{\text{phi2972}}^{+S26S27}$ en la presencia de D2972 permitió la coagulación de la leche hasta el último sub-cultivo sin desarrollar el fago. Además, la variación de impedancia incrementó hasta más de 2500 μS también hasta el último sub-cultivo. Esto demuestra que la cepa variante $WT_{\text{phi2972}}^{+S26S27}$ es más apropiada que la cepa precursora DGCC7710 y aún más apropiada que $WT_{\text{phi2972}}^{+S20}$ para la acidificación de leche en la presencia de fagos. Los experimentos se duplicaron; los resultados se presentan en la Tabla 13-1.

Tabla 13-1. Comparación de la Fermentación de Leche con DGCC7710 o con Cepas Variantes en la presencia del fago D2972				
Cepa	Subcultivos	Coagulación de leche dentro de 6 h	Nivel de fago en la fermentación de la cepa	Variación de impedancia dentro de 6h (en μ S)
DGCC7710	Control	Sí	No	3246
DGCC7710	1	No	Alto	319
Primera prueba	2	No	Alto	23
DGCC7710	1	No	Alto	287
Segunda prueba	2	No	Alto	16
WT _{phi2972} ^{+S20}	1	Sí	No	3281
Primera prueba	2	Sí	Bajo	3148
	3	Sí	Bajo	3251
	4	Sí	Alto	3101
WT _{phi2972} ^{+S20}	1	Sí	Bajo	3371
Segunda prueba	2	Sí	Medio	3265
	3	Sí	Alto	2780
	4	No	Alto	223
WT _{phi2972} ^{+S26S27}	1	Sí	No	3322
Primera prueba	2	Sí	No	3208
	3	Sí	No	3467
	4	Sí	No	3182
WT _{phi2972} ^{+S26S27}	1	Sí	No	3260
Segunda prueba	2	Sí	No	3021
	3	Sí	No	3377
	4	Sí	No	3246

En un segundo conjunto de experimentos, la comparación de la cepa DGCC7710 con la cepa WT_{phi2972}^{+S20} y la cepa WT_{phi2972}^{+S26S27} en la fermentación de leche en la presencia del fago D2972 se reprodujo. Además, la cepa DGCC9836 se estudió. La cepa DGCC9836 es una cepa variante aún más evolucionada de DGCC7710 que es el resultado del enfrentamiento del fago múltiple. Esta cepa presenta 5 espaciadores adicionales en su ubicación CRISPR1 y 3 espaciadores adicionales en su ubicación CRISPR3 (Ver, Ejemplo 11 y Figura 17). DGCC9836 es resistente a todos los fagos probados.

Los experimentos se condujeron como se describen arriba. Los resultados se muestran en la Tabla 13-2. En cuanto al primer conjunto de experimentos, la fermentación de leche con WT_{phi2972}^{+S20} en la presencia de D2972 permitió la coagulación de la leche hasta el 5° sub-cultivo y la evolución lenta del nivel de fago se midió. La variación de impedancia incrementó hasta más de 2500 μ S durante los 5 primeros sub-cultivos, demostrando que la acidificación de la leche no es afectada por los fagos. En el 6° sub-cultivo, el nivel de fago incrementó significativamente y la fermentación de leche se dañó. Para las otras 2 cepas variantes, la fermentación de leche no se afectó a lo largo de los 6 sub-cultivos, los fagos jamás se desarrollaron y la variación registrada de impedancia siempre fue superior a 2500 μ S.

Estos experimentos demuestran que las cepas que contienen al menos una secuencia espaciadora adicional en su ubicación CRISPR1 permiten la fermentación de leche aún en la presencia de fagos. La fermentación de leche es aún más segura cuando las cepas tienen más de una secuencia espaciadora adicional en su ubicación CRISPR.

Tabla 13-2. Comparación de la Fermentación de Leche con Cepas Variantes de DGCC7710 en la presencia del fago D2972				
Cepa	Subcultivos	Coagulación de leche dentro de 6 h	Nivel de fago en la fermentación de la cepa	Variación de impedancia dentro de 6h (en μ S)
WT _{phi2972} ^{+S20}	1	Sí	No	3171
	2	Sí	No	2991
	3	Sí	Bajo	2871
	4	Sí	Medio	2934
	5	Sí	Medio	2906
	6	No	Alto	661
WT _{phi2972} ^{+S26S27}	1	Sí	No	2970
	2	Sí	No	2945
	3	Sí	No	2617
	4	Sí	No	2721
	5	Sí	No	2660
	6	Sí	No	2605
DGCC9836	1	Sí	No	3028
	2	Sí	No	3115
	3	Sí	No	2708
	4	Sí	No	2817
	5	Sí	No	2845
	6	Sí	No	2813

EJEMPLO 14

5 Fagos Combatientes en la Fermentación Usando una Combinación de Cepas Variantes Modificadas con CRISPR

En este Ejemplo, se describen métodos para fagos combatientes en la fermentación a través del uso de una combinación de cepas variantes en lugar de usar una cepa sencilla. De esta manera, este Ejemplo ilustra el uso simultáneo de más de una cepa variante (esto es, una combinación de cepas variantes). Ciertamente, las mezclas de cepas que muestran las mismas funcionales, aún patrones de sensibilidad de fago diferentes encuentran uso en tales aplicaciones. Por ejemplo, 2 ó 3 o aún más cepas variantes como se describen en la presente encuentran uso en tales aplicaciones. Usando una combinación de cepas variantes con secuencias espaciadoras agregadas diferentes en su ubicación CRISPR permite la fermentación para resistir más fácilmente cualquiera de los fagos mutantes emergentes.

En algunos experimentos, las comparaciones se hacen entre la cepa WT_{phi2972}^{+S21} sola y una combinación de 3 cepas (particularmente WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21} y WT_{phi2972}^{+S22}) usadas en la fermentación de leche en la presencia del fago D2972. Las cepas WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21} y WT_{phi2972}^{+S22} se describen en la Tabla 7-1 y en el Ejemplo 8. Existen cepas variantes independientes de DGCC7710. Cada cepa variante muestra en su ubicación CRISPR1 una secuencia espaciadora adicional distinta (la cual es originada del fago D2972) como se compara con la cepa DGCC7710.

Varias fermentaciones se realizaron ya sea con la cepa WT_{phi2972}^{+S21} sola o en la combinación de tres cepas. Primero, medio en polvo de leche al 10% (p/v) se sembró con 1% (v/v) de un pre-cultivo de la cepa sola o la combinación de cepas y con 10⁴ pfu/ml del fago D2972. El cultivo se incubó a 42°C durante 6 h. Después de la primera fermentación, una segunda fermentación se estableció. Las mismas condiciones de fermentación se usaron excepto que el volumen al 0.1% del fermentado de la fermentación precedente se agregó (antes de la adición, el fermentado se filtró usando un filtro 0.45 μ m). Luego las fermentaciones sucesivas se realizaron usando las mismas condiciones experimentales como aquellas usadas para la segunda fermentación. Todas las fermentaciones se registraron por impedimetría. Al final de cada una de las fermentaciones la coagulación de la leche se probó y la titulación de fagos se realizó usando métodos conocidos en el arte. Los experimentos se duplicaron; los resultados se proporcionan en la Tabla 14-1.

Tabla 14-1. Comparación de la Fermentación de Leche con WT _{phi2972} ^{+S21} o con una Combinación de Tres Cepas en la Presencia del Fago D2972				
Cepa	Subcultivos	Coagulación de leche dentro de 6 h	Nivel de fago en la fermentación de la cepa	Variación de impedancia dentro de 6h (en μ S)
DGCC7710	Control	No	Alto	46
Primera prueba	WT _{phi2972} ^{+S21} 1	Sí	Bajo	3367
	2	Sí	Alto	3312
	3	No	Alto	555
	4	No	Alto	33
	5	No	Alto	25
Segunda prueba	WT _{phi2972} ^{+S21} 1	Sí	Bajo	3450
	2	Sí	Alto	3293
	3	No	Alto	1071
	4	No	Alto	29
	5	No	Alto	26
Primera prueba	WT _{phi2972} ^{+S20} 1	Sí	Muy bajo	3242
	WT _{phi2972} ^{+S21} 2	Sí	Bajo	3233
	WT _{phi2972} ^{+S22} 3	Sí	Alto	3319
	4	Sí	Alto	3169
	5	Sí	Alto	3261
Segunda prueba	WT _{phi2972} ^{+S20} 1	Sí	Muy bajo	3384
	WT _{phi2972} ^{+S21} 2	Sí	Bajo	3178
	WT _{phi2972} ^{+S22} 3	Sí	Alto	3206
	4	Sí	Alto	3295
	5	Sí	Alto	3209

La fermentación de leche se hace con WT_{phi2972}^{+S21} en la presencia de fagos fallidos en el tercer subcultivo en ambas pruebas. Esto se mostró por una ausencia de coagulación de la leche y por una variación altamente reducida de impedancia después de 6 horas de fermentación. Por el contrario, a pesar de algún desarrollo del fago D2972, las fermentaciones se condujeron exitosamente hasta el quinto subcultivo cuando la mezcla de las tres cepas se usó. La coagulación de la leche se registró en todos los cultivos y la variación de impedancia dentro de 6 horas de incubación jamás fue menor de 3000 μ S.

Estos experimentos demuestran que el uso de la combinación de cepas variantes que tienen al menos una secuencia espaciadora distinta adicional en su ubicación CRISPR1 permite la fermentación de leche en la presencia de fagos, como se compara con el uso de la cepa variante sencilla.

EJEMPLO 15

Fagos Combatientes en la Fermentación Usando una Rotación de Cepas Variantes Modificadas con CRISPR

En experimentos adicionales, las cepas variantes se usaron en rotación. En algunos experimentos, las cepas tienen las mismas funcionalidades, pero diferentes patrones de sensibilidad de fago. De esta manera, en este Ejemplo, se describen experimentos que se conducen en el uso iterativo/posterior de diversas cepas diferentes (esto es, cepas variantes modificadas con CRISPR) secuencialmente en un esquema de rotación.

En algunos experimentos, las comparaciones se hicieron entre la cepa WT_{phi2972}^{+S21} sola y las cepas WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21} y WT_{phi2972}^{+S22} usadas sucesivamente (en rotación) en la fermentación de leche en la presencia del fago D2972. La primera fermentación de leche se condujo con la cepa WT_{phi2972}^{+S20}. Luego, la cepa WT_{phi2972}^{+S22} se usó para la segunda fermentación, y la cepa WT_{phi2972}^{+S21} se usó para la tercera fermentación. La cuarta fermentación luego se hizo nuevamente usando la cepa WT_{phi2972}^{+S20}; seguida por una fermentación con la cepa WT_{phi2972}^{+S22}, luego la cepa WT_{phi2972}^{+S21}, y etc. Las cepas WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21} y WT_{phi2972}^{+S22} se describen arriba en la Tabla 7-1. Existen cepas variantes independientes de la cepa DGCC7710. Cada cepa variante muestra una secuencia espaciadora adicional distinta en su ubicación CRISPR1, como se compara con la cepa DGCC7710, que es originada del fago D2972.

Varias fermentaciones se realizaron usando los mismos métodos experimentales como los que se describen en el Ejemplo 14. Los experimentos se hicieron en triplicado; los resultados se muestran en la Tabla 15-1. Varias fermentaciones inoculadas con WT_{phi2972}^{+S20} solo fueron exitosas hasta el 3er subcultivo, como se muestra por los valores de variación de impedancia superiores a 3000 μ S y por la coagulación de leche. Los siguientes subcultivos fallaron para coagular la leche y los valores de fago altos se registraron. En contraste, varias fermentaciones hechas al inocular la leche en rotación con 3 cepas variantes diferentes (WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21} y WT_{phi2972}^{+S22}) fueron exitosas hasta el décimo subcultivo. Bajo estas condiciones experimentales, los fagos fallan para propagar y permanecer en niveles bajos. Los resultados indican que usando una rotación de cepas variantes proporciona resistencia al fago mejorada durante la fermentación, como se compara con el uso de una cepa variante sencilla.

Tabla 15-1. Fermentación de Leche Sucesiva con Rotación de Cepas en la Presencia del Fago D2972				
Prueba	Fermentación	Coagulación de leche dentro de 6h	Nivel de fago en la fermentación de la cepa	Variación de impedancia dentro de 6h (en μ S)
Control	1ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3464
	2ª fermentación WT _{phi2972} ^{+S20}	Sí	Bajo	3203
	3ª fermentación WT _{phi2972} ^{+S20}	Sí	Medio	3300
	4ª fermentación WT _{phi2972} ^{+S20}	No	Alto	235
	5ª fermentación WT _{phi2972} ^{+S20}	No	Alto	249
Primera prueba	1ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3134
	2ª fermentación WT _{phi2972} ^{+S22}	Sí	No	3253
	3ª fermentación WT _{phi2972} ^{+S21}	Sí	No	3076
	4ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3103
	5ª fermentación WT _{phi2972} ^{+S22}	Sí	No	2930
	6ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	3127
	7ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3160
	8ª fermentación WT _{phi2972} ^{+S22}	Sí	Bajo	2969
	9ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	2967
	10ª fermentación WT _{phi2972} ^{+S20}	Sí	Bajo	3051
Segunda prueba	1ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3168
	2ª fermentación WT _{phi2972} ^{+S22}	Sí	No	3063
	3ª fermentación WT _{phi2972} ^{+S21}	Sí	No	2992
	4ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3186
	5ª fermentación WT _{phi2972} ^{+S22}	Sí	Bajo	3070
	6ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	2932
	7ª fermentación WT _{phi2972} ^{+S20}	Sí	Bajo	3181
	8ª fermentación WT _{phi2972} ^{+S22}	Sí	Bajo	2920
	9ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	2863
	10ª fermentación WT _{phi2972} ^{+S20}	Sí	Bajo	2995
Tercera prueba	1ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3308
	2ª fermentación WT _{phi2972} ^{+S22}	Sí	No	3290
	3ª fermentación WT _{phi2972} ^{+S21}	Sí	No	3023
	4ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3039
	5ª fermentación WT _{phi2972} ^{+S22}	Sí	Bajo	2994
	6ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	2945
	7ª fermentación WT _{phi2972} ^{+S20}	Sí	Bajo	3184
	8ª fermentación WT _{phi2972} ^{+S22}	Sí	Bajo	3048
	9ª fermentación WT _{phi2972} ^{+S21}	Sí	Bajo	3063
	10ª fermentación WT _{phi2972} ^{+S20}	Sí	No	3016

EJEMPLO 16

Reducción y Control de la Población de Fago Usando Cepas Variantes Modificadas con CRISPR

En este Ejemplo, se describen experimentos que se conducen para determinar la capacidad de una cepa modificada con CRISPR para destruir fagos que se han hecho resistentes. En particular, los experimentos se diseñaron para

determinar si una población de fago se disminuirá a un nivel indetectable durante la fermentación de una cepa modificada con CRISPR.

En algunos experimentos, DGCC9836 (descrita en el Ejemplo 11 y en la Figura 17) se usó para realizar la fermentación de leche en la presencia del fago D2972, en comparación con las fermentaciones hechas con su cepa precursora DGCC7710 en la presencia de D2972. El medio en polvo de leche al diez por ciento (p/v) se sembró con alrededor de 10^6 cfu/ml de un pre-cultivo de la cepa probada y con 10^7 pfu/ml del fago D2972. El cultivo se incubó a 42°C durante 24 h. En varios puntos de tiempo, una alícuota se tomó y la población de fago se midió usando placa de agar de doble capa sembrada con DGCC7710 usando métodos estándar conocidos en el arte. Los resultados se presentan en la Figura 20. En la fermentación de leche por DGCC7710, el fago D2972 se desarrolló hasta alcanzar una población de más de 10^8 pfu/ml. En contraste, durante la fermentación con DGCC9836, la población de fago D2972 gradualmente disminuyó a un nivel muy bajo (120 pfu/ml) después de 6 horas de incubación, y fue casi indetectable después de 24 horas de incubación. Este último resultado sugiere que los fagos se destruyeron durante el proceso de fermentación con la cepa variante DGCC9836.

La propiedad de una cepa variante para destruir fagos, así como la de no ser sensible a los fagos representa un beneficio adicional, como se compara con el programa de rotación de cultivo iniciador tradicional para el cual las cepas no son sensibles, pero son inofensivas a los fagos. Ciertamente, al usar cepas variantes, la erradicación de los fagos inactivos ocurrirá a través de la combinación del lavado de los fagos (en cuanto a la rotación usando cultivo iniciador tradicional) y destrucción de los fagos.

En otros experimentos, las cepas variantes que presentan alguna, pero incompleta, resistencia al fago D2972 se asociaron en la fermentación de leche en la presencia de D2972. Las cepas variantes seleccionadas incluidas $\text{WT}_{\text{phi2972}}^{+\text{S20}}$ y $\text{WT}_{\text{phi2972}}^{+\text{S21}}$, como se describen en el Ejemplo 8 y en la Tabla 7-1. Estas cepas muestran reducciones EOP para el fago D2972 de alrededor de 5 logs. Las fermentaciones de leche se realizaron como se describen arriba (velocidad de inoculación bacteriana de 10^6 cfu/ml; velocidad de inoculación de fago de 10^7 pfu/ml). Las fermentaciones de leche se hicieron ya sea con $\text{WT}_{\text{phi2972}}^{+\text{S20}}$ o con $\text{WT}_{\text{phi2972}}^{+\text{S21}}$ o una mezcla de las dos cepas. En varios puntos de tiempo, la población de fagos se registró. Para este propósito, una alícuota se tomó y la población de fago se midió usando placa de agar de doble capa sembrada ya sea con $\text{WT}_{\text{phi2972}}^{+\text{S20}}$ o con $\text{WT}_{\text{phi2972}}^{+\text{S21}}$, usando métodos estándar conocidos en el arte. Los resultados se presentan en la Figura 21, que indican la suma de los fagos detectados en $\text{WT}_{\text{phi2972}}^{+\text{S20}}$ y en $\text{WT}_{\text{phi2972}}^{+\text{S21}}$ para cada una de las fermentaciones de leche. Cuando una cepa sencilla se usó para la fermentación ($\text{WT}_{\text{phi2972}}^{+\text{S20}}$ o $\text{WT}_{\text{phi2972}}^{+\text{S21}}$), el número de fagos detectados en el tiempo de inoculación fue alrededor de 100 pfu/ml (debido a los 5 logs de la reducción EOP). Durante el cultivo, este número de fagos se elevó hasta 10^6 ó 10^7 (respectivamente), correspondiente a una multiplicación de los fagos por 4 hasta 5 logs. El factor de multiplicación de los fagos fue mucho más bajo (2 logs) para la fermentación de leche inoculada con las 2 cepas. Ciertamente, el número de fagos se elevó desde 100 pfu/ml hasta un máximo de alrededor de 10^4 pfu/ml. Estos resultados conclusivamente mostraron que durante un co-cultivo de 2 cepas variantes la velocidad de propagación de los fagos se redujo significativamente comparada con la velocidad de propagación de los fagos en un cultivo hecho con cepas variantes sencillas.

EJEMPLO 17

Inserción de Espaciadores

En este Ejemplo, se describen métodos y composiciones usadas para insertar dos espaciadores en *S. thermophilus* DGCC7710. La cepa DGCC7710 *S. thermophilus* (depositada en la "Colección Nacional de los Cultivos de Microorganismos" francesa con el número CNCM I-2423) posee al menos 3 ubicaciones CRISPR: CRISPR1, CRISPR2, y CRISPR3. En las cepas CNRZ1066 y LMG18311, para las cuales la secuencia del genoma completo se conoce (Ver, Bolotin et al., [2004], supra), la CRISPR1 se localiza en la misma ubicación cromosomal: entre str0660 (o stu0660) y str0661 (o stu0661) (Ver, Figura 18). En la cepa DGCC7710, la CRISPR1 también se localiza en la misma ubicación cromosomal, entre genes altamente similares. La CRISPR1 de la cepa DGCC7710 contiene 33 repeticiones (incluyendo la repetición terminal), y de esta manera 32 espaciadores (Ver, Figura 19). Todos estos espaciadores son diferentes uno del otro. La mayoría de estos espaciadores no se han descrito previamente para estar dentro de la ubicación CRISPR, pero cuatro espaciadores cercanos a la pista CRISPR1 son idénticos a los espaciadores CRISPR1 conocidos. Por ejemplo, el 28° espaciador de DGCC7710 es 100% idéntico al 31° espaciador CRISPR1 de la cepa CNRZ1575 (Número de acceso al Genbank DQ072991); el 30° espaciador de DGCC7710 es 100% idéntico al 27° espaciador CRISPR1 de la cepa CNRZ703 (Número de acceso al Genbank DQ072990); el 31° espaciador de DGCC7710 es 100% idéntico al 28° espaciador CRISPR1 de la cepa CNRZ703 (Número de acceso al Genbank DQ072990); y el 32° espaciador de DGCC7710 es 100% idéntico al 30° espaciador CRISPR1 de la cepa CNRZ703 (Número de acceso al Genbank DQ072990). La secuencia CRISPR1 (5'-3') de la cepa DGCC7710 se muestra en SEQ ID NO:678, a continuación:

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtggttgacagcaaataagattcgaattgt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttattc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgttttgttgcgaaatcct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgttggaacgtattta
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcggtcccaatcctgattaatacttactcg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggagatgatgctatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggtaactgctcaagtg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaaggcatagaaaggagacaacatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaaatcattttcataacttcat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagatcagcaagcaagctgttagttact
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggatagcttaatatcattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcagcagcgttcgagtttaccgtttc
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggccaactaacaattatgct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaattctgttttaggtacattt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatgggtctt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatcccaga
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagttagttgatgacctctacaatggtttat
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
 GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
 Ttgattcaacataaaaagccagttcaattgaacttggttt (SEQ ID NO:678)

- 5 D858, el fago usado en estos experimentos es un bacteriófago que pertenece a la familia de virus *Siphoviridae*. Su secuencia de genoma se ha determinado completamente, aparentemente permanece para publicarse. Este fago es virulento para la cepa DGCC7710 *S. thermophilus*. La cepa DGCC7778 *S. thermophilus* se aisló como un mutante resistente al fago natural usando DGCC7710 como la cepa precursora, y el fago D858 como el fago virulento. La CRISPR1 de la cepa DGCC7778 contiene 35 repeticiones (incluyendo la repetición terminal), y de esta manera 34 espaciadores. Cuando se compara con la secuencia CRISPR1 de DGCC7710, la secuencia CRISPR1 de
- 10 DGCC7778 posee dos espaciadores adicionales, adyacentes, nuevos (y por supuesto dos repeticiones adicionales que flanquean los nuevos espaciadores) en un extremo de la ubicación CRISPR (esto es, cercano al líder). Todos los otros espaciadores de la ubicación CRISPR1 no se cambian. La secuencia CRISPR1 (5'-3') de la cepa DGCC7778 se muestra en la SEQ ID NO:679, a continuación:

caaggacagttattgattttataatcactatgtgggtataaaaaacgtcaaaatttcatttgag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaacacattcaacagattaatgaagaatac
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtccactcacgtacaaatagtgtgagtgtactc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaataagattcgaattgt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttattc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgttggaatcct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaaataatcttaattacaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgttggaacgtattta
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcggtcccaatcctgattaatacttactcg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacaacagcaagacaaggagatgatgctatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggtaactgctcaagtg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggtcatagaaaggagacaacatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatatttaaatacattttcataacttcat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggatatagcttaatatcattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagtttaccgtttc
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgaggtcaactaacaattatgct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcgttcaaattctgttttaggtacattt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatggctct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggttaacagctactttttacagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgctacatctcaaaggatgatcccaga
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatgggttat
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgcccttctttgccccttgactag
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgccattgagt
 GTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
 Ttgattcaacataaaaaagccagttcaattgaacttggcttt (SEQ ID NO:679)

En el caso de DGCC7778, el primer espaciador (5'-caacacattcaacagattaatgaagaatac-3'; SEQ ID NO:680) y el segundo espaciador (5'-tccactcacgtacaaatagtgtgagtgtactc-3'; SEQ ID NO:681) constituye la marca específica de la cepa que identifica esta cepa etiquetada. Se determinó que la secuencia de ambos espaciadores nuevos existe dentro del genoma de fago D858. La secuencia del segundo espaciador nuevo se encuentra entre las posiciones 25471 y 25442 pb (esto es, en la hebra menos) del genoma D858, como una incompatibilidad (96.7% de nucleótidos idénticos sobre 30 nucleótidos):

Espaciador 2 1 tccactcacgtacaaatagtgtgagtgtactc 30 (SEQ ID NO:681)

|||||

D858 25471 tccactcacgtacaaatagtgtgagcgtactc 25442 (SEQ ID NO:686)

La secuencia del primer espaciador se encuentra entre las posiciones 31481 y 31410 pb (esto es, en la hebra más) del genoma D858 (100% de nucleótidos idénticos sobre 30 nucleótidos):

Espaciador 1 1 caacacattcaacagattaatgaagaatac 30 (SEQ ID NO:3)

|||||

D858 31481 caacacattcaacagattaatgaagaatac 31410 (SEQ ID NO:687)

Aunque no se pretende que la presente invención se limite por ningún mecanismo ni teoría particular, se contempla que dos nuevos espaciadores presentes en la ubicación CRISPR1 de DGCC7778 fueron necesarios para conferir a la cepa DGCC7778 una nueva resistencia para fago D858). El espaciador "2" (como se encuentra en DGCC7778) se insertó primero en la ubicación CRISPR1 de DGCC7710 (33 repeticiones y 32 espaciadores), en un extremo de esta ubicación CRISPR, junto con una repetición. Esta inserción dio elevación a un mutante insensible del bacteriófago (cepa intermedia), marcada por este nuevo espaciador adicional (de esta manera portando 34 repeticiones y 33 espaciadores). Este espaciador se derivó del genoma D858, pero un error de replicación o error de transcripción inverso probablemente ocurrió durante el proceso de inserción, llevando a una mutación de punto. Debido a que la alineación imperfecta (esto es, la incompatibilidad 1) entre este espaciador nuevamente adquirido y la secuencia de fago etiquetada, la eficiencia de la resistencia de esta cepa intermedia al fago D858 fue baja. Un segundo evento de la inserción del espaciador ocurrió en esta cepa intermedia (más resistente al fago D858 que la cepa precursora DGCC7710, pero no resistente "completamente" debido a la incompatibilidad), llevando a la inserción de un segundo espaciador nuevo (esto es, espaciador "1" como se encuentra en DGCC7778) en el mismo extremo de la ubicación CRISPR1, junto con una repetición. Esta segunda inserción dio la elevación a un nuevo mutante insensible al bacteriófago, que se aisló y nombró DGCC7778. DGCC7778 es más resistente a D858 que la cepa intermedia, y por supuesto mucho más resistente que la cepa precursora DGCC7710, debido a la presencia del espaciador "1", que es 100% idéntico a la secuencia de fago etiquetada.

EJEMPLO 18

Método para Etiquetar DGCC7710 y Selección de la Cepa DGCC7778 Etiquetada

En este Ejemplo, se describen métodos usados para etiquetar DGCC7710 y la selección de la cepa DGCC7778 etiquetada. La cepa DGCC7710 se infectó/estimuló por el fago D858 al inocular la leche pasteurizada con la cepa DGCC7710 en alrededor de 2.10^6 cfu/ml y con el fago D858 en alrededor de 1.10^5 pfu/ml. La leche inoculada se cultivó durante 12 horas a 35°C. Después de la incubación, las bacterias viables (esto es, aquellas que son probables que sean mutantes insensibles al bacteriófago) se aislaron en medio no selectivo (placas de agar de leche) a 35°C, usando una dilución apropiada del cultivo infectado. Un aislado, nombrado DGCC7778, se propagó en medio líquido de M17-glucosa a 35°C y su ADN se extrajo usando un protocolo de extracción de ADN clásico, como se conoce en el arte.

El extracto de ADN se amplificó usando PCR como se conoce en el arte (Ver por ejemplo, Bolotin et al. [2005], supra) usando la combinación de un cebador delantero (ya sea yc70 y/o SPIDR-arriba [5'-gTCTTTAgAAACTgTgACACC]; SEQ ID NO:674) y de un cebador inverso (ya sea yc31 y/o SPIDR-abajo [5'-TAAACAgCCTCCCTATCC]; SEQ ID NO:675). La secuencia de los productos PCR se determinó y comparó con la de la ubicación CRISPR de DGCC7710.

EJEMPLO 19

Producción de una Segunda Cepa Etiquetada

En este Ejemplo, se describen métodos usados para producir una segunda cepa etiquetada. La cepa DGCC7710-RH1 *S. thermophilus* se aisló como un mutante resistente al fago natural usando DGCC7710 como la cepa precursora y el fago D858 como el fago virulento.

La CRISPR1 de la cepa DGCC7710-RH1 contiene 34 repeticiones (incluyendo la repetición terminal), y de esta manera 33 espaciadores. Cuando se compara con la secuencia CRISPR1 de la cepa DGCC7710 *S. thermophilus*, la secuencia CRISPR1 de la cepa DGCC7710-RH1 *S. thermophilus* posee un nuevo espaciador adicional (esto es, secuencia de etiquetado) (y por supuesto una repetición adicional que flanquea el nuevo espaciador) en un extremo de la ubicación CRISPR (esto es, cerca del líder, en el extremo 5' de la ubicación CRISPR). Todos los otros espaciadores de la ubicación CRISPR1 no se cambiaron. La secuencia CRISPR1 (5'-3') de la cepa DGCC7710-RH1 es:

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataaccactt
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgtttgogaatcct
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACgctggcgaggaaacgaacaaggcctcaaca
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACgagcgcgagctcgaataatcttaattacaag
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgttggaacgtattta
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatcctgattaatacttactcg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaaggagtgatgctatg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggttaactgctcaagtg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaaggagacaacatg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACcgatattttaaatacattttcataacttcat
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataaatttttaaga
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaataatttatgggtatagcttaatatcattg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtgcacgcagcagcttcgagtttaccgtttc
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtctatatcgagggtcaactaacaattatgct
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaatcaatacgacaagagttaaaatggctct
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtaatgctacatctcaaaggatgatccaga
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatggtttat
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatattgattg
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
 GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAGT
 Ttgattcaacataaaaagccagttcaattgaacttggttt (SEQ ID NO:682)

La secuencia líder es 5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' (SEQ ID NO:688). La secuencia integrada (GTTTTTGTA CTCTCAAGATTTAAGTAACTGTACAACtcaacaattgcaacatcttataaccactt; SEQ ID NO:689) se muestra en gris, que comprende una repetición CRISPR (mayúscula) y un espaciador CRISPR (esto es, secuencia de etiquetado), que se muestra en minúscula. La repetición terminal (5' gttttgtactctcaagatttaagtaactgtacagt 3' (SEQ ID NO:3)) secuencia de pista: 5' ttgattcaacataaaaagccagttcaattgaacttggttt3' (SEQ ID NO:691) se muestran.

En consecuencia, en el caso de la cepa DGCC7710-RH1 *S. thermophilus*, el espaciador (5'-tcaacaattgcaacatcttataaccactt-3'; SEQ ID NO:534) constituye la secuencia de etiquetado específica de la cepa que identifica esta cepa mutante (esto es, la bacteria etiquetada). La secuencia del nuevo espaciador (esto es, secuencia de etiquetado) existe dentro del genoma de fago D858. La secuencia del espaciador se encuentra entre las posiciones 31921 y 31950 pb (esto es, en la hebra más) del genoma D858 (y tiene 100% de identidad con la secuencia genómica D858 sobre 30 nucleótidos):

Espaciador	1	tcaacaattgcaacatcttataaccactt	30	(SEQ ID NO:534)
D858	31921	tcaacaattgcaacatcttataaccactt	31950	(SEQ ID NO:534)

El nuevo espaciador (esto es, secuencia de etiquetado) que se integra en la ubicación CRISPR1 de la cepa DGCC7710-RH1 *Streptococcus thermophilus* confiere a esta cepa una nueva resistencia al fago D858.

EJEMPLO 20

Producción de una Tercera Cepa Etiquetada

En este Ejemplo, se describen métodos usados para producir una tercera cepa etiquetada. La cepa DGCC7710-RH2 *S. thermophilus* se aisló como un mutante resistente al fago natural usando la cepa DGCC7710 *S. thermophilus* como la cepa precursora, y el fago D858 como el fago virulento. La CRISPR1 de la cepa DGCC7710-RH2 *S. thermophilus* contiene 34 repeticiones (incluyendo la repetición terminal), y de esta manera 33 espaciadores. Cuando se compara con la secuencia CRISPR1 de la cepa DGCC7710 *S. thermophilus*, la secuencia CRISPR1 de la cepa DGCC7710-RH2 *S. thermophilus* posee un nuevo espaciador adicional (esto es, secuencia de etiquetado) (y por supuesto una repetición adicional que flanquea el nuevo espaciador) en un extremo de la ubicación CRISPR (esto es, cerca del líder, en el extremo 5' de la ubicación CRISPR). Todos los otros espaciadores de la ubicación CRISPR1 no se cambian.

La secuencia CRISPR1 (5'-3') de la cepa DGCC7710-RH2 es:

```

caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaatcaatga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgtttgacagcaaatcaagattcgaattgt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgacgaggagctattggcacaacttaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatttgacaatctgctgaccactgttatc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacacttggcaggcttattactcaacagcga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACctgttccttgttcttttgttgatcttttc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttcattcttccgtttttgttgcgaaatcct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgctggcgaqaaacgaacaaagcctcaaca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcatagagtggaaaactagaaacagattcaa
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataatgccgttgaattacacggcaagggtca
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgagcgagctcgaataatcttaattacaag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgttcgctagcgtcatgtggtaacgtattta
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACggcgtcccaatcctgattaataacttactcg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaacacagcaagacaagaggatgatgctatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgacacaagaacgtatgcaagagttcaag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacaattcttcatccggttaactgctcaagt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattaagggcatagaaaggagacaacatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcgatattttaaatacattttcataacttcat
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcagtatcagcaagcaagctgttagttact
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataaactatgaaattttataatttttaaga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaataatttatggtatagcttaatatcattg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtgcacgagcagcttcgagttaccgtttc
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACtctatatcgagggtcaactaacaattatgct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcggttcaaattctgttttaggtacattt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatcaatagcacaagagttaaaatggctctt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACgcttagctgtccaatccacgaacgtggatg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACcaaccaacggtaacagctactttttacagt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACataactgaaggataggagcttgtaaagtct
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaatgctacatctcaaaggatgatcccaga
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaagtagttgatgacctctacaatgggttat
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACacctagaagcatttgagcgtatatattgattg
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaattttgccccttctttgccccttgactag
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACaccattagcaatcatttgtgcccattgagt
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAGT
Ttgattcaacataaaaagccagttcaattgaacttggcttt (SEQ ID NO:684)

```

La secuencia líder es 5' caaggacagttattgattttataatcactatgtgggtataaaaacgtcaaaatttcatttgag 3' (SEQ ID NO:688).

La secuencia integrada (GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACttacgtttgaaaagaatatcaaatcaatgaj SEQ ID NO:694) se muestra en gris, que comprende una repetición CRISPR (mayúscula) y un espaciador CRISPR (esto es, secuencia de etiquetado), que se muestra en minúscula. La repetición terminal (5' gttttgtactctcaagatttaagtaactgtacagt (SEQ ID NO:3)) secuencia de pista: 5' ttgattcaacataaaaagccagttcaattgaacttggcttt3' (SEQ ID NO:691) se muestran.

De esta manera, en el caso de la cepa DGCC7710-RH2 *Streptococcus thermophilus*, el espaciador (5'-ttacgtttgaaaagaatatcaaatcaatga-3'; SEQ ID NO:697) constituye la marca específica de la cepa que identifica esta

cepa mutante (esto es, bacteria etiquetada). La secuencia del nuevo espaciador se mostró para existir dentro del genoma de fago D858. La secuencia del espaciador se encuentra entre las posiciones 17215 y 17244 pb (esto es, en la hebra más) del genoma D858 (y tiene 100% de identidad con la secuencia genómica D858 sobre 30 nucleótidos):

```

Espaciador      1 ttacgtttgaaaagaatatcaaataaatga 30 (SEQ ID NO:697)
                  |||
D858            17215 ttacgtttgaaaagaatatcaaataaatga 17244 (SEQ ID NO:698)

```

El nuevo espaciador integrado en la ubicación CRISPR1 de la cepa DGCC7710-RH2 *S. thermophilus* confiere una nueva resistencia al fago D858 hasta la cepa DGCC7710-RH2 *S. thermophilus*.

EJEMPLO 21

Construcción de Fago de "Escape CRISPR" de Variantes Bacterianas Resistentes al Fago

En este Ejemplo, se describen métodos para la construcción de fagos de escape CRISPR. Las variantes hospedadoras resistentes al fago se construyeron primero como se describe en los Ejemplos anteriores. En estos experimentos, una cepa precursora "A" se expone al fago "P" y se selecciona una variante resistente al fago (Variante "A 1.0"). La variante A 1.0 se analiza (por ejemplo por PCR, y/o procesado por secuencia de ADN) para confirmar la presencia de un espaciador insertado adicional dentro de ubicación CRISPR. La secuencia de nucleótidos del espaciador adicional (espaciador Sp1.0) luego se determina. Típicamente, el espaciador Sp1.0 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago P, y da resistencia al fago P y fagos relacionados ("fagos relacionados" son aquellos que contienen la secuencia del espaciador en sus genomas, y define un familia de fagos).

Independientemente de la primera exposición del fago, la misma cepa precursora A se expone al mismo fago P y una segunda variante resistente al fago (Variante A2.0) se selecciona. La variante A2.0 se selecciona con objeto también de tener un espaciador adicional insertado (espaciador Sp2.0) dentro de una ubicación CRISPR pero con la secuencia del espaciador Sp2.0 siendo diferente de tal espaciador Sp1.0. Típicamente, el espaciador Sp2.0 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago P, y da resistencia al fago P y fagos relacionados. Similarmente, en algunas modalidades, la variante A3.0 hasta la variante Ax.0 se generan a través de la exposición de la misma cepa A al mismo fago P. Todas las variantes "A" se seleccionan con objeto también de tener un espaciador adicional insertado (espaciador Sp3.0 hasta Spx.0) dentro de una ubicación CRISPR pero con la secuencia de todos los espaciadores "Sp" siendo diferentes unas de otras. Típicamente, los espaciadores "Sp" son fragmentos de aproximadamente 30 nucleótidos en tamaño del fago P, y todos dan resistencia al fago P y fagos relacionados.

Típicamente, puede estimarse que el nivel de resistencia será aproximadamente el de mutación sencilla que ocurre dentro del genoma de fago dentro de la secuencia que corresponde al espaciador (esto es, aproximadamente 10^{-4} hasta 10^{-6}). De esta manera, el fago que escapa de la resistencia mediada por CRISPR es fácil de aislar. El fago mutado se genera a través de la exposición de la variante A1.0 hasta el fago P. Típicamente, el fago de "escape CRISPR" mutado (P1.0) alberga al menos una mutación dentro de su genoma correspondiente con la secuencia del espaciador Sp1.0 (por ejemplo, eliminaciones, mutaciones de punto, etc.), o en algunas modalidades preferidas, la región que flanquea Sp1.0, más o menos 20 pb correspondientes a la porción CRISPR. La variante A1.0 debería ser sensible al fago P1.0. Similarmente, las variantes resistentes al fago P independientemente generadas (Variante A2.0, A3.0, hasta Ax.0) que alberga espaciadores únicos (Sp2.0, Sp3.0, hasta Spx.0, respectivamente) similarmente se enfrentan con el fago P para generar los fagos mutantes correspondientes (P2.0, P3.0, hasta Px.0, respectivamente). Posteriormente, un grupo de fago virulento mutante, cuyos genomas se han mutado específicamente a una secuencia anticipada para ser un espaciador CRISPR, puede generarse.

Ciertamente, el fago D2972 representa un fago de biocontrol completamente virulento contra la cepa DGCC7710 *S. thermophilus* (WT). En contraste, el análisis de la ubicación CRISPR de cepas WT_{phi2972}^{+S6}, WT_{phi2972}^{+S4}, WT_{phi2972}^{+S20}, WT_{phi2972}^{+S21}, y WT_{phi2972}^{+S22} relacionadas, indica la presencia de una secuencia espaciadora que es similar a las secuencias encontradas en el fago D2972 que indican que el fago D2972 tiene virulencia reducida sobre estas cepas. Los datos de la formación en placa (Ver, Tabla 7-1) confirman la virulencia reducida del fago D2972 en estas cepas. Con respecto a la cepa WT_{phi2972}^{+S6}, que se ha caracterizado para ser resistente al fago D2972 debido a la presencia de un espaciador CRISPR correspondiente, separación por exclusión de fagos relacionados con D2972 para la virulencia completamente incrementada que identifica los fagos D4724 y D4733 como agentes candidatos como agentes de biocontrol (Ver, Tabla 7-1).

En experimentos adicionales, la cepa DGCC7710 se expuso al fago D2972 para generar WT_{phi2972}^{+S6} de variante resistente. Cuando la cepa WT_{phi2972}^{+S6} se expuso al fago D2972, fue posible para aislar el fago mutante, tal como D4724. Este fago D4724 se encontró para ser completamente virulento sobre DGCC7710 y WT_{phi2972}^{+S6}. En una segunda iteración, WT_{phi2972}^{+S6} se expuso al fago D4724, para generar WT_{phi2972}^{+S6}_{phi4724}^{+S15} de variante resistente.

Durante la exposición de esta cepa a D4724, los fagos mutantes se identificaron tales como D4733, que son completamente virulentos hacia DGCC7710 y WT_{phi2972}^{+S6}. En algunas modalidades se usan iteraciones sucesivas para generar el fago con el nivel deseado de virulencia.

5 Los ejemplos de mutantes de fago adicionales se proporcionan en la Figura 13. En esta Figura, el fago mutante 858-A y 858-B derivado del fago precursor D858 se muestran. Las mutaciones corresponden al espaciador S1 de WTΦ858+S1S2 enfrentado con el fago D858.

10 Aún en ejemplos adicionales, los mutantes de fago completamente virulentos donde la mutación se identifica en la porción CRISPR se muestran en la Tabla 20-1. En esta Tabla, las secuencias de nucleótidos en fagos de tipo natural y mutante que corresponden a los espaciadores recientemente adquiridos por las cepas *S. thermophilus* se muestran. La porción AGAAW se resalta en gris. Cada mutación está en negrita y subrayada. *, indica una eliminación. Esta Tabla proporciona secuencias para la variante CRISPR resistente al fago y pares mutantes de fago virulento: DGCC7710_{Φ858}^{+S3} / fago 2972.S3C, DGCC7710_{Φ2972}^{+S4} / fago 2972.S4A o fago 2972.S4C, DGCC7710_{Φ2972}^{+S6} / fago 2972.S6A, y DGCC7710_{Φ2972}^{+S4}_{Φ858}^{+S32} / fago 858.S32A o fago 858.S32D. En esta Tabla, el nuevo espaciador corresponde a la SEQ ID NO:535 (DGCC7710_{Φ858}^{+S3}).

Tabla 20-1. Secuencias de nucleótidos en fagos de tipo natural y mutante que corresponden a espaciadores recientemente adquiridos por cepas <i>S. thermophilus</i>		
Fuente	Secuencia	SEQ ID NO:
DGCC7710 _{Φ858} ^{+S3}	TTACGTTTGAAAAGAATATCAAATCAATGA	SEQ ID NO:535
Fago 2972	TTACGTTTGAAAAGAATATCAAATCAATGACGAGAAAGA	SEQ ID NO:725
Fago 2972, S3A	TTACGTTTGAAAAGAATATCAAATTAATGACGAGAAAGA	SEQ ID NO:726
Fago 2972, S3B	TTACGTTTGAAAAGAATATCAAATCAACGACGAGAAAGA	SEQ ID NO:727
Fago 2972, S3C	TTACGTTTGAAAAGAATATCAAATCAATGACGAGAGAGA	SEQ ID NO:728
Fago 2972, S3D	TTACGTTTGAAAAGAATATCAAATCTATGACGAGAAAGA	SEQ ID NO:729

Tabla 20-1. Secuencias de nucleótidos en fagos de tipo natural y mutante que corresponden a espaciadores recientemente adquiridos por cepas <i>S. thermophilus</i>		
Fuente	Secuencia	SEQ ID NO:
Fago 2972, S3E	TTACGTTTGAAGAATATCAAATTCATGACGAGAAAGA	SEQ ID NO:730
Fago 2972, S3F	TTACGTTTGAAGAATATCAAATTAATGGCGAGAAAGA TTACGTTTGAAGAATATCAAATTAATGGCGAGAAAGA	SEQ ID NO:731
Fago 2972, S3G	TTACGTTTGAAGAACATCAAATTAATGACGAGAAAGA TTACGTTTGAAGAACATCAAATTAATGACGAGAAAGA	SEQ ID NO:732
DGCC7710 ₂₉₇₂ ^{+S4}	CTCAGTCGTTACTGGTGAACCAGTTTCAAT	SEQ ID NO:525
Fago 2972	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAGAAAAA	SEQ ID NO:733
Fago 2972, S4A	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAAAAAA	SEQ ID NO:734
Fago 2972, S4B	CTCAGTCGTTACTGGTGAACCAGTTTCGATTGAGAAAAA	SEQ ID NO:735
Fago 2972, S4C	CTCAGTCGTTACTGGTGAACCAGTTTCAATTGAGAGAAA	SEQ ID NO:736
Fago 2972, S4D	CTCAGTCGTTACTGGTGAACCGGTTTCAATTGAAAAAA CTCAGTCGTTACTGGTGAACCGGTTTCAATTGAAAAAA	SEQ ID NO:737
DGCC7710 ₂₉₇₂ ^{+S6}	GCCCTTCTAATTGGATTACCTCCGAGGTG	SEQ ID NO:524
Fago 2972	GCCCTTCTAATTGGATTACCTCCGAGGTGTTAGAATTC	SEQ ID NO:738
Fago 2972, S6A	GCCCTTCTAATTGGATTACCTCCGAGGTGTTAGAGTTC	SEQ ID NO:739
Fago 2972, S6B	GCCCTTCTAATTGGATTACCTCCGATGTGTTAGAATTC	SEQ ID NO:740
Fago 2972, S6C	GCCCTTCTAATTGGATTACCTCCGAGTTGTTAGAATTC	SEQ ID NO:741
Fago 2972, S6D	GCCCTTCTAATTGGATTACCTCCGA*GTGTTAGAATTC	SEQ ID NO:742
DGCC7710 ₂₉₇₂ ^{+S4} ₈₅₈ ^{+S32}	ATTGTCTATTACGACAACATGGAAGATGAT	SEQ ID NO:526
Fago 858	ATTGTCTATTACGACAACATGGAAGATGATGTAGAAATT	SEQ ID NO:743
Fago 858.S32A	ATTGTCTATTACGACAACATGGAAGATGATGTATAAATT	SEQ ID NO:744
Fago 858.S32B	ATTGTCTATTACGACAACATGGAAGATTATGTAGAAATT	SEQ ID NO:745
Fago 858.S32C	*****ATT	SEQ ID NO:746
Fago 858.S32D	ATTGTCTATTACGACAACATGGAAGATGATGTAAAAATT	SEQ ID NO:747
DGCC7710 ₂₉₇₂ ^{+S6} _{2972.S6B} ^{+S20}	TTATATCGAAGAACGACTGAAAGAGCTTGA	SEQ ID NO:706
Fago 2972	TTATATCGAAGAACGACTGAAAGAGCTTGAGAAGAAAAA	SEQ ID NO:748
Fago 2972.S20A	TTATATCGAAGAACGACTGAAAGAGCTTGAGAAATAAAAA	SEQ ID NO:749

EJEMPLO 22

Fago de "Escape de CRISPR" de Segundo Nivel

En este Ejemplo, se describen experimentos para la construcción de fagos de escape de CRISPR de segundo nivel (esto es, con mutaciones múltiples dirigidas a espaciadores múltiples).

A través de un proceso iterativo para crear variantes resistentes al fago mediadas por CRISPR seguido por aislamiento del fago mutado ("escape de CRISPR") capaz de vencer el mecanismo cas-CRISPR, es posible crear el fago que se ha "pre-adaptado" con mutaciones múltiples contra la resistencia mediada por CRISPR potencial.

En algunas modalidades, las variantes de segundo nivel se producen al aislar un fago mutado a través de la exposición de la variante A1.0 hasta el fago P. Típicamente, este fago mutado (fago P1.0) tiene una mutación (eliminación, mutación de punto, etc.) en su genoma dentro de la región que contiene la secuencia de espaciador Sp1.0 o dentro de la región que flanquea a Sp1.0, más o menos 20 pb correspondientes a la porción CRISPR. La variante A1.0 es sensible al fago P1.0. Luego, la variante A1.0 se expone al fago P1.0 y una variante resistente al fago (Variante A1.1) se selecciona (Ver, Figura 15). La variante A1.1 se selecciona también de manera que tiene un espaciador adicional insertado (espaciador Sp1.1) dentro de una ubicación CRISPR pero con la secuencia del espaciador Sp1.1 siendo diferente de la de los espaciadores Sp1.0, Sp2.0 hasta Spx.0. Típicamente, el espaciador Sp1.1 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago P1.0, y dará resistencia al fago P1.0 y fagos relacionados. La variante A1.1 es resistente al fago P1.0 y preferiblemente, tiene una resistencia incrementada para el fago P debido a la acumulación del espaciador Sp1.0 y Sp1.1.

En modalidades adicionales, un fago recientemente mutado (fago P1.1) se genera a través de la exposición de la variante A1.1 hasta el fago P1.0. Luego, durante la exposición de la variante A1.1 hasta el fago P1.1 una nueva variante A1.2 se obtiene que contenga un nuevo espaciador adicional (Sp1.2). Este espaciador da resistencia al fago P1.1 y preferiblemente incrementa la resistencia para el fago P1.0 y P (esto es, debido a la acumulación de espaciadores Sp1.0, Sp1.1, Sp1.2). El fago P1.1 se infecta completamente a la cepa precursora A, así como las variantes A1.0 y A1.1.

Aún en modalidades adicionales, los diferentes espaciadores (por ejemplo, 2, 3 ó 4) se acumulan iterativamente dentro de la cepa A a través de la variante A1, luego la variante A1.1, luego la variante A1.2, etc., para obtener una variante altamente resistente a los fagos (variante A1.n). Todavía en modalidades adicionales, los diferentes espaciadores adicionales se pueden acumular en la misma cepa a través de la variante A2, luego la variante A2.1, luego la variante A2.2, etc. para generar otra variante de la cepa A altamente resistente a los fagos (variante A2.n) en paralelo. La misma estrategia encuentra uso con las variantes A3.0 hasta Ax.0.

Siguiendo un proceso iterativo para crear variantes resistentes al fago CRISPR y el aislamiento del fago de "escape CRISPR" mutante (por ejemplo, la exposición de la variante A1.1 hasta el fago P1.1 crea la nueva variante A1.2 que contiene un nuevo espaciador adicional (Sp1.2) del cual un fago mutante se aísla (P1.2) que es completamente virulento en la variante A1.2, A1.1, A1.0 y cepa precursora A.

En algunas modalidades, las mutaciones combinatorias se acumulan por la construcción iterativa de variantes bacterianas que combinan diferentes espaciadores (por ejemplo, Sp2.0. Sp3.0 hasta Spx.0), la exposición al fago mutante de primer nivel correspondiente (P2.0, P3.0 hasta Px.0), y el aislamiento de los fagos mutantes de segundo nivel.

Un ejemplo de mutaciones combinatorias iterativas que crean variantes resistentes al fago CRISPR y fago de "escape CRISPR" mutante se muestra en la Tabla 22-1. Esta Tabla proporciona una lista de nuevos espaciadores encontrados en CRISPR1 y la región correspondiente en los fagos 2972, 858, o DT1. En esta Tabla, la "a" indica regiones ADN que son 100% idénticas entre los fagos 858 y 2972. La "posición 5'" se refiere a la posición 5' del espaciador proto en el genoma de fago. Los nucleótidos subrayados y sombreados en la secuencia espaciadora proto indican errores entre el fago y el espaciador. Un asterisco (*) indica una eliminación. En la "región flanqueada 3'" indica la secuencia flanqueada 3' en el genoma de fago. Los errores en la porción AGAAW se subrayan y somborean en gris. En la columna designada "Hebra/Módulo", los módulos de transcripción son "E" (genes expresados en estadios iniciales); "M" (genes expresados en estadios intermedios); y "L" (genes expresados en estadios tardíos).

Tabla 20-2. Lista de Nuevos espaciadores encontrados en CRISPR1 y la región correspondiente en los fagos 2972, 858, y DT1.

Nombre del espaciador	Fago	Posición 5'	Longitud del espaciador (pb)	Secuencia espaciadora proto	Región flanqueada 3'	Hebra/ módulo	Función <i>orf</i> en el genoma del fago usado en el enfrentamiento
S1	858	31378	30	CAACACATTCAACAGATTAAATGAAGAATAC (SEQ ID NO:680)	AAAGAAAAAA (SEQ ID NO:750)	(+)/E	orf40/primasa
S2	2972 *	25432	30	TCCACTCACGTACAAATAGTGAGGGTACTC (SEQ ID NO:686)	CTAAAAAGGAT (SEQ ID NO:751)	(-)/L	orf27/desconocido
S3	2972 *	17202	30	TTACGTTTGAAAAAGAAATATCAAAATCAATGA (SEQ ID NO:697)	CGAGAAAGAT (SEQ ID NO:752)	(+)/L	orf20/proteína de enlace al receptor
S4	2972	31582	30	CTCAGTCGTTACTGTTGTAACCAAGTTTCAAT (SEQ ID NO:525)	TCAGAAAAAA (SEQ ID NO:753)	(+)/E	orf38/primasa
S5	2972	22075	30	AGTTCTTTGTCAGACTCTAACACAGCCGC (SEQ ID NO:734)	TCAGAAAGTT (SEQ ID NO:755)	(+)/L	orf21/proteína de cola
S6	2972 *	34521	30	GCCCTTCTAAATGGATTACCTTCCGAGGTG (SEQ ID NO:524)	TTAGAAATCC (SEQ ID NO:756)	(-)/E	orf44/desconocido
S7	2972 *	10299	30	AAGCAAGTTGATATATTTCTCTTCTTTAT (SEQ ID NO:757)	TAAGAAAAACG (SEQ ID NO:758)	(-)/L	orf17/desconocido
S8	2972	30016	29	CGTTTCAAGTCATTGGTGGTTTGTCAGCG (SEQ ID NO:759)	AAAGAAATAA (SEQ ID NO:760)	(-)/E	orf37/replicación
S9	2972 *	7874	30	TTACTAGAGCGGTGCTGTTAACCACTTTAAA (SEQ ID NO:528)	TCAGAAATG (SEQ ID NO:761)	(+)/M	orf11/desconocido
S10	2972 *	20650	30	TTCGTTAAAGTCACCTCGTGTAGCGTTGC (SEQ ID NO:529)	ATAGAAAGTT (SEQ ID NO:772)	(-)/L	orf20/proteína de enlace al receptor
S11	2972 *	8360	30	ATAACGGTAGCAAAATATAAACCTGTTACTG (SEQ ID NO:530)	TCAGAAAGCTA (SEQ ID NO:773)	(+)/M	orf12/desconocido
S12	2972 *	18998	30	GAAGTAGCCATACAAGAAGATGGATCAGCA (SEQ ID NO:531)	CCAGAAATTG (SEQ ID NO:774)	(+)/L	orf20/proteína de enlace al receptor
S13	2972 *	33602	30	GATGCTAGTGTCTAAGCAATTGCGTAC (SEQ ID NO:776)	GAGGAAATCA (SEQ ID NO:775)	(+)/E	orf42/Enlace de ADN
S14	2972 *	4830	30	TGAATAAGCAGTTCTTGACGACCAACCGAC (SEQ ID NO:533)	ATAGAAAGT (SEQ ID NO:778)	(-)/M	orf6/proteína cápsida
S15	2972 *	34444	29	CAATTAACACAGCAATTAACACAGTATAT (SEQ ID NO:779)	ACAGAAATTG (SEQ ID NO:780)	(+)/E	orf44/desconocido
S16	2972 *	6799	30	ATGCCATTCTTTAAAGAGGCTTTACTCGTT (SEQ ID NO:781)	AAAGAAAAACG (SEQ ID NO:782)	(+)/M	orf9/proteína cápsida
S17	2972	30547	30	GTTGGCGGACTACTCTTCGAGGGTTGAT (SEQ ID NO:783)	CCAGAAATTA (SEQ ID NO:784)	(+)/E	orf37/replicación
S18	2972	30370	29	GAAGCACCTCTTCGCTTGATAAAAGTATT (SEQ ID NO:785)	GCAGAAATG (SEQ ID NO:786)	(+)/E	orf37/replicación
S19	2972	31709	29	ACATATCGACGTATCGTGATTATCCCAATT (SEQ ID NO:787)	CAAGAAAAACA (SEQ ID NO:788)	(+)/E	orf38/primasa

S20	2972 *	1113	30	TTATATCGAAGAACGACTGAAAGAGCTTGA (SEQ ID NO: 766)	GAAGAAAAA (SEQ ID NO: 789)	(+)/M	orf2/terminasa pequeña
S21	2972 *	19188	30	AAATCAACGTACATCCCGATATAGGCACGA (SEQ ID NO: 762)	TTAGAATCAG (SEQ ID NO: 763)	(-)/L	orf20/proteína de enlace al receptor
S22	2972	31708	30	GACATATCGACGTATCGTGATTTATCCCAT (SEQ ID NO: 764)	CAAGAAAAA (SEQ ID NO: 765)	(+)/E	orf38/primasa
S23	2972	26529	31	TGAAGTATTAGGTCTCTCAAAAGATGATATT (SEQ ID NO: 766)	GTAGAATACT (SEQ ID NO: 767)	(+)/E	orf31/represor tipo Cro
S24	2972	29923	30	AGTTGATTGCGTAATCAACCATCTCCATAA (SEQ ID NO: 769)	TTAGAATGGA (SEQ ID NO: 768)	(-)/E	orf37/replicación
S25	2972 *	441	30	GCAACACTCAACGTTGCAACGCAAGCTT (SEQ ID NO: 769)	CGAGAATATC (SEQ ID NO: 770)	(+)/E	orf11/desconocido
S26	2972	31606	31	CTCAGTCGTTACTGGTGAACCAAGTTTCAAT (SEQ ID NO: 771)	TGAGAAAAA (SEQ ID NO: 753)	(+)/E	orf38/primasa
S27	2972 *	27032	30	TTCATCGTCAATTTCCATGTTATAAATCT (SEQ ID NO: 790)	CTAGAACTG (SEQ ID NO: 791)	(-)/E	orf33/desconocido
S28	2972	26530	30	GAAGTATTAGGTCTCTCAAAAGATGATATT (SEQ ID NO: 715)	GTAGAATACT (SEQ ID NO: 792)	(+)/E	orf31/represor tipo Cro
S29	2972	32136	29	ATTGGCATGATTTCAAATTTTAAATGGGAT (SEQ ID NO: 793)	GTAGAAAAAG (SEQ ID NO: 794)	(+)/E	orf38/primasa
S30	2972 *	33968	30	TCCAAAGTTATTGAGGAGGTTATTAAGACAT (SEQ ID NO: 707)	GAAGAAATAT (SEQ ID NO: 795)	(+)/E	orf43/desconocido
S31	2972	30803	30	TACCGAAACGACTGGTTTGAAAAAATTCAG (SEQ ID NO: 708)	GAAGAAATC (SEQ ID NO: 796)	(+)/E	orf38/primasa
S32	2972 *	33044	30	ATTGTCTATTACGACAAACATGGAAGATGAT (SEQ ID NO: 526)	GTAGAAATTT (SEQ ID NO: 797)	(+)/E	orf41/desconocido
S33	858	30335	30	CTTCAATGTACTGCAAGGCTGCAAAAGTA (SEQ ID NO: 710)	CCAGAAATA (SEQ ID NO: 798)	(+)/E	orf38/desconocido

DGCC7710 se expuso al fago 2972 para crear variante DGCC7710_{Φ2972}^{+S6} resistente al fago CRISPR de la cual se generó el fago mutante 2972.S6B de escape de CRISPR. La exposición de DGCC7710_{Φ2972}^{+S6} al fago 2972.S6B creó la variante DGCC7710_{Φ2972}^{+S6}_{Φ2972.S6B}^{+S20} resistente al fago CRISPR de la cual se aisló el fago mutante 2972.S20A de escape de CRISPR.

En algunas modalidades, se proporcionan cepas que son resistentes a más de una familia de fagos. Como una cepa dada puede ser sensible a más de una familia de fagos, en algunas modalidades, se desea para agrandar la resistencia de la cepa a familias de fago múltiples al introducir espaciadores adicionales dentro de una ubicación CRISPR que se origina de las otras familias de fagos (Ver, Figura 16). Por ejemplo, los fagos P, Q, y R son fagos representativos de tres familias de fagos capaces de infectar la cepa A. Usando el método resumido arriba y en la presente, las variantes resistentes a las tres familias de fagos se producen. En algunas modalidades, el fago P se usa para generar la variante A1^P (que contiene el espaciador Sp1) que es resistente al fago P. Luego, la variante A1^P se expone al fago Q y una variante resistente al fago (Variante A1^{PQ}) se selecciona. La variante A1^{PQ} tiene un espaciador adicional (Sq1) insertado dentro de una ubicación CRISPR. Típicamente, el espaciador Sq1 es un

fragmento de aproximadamente 30 nucleótidos en tamaño del fago Q, y da resistencia al fago Q y fagos relacionados. La variante A1^{pq} es resistente a tanto fagos P como Q. Después, la variante A1^{pq} se expone al fago R y una variante resistente al fago (Variante A1^{pqr}) se selecciona. La variante A1^{pqr} tiene un tercer espaciador adicional (Sr1) insertado dentro de una ubicación CRISPR. Típicamente, Sr1 es un fragmento de aproximadamente 30 nucleótidos en tamaño del fago R, y también da resistencia al fago R y fagos relacionados. La variante A1^{pqr} es resistente a los tres fagos. En algunas modalidades particularmente preferidas, la variante también es resistente a los fagos relacionados.

Estos fagos de escape de CRISPR encuentran uso como fagos de biocontrol/terapéuticos. Como se describe arriba, a través del proceso para crear variantes resistentes al fago mediado por CRISPR, se genera la exposición al fago y aislamiento del fago de “escape de CRISPR” virulento, una mezcla de la especie de fago que alberga mutaciones sencillas y/o múltiples dirigidas contra secuencias del genoma de fago sencillas y/o múltiples que son objetivos del espaciador CRISPR potenciales. Como las bacterias hospedadoras objetivo pueden volverse resistentes al fago a través de la incorporación de espaciadores sencillos o múltiples y que el mecanismo Cas-CRISPR puede vencerse a través de una mutación dentro del genoma de fago correspondiente a tales espaciadores, el uso de una mezcla de fago alberga varias mutaciones que reducen la velocidad de una bacteria individual para adquirir exitosamente nuevos espaciadores y proliferan.

En una modalidad adicional, el análisis del protoespaciador y regiones flanqueadas, como se determina de los espaciadores en las variantes resistentes al fago CRISPR correspondientes, facilita la identificación de la porción CRISPR para un CRISPR específico. En el ejemplo de las variantes resistentes al fago DGCC7710 CRISPR 1 que contienen espaciadores S1-S33, se generaron después del enfrentamiento con el fago 2972 u 858. El alineado del protoespaciador y regiones flanqueadas, del genoma de los fagos 2972 u 858 que corresponde a los espaciadores S1-S33, usando el programa de software Clustal X, identificó la porción CRISPR 1 como NNAGAAW (SEQ ID NO:696), y se visualiza usando WebLogo (Figura 22).

En un ejemplo adicional, las variantes resistentes al fago CRISPR 3 se derivaron de DGCC7710 después del enfrentamiento con los fagos 858 y 3821, y LMD-9 después del enfrentamiento con el fago 4241. El alineado de los protoespaciadores y la región flanqueada de los genomas de fago respectivos con los espaciadores correspondientes de las variantes resistentes al fago CRISPR 3 respectivas, identificó la porción CRISPR 3 como NGGNG (SEQ ID NO:723) (Figura 23).

El análisis para la presencia de una porción CRISPR específica proporciona medios para identificar la ubicación de protoespaciadores supuestos dentro de un genoma u otra secuencia específica (por ejemplo, un plásmido u otro elemento genético móvil). En el ejemplo de los fagos secuenciados 858, 2972, y DT1, el análisis para la distribución de la porción AGAAW CRISPR 1 identificó la ubicación de protoespaciadores potenciales dentro de los genomas respectivos. Utilizando la degeneración del código genético y/o el uso de sustituciones de aminoácidos conservadoras, cada porción AGAAW se eliminó en el proceso de sintetizar químicamente un genoma como se describe para el fago ØX174, como se conocen en el arte. De esta manera, el fago se vuelve insensible al sistema de resistencia Cas-CRISPR 1. De esta manera, una molécula de ADN, que carece de porciones CRISPR específicas es insensible al sistema Cas-CRISPR correspondiente.

Estos fagos y “cocteles” de tipos de fagos múltiples encuentran uso en las estrategias de rotación (por ejemplo, administración secuencial definida de fago). Como una extensión para el uso de un coctel sencillo, compuesto de fago que alberga diferentes mutaciones espaciadoras, en algunas modalidades, se usan fagos virulentos múltiples, donde cada uno alberga una mutación espaciadora diferente en una manera secuencial definida. Por ejemplo, usando un conjunto de fagos de “escape de CRISPR” (P1.0, P2.0, y P3.0, o P1.0, P1.1, P1.2, o alguna combinación de los mismos), cada fago se aplica individualmente y en una secuencia y rotación definidas (P.10 > P2.0 > P3.0 > P1.0, P2.0 > etc) para minimizar la probabilidad de que las bacterias objetivo desarrollen resistencia mediada por CRISPR al fago. Similarmente, un conjunto de cocteles de fago (esto es, cada fago dentro del coctel así como cada coctel posee una combinación única de mutaciones) encuentra uso en secuencia y rotación. En algunas modalidades, el fago y/o coctel comprende una familia de fago sencilla, mientras que en otras modalidades, el fago y/o coctel comprende familias de fago múltiples.

EJEMPLO 23

Combinaciones funcionales

Este Ejemplo proporciona varias combinaciones funcionales que encuentran uso en la presente invención. A manera de ejemplo solamente, las siguientes combinaciones funcionales se pueden usar de acuerdo con la presente invención.

Combinación funcional #1:

Secuencias cas: SEQ ID NO:461 hasta SEQ ID NO:465 y SEQ ID NO:473 hasta SEQ ID NO:477 (todas las cuales son secuencias *S. thermophilus*), como se establecen a continuación:

SEQ ID NO:461:

ATGAGTGACTTAGTTTTAGGACTTGATATCGGTATAGGTTCTGTTGGTGTAGGTATCCTTAAC
 AAAGTGACAGGAGAAATTATCCATAAAAACTCACGCATCTTCCCAGCAGCTCAAGCAGAAA
 ATAACCTAGTACGTAGAACGAATCGTCAAGGAAGACGCTTGACACGACGTAAAAAACATCG
 TATAGTTCGTTTTAAATCGTCTATTTGAGGAAAGTGGATTAATCACCGATTTTACGAAGATTT
 CAATTAATCTTAACCCATATCAATTACGAGTTAAGGGCTTGACCGATGAATTGTCTAATGAA
 GAACTGTTTATCGCTCTTAAAAATATGGTGAAACACCGTGGGATTAGTTACCTCGATGATGC
 TAGTGATGACGGAAATTCATCAGTAGGAGACTATGCACAAATTGTAAAGGAAAAATAGTAAA
 CAATTAGAACTAAGACACCGGGACAGATACAGTTGGAACGCTACCAAAACATATGGTCAAT
 TACGTGGTGAATTTTACTGTTGAGAAAGATGGCAAAAAACATCGCTTGATTAATGTCTTTCCA
 ACATCAGCTTATCGTTTCAAGGCCTTAAGGATACTGCAAACTCAACAAGAATTTAATCCACA
 GATTACAGATGAATTTATTAATCGTTATCTCGAAATTTTAACTGGAAAACGGAAATATTATC
 ATGGACCCGGAAATGAAAAGTACGCGACTGATTATGGTCGTTACAGAACGAGTGGAGAAAC
 TTTAGACAATATTTTGGAAATCTAATTGGGAAATGTACATTTTATCCAGAAGAGTTTAGAG
 CAGCAAAAGCTTCTACACGGCTCAAGAATTCAATTTGCTAAATGATTTGAACAATCTAACA
 GTTCTACTGAAACCAAAAAAGTTGAGCAAAAGAACAGAAGAATCAAATCATTAAATTATGTCA
 AAAATGAAAAGGCAATGGGGCCAGCGAAACTTTTTAAATATATCGCTAAGTTACTTTCTTGT
 GATGTTGCAGATATCAAGGGATACCGTATCGACAAATCAGGTAAGGCTGAGATTCATACTTT
 CGAAGCCTATCGAAAAATGAAAACGCTTGAAACCTTAGATATTGAACAAATGGATAGAGAA
 ACGCTTGATAAATTAGCCTATGTCTTAACATTAACACTGAGAGGGAAGGTATTCAAGAAG
 CCTTAGAACATGAATTTGCTGATGGTAGCTTTAGCCAGAAGCAAGTTGACGAATTTGGTTCAA
 TTCCGCAAAGCAAATAGTTCCATTTTTGGAAAAGGATGGCATAATTTTTCTGTCAAACGTGAT
 GATGGAGTTAATTCAGAAATTGTATGAGACGTCAGAAGAGCAAATGACTATCCTGACACGA
 CTTGGAAAACAAAAACGACTTCGTCTTCAAATAAAACAAAATATTTCAAATAAAACAAAAT
 ATATAGATGAGAACTATTAACCTGAAGAAATCTATAATCCTGTTGTTGCTAAGTCTGTTTCGC
 CAGGCTATAAAAAATCGTAAATGCGGCGATTAAAGAATACGGAGACTTTGACAATATTGTCA
 TCGAATGGCTCGTGAAACAAATGAAGATGATGAAAAGAAAGCTATTCAAAAGATTCAAAA
 AGCCAACAAAGATGAAAAAGATGACAGCAATGCTTAAGGCTGCTAACCAATATAATGGAAA
 GGCTGAATTACCACATAGTGTTTTCCACGGTCATAAGCAATTAGCGACTAAAATCCGCCTTT
 GGCATCAGCAAGGAGAACGTTGCCTTTATACTGGTAAGACAATCTCAATCCATGATTTGATA
 AATAATCCTAATCAGTTTGAAGTAGATCATATTTTACCTCTTTCTATCACATTCGATGATAGC
 CTTGCAAATAAGGTTTTGGTTTATGCAACTGCTAACCAAGAAAAAAGGACAACGAACACCTT
 ATCAGGCTTTAGATAGTATGGATGATGCGTGGTCTTTCCGTGAATTAAGGCTTTTGTACGT
 GAGTCAAAAACACTTTCAAACAAGAAAAAAGAATACCTCCTTACAGAAGAAAGATATTTCAA
 AGTTTGATGTTTCGAAAGAAATTTATTGAACGAAATCTTGATAGATACAAGATACGCTTCAAGA
 GTTGTCTCAATGCCCTTCAAGAACACTTTAGAGCTCACAAGATTGATACAAAAGTTTCCGT
 GGTTTCGTGGCCAATTTACATCTCAATTGAGACGCCATTGGGGAATTGAGAAGACTCGTGATA
 CTTATCATCACCATGCTGTGATGCATTGATTATTGCCGCCTCAAGTCAGTTGAATTTGTGGA
 AAAAAACAAAAGAATACCCTTGTAAGTTATTCAGAAGAACAACTCCTTGATATTGAAACAGG
 TGAACCTTATTAGTGATGATGAGTACAAGGAATCTGTGTTCAAAGCCCCTTATCAACATTTTG
 TTGATACATTGAAGAGTAAAGAATTTGAAGACAGTATCTTATTCTCATATCAAGTGGATTCT
 AAGTTTAATCGTAAAATATCAGATGCCACTATTTATGCGACAAGACAGGCTAAAGTGGGAA
 AAGATAAGAAGGATGAACTTATGTCTTAGGGGAAAATCAAAGATATCTATACTCAGGATGG
 TTATGATGCCTTTATGAAGATTTATAAGAAGGATAAGTCAAAATTCCTCATGTATCGTCACG
 ACCACAAACCTTTGAGAAAGTTATCGAGCCAAATTTAGAGAACTATCCTAATAAGCAAAT
 GAATGAAAAAGGAAAAGAGGTACCATGTAATCCTTTTCTTAAATATAAAGAAGAACATGGC
 TATATTCGTAAATATAGTAAAAAAGGCAATGGTCCTGAAATCAAGAGTCTTAAATACTATG
 ATAGTAAGCTTTTAGGTAATCCTATTGATATTACTCCAGAGAATAGTAAAAATAAAGTTGTC
 TTACAGTCATTAAAACCTTGGAGAACAGATGTCTATTTCAATAAGGCTACTGGAAAATACGA
 AATCCTTGGATTAAAATATGCTGATCTACAATTTGAGAAAGGGACAGGAACATATAAGATT
 TCCCAGGAAAAATACAATGACATTAAGAAAAAAGAGGGTGTAGATTCTGATTGAGAATTCA
 AGTTTACACTTTATAAAAAATGATTTGTTACTCGTTAAAGATACAGAAACAAAAGAACAACA
 GCTTTTCCGTTTTCTCTCGAACTTTACCTAAACAAAGCATTATGTTGAATTTAAACCTTA
 TGATAAACAGAAATTTGAAGGAGGTGAGGCGTTAATTAAAGTGTGGGTAACGTTGCTAAT
 GGTGGTCAATGCATAAAAGGACTAGCAAAATCAAATATTTCTATTTATAAAGTAAGAACAG

ATGTCCTAGGAAATCAGCATATCATCAAAAATGAGGGTGATAAGCCTAAGCTAGATTTTTTA
 ATATTAATTGTTAGAAAAGTGTTGCAATTATAGTTATCATATGCTATAATAATCGTGTAAGGG
 ACGCCTTACACAGTTACTTAAATCTTGCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATC
 AACACCCTGTCATTTTATGGCAGGGTGTTTTCGTTATTTAAAGAGGAGAAGAAATGACTTGG
 AGAGTTGTACATGTCAGTCAAAGTGAGAAGATGCGCTTAAAGCTTGATAACTTATTAGTGCA
 AAAAATGGGACAAGAGTTTACGGTGCCACTAAGTGATATTTGATAATCGTTGCAGAAGGT
 GGGGATACAGTTGTTACCCCTTCGTCTATTAAGTGCCTTAAAGTAAATATAATATTGCCTTGGTC
 GTTTGTGATAACGAACATTTACCAACAGGAATTTATCACTCACAAAATGGGCACTTTAGAGC
 GTACAAGCGCTTGAAAGAACAGCTGGATTGGTCTCAGAAACAAAAGGACAAGGCATGGCA
 GATTGTAACCTTATTATAAAATCAATAACCAAGAGGATGTTCTAGCCATGTTTGAAAAAAGTC
 TGGACAACATTAGATTACTTTCAGACTATAAAGAGCAGATAGAACCTGGTGATAGAACGAA
 TAGAGAGGGACATGCTGCCAAGGTCTACTTTAATGAGCTCTTTGGTAAACAATTTGTCAGAG
 TAACTCAGCAAGAAGCTGATGTCATCAATGCTGGTTTAACTATGGCTATGCTATCATGAGG
 GCTCAGATGGCTAGAATAGTGCGGGGTTATGGTTTAAATGGCCTATTAGGAATCTTCCATAA
 AAATGAATACAATCAGTTTAAATTTGGTTGACGATTTGATGGAGCCATTTAGACAGATTGTAG
 ATGTTTGGGTATATGATAATCTACGAGATCAGGAATTCCTTAAAGTATGAGTATAGGTTGGGA
 TTGACAGATTACTCAATGCTAAAATCAAATAACCAATATGGCAAAGAGACTTGCTCAGTGACAGTG
 TATGGACAAATATGTCAAAGGCTTTATCAAATATATTTTCGGAAAAAGATAGTAGTAAATTTT
 ACTGCCCAGTGGTATCAAGTTTAGAGTGGAGAAAAATAAGATGAGGTATGAAGCATTGAGAT
 TATTATGTTTTTTTGGATTACCAATGGAATCCAAGGATGAAAAAAGAATATATCGTAATTTT
 CGTAAAGAATTAATTTCAAATGGGTTTGAAATGTTACAATTTTCGGTCTACTATCGCAGCTTGT
 CCTAATAGAAGCTTTGCAAATAAATTTTATAAGAAGTTAAAGATTAGCAATCTTCTGCTGG
 GAATGTGAGACTTTTGGCAGTTACTGAAAAACAATTTTCAGAGATGACATTAATTATAGGTG
 GTAAACTAAGCAAGAAGAAATCGTCAGTGATAATAAGTTGGTGGTTATATGAAATATTTT
 GTACAACATCCTTACAAAGAACGTATTGAATTAATATTGGTGCAATCACACAAATTTGTTGG
 TCAGAAATAAGAACTCAAATATTATATTGGCAAATTTTGAGCTGGTATTTTGGCGGAAAAA
 AATACTCAAGTGAGGACTTAAGTATTTTGGATTATGAGGAACCTACTATACTTGATGAGTCT
 GGAGAAATAGTGAAGCGAAGTAGCTATCACTATATCGACATTTCAAGTTTAAAGGATTTACT
 GGAGCAGATGGAATACAAGAAAGGAACACTTGCTCAGGGTTACCTTAGTAAAATTCTCAAT
 CAGGTTGATATTGTAGGCCATTTGGAGAAAAATTAATGAACAAGTAGAGCTTATAGAAGGAG
 CAATGAATCAGCATATAAACTTAACTGTGGTCAGGTGGAGTACCATTTGGAGAATCACCCCT
 CTAACACTAGACCAATTACTTTCAAAAAATTTTAGTCCCTTTTGTCTATCGAGAATAAGAA
 TTTATCTTTTGAATGGGTTTCAAATACTGATAAACTTTCTCTCTTTCTAGAAATGTTAGACCG
 CCTTCTGTCAACAAACAGAGAAGTATCTCATTGTGCTAAAAAATATTGATGGCTTTATCT
 CAGAAGAATCTTATACTATTTTTTATAGGCAAATCTGTCATCTGGTCAAGAAGTATCCAAAT
 CTAACCTTTATTTGTTTCTAGTGACCAAGGCTATTTAAAAATTGATGAAGAAAATAGTAG
 GTTCGTCAATATTTTATCTGACCAGGTGGAGCATTTGTATGATGTTGAGTTTATGTATGAAA
 GAGTAATGAAATATTATCCAAGTAATGATTTCCGACGAGAGAAGGTTTATAGGATGTCTTTA
 GAACTGTGACACCTTATTTATTGACAAAAATGCTGAGACAACCTAGTCTCTCACTTGTGTA
 TTCAGTAATATTGAATATCCTAAATCAGTTGTTTCATTTAGTTACCGTATAAGATATTCTCA
 GACACCTGATAAGGAACTATTACATAAATTTTAGAAAGTAAGGATTGA (SEQ ID NO:461)

SEQ ID NO:462:

ATGAGTGACTTAGTTTTAGGACTTGATATCGGTATAGGTTCTGTTGGTGTAGGTATCCTTAAC
 AAAGTGACAGGAGAAATTATCCATAAAAACTCACGCATCTTCCAGCAGCTCAAGCAGAAA
 ATAACCTAGTACGTAGAACGAATCGTCAAGGAAGACGCTTGACACGACGTAAAAAACATCG
 TATAGTTGTTTTAAATCGTCTATTTGAGGAAAGTGATTAAATCACCGATTTTACGAAGATTT
 CAATTAATCTTAACCCATATCAATTACGAGTTAAGGGCTTGACCGATGAATTGTCTAATGAA
 GAACTGTTTATCGCTCTTAAAAATATGGTGAAACACCGTGGGATTAGTTACCTCGATGATGC
 TAGTGATGACGGAAATTCATCAGTAGGAGACTATGCACAAATTGTTAAGGAAAAATAGTAAA
 CAATTAGAACTAAGACACCGGGACAGATACAGTTGGAACGCTACCAAACATATGGTCAAT
 TACGTGGTGATTTTACTGTTGAGAAAGATGGCAAAAAACATCGCTTGATTAATGTCTTTCCA
 ACATCAGCTTATCGTTTCAGAAAGCCTTAAGGATACTGCAAACTCAACAAGAATTTAATCCACA
 GATTACAGATGAATTTATTAATCGTTATCTCGAAATTTTAACTGGAAAACGGAAATATTATC
 ATGGACCCGGAAATGAAAAGTCACGGACTGATTATGGTCGTTACAGAACGAGTGGAGAAAC

TTTAGACAATATTTTTGGAATTCTAATTGGGAAATGTACATTTTATCCAGAAGAGTTTAGAG
 CAGCAAAAGCTTCCTACACGGCTCAAGAATTCAATTTGCTAAATGATTTGAACAATCTAACA
 GTTCCTACTGAAACCAAAAAGTTGAGCAAAGAACAGAAGAATCAAATCATTAAATTATGTCA
 AAAATGAAAAGGCAATGGGGCCAGCGAAACTTTTTAAATATATCGCTAAGTTACTTTCTTGT
 GATGTTGCAGATATCAAGGGATACCGTATCGACAAATCAGGTAAGGCTGAGATTCTACTTT
 CGAAGCCTATCGAAAAATGAAAACGCTTGAAACCTTAGATATTGAACAAATGGATAGAGAA
 ACGCTTGATAAATTAGCCTATGTCTTAACATTAACACTGAGAGGGAAGGTATTCAAGAAG
 CCTTAGAACATGAATTTGCTGATGGTAGCTTTAGCCAGAAGCAAGTTGACGAATTGGTTCAA
 TTCCGCAAAGCAAATAGTTCCATTTTTGGAAAAGGATGGCATAATTTTTCTGTCAAACCTGAT
 GATGGAGTTAATTCAGAAATTGTATGAGACGTCAGAAGAGCAAATGACTATCCTGACACGA
 CTTGGAAAACAAAAACGACTTCGTCTTCAAATAAAAACAAAATATTTCAAATAAAAACAAAAT
 ATATAGATGAGAACTATTAACCTGAAGAAATCTATAATCCTGTTGTTGCTAAGTCTGTTCCG
 CAGGCTATAAAAATCGTAAATGCGGCGATTAAAGAATACGGAGACTTTGACAAATATTGTCA
 TCGAAATGGCTCGTGAAACAAATGAAGATGATGAAAAGAAAGCTATTCAAAAGATTCAAAA
 AGCCAACAAAGATGAAAAAGATGCAGCAATGCTTAAGGCTGCTAACCAATATAATGGAAA
 GGCTGAATTACCACATAGTGTTTTCCACGGTCATAAGCAATTAGCGACTAAAAATCCGCTTT
 GGCATCAGCAAGGAGAACGTTGCCTTTATACTGGTAAGACAATCTCAATCCATGATTTGATA
 AATAATCCTAATCAGTTTGAAGTAGATCATATTTTACCTCTTTCTATCACATTCGATGATAGC
 CTTGCAAATAAGGTTTTGGTTTTATGCAACTGCTAACCAAGAAAAAGGACAACGAACACCTT
 ATCAGGCTTTAGATAGTATGGATGATGCGTGGTCTTTCCGTGAATTAAGGCTTTTGTACGT
 GAGTCAAAAACACTTTCAAACAAGAAAAAGAATACCTCCTTACAGAAGAAGATATTTCAA
 AGTTTGATGTTGAAAGAAATTTATTGAACGAAATCTTGTAGATACAAGATACGCTTCAAGA
 GTTGTCTCAATGCCCTTCAAGAACCTTTAGAGCTCACAAGATTGATACAAAAGTTTCCGT
 GGTTCGTGGCCAATTTACATCTCAATTGAGACGCCATTGGGGAATTGAGAAGACTCGTGATA
 CTTATCATCACCATGCTGTTCGATGCATTGATTATTGCCGCCTCAAGTCAGTTGAATTTGTGGA
 AAAAAACAAAAGAATACCTTTGTAAGTTATTCAGAAGAACAACCTCCTTGATATTGAAACAGG
 TGAACCTTATTAGTGATGATGAGTACAAGGAATCTGTGTTCAAAGCCCCCTTATCAACATTTTG
 TTGATACATTGAAGAGTAAAGAATTTGAAGACAGTATCTTATTCTCATATCAAGTGGATTCT
 AAGTTTAATCGTAAAAATATCAGATGCCACTATTTATGCGACAAGACAGGCTAAAGTGGGAA
 AAGATAAGAAGGATGAAACTTATGTCTTAGGGAAAAATCAAAGATATCTATACTCAGGATGG
 TTATGATGCCTTTATGAAGATTTATAAGAAGGATAAGTCAAAATTCCTCATGTATCGTCACG
 ACCCACAAACCTTTGAGAAAGTTATCGAGCCAATTTTAGAGAACTATCCTAATAAGCAAAT
 GAATGAAAAAGGAAAAAGAGGTACCATGTAATCCTTTCTAAAATATAAAGAAGAACATGGC
 TATATTCGTAAATATAGTAAAAAAGGCAATGGTCCTGAAATCAAGAGTCTTAAATACTATG
 ATAGTAAGCTTTTAGGTAATCCTATTGATATTACTCCAGAGAATAGTAAAAATAAAGTTGTC
 TTACAGTCATTAACCTTTGGAGAACAGATGTCTATTTCAATAAGGCTACTGGAAAAATACGA
 AATCCTTGGATTAAAAATATGCTGATCTACAATTTGAGAAAGGGACAGGAACATATAAGATT
 TCCCAGGAAAAATACAATGACATTAAGAAAAAAGAGGGGTGTAGATTCTGATTTCAGAATTCA
 AGTTTACACTTTATAAAAAATGATTTGTTACTCGTTAAAGATACAGAAACAAAAGAACAACA
 GCTTTTCCGTTTTCTTCTCGAACTTTACCTAAACAAAAGCATTATGTTGAATTAACCTTA
 TGATAAACAGAAATTTGAAGGAGGTGAGGCGTTAATTAAGTGTTGGGTAACGTTGCTAAT
 GGTGGTCAATGCATAAAAGGACTAGCAAAATCAAATATTTCTATTTATAAAGTAAGAACAG
 ATGTCCTAGGAAATCAGCATATCATCAAAAATGAGGGTGATAAGCCTAAGCTAGATTTTTTA
 A (SEQ ID NO:462)

SEQ ID NO:463:

ATGACTTGGAGAGTTGTACATGTCAGTCAAAGTGAGAAGATGCGCTTAAAGCTTGATAACTT
 ATTAGTGCAAAAAATGGGACAAGAGTTTACGGTGCCACTAAGTGATATTTGATAATCGTTG
 CAGAAGGTGGGGATACAGTTGTTACCTTCGTCTATTAAGTGCCTTAAGTAAATATAATATT
 GCCTTGGTCGTTTGTGATAACGAACATTTACCAACAGGAATTTATCACTCACAAAATGGGCA
 CTTTAGAGCGTACAAGCGCTTGAAAGAACAAGTGGATTGGTCTCAGAAACAAAAGGACAAG
 GCATGGCAGATTGTAACCTATTATAAAATCAATAACCAAGAGGATGTTCTAGCCATTGTTGA
 AAAAAAGTCTGGACAACATTAGATTACTTTAGACTATAAAGAGCAGATAGAACCTGGTGAT
 AGAACGAATAGAGAGGGACATGCTGCCAAGGTCTACTTTAATGAGCTCTTTGGTAAACAAT
 TTGTCAGAGTAACTCAGCAAGAAGCTGATGTCATCAATGCTGGTTTAACTATGGCTATGCT

ATCATGAGGGCTCAGATGGCTAGAAATAGTGGCGGGTTATGGTTTAAATGGCCTATTAGGAA
TCTTCCATAAAAAATGAATACAATCAGTTTAAATTTGGTTGACGATTTGATGGAGCCATTTAGA
CAGATTGTAGATGTTTGGGTATATGATAATCTACGAGATCAGGAATTCCTTAAGTATGAGTA
TAGGTTGGGATTGACAGATTTACTCAATGCTAAAATCAAATATGGCAAAGAGACTTGCTCA
GTGACAGTTGCTATGGACAAATATGTCAAAGGCTTTATCAAATATATTTTCGGAAAAAGATA
GTAGTAAATTTCACTGCCCAGTGGTATCAAGTTTAGAGTGGAGAAAAATAA (SEQ ID NO:463)

SEQ ID NO:464:

ATGAGGTATGAAGCATTGAGATTATTATGTTTTTTTGGATTACCAATGGAATCCAAGGATGA
AAAAAGAATATATCGTAATTTTCGTAAAGAATTAATTTCAAATGGGTTTGAAATGTTACAAT
TTTCGGTCTACTATCGCACTTGTCTAATAGAAGCTTTGCAAATAAATTTTATAAGAAGTTA
AAGATTAGCAATCTTCCTGCTGGGAATGTGAGACTTTTGGCAGTTACTGAAAAACAATTTTC
AGAGATGACATTAATTATAGGTGGTAAACTAAGCAAGAAGAAATCGTCAGTGATAATAAG
TTGGTGGTTATATGA (SEQ ID NO:464)

SEQ ID NO:465:

ATGAAATATTTTGTACAACATCCTTACAAAGAACGTATTGAATTAATATTGGTGCAATCAC
ACAAATTGTTGGTCAGAATAAAGAACTCAAATATTATTTGGCAAATTTTGAGCTGGTATT
TTGGCGGAAAAAAATACTCAAGTGAGGACTTAAGTATTTTGGATTATGAGGAACCTACTATA
CTTGATGAGTCTGGAGAAATAGTGAAGCGAAGTAGCTATCACTATATCGACATTTCAAGTTT
TAAGGATTTACTGGAGCAGATGGAATACAAGAAAGGAACACTTGCTCAGGGTTACCTTAGT
AAAATTCTCAATCAGGTTGATATTGTAGGCCATTTGGAGAAAATTAATGAACAAGTAGAGC
TTATAGAAGGAGCAATGAATCAGCATATAAACTTAACTGTGGTCAGGTGGAGTACCATT
GGAGAATCACCTCTAACACTAGACCAATTACTTTCAAAAAATTTTAGTCCCTTTTTTGCTAT
CGAGAATAAGAATTTATCTTTTGAATGGGTTTCAAATACTGATAAACTTTCTCTCTTTCTAGA
AATGTTAGACCGCCTTCTGTCAACAAACAGAGAAGTATCTCATTGTGCTAAAAAATATTG
ATGGCTTTATCTCAGAAGAATCTTATACTATTTTATAGGCAAATCTGTCATCTGGTCAAGA
AGTATCCAAATCTAACCTTTATTTTGTTCCTAGTGACCAAGGCTATTTAAAAATTGATGAA
GAAATAGTAGGTTCTGTAATTTTATCTGACCAAGGTGGAGCATTGTGATGATGTTGAGTT
TATGTATGAAAGAGTAATGAAATATTATCCAAGTAATGATTTTCCGACGAGAGAAGGTTTAA
GGATGTCTTTAGAACTGTGACACCTTATTTATTGACAAAAATGCTGAGACAACCTAGTCTC
TCACTTGTGATTACAGTAATATTGAATATCCTAAATCAGTTGTTTCATTTTAGTTACCGTATA
AGATATTCTCAGACACCTGATAAGGAACTATTACATAAATTTTAGAAAGTAAGGATTGA
(SEQ ID NO:465)

SEQ ID NO:473:

ATGAGTGACTTAGTTTTAGGACTTGATATCGGTATAGGTTCTGTTGGTGTAGGTATCCTTAAC
AAAGTGACAGGAGAAATTATCCATAAAAACTCACGCATCTTCCCAGCAGCTCAAGCAGAAA
ATAACCTAGTACGTAGAACGAATCGTCAAGGAAGACGCTTGACACGACGTAAAAAACATCG
TATAGTTCTGTTTAAATCGTCTATTTGAGGAAAGTGGATTAATCACCGATTTTACGAAGATTT
CAATTAATCTTAACCCATATCAATTACGAGTTAAGGGCTTGACCGATGAATTGTCTAATGAA
GAACTGTTTATCGCTCTTAAAAATATGGTGAAACACCGTGGGATTAGTTACCTCGATGATGC
TAGTGATGACGGAATTCATCAGTAGGAGACTATGCACAAATTGTTAAGGAAAAATAGTAAA
CAATTAGAACTAAGACACCGGGACAGATACAGTTGGAACGCTACCAACATATGGTCAAT
TACGTGGTGAATTTACTGTTGAGAAAGATGGCAAAAAACATCGCTTGATTAATGTCTTTCCA
ACATCAGCTTATCGTTTCAAGGCCTTAAGGATACTGCAAACTCAACAAGAATTTAATTCACA
GATTACAGATGAATTTATTAATCGTTATCTCGAAATTTAACTGGAAAACGGAAATATTATC
ATGGACCCGGAAATGAAAAGTCAAGGACTGATTATGGTCGTTACAGAACGAATGGAGAAAC
TTTAGACAATATTTTTGGAATTCTAATTGGGAAATGTACATTTTATCCAGACGAGTTAGAG
CAGCAAAAGCTTCTACACGGCTCAAGAATTCAATTTGCTAAATGATTTGAACAATCTAACA
GTTCTACTGAAACCAAAAAAGTTGAGCAAAGAAGCAAGAATCAAATCATTAAATTATGTCA
AAAATGAAAAGGTAATGGGGCCAGCGAACTTTTAAATATATCGCTAAATTACTTTCTTGT
GATGTTGCAGATATCAAGGGACACCGTATCGACAAATCAGGTAAGGCTGAGATTCACTTT
TCGAAGCCTATCGAAAAATGAAAACGCTTGAAACCTTAGATATTGAGCAAATGGATAGAGA

AACGCTTGATAAATTAGCCTATGTCTTAACATTAAACACTGAGAGGGAAGGTATTCAAGAA
GCTTTAGAACATGAATTTGCTGATGGTAGCTTTAGCCAGAAGCAAGTTGACGAATTGGTTCA
ATTCCGCAAAGCAAATAGTTCCATTTTTGGAAAAGGATGGCATAATTTTTCTGTCAAACCTGA
TGATGGAGTTAATTCCAGAATTGTATGAGACGTCAGAAGAGCAAATGACTATCCTGACACG
ACTTGGAACAAAAACAACCTTCGTCTTCAAATAAAAACAAAATATATAGATGAGAACTA
TTAACTGAAGAAATCTATAATCCTGTTGTTGCTAAGTCTGTTCCGCCAGGCTATAAAAAATCGT
AAATGCGGCGATTAAAGAATACGGAGACTTTGACAATATTGTCATCGAAATGGCTCGTGAA
ACAAATGAAGATGATGAAAAGAAAGCTATTCAAAAAGATTCAAAAAGCCAACAAAGATGAA
AAAGATGCAGCAATGCTTAAGGCTGCTAACCAATATAATGGAAAGGCTGAATTACCACATA
GTGTTTTCCACGGTCATAAGCAATTAGCGACTAAAATCCGCCTTTGGCATCAGCAAGGAGAA
CGTTGCCTTTATACTGGTAAGACAATCTCAATCCATGATTTGATAAATAATCCTAATCAGTTT
GAAGTAGATCATATTTTACCTCTTTCTATCACATTCGATGATAGCCTTGCAAATAAGGTTTTG
GTTTATGCAACTGCTAACCAAGAAAAAGGACAACGAACACCTTATCAGGCTTTAGATAGTA
TGGATGATGCGTGGTCTTTCCGTGAATTAAGGCTTTTGTACGTGAGTCAAAAACACTTTCA
AACAAAGAAAAAGAATACCTCCTTACAGAAGAAGATATTTCAAAGTTTGATGTTGAAAGA
AATTTATTGAACGAAATCCTGTAGATACAAGATACGCTTCAAGAGTTGTCCTCAATGCCCTT
CAAGAACACTTTAGAGCTCACAAGATTGATACAAAAGTTTCCGTGGTTCGTGGCCAAATTTAC
ATCTCAATTGAGACGCCATTGGGGAATTGAGAAGACTCGTGATACTTATCATCACCATGCTG
TCGATGCATTGATTATTGCCGCCTCAAGTCAGTTGAATTTGTGGAAAAACAAAAGAATACC
CTTGTAAGTTATTGAGAAGAACAACTCCTTGATTTGAAACAGGTGAACATTATTAGTGATGA
TGAGTACAAGGAATCTGTGTTCAAAGCCCTTATCAACATTTTGTGATACATTGAAGAGTA
AAGAATTTGAAGACAGTATCTTATTCTCATATCAAGTGGATTCTAAGTTTAATCGTAAAATA
TCAGATGCCACTATTTATGCGACAAGACAGGCTAAAGTGGGAAAAGATAAGAAGGATGAA
ACTTATGTCTTAGGGAAAATCAAAGATATCTATACTCAGGATGGTTATGATGCCTTTATGAA
GATTTATAAGAAGGATAAGTCAAAATTCCTCATGTATCGTCACGACCCACAAACCTTTGAGA
AAGTTATCGAGCCAATTTTAGAGAACTATCCTAATAAGGAAATGAATGAAAAAGGGAAAGA
AGTACCATGTAATCCTTTCTTAAATATAAAGAAGAACATGGCTATATTCGTAAATATAGTA
AAAAAGGCAATGGTCTGAAATCAAGAGTCTTAAATACTATGATAGTAAGCTTTTAGGTAA
TCCTATTGATATTACTCCAGAGAATAGTAAAAATAAAGTTGTCTTACAGTCATTAAACCTT
GGAGAACAGATGTCTATTTCAATAAAAATACTGGTAAATATGAAATTTTAGGACTGAAATA
TGCTGATTTACAATTTGAAAAGAAGACAGGAACATATAAGATTTCCAGGAAAAATACAAT
GGCATTATGAAAGAAGAGGGGTGTAGATTCTGATTGAGAATTCAAGTTTACACTTTATAAAAA
TGATTTGTTACTCGTTAAAGATACAGAAACAAAAGAACAACAGCTTTTCCGTTTTCTTTCTC
GAACATATGCCTAATGTGAAATATTATGTAGAGTTAAAGCCTTATTCAAAAAGATAAATTTGAG
AAGAATGAGTCACTTATTGAAATTTTAGGTTCTGCAGATAAGTCAGGACGATGTATAAAAG
GGCTAGGAAAATCAAATATTTCTATTTATAAGGTAAGAACAGATGTCTAGGAAATCAGCA
TATCATCAAAAATGAGGGTGATAAGCCTAAGCTAGATTTTTAATATTAATTGTTAAAAAAGT
GTTGCAATTATAGTTATCATATGCTATAATAATCGTGTAAGGGACGCCTTACACAGTTACTT
AAATCTTGCAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCAATTTATG
GCAGGGTGTTTTCGTTATTTAAAGAGGAGAAGAAATGACTTGGAGAGTTGTACATGTCAGTC
AAAGTGAGAAGATGCGCTTAAAGCTTGATAACTTATTAGTGCAAAAGATGGGACAAGAGTT
TACGGTGCCACTAAGTGATATTTGATAATCGTTGCAGAAGGTGGGGATACAGTTGTTACCC
TTCGTCTATTAAGTGCCTTAAGTAAATATAATATTGCCTTGGTCGTTTGTGATAACGAACATT
TACCAACAGGAATTTATCACTCACAAAATGGGCACCTTAGAGCGTACAAGCGCTTGAAAGA
ACAGCTGGATTGGTCTCAGAAACAAAAGGAAAGGCATGGCAGATTGTAACCTTATTATAAA
ATCAATAACCAAGAGGATGTCTAGCCATGTTTGAAAAAAGTCTGGACAACATTAGATTAC
TTTCAGACTATAAAGAGCAGATAGAACCTGGTGATAGAACGAATAGAGAGGGACATGCTGC
CAAGGTCTACTTTAATGAGCTCTTTGGTAAACAATTTGTGAGAGTAACTCAGCAAGAAGCTG
ATGTCATCAATGCTGGTTTAAACTATGGCTATGCTATCATGAGGGCTCAGATGGCTAGAATA
GTGGCGGGTTATGGTTTAAATGGCCTATTAGGAATCTTCATAAAAAATGAATACAATCAGTT
TAATTTGGTTGACGATTTGATGGAGCCATTTAGACAGATTGTAGATGTTTGGGTATATGATA
ATCTACGAGATCAGGAATTCCTTAAGTATGAGTATAGGTTGGGATTGACAGATTTACTCAAT
GCTAAAATCAAATATGGCAAAGAGACTTGCTCAGTGACAGTTGCTATGGACAAATATGTCA
AAGGCTTTATCAAATATATTTCCGAAAAAGATAGTAGTAAATTTCACTGCCAGTGTTATCA
AGTTTAGAGTGGAGAAAATAAGATGAGGTATGAAGCATTGAGATTATTATGTTTTTTTGATT

TACCAATGGAATCCAAGGATGAAAAAGAATATATCGTAATTTTCGTAAAGAATTAATTTTC
 AAATGGGTTTGAAATGTTACAATTTTCGGTCTACTATCGCACTTGTCTAATAGAAGCTTTGC
 AAATAAATTTTATAAGAAGTTAAAGATGAGCAATCTTCCTGCTGGGAATGTGAGACTTTTGG
 CAGTTACTGAAAAACAATTTTCAGAGATGACATTAATTATAGGTGGTAAACTAAGCAAGA
 AGAAATCGTCAGTGATAATAAGTTGGTGATCATATGAAATTTTTGTACAACATCCTTACAA
 AGAACGTATTGAATTAATATTGGTGCAATCACACAAATTGTTGGTCAGAATAATGAACTCA
 AATATTACTTGGCAGATTTTGGCTGGTATTTTGGTGGAATAATACTCAAGTGAGGAC
 TTAAGTATTTTGTATTATGAGGAGCCTACCATACTTGATGAGGCCAGAGAAATAGTGAAACG
 AAGTAGCTATCACTATATCGACATTTCAAGTTTTAAGGATTTACTGGAGCAGATGGAATACA
 AGAAAGGAACACTTGCTCAGGGTTACCTTCGTAAAATTGTCAATCAAGTTGATATTGTAGGC
 CATTTGGAGAAAATTAATGAACAAGTAGAGCTTATTGAAGAAGCTATGAATCGGCATATAA
 ACTTAACTGTGGACAGGTAGAATACCATTTGGAGAATCTCCCTCTAACACTAGACCAACTA
 CTCACAAAAAATTTTAGCCCATTTTGGCCATTGAGAACAGAATCTATCTTTGAATGGGTT
 TCTAATATTGATAAACTATCCCTCTTTTAGAAATGTTAGACCATCTTCTTACAAAACAACA
 GAGAAGTATCTCATTGTGCTAAAAAATATTGATGGCTTTATCTCAGAAGAATCTTATACTAT
 TTTTATAGGCAAATCTGTCATCTGGTCAAGAAGTATCCAAATCTAACCTTTATTTTGTTC
 TAGTGACCAAGGCTATTTAAAAATTTGATGAAGAAAATAGTAGGTTCTGCAATATTTTATCTG
 ACCAGGTGGAACATTTGTATGATGTTGAGTTTATGTATGAAAGGGTAATGAAATATTATCCA
 AGTAATGATTTTCCGACGAGAGAAGGTTTTAGGATGTCTTTAGAACTGTGACACCTTATTT
 ATTGACAAAAATGCTGAGACAACCTAGTCTCTCACTTGTGATTGAGTAATATTGAATATCC
 TAAATCAGCTGTTTCATTTAGTTACCGTATAAGATGTTCTCAGACACCTGATAAGGAACTA
 TTACAGAAATTTTGTAGAAAGTAAGGATTGA (SEQ ID NO:473)

SEQ ID NO:474:

ATGAGTGACTTAGTTTTAGGACTTGATATCGGTATAGGTTCTGTTGGTGTAGGTATCCTTAAC
 AAAGTGACAGGAGAAATTATCCATAAAAACTCACGCATCTTCCAGCAGCTCAAGCAGAAA
 ATAACCTAGTACGTAGAACGAATCGTCAAGGAAGACGCTTGACACGACGTAAAAAACATCG
 TATAGTTCTGTTTAAATCGTCTATTTGAGGAAAGTGGATTAATCACCGATTTTACGAAGATTT
 CAATTAATCTTAACCCATATCAATTACGAGTTAAGGGCTTGACCGATGAATTGTCTAATGAA
 GAACTGTTTATCGCTCTTAAAAATATGGTGAACACCGTGGGATTAGTTACCTCGATGATGC
 TAGTGATGACGGAATTCATCAGTAGGAGACTATGCACAAATTGTTAAGGAAAATAGTAAA
 CAATTAGAACTAAGACACCGGGACAGATACAGTTGGAACGCTACCAAAACATATGGTCAAT
 TACGTGGTGAATTTTACTGTTGAGAAAAGATGGCAAAAAACATCGCTTGATTAATGTCTTTCCA
 ACATCAGCTTATCGTTTCAGAAAGCCTTAAGGATACTGCAAACTCAACAAGAATTTAATTCACA
 GATTACAGATGAATTTATTAATCGTTATCTCGAAAATTTAACTGGAAAACGGAAATATTATC
 ATGGACCCGGAATGAAAAGTCAAGGACTGATTATGGTCGTTACAGAACGAATGGAGAAAC
 TTTAGACAATATTTTGGAAATCTAATTGGGAAATGTACATTTTATCCAGACGAGTTTAGAG
 CAGCAAAAGCTTCCTACACGGCTCAAGAATTCAATTTGCTAAATGATTTGAACAATCTAACA
 GTTCCTACTGAAACCAAAAAGTTGAGCAAAAGAACAGAAGAATCAAAATCATTAAATTATGTCA
 AAAATGAAAAGGTAATGGGGCCAGCGAACTTTTTAAATATATCGCTAAATTACTTTCTTGT
 GATGTTGCAGATATCAAGGGACACCGTATCGACAAATCAGGTAAGGCTGAGATTCATACTT
 TCGAAGCCTATCGAAAAATGAAAACGCTTGAACCTTAGATATTGAGCAATGGATAGAGA
 AACGCTTGATAAATTAGCCTATGTCTTAACATTAAACACTGAGAGGGAAGGTATTCAAGAA
 GCTTTAGAACATGAATTTGCTGATGGTAGCTTTAGCCAGAAGCAAGTTGACGAATTGGTTCA
 ATTCCGCAAAGCAAATAGTTCCATTTTGGAAAAGGATGGCATAATTTTCTGTCAAACCTGA
 TGATGGAGTTAATTCAGAAATTGTATGAGACGTCAGAAGAGCAAATGACTATCCTGACACG
 ACTTGGAAAACAAAAACAACCTTCGTCTTCAAATAAAAACAAAAATATATAGATGAGAACTA
 TTAACCTGAAGAAATCTATAATCCTGTTGTTGCTAAGTCTGTTCCGAGGCTATAAAAAATCGT
 AAATGCGGCGATTAAAGAATACGGAGACTTTGACAATATTGTCATCGAAATGGCTCGTGAA
 ACAAATGAAGATGATGAAAAGAAAGCTATTCAAAAGATTCAAAAAGCCAACAAAGATGAA
 AAAGATGCAGCAATGCTTAAGGCTGCTAACCAATATAATGGAAAGGCTGAATTACCACATA
 GTGTTTTCCACGGTCATAAGCAATTAGCGACTAAAATCCGCTTTGGCATCAGCAAGGAGAA
 CGTTGCCTTTACTGGTAAGACAATCTCAATCCATGATTTGATAAATAATCCTAATCAGTTT
 GAAGTAGATCATATTTTACCTCTTTTATCATCATTCGATGATAGCCTTGCAATAAGGTTTTG
 GTTTATGCAACTGCTAACCAAGAAAAAGGACAACGAACACCTTATCAGGCTTTAGATAGTA

TGGATGATGCGTGGTCTTTCCGTGAATTAAAAGCTTTTGTACGTGAGTCAAAAACACTTTCA
 AACAAGAAAAAGAATACCTCCTTACAGAAGAAGATATTTCAAAGTTTGATGTTGAAAGA
 AATTTATTGAACGAAATCTTGTAGATACAAGATACGCTTCAAGAGTTGTCCTCAATGCCCTT
 CAAGAACACTTTAGAGCTCACAAGATTGATACAAAAGTTTCCGTGGTTCGTGGCCAATTTAC
 ATCTCAATTGAGACGCCATTGGGGAATTGAGAAGACTCGTGATACTTATCATCACCATGCTG
 TCGATGCATTGATTATTGCCGCTCAAGTCAGTTGAATTTGTGGAAAAACAAAAGAATACC
 CTTGTAAGTTATTCAGAAGAACAACCTCCTTGATATTGAAACAGGTGAACTTATTAGTGATGA
 TGAGTACAAGGAATCTGTGTTCAAAGCCCCCTTATCAACATTTTGTGATACATTGAAGAGTA
 AAGAATTTGAAGACAGTATCTTATTCTCATATCAAGTGGATTCTAAGTTTAATCGTAAAATA
 TCAGATGCCACTATTTATGCGACAAGACAGGCTAAAGTGGGAAAAGATAAGAAGGATGAA
 ACTTATGTCTTAGGGAAAATCAAAGATATCTATACTCAGGATGGTTATGATGCCTTTATGAA
 GATTTATAAGAAGGATAAGTCAAAATTCCTCATGTATCGTCACGACCCACAAACCTTTGAGA
 AAGTTATCGAGCCAATTTTAGAGAATACTCTAATAAGGAAATGAATGAAAAAGGGAAAGA
 AGTACCATGTAATCCTTTCCTAAAATATAAAGAAGAACATGGCTATATTTCGTAAATATAGTA
 AAAAAGGCAATGGTCTGAAATCAAGAGTCTTAAATACTATGATAGTAAGCTTTTAGGTAA
 TCCTATTGATATTACTCCAGAGAATAGTAAAAATAAAGTTGTCTTACAGTCATTAAAACCTT
 GGAGAACAGATGTCTATTTCAATAAAAAATACTGGTAAATATGAAATTTTAGGACTGAAATA
 TGCTGATTTACAATTTGAAAAGAAGACAGGAACATATAAGATTTCCAGGAAAAATACAAT
 GGCATTATGAAAGAAGAGGGTGTAGATTCTGATTCAGAATTCAAGTTTACACTTTATAAAAA
 TGATTTGTTACTCGTTAAAGATACAGAAACAAAAGAACAACAGCTTTTCCGTTTTCTTCTC
 GAACTATGCCTAATGTGAAATATTATGTAGAGTTAAAGCCTTATTCAAAAGATAAATTTGAG
 AAGAATGAGTCACTTATTGAAATTTTAGGTTCTGCAGATAAGTCAGGACGATGTATAAAAG
 GGCTAGGAAAATCAAATATTTCTATTTATAAGGTAAGAACAGATGTCCTAGGAAATCAGCA
 TATCATCAAAAATGAGGGTGATAAGCCTAAGCTAGATTTTAA (SEQ ID NO:474)

SEQ ID NO:475:

ATGACTTGGAGAGTTGTACATGTCAGTCAAAGTGAGAAGATGCGCTTAAAGCTTGATAACTT
 ATTAGTGCAAAAGATGGGACAAGAGTTTACGGTGCCACTAAGTGATATTTGATAATCGTTG
 CAGAAGGTGGGGATACAGTTGTTACCCTTCGTCTATTAAGTGCCTTAAGTAAATATAATATT
 GCCTTGGTTCGTTTGTGATAACGAACATTTACCAACAGGAATTTATCACTCACAAAATGGGCA
 CTTTAGAGCGTACAAGCGCTTGAAAGAACAGCTGGATTGGTCTCAGAAACAAAAGGAAAAG
 GCATGGCAGATTGTAACCTTATTATAAAATCAATAACCAAGAGGATGTCCTAGCCATGTTTGA
 AAAAAGTCTGGACAACATTAGATTACTTTAGACTATAAAGAGCAGATAGAACCTGGTGAT
 AGAACGAATAGAGAGGGACATGCTGCCAAGGTCTACTTTAATGAGCTCTTTGGTAAACAAT
 TTGTCAGAGTAACTCAGCAAGAAGCTGATGTCATCAATGCTGGTTTAAACTATGGCTATGCT
 ATCATGAGGGCTCAGATGGCTAGAATAGTGGCGGGTTATGGTTTAAATGGCCTATTAGGAA
 TCTTCCATAAAAAATGAATACAATCAGTTTAAATTTGGTTGACGATTTGATGGAGCCATTAGA
 CAGATTGTAGATGTTTGGGTATATGATAATCTACGAGATCAGGAATTCCTTAAGTATAGTA
 TAGGTTGGGATTGACAGATTTACTCAATGCTAAATCAAAATATGGCAAAAGAGACTTGCTCA
 GTGACAGTTGCTATGGACAAATATGTCAAAGGCTTTATCAAATATATTTTCGAAAAAGATA
 GTAGTAAATTTCACTGCCCAGTGGTATCAAGTTTAGAGTGGAGAAAATAA (SEQ ID NO:475)

SEQ ID NO:476:

ATGAGGTATGAAGCATTGAGATTATTATGTTTTTTGATTTACCAATGGAATCCAAGGATGA
 AAAAAGAATATATCGTAATTTTCGTAAAGAATTAATTTCAAATGGGTTTGAAATGTTACAAT
 TTTCCGTCTACTATCGCACTTGTCTAATAGAAGCTTTGCAAATAAATTTTATAAGAAGTTA
 AAGATGAGCAATCTTCCTGCTGGGAATGTGAGACTTTTGGCAGTTACTGAAAAACAATTTTC
 AGAGATGACATTAATTATAGGTGGTAAACTAAGCAAGAAGAAATCGTCAGTGATAATAAG
 TTGGTGATCATATGA (SEQ ID NO:476)

SEQ ID NO:477:

ATGAAATTTTTGTACAACATCCTTACAAAGAACGTATTGAATTAATATTGGTGCAATCAC
 ACAAATTTGTTGGTCAGAATAATGAACTCAAATATTATACTTGGCAGATTTTGAGCTGGTATT

TTGGTGGAAAAAATACTCAAGTGAGGACTTAAGTATTTTTGATTATGAGGAGCCTACCATA
 CTTGATGAGGCCAGAGAAATAGTGAAACGAAGTAGCTATCACTATATCGACATTTCAAGTTT
 TAAGGATTTACTGGAGCAGATGGAATACAAGAAAGGAACACTTGCTCAGGGTTACCTTCGT
 AAAATTGTCAATCAAGTTGATATTGTAGGCCATTTGGAGAAAATTAATGAACAAGTAGAGC
 TTATTGAAGAAGCTATGAATCGGCATATAAACTTAACTGTGGACAGGTTAGAATACCATTG
 GAGAATCTCCCTCTAACACTAGACCAACTACTCACAAAAAATTTTAGCCCCATTTTTGCCAT
 TGAGAACAAAGAATCTATCTTTTGAATGGGTTTCTAATATTGATAAACTATCCCTCTTTTTAGA
 AATGTTAGACCATCTTCTTTCACAAACAACAGAGAAGTATCTCATTGTGCTAAAAAATATTG
 ATGGCTTTATCTCAGAAGAATCTTATACTATTTTTTATAGGCAAATCTGTCATCTGGTCAAGA
 AGTATCCAAATCTAACCTTTATTTTGTTCCTAGTGACCAAGGCTATTTAAAAATTGATGAA
 GAAAATAGTAGGTTCTGCAATATTTTATCTGACCAGGTGGAACATTTGTATGATGTTGAGTT
 TATGTATGAAAGGGTAATGAAATATTATCCAAGTAATGATTTTCCGACGAGAGAAGGTTTTA
 GGATGTCTTTAGAACTGTGACACCTTATTTATTGACAAAAATGCTGAGACAACCTAGTCTC
 TCACTTGTGATTAGTAATATTGAATATCCTAAATCAGCTGTTTCATTTTAGTTACCGTATA
 AGATGTTCTCAGACACCTGATAAGGAATATTACAGAAATTTTAGAAAGTAAGGATTGA
 (SEQ ID NO:477)

con secuencias repetidas: SEQ ID NO: 1 hasta SEQ ID NO: 10

Combinación funcional #2:

- 5 secuencias cas: SEQ ID NO:466 hasta SEQ ID NO:472, y SEQ ID NO:478 hasta SEQ ID NO:487 (todas las cuales son secuencias *S. thermophilus*), como se muestran a continuación:

SEQ ID NO:466:

ATGAGCGATTTATATAGTCAAAGGTCCAATTATTACCTGTCCTTATCTGAACAAAGAATTAT
 CATTAAAAATGATAATAAAGAGATTGTCAAAGAAGTGTCCATTTCACTCGTTGATAATGTAT
 TACTTTTTGGTAATGCACAACCTGACCACCCAACCTCATCAAAGCCTTGTCAAAGAACAAAGGTG
 AATGTTTACTATTTCTCAAATGTTGGTCAATTTATTTCTAGTATTGAAACCCACAGGCAGGAC
 GAATTCCAAAAGCAAGAGTTGCAAGCAAAGGCTTATTTTGAAGAGGATTTCCGTTTAGAGG
 TTGCGAGGAGTATTGCTACGACCAAGGTGAGGCACCCAATTGCCTTACTTAGAGAGTTTGAT
 ACGGATGGTCTACTAGATACCTCAGATTATTCTAGGTTTGAAGATAGTGTCAATGATATTCA
 GAAAGCTTATTTCCATTACAGAAATTATGGGTTACGAAGGTGCGCTTGCGAAATCCTATTTTT
 ACTATCTGAATTTACTCGTTCCTAATGACTTTTCATTTTAAATGGTAGGAGTAGACGGCCTGGG
 GAGGATTGTTTTAACAGTGCCCTCAATTTTGGCTATAGTATCTTATATTCTTGCTTAATGGGC
 TGATTAAGAAAAACGGGCTAAGCTTGGGATTTGGGGTAATTCACAAGCATCATCAGCATCA
 TGCGACCTTGGCCAGTGATTTAATGGAAGAATGGAGACCTATCATCGTCGATAATACGCTTA
 TGGAGTTGGTACGAAATGGTAAACTTCTTTTAAAGTCATTTTGAAAAATAAGGATCAAGACTTC
 ATACTACCCATGAAGGCAGAGAAATCTTTGCACGGGCTTTACGTTCAAGAATATTAGAAGT
 CCATCAGTATATTGAGTTAGATAAAAAACGCTATTCTTTTCTTTATACAGCAGATAGGCAAA
 TCAAGAGTTTGATTAGGGCTTTTAGAGAACTTGACCCTAGTCTCTATGAGACAAGTTACACA
 GGAGGGCATTAAATGGGACTTTACTTTAACCTCAGCGAAGAAGAGCGTGAGTTTGCCAAACA
 AAAAAACCATGTTTTGTCTGATTATTTATGATATTCTGAAGTAACAAACGTAGACTTAAACTC
 TCGAAATTACTTGAGGGTTATGGCGTGAGGGTGCAAAAATCCTGTTTTCGAAGTCGACCTGTC
 AAGAAATGATTATCAGTCTCTCCTTAAGGATATCGAGGGCTTCTCCAAGGCTGATGAAGAA
 GACAGCATAATAGTGATGTGCCAACCAAGAAGAGGTGACTAGTTTTAGCCCCCTACCATA
 GTGCTGAAAAATTAGATGACATTCTCTTCCCCTAAGCCTTTATAGACCTTTAATCATATGGTA
 CACTATAGATAGTGTTCAGAGGCTCTTAAGGAAATCAAAGATAGAGAGACACTTCAAAG
 ATTTTGTAGATATATGGAAGCATTAGTAGCCTATTTCAAGTTTTATGGAGGTAAAGATTAAT
 GACATTCGCTAAGATTAAATTTTCAGCTCAAATTCGTTTATAGAGACAGGCCTCCATATTGGTG
 GAAGCGATGCTTTTGCAGCCATTGGTGCAATCGATTGCGCTGTTATTAAGATCCTATTACC
 AACCTACCGATCATTCCTGTTCAAGTCTCAAAGGAAAAATGAGAACGCTTCTTGCCAAGGT
 TTATAATGAAAAGGTAGCTGAGAAACCAAGCGATGACAGTGATATTCTTAGCCGTTTATTTG

GGAATAGTAAAGATAAACGATTCAAAATGGGACGCTTGATTTTTCGTGATGCCTTCTTGTC
AACGCTGATGAGCTAGACTCTCTTGGGGTAAGAAGTTATACAGAAGTAAAATTTGAAAATA
CAATTGACCGTATCACTGCCGAAGCTAATCCAAGACAAATTGAACGTGCTATTCGTACCAAGT
ACTTTTGATTTTCGAGTTGATTTATGAAATTACAGATGAGAATGAAAATCAAGTCGAAGAAG
ATTCCAAAGTGATTCGAGATGGTTTAAACTGCTTGAACCTTGATTATCTTGGTGGTTCTGGA
TCTCGAGGTTACGGTAAGGTTGCTTTTGAACCTCAAAGCTACTACCGTATTTGGTAATTA
TGATGTTAAACATTAAATGAACTTTAACTGCGGAGGTCTAATATGACCTATAAACTGTAT
ATTATGACCTTTCAGAATGCTCATTTTGGTTTCGGGCACTCTTGATAGCTCAAAATTAACATTC
TCAGCAGACCGTATCTTCTCAGCACTAGTGCTAGAATCCCTAAAAATGGGAAAACCTCGATGC
ATTTCTTGCGGAAGCTAACCAAGACAAGTTCACGCTCACAGATGCCTTTCATTTCATTTTG
GTCCCTTTTTGCCGAAACCTATTGGTTATCCCAAACATGACCAAATAGATCAATCAGTTGAT
GTCAAAGAGGTTTCGCCGTCAAGCAAAATTGTCTAAGAACTGCAATTTCTTGCTCTAGAAAA
TGTTGACGATTATATCAATGGAGAGTTATTTGAAAATGAAGAGCATGCAGTCATCGATACTG
TGACAAAAAATCAACCACATAAGGACGGCAATCTTTATCAGGTAGCTACAACCAGATTTTC
AAATGATACGTCGCTTTACGTCATCGCAAACGAATCTGATTTGCTTAATGAGTTGATGTCTA
GTCTTCAGTATTCAGGTCTTGGTGGAAAGCGTTCAAGTGGTTTTGGTCGTTTTGAGTTAGATA
TTCAAAATATCCCACTAGAATTGTCAGATAGACTGACTAAGAATCATTACAGATAAAGTGATG
AGTCTTACGACAGCACTTCCTGTAGATGCTGACCTTGAAGAAGCAATGGAAGATGGACATT
ACTTATTAATAAATCAAGTGGTTTTGCATTTAGTCATGCCACCAATGAGAATTATCGTAAG
CAGGATCTTTACAAATTTGCTTCTGGTTCAACTTTTAGTAAAACATTTGAAGGTCAGATTGTT
GATGTGAGACCACTTGATTTCCCTCATGCTGTTTTAAATTATGCTAAACCACTCTTCTTTAA
TTGGAGGTATAAAAAATGAAAAATGACTATAGAACATTTAAATTAAGCCTCCTGACACTTGCT
CCAATTCATATTGGTAATGGAGAGAAGTATACCTCTAGAGAATTTATCTATGAAAATAAAA
AGTTTTACTTTCTGACATGGGGAAATTCTATAATAAAATGGTGGAGAAGAGGCTTGCTGAA
AAGTTTGAAGCATTCTAATTCAAACCTCGTCCAAATGCACGTAATAATCGTCTTATTTCTTC
TTAAATGATAACCGAATTGCAGAGCGTTCTTTTGGAGGTTATAGTATCTCTGAAACAGGTTT
AGAATCGGACAAAAATCCTGATTCAACCGGAGCTATTAACGAAGTTAATAAATTTATTCTGA
GATGCTTTTGGAAATCCCTACATTCCTGGTAGCTCACTAAAAGGTGCTATTCGTACCATTTTA
ATGAATACTACCCCTAAGTGGATAATGAAAATGCTGTAAATGACTTTGGAAGATTTCCGA
AAGAGAATAAGAACCCTTATCCCTTGGGGACCAAAAAAGGGAAAAGAATACGATGATTTGTT
TAACGCAATTCGTGTGAGTGATAGTAAGCCTTTTGATAATAAGAGTCTTATCTTAGTGACAG
AATGGGATTATTCAGCGAAAACAAATAAAGCTAAACCACTTCCCTTGATAGAGAATCAAT
CTCTCCATTAACAAAAATTGAATTTGAGATTACAACAACCACTGATGAAGCTGGAAGATTG
ATTGAAGAATTAGGTAAGAGAGCACAAGCGTTTTATAAAGACTATAAGGCATTTTCTCTATC
TGAATTTCTGATGATAAGATTCAAGCCAATCTACAATACCCAATTTATTTAGGTGCGGGGA
GCGGTGCTTGGACAAAGACTCTATTTAAGCAAGCTGATGGTATTTTACAAAGACGATACAGT
CGAATGAAAACATAAAATGGTTAAAAAAGGAGTTCTTAAGCTCACAAAAGCACCTCTTAAAA
CAGTTAAGATTCCATCTGGTAATCATTATTAGTCAAGAACCACGAGTCCTTTTATGAAATG
GGAAAAGCTAATTTTCATGATTAAGGAGATTGATAAATGA (SEQ ID NO:466)

SEQ ID NO:467:

ATGAGCGATTTATATAGTCAAAGGTCCAATTATTACCTGTCCTTATCTGAACAAAGAATTAT
CATTAAAAATGATAATAAAGAGATTGTCAAAGAAGTGTCATTTCACTCGTTGATAATGTAT
TACTTTTTGGTAATGCACAACCTGACCACCAACTCATCAAAGCCTTGTCAAAGAACAAGGTG
AATGTTTACTATTTCTCAAATGTTGGTCAATTTATTTCTAGTATTGAAACCCACAGGCAGGAC
GAATTCAAAAAGCAAGAGTTGCAAGCAAAGGCTTATTTTGAAGAGGATTTCCGTTTAGAGG
TTGCGAGGAGTATTGCTACGACCAAGGTGAGGCACCAATTGCCTTACTTAGAGAGTTTGAT
ACGGATGGTCTACTAGATACCTCAGATTATTCTAGGTTTGAAGATAGTGTCAATGATATTCA
GAAAGCTTATTCCATTACAGAAATTATGGGTTACGAAGGTGCGCTTGCGAAATCCTATTTT
ACTATCTGAATTTACTCGTTCCTAATGACTTTTCAATTTAATGGTAGGAGTAGACGGCCTGGG
GAGGATTGTTTTAACAGTGCCCTCAATTTTGGCTATAGTATCTTATATTCTTGCTTAATGGGC
TGA (SEQ ID NO:467)

SEQ ID NO:468:

TTGCTTAATGGGCTGATTAAGAAAAACGGGCTAAGCTTGGGATTTGGGGTAATTCACAAGC
ATCATCAGCATCATGCGACCTTGGCCAGTGATTTAATGGAAGAATGGAGACCTATCATCGTC
GATAATACGCTTATGGAGTTGGTACGAAATGGTAACTTCTTTTAAGTCATTTTGAAAATAA
GGATCAAGACTTCATACTACCCATGAAGGCAGAGAAATCTTGCACGGGCTTTACGTTCAA
GAATATTAGAAGTCCATCAGTATATTGAGTTAGATAAAAAACGCTATTCTTTCTTTATACA
GCAGATAGGCAAATCAAGAGTTTGATTAGGGCTTTTAGAGAACTTGACCCTAGTCTCTATGA
GACAAGTTACACAGGAGGGCATTAA (SEQ ID NO:468)

SEQ ID NO:469:

ATGTTTTGTCTGATTATTTATGATATTCGAAGTAACAAACGTAGACTTAAACTCTCGAAATT
ACTTGAGGGTTATGGCGTGAGGGTGCAAAAATCCTGTTTCGAAGTCGACCTGTCAAGAAAT
GATTATCAGTCTCTCCTTAAGGATATCGAGGGCTTCTCCAAGGCTGATGAAGAAGACAGCAT
AATAGTGTATGTGCCAACCAAAGAAGAGGTGACTAGTTTTAGCCCCTACCATAGTGCTGAA
AAATTAGATGACATTCTCTCCCCTAA (SEQ ID NO:469)

SEQ ID NO:470:

ATGACATTCGCTAAGATTAAATTTTCAGCTCAAATTCGTTTAGAGACAGGCCTCCATATTGG
TGGAAGCGATGCTTTTGCAGCCATTGGTGCAATCGATTGCCTGTTATTAAAGATCCTATTA
CCAACCTACCGATCATTCTGTTCAAGTCTCAAAGGAAAAATGAGAACGCTTCTTGCCAAG
GTTTATAATGAAAAGGTAGCTGAGAAACCAAGCGATGACAGTGATATTCTTAGCCGTTTATT
TGGAATAGTAAAGATAAACGATTCAAATGGGACGCTTGATTTTTCGTGATGCCTTCTTGT
CAAACGCTGATGAGCTAGACTCTCTTGGGGTAAGAAGTTATACAGAAGTAAATTTGAAAA
TACAATTGACCGTATCACTGCCGAAGCTAATCCAAGACAAATTGAACGTGCTATTCGTACCA
GTACTTTTGATTTGAGTTGATTTATGAAATTACAGATGAGAATGAAAATCAAGTCGAAGAA
GATTCAAAGTGATTGAGATGGTTTAAACTGCTTGAACCTTGATTATCTTGGTGGTTCTGG
ATCTCGAGGTTACGGTAAGGTTGCTTTTGAAAACCTCAAAGCTACTACCGTATTTGGTAATT
ATGATGTTAAACATTAAATGAACCTTTAACTGCGGAGGTCTAA (SEQ ID NO:470)

SEQ ID NO:471:

ATGACCTATAAACTGTATATTATGACCTTTCAGAATGCTCATTTTGGTTCCGGGCACTCTTGAT
AGCTCAAAATTAACATTCTCAGCAGACCGTATCTTCTCAGCACTAGTGCTAGAATCCCTAAA
AATGGGAAAACCTCGATGCATTTCTTGCAGGAAGCTAACCAAGACAAGTTCACGCTCACAGAT
GCCTTTCCATTTCAATTTGGTCCCTTTTGCAGAAACCTATTGGTTATCCCAAACATGACCAA
ATAGATCAATCAGTTGATGTCAAAGAGGTTTCGCCGTCAAGCAAAATTGTCTAAGAACTGC
AATTTCTTGCTCTAGAAAATGTTGACGATTATATCAATGGAGAGTTATTTGAAAATGAAGAG
CATGCAGTCATCGATACTGTGACAAAAAATCAACCACATAAGGACGGCAATCTTTATCAGG
TAGCTACAACCAGATTTTCAAATGATACGTCGCTTACGTCATCGCAAACGAATCTGATTTG
CTTAATGAGTTGATGTCTAGTCTTCAGTATTCAGGTCTTGGTGGAAAGCGTTCAAGTGGTTTT
GGTCGTTTTGAGTTAGATATTCAAATATCCCACTAGAATTGTCAGATAGACTGACTAAGAA
TCATTGAGATAAAGTGATGAGTCTTACGACAGCACTTCTGTAGATGCTGACCTTGAAGAAG
CAATGGAAGATGGACATTACTTATTAATAAATCAAGTGGTTTTGCATTTAGTCATGCCACC
AATGAGAATTATCGTAAGCAGGATCTTTACAAATTTGCTTCTGGTTCAACTTTTAGTAAAAC
ATTTGAAGGTCAGATTGTTGATGTGAGACCACTTGATTTCCCTCATGCTGTTTTAAATTATGC
TAAACCACTCTTCTTAAATTGGAGGTATAA (SEQ ID NO:471)

SEQ ID NO:472:

ATGAAAAATGACTATAGAACATTTAAATTAAGCCTCCTGACACTTGCTCCAATTCATATTGG
TAATGGAGAGAAGTATACCTCTAGAGAATTTATCTATGAAAATAAAAAGTTTTACTTTCTG
ACATGGGGAAATTTCTATAATAAAATGGTGGAGAAGAGGCTTGCTGAAAAGTTTGAAGCATT
TCTAATTCAAACTCGTCCAAATGCACGTAATAATCGTCTTATTTCTTCTTAAATGATAACCG
AATTGCAGAGCGTTCTTTGGAGGTTATAGTATCTCTGAAACAGGTTTAGAATCGGACAAAA
ATCCTGATTCAACCGGAGCTATTAACGAAGTTAATAAATTTATTCGAGATGCTTTTGGAAAT

CCCTACATTCCCTGGTAGCTCACTAAAAGGTGCTATTCGTACCATTTTAATGAATACTACCCCT
 AAGTGGAATAATGAAAATGCTGTAATGACTTTGGAAGATTTCCGAAAGAGAATAAGAACC
 TTATCCCTTGGGGACCAAAAAAGGGAAAAGAATACGATGATTTGTTAACGCAATTCGTGT
 GAGTGATAGTAAGCCTTTTGATAATAAGAGTCTTATCTTAGTGCAGAAATGGGATTATTCAG
 CGAAAACAAATAAAGCTAAACCACTTCCCTTGTATAGAGAATCAATCTCTCCATTAACAAA
 AATTGAATTTGAGATTACAACAACCACTGATGAAGCTGGAAGATTGATTGAAGAATTAGGT
 AAGAGAGCACAAGCGTTTTATAAAGACTATAAGGCATTTTCCTATCTGAATTTCTGATGA
 TAAGATTCAAGCCAATCTACAATACCCAATTTATTTAGGTGCGGGGAGCGGTGCTTGGACAA
 AGACTCTATTTAAGCAAGCTGATGGTATTTTACAAAGACGATACAGTCGAATGAAAACATAA
 ATGGTTAAAAAAGGAGTTCTTAAGCTCACAAAAGCACCTCTTAAACAGTTAAGATTCCA
 TCTGGTAATCATTATTAGTCAAGAACACGAGTCTTTTATGAAATGGGAAAAGCTAATTT
 CATGATTAAGGAGATTGATAAATGA (SEQ ID NO:472)

SEQ ID NO:478:

ATGAGCGATTTATATAGTCAAAGGTCCAATTATTACCTGTCCTTATCTGAACAAAGAATTAT
 CATTAAAAATGATAATAAAGAGATTGTCAAAGAAGTGTCCATTTCACTCGTTGATAATGTAT
 TACTTTTTGGTAATGCACAACCTGACCACCAACTCATCAAAGCCTTGTCAAAGAACAAGGTG
 AATGTTTACTATTTCTCAAATGTTGGTCAATTTATTTCTAGTATTGAAACCCACAGGCAGGAC
 GAATTCCAAAGCAAGAGTTGCAAGCAAAGGCTATTTTGAAGAGGATTTCCGTTTAGAGT
 TTGCGAGGAGTATTGCTACGACCAAGGTGAGGCACCCAATTGCCTTACTTAGAGAGTTTGAT
 ACGGATGGTCTACTAGATACCTCAGATTATTCTAGGTTTGAAGATAGTGTCAATGATATTCA
 GAAAGCTTATTCCATTACAGAAATTATGGGTTACGAAGGTGCGCTTGCAGAAATCCTATTTTT
 ACTATCTGAATTTACTCGTTCCTAATGACTTTTCATTTTAATGGTAGGAGTAGACGGCCTGGG
 GAGGATTGTTTTAACAGTGCCTCAATTTTGGCTATAGTATCTTATATTCTTGCTTAATGGGC
 TGATTAAGAAAAACGGGCTAAGCTTGGGATTGGGGTAATTCACAAGCATCATCAGCATCA
 TGCAGCTTGGCCAGTGATTTAATGGAAGAATGGAGACCTATCATCGTCGATAATACGCTTA
 TGGAGTTGGTACGAAATGGTAACTTCTTTTAAGTCATTTTGAATAAAGGATCAAGACTTC
 ATACTACCCATGAAGGCAGAGAAATCTTTGCACGGGCTTTACGTTCAAGAATATTAGAAGT
 CCATCAGTATATTGAGTTAGATAAAAAACGCTATTCTTTCTTTATACAGCAGATAGGCAAA
 TCAAGAGTTTGATTAGGGCTTTTAGAGAACTTGACCCTAGTCTCTATGAGACAAGTTACACA
 GGAGGGCATTAAATGGGACTTTACTTTAACCTCAGCGAAGAAGAGCGTGAGTTGCCAAACA
 AAAAACCATGTTTTGTCTGATTATTTATGATATTCGAAGTAACAAACGTAGACTTAACTCT
 CGAAATTACTTGAGGGTTATGGCGTGAGGGTGCAAAAATCCTGTTTCGAAGTCAACCTGTCA
 AGAAATGATTATCAGTCTCTCCTTAAGGATATCGAGGGCTTCTACAAGGCTGATGAAGAAG
 ACAGCATAATAGTGTATGTGACAACCAAAGAAGAGGTGACTAGTTTTAGCCCTACCATAG
 TGCTGAAAAATTAGATGACATTCTTCTTCTAAGCCTTTATAGACCTTTAATCATATGGTAC
 ACTATAGATAGTGTTCAGTAGGTCTTACATCTTGTGCTCTAGCACTAGAGCACA
 AGATATGGGGATATAAACCTAATTACCTCGAGAGGGGACGGAACGCTTTCTAGCTCGCTA
 TAATTACCATTCCTAGAAAGATATAAACCTAATTACCTCGAGAGGGGACGGAACCTTTGA
 ATAGTCTTTGAATCGCATTTGAACCATATAGATATAAACCTAATTACCTCGAGAGGGGACGG
 AAACAGGTTTTTTGCCATAGATTTTCCAAGACCTTCCCAACTGATATAAACCTAATTACCTC
 GAGAGGGGACGGAACGCTTTCTAGCTCGCTATAATTACCCATTCCTAGAAAGATATAAAC
 CTAATTACCTCGAGAGGGGACTTTTTTGAATAATTTTGAACAGTATTGATACCGCTTCCAG
 AAAGTGTTAGACTAAAAGCACATTAAGGGCGCCCCAATGAGTTGAAAAGTACTTTCAGCTT
 TTGGGGTTTTTTCATACAAAGATGAAGGAGTCGAATGAAAAAATTAGTATTTACTTTTAAAA
 GGATCGACCATCCTGCACAAGATTTGGCTGTAAATTTTCATGGCTTCTTGATGGAGCAGTTG
 GATAGTGACTATGTTGATTATCTGCATCAGCAGCAACAAATCCCTATGCGACCAAGGTAAT
 CCAAGGGAAAGAAAACACGCAGTGGGTGTACATCTGCTCACAGACGACATCGAGGATAAG
 GTTTTTATGACCTTATTACAGATTAAAGAGGTGTCCTTAAACGATCTGCCTAAACTCAGTGT
 CGAAAAAGTTGAGATTGAGGAGTTGGGGGACAGATAAACTGTTAGAGATTTTCAATAGTGAG
 GAAAAATCAAACCTATTTTTCAATTATTTTTGAGACTCCAACAGGTTTTAAATCTCAAGGTTCC
 TACGTATCTTCCCGTCTATGCGTTTGATTTTTCAAAGTTTGATGCAAAAGTATGGAAGGTTG
 GTTGAAAATCAACCTGAAATTGAAGAGGATACCTTAGATTACCTATCTGAACACAGCACTAT
 CACGAATTATCGCTTGGAGACGAGTTATTTCAAGGTGCACAGGCAACGAATTCCTGCCTTA
 GAGGAAAGTTAACCTTTAAAGTACAAGGCGCCCAAACTCTAAAAGCTTATGTCAAAATGCT

TCTAACATTTCGGTGAATATTCAGGTCTTGGCATGAAAACGAGTCTCGGTATGGGAGGGATA
 AAGCTTGAAGAAAGAAAAGATTGATTTATTTTACGGAGCTCTTTTGCATGATATCGGTAAAGG
 TCATTCAAAGGGCGACAGGAGAACGAAAAAACACGCCCTTGGTAGGCGCGGATTGGTTTGA
 TGAGATTGCTGATAATCAAGTTATTTCCGATCAAATTAGATATCACATGGCTAACTACCAGA
 GTGATAAACTTGGAAATGACCATCTTGCTTACATAACTTATATCGCTGATAACATTGCCTCT
 GGTGTGACAGAAAGACAGTCAAATGAGGAGAGTGACGAGGATACATCAGCTAAGATTTGG
 GATACCTATACAAACCAGGCTGATATTTTAAACGTTTTTGGGGCACAAACGGATAAACGCTA
 CTTTAAACCGACGGTTCTAAACTTGAAATCTAAACCTAACTTTGCGTCGGCAACATATGAAC
 CTTTCTCAAAGGTGATTATGCGGCAATTGCGACTCGTATCAAAAATGAATTGGCAGAAATTT
 GAGTTTAAATCAAGTACAAATTGACTCTTTGTTAAATCTGTTTGAAGCAACCCTCTCTTTTGTG
 CCTTCTTCGACTAATACTAAAGAAATCGCTGATATTTCACTTGCTGATCATAGTCGTCTGACA
 GCAGCTTTTGTCTAGCCATCTATGATTACTTGGGAAGACAAAGGTCGTACATAACTATAAGGA
 GGACTTGTTTACTAAAGCATCAGCCTTTTATGAGGAAGAAGCTTTTCTCCTAGCTAGCTTTG
 ACTTATCAGGGATTCAAGACTTTATCTATAATATTAATATTGCGACGAATGGTGTCTGCTAAA
 CAATTGAAGGCTAGATCTTTATATCTTGACTTTATGAGCGAGTATATAGCAGACAGTTTACT
 TGATAAACTAGGCCTCAATCGGGCTAATATGCTCTATGTCTGGTGGGGACATGCTTACTTTG
 TCCTAGCCAATACTGAAAAACGGTAGAAACACTCGTTCAATTTGAAAAAGATTTCAATCA
 ATTTTATTGGCAAATTTCCAAACCAGATTATATGTTGCTTTGGTTGGGGAAGCTTTGCGGC
 TAAGGATATCATGAGCGAACTGAACTCACTGAAAGCTATAGACAGGTCTATCAAAAGGCT
 AGTCGCATGATTTCTGAGAAAAAATCTCAAGGTATGATTATCAAACCCTTATGTTGTTGAA
 CAGGGGCGGTAAATCTTCTGAAAGAGAGTGCGAGATTTGTCAATTCGTTGAGAATTTAGTTG
 CTTATCATGACCAAAAAGTGTGTGACATTTGTGCGAGGCTTGTATCAATTTTCTAAAGAGATT
 GCCCATGACCATTTCAATTATCACTGAAAATGAAGGGCTTCTTATTGGTCCGAACGCATGTCT
 TAAGGGTGTTCATTTGAAAAGCTGAGCCAAGAAGCTTTTCCCGTGTCTATGTCAAAAATG
 ACTATAAGGCTGGTACAGTTAAGGCAACCCATGTTTTTGTGAGATTACCAGTATGATGAA
 ATATACAATTATGCTGCCTTATCTAAAAACGAAAATGGGTTAGGTATTAACGTTTAGCTGT
 TGTACGTCTTGACGTGGATGATTGGGAGCAGCCTTTATGGCTGGCTTCTCCCAACAAGGAA
 ATGGGCAATATAGTACTCTATCACGCTCAGCCACTTTCTCTCGAAGCATGAGTCTTTTCTTCA
 AGGTTTATATTAACCAGTTTGCTAGTGATAAGAAGCTCTCTATCATCTATGCCGGTGGGGAT
 GATGTTTTTGTCTATTGGCTCTTGGAAGATATTATTGCCTTACTGTTGAACCTTCGTGAGAAC
 TTCATTAATGGACAAATGGAAAACAACTATCAGCTGGTATCGGTCTGTTTGCTGATAA
 GACCCCTATTAGCTTAATGGCACATCAAACAGGGGAGCTAGAAGAAACAGCTAAAGGCAAT
 GAGAAAGATAGTATTTCACTCTTTAGTTCCGACTATACCTTTAAATTTGATCGGTTTATCACT
 AATGTTTACGACGATAAGTTAGAGCAGATTTCGCTATTTCTTTAATCACCAAGATGAACGAGG
 CAAGAATTTCAATTATAAATTGATTGAATTGCTTCGAAATTATGATCGTATGAATATGGCAC
 GTTAGCTTATTATTTAACACGACTTGAAGAATTGACGCGTGAAACAGACAGGGATAAATTT
 AAAACATTTAAAAATTTATTCTATTCTTGGTACACAAATAAGGATGATAAGGATAGAAAAG
 AAGCAGAGTTAGCCTTGCTTCTATATCTATGAGATTAGAAAGGATTAGGATATGACAATC
 TTGACTGATGAGAATTACGTTGATATTGCAGAAAAAGCAATTCTAAACTAGAAAGAAATA
 CTAGGAACAGAAAGAATCCTGATGCCTTCTTCTTACAACAAGTAAGCTCAGAAACTTGCTG
 AGCTTAACTAGTACACTTTTTGATGAGAGTAAGGTCAAAGAATATGATGCTCTCCTTGATCG
 TATTGCTTATTTAAGAGTACAATTTGTCTACCAAGCAGGTAGAGAGATTGCAGTAAAAGATC
 TGATAGAAAAGGCTCAAATTTCTGAGGCTCTTAAGGAAATCAAAGATAGAGAGACACTTCA
 AAGATTTTGTAGATATATGGAAGCATTAGTAGCCTATTTCAAGTTTTATGGAGGTAAGATT
 AATGACATTCGCTAAGATTAAATTTTCAGCTCAAATTCGTTTAGAGACAGGCCTCCATATTG
 GTGGAAGCGATGCTTTTGCAGCCATTGGTGCAATCGATTGCGCTGTTATTAAGATCCTATT
 ACCAACCTACCGATCATTCTGTTCAAGTCTCAAAGGAAAAATGAGAACGCTTCTTGCCAA
 GGTTTATAATGAAAAGGTAGCTGAGAAACCAAGCGATGACAGTGATATTCTTAGCCGTTTAT
 TTGGGAATAGTAAAGATAAACGATTCAAAATGGGACGCTTGATTTTTCTGTGATGCCTTCTTG
 TCAAACGCTGATGAGCTAGACTCTCTTGGGGTAAGAAGTTATACAGAAGTAAAATTTGAAA
 ATACAATTGACCGTATCACTGCCGAAGCTAATCCAAGACAAATTGAACGTGCTATTCGTACC
 AGTACTTTTGATTTTCGAGTTGATTTATGAAATTACAGATGAGAATGAAAATCAAGTCGAAGA
 AGATTTCAAAGTGATTCGAGATGGTTTAAACTGCTTGAACCTTGATTATCTTGGTGGTTCTG
 GATCTCGAGGTTACGGTAAGGTTGCTTTTGA AAAACTCAAAGCTACTACCGTATTTGGTAAT
 TATGATGTTAAACATTAAATGAACCTTTTAACTGCGGAGGTCTAATATGACCTATAAACTGT

ATATTATGACCTTTCAGAATGCTCATTTTGGTTCGGGCACTCTTGATAGCTCAAAATTAACAT
TCTCAGCAGACCGTATCTTCTCAGCACTAGTGCTAGAATCCCTAAAAATGGGAAAACCTCGAT
GCATTTCTTGCGGAAGCTAACCAAGACAAGTTCACGCTCACAGATGCCTTTCCATTTCAATT
TGGTCCCTTTTGGCGAAACCGATTGGTTATCCCAAACATGACCAAATAGATCAATCAGTTG
ATGTCAAAGAGGTTTCGCCGTCAAGCAAAATTGTCTAAGAACTGCAATTTCTTGCTCTAGAA
AATGTTGACGATTATCTCAATGGAGAGTTATTTGAAAATGAAGAGCATGCAGTCATCGATAC
TGTGACAAAAAATCAACCACATAAGGACGGCAATCTTTATCAGGTAGCTACAACCAGATTT
TCAAATGATACGTGCTTTACGTTCATCGCAAACGAATCTGATTTGCTTAATGAGTTGATGTC
TAGTCTTCAGTATTCAGGTCTTGGTGGAAAGCGTTCAAGTGGTTTTGGTCTTTTGAGTTAGA
TATTCAAAATATCCCACTAGAAATTGTCAGATAGACTGACTAAGAATCATTGAGATAAAGTGA
TGAGTCTTACGACAGCACTTCCTGTAGATGCTGACCTTGAAGAAGCAATGGAAGATGGACA
TTACTTATTAATAAATCAAGTGGTTTTGCATTTAGTCATGCTACCAATGAGAATTATCGTAA
GCAGGATCTTTACAAATTTGCTTCTGGTTCAACTTTTAGTAAAACATTTGAAGGTCAGATTGT
TGATGTGAGACCACTTGATTTCCCTCATGCTGTTTTAAATTATGCTAAACCACTCTTCTTTAA
ATTGGAGGTATAAAAAATGAAAAATGACTATAGAACATTTAAATTAAGCCTCCTGACACTTG
CTCCAATTCATATTGGTAATGGAGAGAAGTATACCTCTAGAGAATTTATCTATGAAAATAAG
AAGTTTTACTTTCTGACATGGGGAAATTCTATAATAAAATGGTGGAGAAGAGGCTTGCTGA
AAAGTTTGAAGCATTCTAATTCAAACCTCGTCCAAATGCACGTAATAATCGTCTTATTTCTT
CTTAAATGATAACCGAATTGCAGAGCGTTCTTTTGGAGGTTATAGTATCTCTGAAACAGGTT
TAGAATCGGACAAAAATCCTGATTCAGCGGAGCTATTAACGAAGTTAATAAATTTATTCTGA
GATGCTTTTGGAAATCCCTACATTCCTGGTAGCTCACTAAAAGGTGCTATTTCGTACCATTTTA
ATGAATACTACCCCTAAGTGGAATAATGAAAATGCTGTAAATGACTTTGGAAGATTTCCGA
AAGAGAATAAGAACCTTATCCCTTGGGGACCAAAAAAGGGAAAAGAATACGATGATTTGTT
TAACGCAATTCGTGTGAGTGATAGTAAGCCTTTTGATAATAAGAGTCTTATCTTAGTGACAGA
AATGGGATTATTCAGCGAAAAACAAATAAAGCTAAACCACTTCCCTTGTATAGAGAATCAAT
CTCTCCATTAACAAAAATTGAATTTGAGATTACAACAACCACTGATGAAGCTGGAAGATTG
ATTGAAGAATTAGGTAAGAGAGCACAAAGCGTTTTATAAAGACTATAAGGCATTTTCTCTATC
TGAATTTCTGATGATAAGATTCAAGCCAATCTACAATACCCAATTTATTTAGGTGCGGGGA
CGGTGCTTGGACAAAGACTCTATTTAAGCAAGCTGATGGTATTTTACAAAGACGATACAGT
CGAATGAAAACATAAATGGTTAAAAAAGGAGTTCTTAAGCTCACAAAAGCACCTCTTAAAA
CAGTTAAGATTCCATCTGGTAATCATTCAATTAGTCAAGAACCACGAGTCCTTTTATGAAATG
GGAAAAGCTAATTTTCATGATTAAGGAGATTGATAAATGA (SEQ ID NO:478)

SEQ ID NO:479:

ATGAGCGATTTATATAGTCAAAGGTCCAATTATTACCTGTCCTTATCTGAACAAAGAATTAT
CATTAATAAATGATAATAAAGAGATTGTCAAAGAAGTGTCCATTTCACTCGTTGATAATGTAT
TACTTTTTGGTAATGCACAACTGACCACCCAATCATCAAAGCCTTGTCAAAGAACAAGGTG
AATGTTTACTATTTCTCAAATGTTGGTCAATTTATTTCTAGTATTGAAACCCACAGGCAGGAC
GAATTCAAAAGCAAGAGTTGCAAGCAAAGGCTTATTTTGAAGAGGATTTCCGTTTAGAGG
TTGCGAGGAGTATTGCTACGACCAAGGTGAGGCACCCAATTGCCTTACTTAGAGAGTTTGAT
ACGGATGGTCTACTAGATACCTCAGATTATTCTAGGTTTGAAGATAGTGTCAATGATATTCA
GAAAGCTTATTCATTACAGAAATTATGGGTTACGAAGGTCGCCTTGCGAAATCCTATTTTT
ACTATCTGAATTTACTCGTTCCTAATGACTTTTCAATTTAATGGTAGGAGTAGACGGCCTGGG
GAGGATTGTTTTAACAGTGCCCTCAATTTTGGCTATAGTATCTTATATTCTTGCTTAATGGGC
TGA (SEQ ID NO:479)

SEQ ID NO:480:

TTGCTTAATGGGCTGATTAAGAAAAACGGGCTAAGCTTGGGATTTGGGGTAATTCACAAGC
ATCATCAGCATCATGCGACCTTGGCCAGTGATTTAATGGAAGAATGGAGACCTATCATCGTC
GATAATACGCTTATGGAGTTGGTACGAAATGGTAACTTCTTTAAGTCATTTTGAAAATAA
GGATCAAGACTTCATACTACCCATGAAGGCAGAGAAATCTTTCACGGGCTTTACGTTCAA
GAATATTAGAAGTCCATCAGTATATTGAGTTAGATAAAAAACGCTATTCTTTTCTTTATACA
GCAGATAGGCAAATCAAGAGTTTGATTAGGGCTTTAGAGAACTTGACCCTAGTCTCTATGA
GACAAGTTACACAGGAGGGCATTAA (SEQ ID NO:480)

SEQ ID NO:481:

ATGGGACTTTACTTTAACCTCAGCGAAGAAGAGCGTGAGTTTGCCAAACAAAAAACCATGT
 TTTGTCTGATTATTTATGATATTCGAAGTAACAAACGTAGACTTAAACTCTCGAAATTACTTG
 AGGGTTATGGCGTGAGGGTGCAAAAATCCTGTTTCGAAGTCAACCTGTCAAGAAATGATTA
 TCAGTCTCTCCTTAAGGATATCGAGGGCTTCTACAAGGCTGATGAAGAAGACAGCATAATA
 GTGTATGTGACAACCAAGAAGAGGTGACTAGTTTTAGCCCCTACCATAGTGCTGAAAAAT
 TAGATGACATTCTCTTCTTCTAA (SEQ ID NO:481)

SEQ ID NO:482:

ATGAAAAAATTAGTATTTACTTTTAAAAGGATCGACCATCCTGCACAAGATTTGGCTGTAA
 ATTTTCATGGCTTCTTGATGGAGCAGTTGGATAGTGACTATGTTGATTATCTGCATCAGCAGC
 AAACAAATCCCTATGCGACCAAGGTAATCCAAGGGAAAGAAAACACGCAGTGGGTTGTACA
 TCTGCTCACAGACGACATCGAGGATAAGGTTTTTATGACCTTATTACAGATTAAAGAGGTGT
 CCTTAAACGATCTGCCTAAACTCAGTGTGCAAAAAGTTGAGATTACAGGAGTTGGGGGCAGA
 TAAACTGTTAGAGATTTTCAATAGTGAGGAAAATCAAACCTATTTTTCAATTATTTTTGAGA
 CTCCAACAGGTTTTAAATCTCAAGGTTTCTACGTCATCTTCCCGTCTATGCGTTTTGATTTTT
 AAAGTTTGATGCAAAAGTATGGAAGGTTGGTTGAAAATCAAACCTGAAATTGAAGAGGATAC
 CTTAGATTACCTATCTGAACACAGCACTATCACGAATTATCGCTTGGAGACGAGTTATTTCA
 GGGTGACAGGCAACGAATTCCTGCCTTTAGAGGAAAGTTAACCTTTAAAGTACAAGGCGC
 CCAAACCTCTAAAAGCTTATGTCAAAATGCTTCTAACATTTCGGTGAATATTCAGGTCTTGGCA
 TGAACACGAGTCTCGGTATGGGAGGGATAAAGCTTGAAGAAAGAAAAGATTGA (SEQ ID
 NO:482)

SEQ ID NO:483:

TTGAAGAAAGAAAAGATTGATTTATTTTACGGAGCTCTTTTGCATGATATCGGTAAGGTCAT
 TCAAAGGGCGACAGGAGAACGAAAAAACACGCCTTGGTAGGCGCGGATTGGTTTGATGA
 GATTGCTGATAATCAAGTTATTTCCGATCAAATTAGATATCACATGGCTAACTACCAGAGTG
 ATAACTTGGAATGACCATCTTGCTTACATAACTTATATCGCTGATAACATTGCCTCTGGT
 GTCGACAGAAGACAGTCAAATGAGGAGAGTGACGAGGATACATCAGCTAAGATTTGGGAT
 ACCTATACAAACCAGGCTGATATTTTAAACGTTTTTGGGGCACAAACGGATAAACGCTACTT
 TAAACCGACGGTTCTAAACTTGAAATCTAAACCTAACTTTGCGTCGGCAACATATGAACCTT
 TCTCAAAGGTGATTATGCGGCAATTGCGACTCGTATCAAAAATGAATTGGCAGAATTTGA
 GTTTAATCAAGTACAAATTGACTCTTTGTTAAATCTGTTCTGAAGCAACCCCTCTCTTTGTGCC
 TTCTTCGACTAATACTAAAGAAATCGCTGATATTTCACTTGCTGATCATAGTCGCTGACAG
 CAGCTTTTGCTCTAGCCATCTATGATTACTTGGAAGACAAAGGTCGTCATAACTATAAGGAG
 GACTTGTTTACTAAAGCATCAGCCTTTTATGAGGAAGAAGCTTTTCTCCTAGCTAGCTTTGAC
 TTATCAGGGATTCAAGACTTTATCTATAATATTAATATTGCGACGAATGGTGCTGCTAAACA
 ATTGAAGGCTAGATCTTTATATCTTGACTTTATGAGCGAGTATATAGCAGACAGTTTACTTG
 ATAACTAGGCCTCAATCGGGCTAATATGCTCTATGTGCGTGGGGGACATGCTTACTTTGTC
 CTAGCCAATACTGAAAAACGGTAGAAACACTCGTTCAATTTGAAAAAGATTTCAATCAAT
 TTTTATTGGCAAATTTCCAAACCAGATTATATGTTGCCTTTGGTTGGGGAAGCTTTGCGGCTA
 AGGATATCATGAGCGAACTGAACTCACCTGAAAGCTATAGACAGGTCTATCAAAAGGCTAG
 TCGCATGATTTCTGAGAAAAAATCTCAAGGTATGATTATCAAAACCCCTATGTTGTTGAACA
 GGGGCGGTAAATCTTCTGAAAGAGAGTGCGAGATTTGTCATTCCGTTGAGAATTTAGTTGCT
 TATCATGACCAAAAAGTGTGTGACATTTGTCGAGGCTTGATCAATTTTCTAAAGAGATTGC
 CCATGACCATTTTATTATCACTGAAAATGAAGGGCTTCTTATTGGTCCGAACGCATGTCTTA
 AGGGTGTGCAATTTGAAAAGCTGAGCCAAGAAGCTTTTCCCGTGTCTATGTCAAAAATGAC
 TATAAGGCTGGTACAGTTAAGGCAACCCATGTTTTTGTGGAGATTACAGTATGATGAAAT
 ATACAATTATGCTGCCTTATCTAAAAACGAAAATGGGTTAGGTATTAACGTTTAGCTGTG
 TACGTCTTGACGTGGATGATTTGGGAGCAGCCTTTATGGCTGGCTTCTCCCAACAAGGAAAT
 GGGCAATATAGTACTCTATCACGCTCAGCCACTTCTCTCGAAGCATGAGTCTTTTCTTCAAG
 GTTTATATTAACCAGTTTGCTAGTGATAAGAAGCTCTCTATCATCTATGCCGGTGGGGATGA
 TGTTTTTGCTATTGGCTCTTGGCAAGATATTATTGCCTTTACTGTTGAACTTCGTGAGAACTT

CATTAAATGGACAAATGGAAAACCTAACACTATCAGCTGGTATCGGTCTGTTTGCTGATAAGA
 CCCCTATTAGCTTAATGGCACATCAAACAGGGGAGCTAGAAGAAACAGCTAAAGGCAATGA
 GAAAGATAGTATTTCACTCTTTAGTTCCGACTATACCTTTAAATTTGATCGGTTTATCACTAA
 TGTTTACGACGATAAGTTAGAGCAGATTTCGCTATTTCTTTAATCACCAAGATGAACGAGGCA
 AGAATTTCAATTTATAAATTGATTGAATTGCTTCGAAATTATGATCGTATGAATATGGCACGT
 TTAGCTTATTATTTAACACGACTTGAAGAATTGACGCGTGAAACAGACAGGGATAAATTTAA
 AACATTTAAAAATTTATTCTATTCTTGGTACACAAATAAGGATGATAAGGATAGAAAAGAA
 GCAGAGTTAGCCTTGCTTCTCTATATCTATGAGATTAGAAAGGATTAG (SEQ ID NO:483)

_ SEQ ID NO:484:

ATGACAATCTTGACTGATGAGAATTACGTTGATATTGCAGAAAAAGCAATTCTAAAACCTAG
 AAAGAAATACTAGGAACAGAAAGAATCCTGATGCCTTCTTTCTTACAACAAGTAAGCTCAG
 AACTTGCTGAGCTTAAGTAGTACACTTTTGTATGAGAGTAAGGTCAAAGAATATGATGCTC
 TCCTTGATCGTATTGCTTATTTAAGAGTACAATTTGTCTACCAAGCAGGTAGAGAGATTGCA
 GTAAAAGATCTGATAGAAAAGGCTCAAATTTCTTGAGGCTCTTAAGGAAATCAAAGATAGAG
 AGACACTTCAAAGATTTTGTAGATATATGGAAGCATTAGTAGCCTATTTCAAGTTTATGGA
 GGTAAAGATTAA (SEQ ID NO:484)

SEQ ID NO:485:

ATGACATTTCGCTAAGATTAAATTTTCAGCTCAAATTCGTTTAGAGACAGGCCTCCATATTGG
 TGAAGCGATGCTTTTGCAGCCATTGGTGCAATCGATTTCGCTGTTATTAAAGATCCTATTA
 CCAACCTACCGATCATTCTCGGTTCAAGTCTCAAAGGAAAAATGAGAACGCTTCTTGCCAAG
 GTTTATAATGAAAAGGTAGCTGAGAAACCAAGCGATGACAGTGATATTCTTAGCCGTTTATT
 TGGAATAGTAAAGATAAACGATTCAAATGGGACGCTTGATTTTTCTGTATGCCTTCTTGT
 CAAACGCTGATGAGCTAGACTCTTGGGGTAAGAAGTTATACAGAAGTAAAATTTGAAAA
 TACAATTGACCGTATCACTGCCGAAGCTAATCCAAGACAAATTGAACGTGCTATTCGTACCA
 GTACTTTTGATTTTCGAGTTGATTTATGAAATTACAGATGAGAATGAAAATCAAGTCGAAGAA
 GATTTCAAAGTGATTCGAGATGGTTTAAAACCTGCTTGAACCTGATTATCTTGGTGGTTCTGG
 ATCTCGAGGTTACGGTAAGGTTGCTTTTGA AAAACTCAAAGCTACTACCGTATTTGGTAATT
 ATGATGTTAAACATTAAATGAACTTTTAACTGCGGAGGTCTAA (SEQ ID NO:485)

SEQ ID NO:486:

ATGACCTATAAACTGTATATTATGACCTTTCAGAATGCTCATTTTGGTTTCGGGCACTCTTGAT
 AGCTCAAAATTAACATTCTCAGCAGACCGTATCTTCTCAGCACTAGTGCTAGAATCCCTAAA
 AATGGGAAAACCTCGATGCATTTCTTGCGGAAGCTAACCAAGACAAGTTCACGCTCACAGAT
 GCCTTTCCATTTCAATTTGGTCCCTTTTTGCCGAAACCGATTGGTTATCCCAAACATGACCAA
 ATAGATCAATCAGTTGATGTCAAAGAGGTTTCGCCGTCAAGCAAAATTGTCTAAGAACTGC
 AATTTCTTGCTCTAGAAAATGTTGACGATTATCTCAATGGAGAGTTATTTGAAAATGAAGAG
 CATGCAGTCATCGATACTGTGACAAAAAATCAACCACATAAGGACGGCAATCTTTATCAGG
 TAGCTACAACCAGATTTTCAAATGATACGTCGCTTTACGTCATCGCAAACGAATCTGATTTG
 CTTAATGAGTTGATGTCTAGTCTTCAGTATTCAGGTCTTGGTGGAAAGCGTTCAAGTGGTTTT
 GGTCGTTTTGAGTTAGATATTCAAAATATCCCACTAGAATTGTCAGATAGACTGACTAAGAA
 TCATTGAGATAAAGTGATGAGTCTTACGACAGCACTTCCTGTAGATGCTGACCTTGAAGAAG
 CAATGGAAGATGGACATTACTTATTAATAAATCAAGTGGTTTTGCATTTAGTCATGCTACC
 AATGAGAATTATCGTAAGCAGGATCTTTACAAATTTGCTTCTGGTTCAACTTTTAGTAAAC
 ATTTGAAGGTCAGATTGTTGATGTGAGACCACTTGATTTCCCTCATGCTGTTTTAAATTATGC
 TAAACCACTCTTCTTTAAATTGGAGGTATAA (SEQ ID NO:486)

SEQ ID NO:487

ATGAAAAATGACTATAGAACATTTAAATTAAGCCTCCTGACACTTGCTCCAATTCATATTGG
 TAATGGAGAGAAGTATACCTCTAGAGAATTTATCTATGAAAATAAGAAGTTTACTTTCTTG
 ACATGGGGAAATTTCTATAATAAAATGGTGGAGAAGAGGCTTGCTGAAAAGTTTGAAGCATT
 TCTAATTCAAACTCGTCCAAATGCACGTAATAATCGTCTTATTTCTTCTTAAATGATAACCG

AATTGCAGAGCGTTCTTTTGGAGGTTATAGTATCTCTGAAACAGGTTTAGAATCGGACAAAA
ATCCTGATTACGCCGGAGCTATTAACGAAGTTAATAAATTTATTCGAGATGCTTTTGGAAAT
CCCTACATTCCTGGTAGCTCACTAAAAGGTGCTATTCGTACCATTTTAATGAATACTACCCCT
AAGTGGAAATAATGAAAAAGCTGTAAGTACTTTGGAAGATTCCGAAAGAGAATAAGAACC
TTATCCCTTGGGGACCAAAAAAGGGAAAAAGAAATACGATGATTTGTTTAACGCAATTCGTGT
GAGTGATAGTAAGCCTTTTGATAATAAGAGTCTTATCTTAGTGACAGAAATGGGATTATTCAG
CGAAAACAAATAAAGCTAAACCACTTCCCTTGTATAGAGAATCAATCTCTCCATTAACAAA
AATTGAATTTGAGATTACAACAACCACTGATGAAGCTGGAAGATTGATTGAAGAATTAGGT
AAGAGAGCACAAGCGTTTATAAAGACTATAAGGCATTTTCTATCTGAATTTCTCTGATGA
TAAGATTCAAGCCAATCTACAATACCCAATTTATTTAGGTGCGGGGAGCGGTGCTTGGACAA
AGACTCTATTTAAGCAAGCTGATGGTATTTACAAAGACGATACAGTCGAATGAAAATAA
AATGGTTAAAAAAGGAGTTCTTAAGCTCACAAAAGCACCTCTTAAACAGTTAAGATTCCA
TCTGGTAATCATTCAATTAGTCAAGAACCACGAGTCCTTTTATGAAATGGGAAAAGCTAATT
CATGATTAAGGAGATTGATAAATGA (SEQ ID NO:487)

con secuencias repetidas: SEQ ID NO: 11 y/o SEQ ID NO: 12

Combinación funcional #3:

5

secuencias cas: SEQ ID NO:488 hasta SEQ ID NO:508, y SEQ ID NO:517 hasta SEQ ID NO:521, como se muestra a continuación. Las SEQ ID NOS:488-497 son de *S. agalactiae*, mientras que las SEQ ID NOS:498-503 son de *S. mutans*, y las SEQ ID NOS: 504-508, 517-521 son de *S. pyogenes*.

10

SEQ ID NO:488:

ATGAATAAGCCATATTCAATAGGCCTTGACATCGGTACTAATTCGTCGGATGGAGCATTAT
TACAGATGATTATAAAGTACCTGCTAAGAAGATGAGAGTTTTAGGGAACACTGATAAAGAA
TATATTAAGAAGAATCTCATAGGTGCTCTGCTTTTTGATGGCGGGAATACTGCTGCAGATAG
ACGCTTGAAGCGAACTGCTCGTCGTCGTTATACACGTCGTAGAAATCGTATTCTATATTTAC
AAGAAATTTTTGCAGAGGAAATGAGTAAAGTTGATGATAGTTTCTTTTCATCGATTAGAGGAT
TCTTTTCTAGTTGAGGAAGATAAGAGAGGGAGCAAGTATCCTATCTTTGCAACATTGCAGGA
AGAGAAAGATTATCATGAAAAATTTTCGACAATCTATCATTGAGAAAAGAATTAGCTGAC
AAGAAAGAAAAAGCAGACCTTCGTCCTATTTATATTGCTCTAGCTCATATCATTAAATTTAG
AGGGCATTTCCTAATTGAGGATGATAGCTTTGATGTCAGGAATACAGACATTTCAAAACAAT
ATCAAGATTTTTTAGAAATCTTTAATAACAACTTTTGAAAATAATGATTTGTTATCTCAAAACG
TTGACGTAGAGGCAATACTAACAGATAAGATTAGCAAGTCTGCGAAGAAAGATCGTATTTT
AGCGCAGTATCCTAACCAAAAAATCTACTGGCATTTTTTGCAGAATTTTTGAAATTGATTGTCG
GAAATCAAGCTGACTTCAAGAAATATTTCAATTTGGAGGATAAAACGCCGCTTCAATTCGCT
AAGGATAGCTACGATGAAGATTTAGAAAATCTTCTTGGACAGATTGGTGTATGAATTTGCAG
ACTTATTCTCAGCAGCGAAAAAGTTATATGATAGTGTCTTTTGTCTGGCATTCTTACAGTAA
TCGACCTCAGTACCAAGGCGCCACTTTCAGCTTCTATGATTACGCGTTATGATGAACATAGA
GAGGACTTGAAACAGTTAAAACAATTCGTAAAAGCTTCATTGCCGGAATAATCAAGAAA
TATTTGCTGATTCATCAAAAGATGGCTACGCTGGTTATATTGAAGGTAAACTAATCAAGAA
GCTTTTTATAAATACCTGTCAAAATTGTTGACCAAGCAAGAAGATAGCGAGAATTTTCTTGA
AAAAATCAAGAATGAAGATTTCTTGAGAAAACAAAGGACCTTTGATAATGGCTCAATTCCA
CACCAAGTCCATTTGACAGAGCTGAAAGCTATTATCCGCCGTCAATCAGAATACTATCCCTT
CTTGAAAGAGAATCAAGATAGGATTGAAAAAATCCTTACCTTTAGAATTCCTTATTATATCG
GGCCACTAGCACGTGAGAAGAGTGATTTTGCATGGATGACTCGCAAAACAGATGACAGTAT
TCGACCTTGGAATTTTGAAGACTTGGTTGATAAAGAAAAATCTGCGGAAGCTTTTATCCATC
GTATGACCAACAATGATTTTTATCTTCTGAAGAAAAAGTTTTACCAAGCATAGTCTTATT
TATGAAAAATTTACGGTCTATAATGAGTTGACTAAGGTTAGATATAAAAAATGAGCAAGGTG
AGACTTATTTTTTTGATAGCAATATTAACAAGAAAATCTTTGATGGAGTATTCAAGGAACAT
CGTAAGGTATCCAAGAAGAAGTTGCTAGATTTTCTGGCTAAAGAATATGAGGAGTTTAGGA
TAGTAGATGTTATTGGTCTAGATAAAGAAAAATAAAGCTTTCAACGCCCTCATTGGGAACCTTAC
CACGATCTCGAAAAAATACTAGACAAAGATTTTCTAGATAATCCAGATAATGAGTCTATTCT
GGAAGATATCGTCCAAACTCTAACATTATTTGAAGACAGAGAAATGATTAAGAAGCGTCTT

GAAAAC TATAAAGATCTTTTTACAGAGTCACAAC TAAAAAACTCTATCGTCGTCACTATAC
 TGGCTGGGGACGATTGTCTGCTAAGTTAATCAATGGTATTTCGAGATAAAGAGAGTCAAAAA
 ACAATCTTGGACTATCTTATTGATGATGGTAGATCTAATCGCAACTTTATGCAGTTGATAAA
 TGATGATGGTCTATCTTTCAAATCAATTATCAGTAAGGCACAGGCTGGTAGTCATTGAGATA
 ATCTAAAAGAAGTTGTAGGTGAGCTTGCAGGTAGCCCTGCTATTAAAAAGGGAATTCTACA
 AAGTTTGAAAATTGTTGATGAGCTTGTAAAGTCATGGGATACGAACCTGAACAAATTGTGG
 TTGAGATGGCGCGTGAGAATCAAACAACAAATCAAGGTCGTCGTAACCTCTCGACAACGCTA
 TAAACTTCTTGATGATGGCGTTAAGAATCTAGCTAGTGACTTGAATGGCAATATTTTGAAAG
 AATATCCTACGGATAATCAAGCGTTGCAAAATGAAAGACTTTTCTTTACTACTTACAAAAC
 GGAAGAGATATGTATACAGGGGAAGCTCTAGATATTGACAATTTAAGTCAATATGATATTG
 ACCACATTATTCCTCAAGCTTTCATAAAAGATGATTCTATTGATAATCGTGTTTTGGTATCAT
 CTGCTAAAAATCGTGGAAGTCAGATGATGTTTCTAGCCTTGAAATTGTAAAAGATTGTAAA
 GTTTTCTGGAAAAAATTACTTGATGCTAAGTTAATGAGTCAGCGTAAGTATGATAATTTGAC
 TAAGGCAGAGCGCGGAGGCCTAACTTCCGATGATAAGGCAAGATTTATCCAACGTCAGTTG
 GTTGAGACACGACAAATTACCAAGCATGTTGCCCGTATCTTGATGAACGCTTTAATAATGA
 GCTTGATAGTAAAGGTAGAAAGGATCCGCAAAGTTAAATTTGTAACCTTGAAGTCAAATTTG
 GTTTCAAATTTCCGAAAAGAATTTGGATTCTATAAAATTCGTGAAGTTAACAATTATCACCA
 TGCACATGATGCCTATCTTAATGCAGTAGTTGCTAAAGCTATTCTAACCAAAATATCCTCAGT
 TAGAGCCAGAATTTGTCTACGGCGACTATCCAAAATATAATAGTTACAAAACGCGTAAATC
 CGCTACAGAAAAGCTATTTTTCTATTCAAATATTATGAACCTCTTTAAAACTAAGGTAACCTT
 AGCGGATGGAACCGTTGTTGTAAAAGATGATATTGAAGTTAATAATGATACGGGTGAAATT
 GTTTGGGATAAAAAAGAAACACTTTGCGACAGTTAGAAAAGTCTTGTACATACCCTCAGAACA
 ATATCGTGAAGAAGACAGAGATTCAGACAGGTGGTTTCTCTAAGGAATCAATCTTGGCGCA
 TGGTAACTCAGATAAGTTGATTCCAAGAAAAACGAAGGATATTTATTTAGATCCTAAGAAA
 TATGGAGGTTTTGATAGTCCGATAGTAGCTTACTCTGTTTTAGTTGTAGCTGATATCAAAAA
 GGGTAAAGCACAAAAACTAAAAACAGTTACGGAACCTTTAGGAATTACCATCATGGAGAGG
 TCCAGATTTGAGAAAAATCCATCAGCTTTCTTGAATCAAAAGGCTATTTAAATATTAGGGC
 TGATAAACTAATTATTTTGCCCAAGTATAGTCTGTTTGAATTAGAAAAATGGGCGTCGTCGAT
 TACTTGCTAGTGCTGGTGAATTACAAAAAGGTAATGAGCTAGCCTTACCAACACAATTTATG
 AAGTTCTTATACCTTGCAAGTCGTTATAATGAGTCAAAAGGTAAACCGAGGAGATTGAGA
 AGAAACAAGAATTTGTAAATCAACATGTCTCTTATTTTGATGACATCCTTCAATTAATTAAT
 GATTTTTCAAAACGAGTTATTCTAGCAGATGCTAATTTAGAGAAAAATCAATAAGCTTTACCA
 AGATAATAAGGAAAAATATATCAGTAGATGAACCTTGCTAATAATATTATCAATCTATTACTT
 TTACCAGTCTAGGAGCTCCAGCAGCTTTTAAATTTTTTGATAAAATAGTTGATAGAAAACGC
 TATACATCAACTAAAGAAGTACTTAATTCTACCCTAATTCATCAATCTATTACTGGACTTTAT
 GAAACACGTATTGATTTGGGTAAAGTTAGGAGAAGATTGATATGGCAGGTTGGCGAACCGTT
 GTTGTAATAACACATTCTAAGCTCTCTTATAAAAAATAATCATCTGATTTTTAAAGATTCTTAT
 CAGACGGAAATGATTCATCTATCAGAGATTGACATTCTAATCATGGAAACAACAGATATCG
 TTTTGTGACCATGCTGATTAAACGTTTGGTTGATGAAAATATTTTAGTTATATTTTGTGACG
 ATAAACGCTTGCCAACAGCTATGTTAATGCCGTACTATGCCAGACATGATTCGAGTTTACAA
 TTATCTAGGCAGATGTCATGGATTGAAGATGTCAAAGCAGATGTTTGGACATCAATTATTGC
 ACAAAAAATTTTGAATCAGTCTTTTTATCTCGGTGAGTGTTCTTTCTTTGAAAAATCCCAGTC
 TATTATGAATCTCTACCATGACTTAGAACCTTTTGATCCTTCTAATCGTGAGGGGCATGCTGC
 TAGGATTTATTTCAATACACTTTTTGGAAATGATTTTTCAAGAGAGCAGGATAATCCAATAA
 ATGCTGGTTTAGACTACGGATATTCATTGCTTTTGTAGTATGTTTGGCGGTGAAGTTGTTAAGT
 GTGGTTGCATGACACAATTTGGCTTGAAGCATGCTAATCAATTTAATCAGTTCAACCTAGCA
 AGCGATATTATGGAACCATTTCCGCCAATCGTTGATAGGATTATTTATGAAAATAGGCAGAG
 TGATTTTGTCAAATGAAAAGAGAAGTCTTTTCTATGTTTTAGAGACATACAGCTACAATG
 GTAAAGAAATGTATCTCTCAAATATTGTCTAGCGACTATACCAAAAAAGTTATTAAGTCGCTA
 AATAGTGATGGGAATGGAATTCGGGAGTTTAGGATATGAGTTATCGGTATATGCGAATGATT
 TTAATGTTTGATATGCCTACTGAAACAGCAGAAGAACGGAAGGCGTATCGTAAGTTTAGAA
 AGTTTCTCTTGAGCGAAGGCTTTATCATGCATCAGTTTTCTGTTTATAGTAAATTATTACTCA
 ATAATACAGCTAATAATGCTATGATAGGTCGGCTTAAAGTGAATAATCCTAAAAAGGGTAA
 TATCACACTCTTAACAGTTACGGAAAAACAATTTGCGAGAATGGTTTACCTCATGGAGAAC
 GCAACACAAGTGTTGCCAACTCTGATAGTCGCTTGGTTTTCTAGGAGATTCTTATGATCAA

GATTAATTTTCCAATTTTAGATGAACCATTAGTGTTAAGTAATGCTACGATTTTAACGATAG
 AAGATGTTTCAGTTTATTCTTCATTGGTGAAACATTTTATCAATATGACGTAGATGAACATT
 TGAAATTATTTGATGATAAGCAGAAAAGTCTGAAGGCAACAGAGTTAATGCTGGTTACAGA
 TATCTTAGGATACGATGTCAACTCAGCACCTATTCTAAAGTTGATACATGGTGACTTAGAAA
 ATCAATTCAACGAAAAGCCAGAAGTGAAATCAATGGTAGAAAAATTAGCAGCTACTATTAC
 AGAACTTATCGCATTTGAGTGTCTAGAGAATGAGCTTGATTTAGAATACGATGAAATTAAGA
 TTTTAGAACTCATTAAGGCACTGGGAGTCAAAATTGAGACACAGAGCGACACTATCTTTGA
 AAAATGTTTTGAAATTATACAAGTTTACCATTATTTAACGAAAAAGAATCTCTTGGTTTTTGT
 TAATAGCGGAGCTTATCTTACCAAAGATGAAGTTATAAAATTATGTGAATACATCAATTTAA
 TGCAAAAGTCAGTACTCTTTCTAGAACCTAGAAGACTCTATGATTTACCGCAATATGTTATT
 GATAAGGATTATTTCTTGATAGGCGAAAATATGGTATAA (SEQ ID NO:488)

SEQ ID NO:489:

ATGAATAAGCCATATTCAATAGGCCTTGACATCGGTACTAATCCGTCGGATGGAGCATTAT
 TACAGATGATTATAAAGTACCTGCTAAGAAGATGAGAGTTTTAGGGAACACTGATAAAGAA
 TATATTAAGAAGAATCTCATAGGTGCTCTGCTTTTTGATGGCGGGAATACTGCTGCAGATAG
 ACGCTTGAAGCGAACTGCTCGTCGCTTATACACGTCGTAGAAATCGTATTCTATATTTAC
 AAGAAATTTTTGCAGAGGAAATGAGTAAAGTTGATGATAGTTTCTTTCATCGATTAGAGGAT
 TCTTTTCTAGTTGAGGAAGATAAGAGAGGGAGCAAGTATCCTATCTTTGCAACATTGCAGGA
 AGAGAAAGATTATCATGAAAAATTTTCGACAATCTATCATTTGAGAAAAGAATTAGCTGAC
 AAGAAAGAAAAAGCAGACCTTCGTCTTATTTATATTGCTCTAGCTCATATCATTAATTTAG
 AGGGCATTTCCTAATTGAGGATGATAGCTTTGATGTCAGGAATACAGACATTTCAAAACAAT
 ATCAAGATTTTTTAGAAATCTTTAATACAACCTTTGAAAATAATGATTTGTTATCTCAAAACG
 TTGACGTAGAGGCAATACTAACAGATAAGATTAGCAAGTCTGCGAAGAAAGATCGTATTTT
 AGCGCAGTATCCTAACCAAAAAATCTACTGGCATTTTTGCAGAAATTTTTGAAATTGATTGTCG
 GAAATCAAGCTGACTTCAAGAAATATTTCAATTTGGAGGATAAAACGCCGCTTCAATTCGCT
 AAGGATAGCTACGATGAAGATTTAGAAAATCTTCTTGGACAGATTGGTGATGAATTTGCAG
 ACTTATTCTCAGCAGCGAAAAAGTTATATGATAGTGTCTTTTGTCTGGCATTCTTACAGTAA
 TCGACCTCAGTACCAAGGCGCCACTTTCAGCTTCTATGATTTCAGCGTTATGATGAACATAGA
 GAGGACTTGAAACAGTTAAACAATTCGTAAAGCTTCATTGCCGGAATAATCAAGAAA
 TATTTGCTGATTATCAAAAGATGGCTACGCTGGTTATATTGAAGGTAAAACTAATCAAGAA
 GCTTTTTATAAATACCTGTCAAAATTGTTGACCAAGCAAGAAGATAGCGAGAATTTTCTTGA
 AAAAATCAAGAATGAAGATTTCTTGAGAAAACAAAGGACCTTTGATAATGGCTCAATTCCA
 CACCAAGTCCATTTGACAGAGCTGAAAGCTATTATCCGCCGTCAATCAGAATACTATCCCTT
 CTTGAAAGAGAATCAAGATAGGATTGAAAAAATCCTTACCTTTAGAATTCTTATTATATCG
 GGCCACTAGCACGTGAGAAGAGTGATTTTGCATGGATGACTCGCAAAACAGATGACAGTAT
 TCGACCTTGGAATTTTGAAGACTTGGTTGATAAAGAAAAATCTGCGGAAGCTTTTATCCATC
 GTATGACCAACAATGATTTTTATCTTCTGAAGAAAAAGTTTTACCAAAGCATAGTCTTATT
 TATGAAAAATTTACGGTCTATAATGAGTTGACTAAGGTTAGATATAAAAAATGAGCAAGGTG
 AGACTTATTTTTTTGATAGCAATATTAACAAGAAATCTTTGATGGAGTATTCAAGGAACAT
 CGTAAGGTATCCAAGAAGAAGTTGCTAGATTTTCTGGCTAAAGAATATGAGGAGTTTAGGA
 TAGTAGATGTTATTGGTCTAGATAAAGAAAAATAAGCTTTCAACGCCTCATTGGGAACCTTAC
 CAGGATCTCGAAAAAATACTAGACAAAGATTTTCTAGATAATCCAGATAATGAGTCTATTCT
 GGAAGATATCGTCCAAACTCTAACATTATTTGAAGACAGAGAAATGATTAAGAAGCGTCTT
 GAAAACTATAAAGATCTTTTTACAGAGTCACTAACTAAAAAACTCTATCGTCGTCCTATAC
 TGGCTGGGGACGATTGTCTGCTAAGTTAATCAATGGTATTCGAGATAAAGAGAGTCAAAAA
 ACAATCTTGGACTATCTTATTGATGATGGTAGATCTAATCGCAACTTTATGCAGTTGATAAA
 TGATGATGGTCTATCTTTCAAATCAATTATCAGTAAGGCACAGGCTGGTAGTCATTACAGATA
 ATCTAAAAGAAGTTGTAGGTGAGCTTGCAGGTAGCCCTGCTATTAAGGAAGGAATTCTACA
 AAGTTTGAAAATTGTTGATGAGCTTGTAAAGTCATGGGATACGAACCTGAACAAATTGTGG
 TTGAGATGGCGCGTGAGAATCAAACAACAAATCAAGGTGCTCGTAACCTCTCGACAACGCTA
 TAAACTTCTTGATGATGGCGTTAAGAATCTAGCTAGTGAATGGCAATATTTTGAAAG
 AATATCCTACGGATAATCAAGCGTTGCAAAATGAAAGACTTTTCTTTACTACTTACAAAAC
 GGAAGAGATATGTATACAGGGGAAGCTCTAGATATTGACAATTTAAGTCAATATGATATTG
 ACCACATTATTCCTCAAGCTTTCATAAAAGATGATTCTATTGATAATCGTGTTTTGGTATCAT

CTGCTAAAAATCGTGGAAGTCAGATGATGTTCTAGCCTTGAAATTGTAAAAGATTGTAAA
 GTTTTCTGGAAAAAATTACTTGATGCTAAGTTAATGAGTCAGCGTAAGTATGATAATTTGAC
 TAAGGCAGAGCGCGGAGGCCTAACTTCCGATGATAAGGCAAGATTTATCCAACGTCAGTTG
 GTTGAGACACGACAAATTACCAAGCATGTTGCCCGTATCTTGGATGAACGCTTTAATAATGA
 GCTTGATAGTAAAGGTAGAAGGATCCGCAAAGTTAAAATTGTAACTTGAAGTCAAATTTG
 GTTTCAAAATTTCCGAAAAGAATTTGGATTCTATAAAATTCGTGAAGTTAACAATTATCACCA
 TGCACATGATGCCTATCTTAATGCAGTAGTTGCTAAAGCTATTCTAACCATAATCCTCAGT
 TAGAGCCAGAATTTGTCTACGGCGACTATCCAAAATATAATAGTTACAAAACGCGTAAATC
 CGCTACAGAAAAGCTATTTTTCTATTCAAATATTATGAACTTCTTTAAACTAAGGTAACTTT
 AGCGGATGGAACCGTTGTTGTAAAAGATGATATTGAAGTTAATAATGATACGGGTGAAATT
 GTTTGGGATAAAAAGAAACACTTTGCGACAGTTAGAAAAGTCTTGTACATACCCTCAGAACCA
 ATATCGTGAAGAAGACAGAGATTCAGACAGGTGGTTTCTCTAAGGAATCAATCTTGGCGCA
 TGGTAACTCAGATAAGTTGATTCCAAGAAAAACGAAGGATATTTATTTAGATCCTAAGAAA
 TATGGAGGTTTTGATAGTCCGATAGTAGCTTACTCTGTTTTAGTTGTAGCTGATATCAAAAA
 GGGTAAAGCACAAAAACTAAAAACAGTTACGGAACCTTTAGGAATTACCATCATGGAGAGG
 TCCAGATTTGAGAAAAATCCATCAGCTTTTCTTGAATCAAAAAGGCTATTTAAATATTAGGGC
 TGATAAACTAATTATTTTGCCCAAGTATAGTCTGTTTGAATTAGAAAATGGGCGTCGTCGAT
 TACTTGCTAGTGCTGGTGAATTACAAAAAGGTAATGAGCTAGCCTTACCAACACAATTTATG
 AAGTTCTTATACCTTGCAAGTCGTTATAATGAGTCAAAAAGGTAAACCAGAGGAGATTGAGA
 AGAAACAAGAATTTGTAAATCAACATGTCTCTTATTTTGATGACATCCTTCAATTAATTAAT
 GATTTTTCAAAACGAGTTATTCTAGCAGATGCTAATTTAGAGAAAATCAATAAGCTTTACCA
 AGATAATAAGGAAAATATATCAGTAGATGAACTTGCTAATAATATTATCAATCTATTTACTT
 TTACCAGTCTAGGAGCTCCAGCAGCTTTTAAATTTTTTGATAAAATAGTTGATAGAAAACGC
 TATACATCAACTAAAGAAGTACTTAATTCTACCCTAATTCATCAATCTATTACTGGACTTTAT
 GAAACACGTATTGATTTGGGTAAAGTTAGGAGAAGATTGA (SEQ ID NO:489)

SEQ ID NO:490:

ATGGCAGGTTGGCGAACCGTTGTTGTAAATACACATTCTAAGCTCTCTTATAAAAAATAATCA
 TCTGATTTTTAAAGATTCTTATCAGACGGAAATGATTCATCTATCAGAGATTGACATTCTAAT
 CATGGAAACAACAGATATCGTTTTGTCGACCATGCTGATTAAACGTTTGGTTGATGAAAATA
 TTTTAGTTATATTTTGTGACGATAAACGCTTGCCAACAGCTATGTTAATGCCGTACTATGCCA
 GACATGATTCGAGTTTACAATTATCTAGGCAGATGTCATGGATTGAAGATGTCAAAGCAGAT
 GTTTGGACATCAATTATTGCACAAAAAATTTTGAATCAGTCTTTTTATCTCGGTGAGTGTCT
 TTCTTTGAAAAATCCCAGTCTATTATGAATCTCTACCATGACTTAGAACCTTTTGATCCTTCT
 AATCGTGAGGGGCATGCTGCTAGGATTTATTTCAATACACTTTTTGGAAATGATTTTTCAAG
 AGAGCAGGATAATCCAATAAATGCTGGTTTAGACTACGGATATTCATTGCTTTTGAGTATGT
 TTGCGCGTGAAGTTGTTAAGTGTGGTTGCATGACACAATTTGGCTTGAAGCATGCTAATCAA
 TTTAATCAGTTCAACCTAGCAAGCGATATTATGGAACCATTTGCCCCAATCGTTGATAGGAT
 TATTTATGAAAATAGGCAGAGTGATTTTGTCAAAATGAAAAGAGAACTCTTTTCTATGTTTT
 CAGAGACATACAGCTACAATGGTAAAGAAATGTATCTCTCAAATATTGTCAGCGACTATAC
 CAAAAAAGTTATTAAGTCGCTAAATAGTGATGGGAATGGAATTCCGGAGTTTAGGATATGA
 (SEQ ID NO:490)

SEQ ID NO:491:

ATGCGAATGATTTTAATGTTTGATATGCCTACTGAAACAGCAGAAGAACGGAAGGCGTATC
 GTAAGTTTAGAAAGTTTCTCTTGAGCGAAGGCTTTATCATGCATCAGTTTCTGTTTATAGTA
 AATTATTACTCAATAATACAGCTAATAATGCTATGATAGGTGCGCTTAAAGTGAATAATCCT
 AAAAAGGGTAATATCACACTCTTAACAGTTACGGAAAAACAATTTGCGAGAATGGTTTACC
 TCCATGGAGAACGCAACACAAGTGTGCGCAACTCTGATAGTCGCTTGGTTTTCTAGGAGAT
 TCTTATGATCAAGATTAA (SEQ ID NO:491)

SEQ ID NO:492:

ATGATCAAGATTAATTTTCCAATTTTAGATGAACCATTAGTGTTAAGTAATGCTACGATTTTA
 ACGATAGAAGATGTTTCAGTTTATTCTTCATTGGTGAAACATTTTTATCAATATGACGTAGAT

GAACATTTGAAATTATTTGATGATAAGCAGAAAAGTCTGAAGGCAACAGAGTTAATGCTGG
 TTACAGATATCTTAGGATACGATGTCAACTCAGCACCTATTCTAAAGTTGATACATGGTGAC
 TTAGAAAATCAATTCAACGAAAAGCCAGAAAGTGAAATCAATGGTAGAAAAATTAGCAGCTA
 CTATTACAGAACTTATCGCATTTGAGTGTCTAGAGAATGAGCTTGATTTAGAATACGATGAA
 ATTAAGATTTTAGAACTCATTAAGGCACTGGGAGTCAAAATTGAGACACAGAGCGACACTA
 TCTTTGAAAAATGTTTTGAAATTATACAAGTTTACCATTATTTAACGAAAAAGAATCTCTTG
 GTTTTGTTAATAGCGGAGCTTATCTTACCAAAGATGAAGTTATAAAATTATGTGAATACAT
 CAATTTAATGCAAAAGTCAGTACTCTTTCTAGAACCTAGAAGACTCTATGATTTACCGCAAT
 ATGTTATTGATAAGGATTATTTCTTGATAGGCGAAAAATATGGTATAA (SEQ ID NO:492)

SEQ ID NO:493:

ATGAATAAGCCATATTCAATAGGCCTTGACATCGGTACTAATTCCGTCGGATGGAGCATTAT
 TACAGATGATTATAAAGTACCTGCTAAGAAGATGAGAGTTTTAGGGAACACTGATAAAGAA
 TATATTAAGAAGAATCTCATAGGTGCTCTGCTTTTTGATGGCGGAATACTGCTGCAGATAG
 ACGCTTGAAGCGAACTGCTCGTCGTCGTTATACACGTCGTAGAAATCGTATTCTATATTTAC
 AAGAAATTTTTGCAGAGGAAATGAGTAAAGTTGATGATAGTTTCTTTCATCGATTAGAGGAT
 TCTTTTCTAGTTGAGGAAGATAAGAGAGGGTAGCAAGTATCCTATCTTTGCAACAATGCAGGA
 GGAGAAATATTATCATGAAAAATTTCCGACAATCTATCATTTGAGAAAAGAATTGGCTGAC
 AAGAAAGAAAAAGCAGACCTTCGTCTTGTTTATCTGGCTCTAGCTCATATCATTTAAATTCAG
 AGGGCATTTCCTAATTGAGGATGATAGATTTGATGTGAGGAATACCGATATTCAAAAACAA
 TATCAAGCCTTTTTAGAAATTTTTGATACTACCTTTGAAAATAATCATTTGTTATCTCAAAAT
 GTAGATGTAGAAGCAATTCTAACAGATAAGATTAGCAAGTCTGCGAAGAAGGATCGCATCT
 TAGCGCAGTATCCTAACCAAAAAATCTACTGGTATTTTTGCAGAAATTTTTGAAATTGATTGTC
 GGAAATCAAGCTGACTTCAAGAAACATTTCAATTTGGAGGATAAAACACCGCTTCAATTCG
 CTAAGGATAGCTACGATGAAGATTTAGAAAATCTTCTTGACAGATTGGTGATGAATTTGCA
 GACTTATTCTCAGTAGCGAAAAAGCTATATGATAGTGTTCTTTTATCTGGCATTCTTACAGTA
 ACTGATCTCAGTACCAAGGCGCCACTTTCTGCCTCTATGATTCAGCGTTATGATGAACATCA
 TGAGGACTTAAAGCATCTAAAACAATTCGTAAAAGCTTCATTACCTGAAAATTATCGGGAA
 GTATTTGCTGATTCATCAAAAGATGGCTACGCTGGCTATATTGAAGGCCAAACTAATCAAGA
 AGCTTTTTATAAATATCTGTATAAATTTGTTGACCAAACAAGAAGGTAGCGAGTATTTTCTTG
 AGAAAATTAAGAATGAAGATTTTTTGAGAAAACAGAGAACCCTTGATAATGGCTCAATCCC
 GCATCAAGTCCATTTGACAGAATTGAGGGCTATTATTCGACGTCAATCAGAATACTATCCAT
 TCTTGAAAGAGAATCAAGATAGGATTGAAAAAATCCTTACCTTTAGAATTCCTTATTATGTC
 GGGCCACTAGCACGTGAGAAGAGTGATTTTGCATGGATGACTCGCAAAACAGATGACAGTA
 TTCGACCTTGGAATTTGAAGACTTGGTTGATAAAGAAAAATCTGCGGAAGCTTTTATCCAT
 CGCATGACCAACAATGACCTCTATCTCCAGAAAGAAAAAGTTTTACCAAAGCATAGTCTTAT
 TTATGAAAAATTTACTGTTTACAATGAATTAACGAAGGTTAGATTTTTTGGCAGAAGGCTTTA
 AAGATTTTCAATTTTAAATAGGAAGCAAAAAGAACTATCTTTAACAGCTTGTTTAAGGAA
 AAACGTAAGTAAGTGAAGGATATTATTAGTTTTTTGAATAAAGTTGATGGATATGAAG
 GAATTGCAATCAAAGGAATTGAGAAACAGTTTAACGCTAGCCTTTCAACCTATCATGATCTT
 AAAAAAATACTTGGAAGGATTTCCCTTGATAATACAGATAACGAGCTTATTTTGGAAGATAT
 CGTCCAAACTCTAACCTTATTTGAAGATAGAGAAATGATTAAGAAGTGCTTGACATCTATA
 AAGATTTTTTTACAGAGTCACAGCTTAAAAAGCTCTATCGCCGTCCTATACTGGCTGGGGA
 CGATTGTCTGCTAAGCTAATAAATGGCATCCGAAATAAAGAGAATCAAAAAACAATCTTGG
 ACTATCTTATTGATGATGGAAGTGCAAACCGAAACTTCATGCAGTTGATAAATGATGATGAT
 CTATCATTTAAACCAATTATTGACAAGGCACGAACTGGTAGTCATTCCGGATAATCTGAAAGA
 AGTTGTAGGTGAACCTTGCTGGTAGCCCTGCTATTAATAAAGGGATTCTACAAAGTTTGAAAA
 TAGTTGATGAGCTGGTTAAAGTCATGGGCTATGAACCTGAACAAATCGTGGTTGAAATGGC
 ACGTGAGAACCAACGACAGCAAAAGGATTAAGTCGTTACGACAACGCTTGACAACCTTG
 AGAGAATCTCTTGCTAATTTGAAGAGTAATATTTTGGAAGAGAAAAAGCCTAAGTATGTGA
 AAGATCAAGTTGAAAATCATCATTTATCTGATGACCGTCTTTTCTTTACTACTTACAAAACG
 GAAGAGATATGTATACAAAAAAGGCTCTGGATATTGATAATTTAAGTCAATATGATATTGA
 CCACATTATTCCTCAAGCTTTCATAAAAGATGATTCTATTGATAATCGTGTTTTGGTATCATC
 TGCTAAAAATCGTGGAATAATCAGATGATGTTCTAGCATTGAAATTGTAAAAGCTCGCAAA
 ATGTTCTGGAATAATTTACTGGATGCTAAGTTAATGAGTCAGCGTAAGTATGATAATTTGAC

TAAGGCAGAGCGCGGAGGCCTAACTTCCGATGATAAGGCAAGATTTATCCAACGTCAGTTG
 GTTGAGACTCGACAAATTACCAAGCATGTAGCTCGTATCTTGATGAACGCTTCAATAATGA
 AGTTGATAATGGTAAAAAGATTTGCAAGGTAAAAATTGTAACCTTGAAGTCAAATTTGGTTT
 CAAATTTCCGAAAAGAATTTGGATTCTATAAAATTCGTGAAGTTAATGATTATCACCATGCA
 CACGATGCTTATCTTAATGCAGTAGTTGCCAAAGCTATTCTAACCAATATCCACAGTTAGA
 GCCAGAGTTTGTCTACGGAATGTATAGACAGAAAAAACTTTGCAAAAATCGTTCATGAGGAT
 AAGGAAGAAAAATATAGTGAAGCAACCAGGAAAAATGTTTTTCTACTCCAACCTTGATGAATA
 TGTTCAAAAGAGTTGTGAGGTTAGCAGATGGTTCTATTGTTGTAAGACCAGTAATAGAACT
 GGTAGATATATGAGAAAACTGCATGGGATAAAAAAGAAACACTTTGCGACAGTTAGAAAA
 GTCTTGTCATACCCTCAGAACAAATATCGTGAAGAAGACAGAGATTCAGACAGGTGGTTTCTC
 TAAGGAATCAATCTTGCGCATGGTAACTCAGATAAGTTGATTCCAAGAAAAACGAAGGAT
 ATTTATTTAGATCCTAAGAAATATGGAGGTTTTGATAGTCCGATAGTAGCTTACTCTGTTTTA
 GTTGTAGCTGATATCAAAAAAGGTAAAGCACAAAAACTAAAAACAGTTACGGAACTTTTAG
 GAATTACCATCATGGAGAGGTCCAGATTTGAGAAAAATCCATCAGCTTTCCTTGAATCAAAA
 GGTATTTAAATATTAGGGACGATAAATTAATGATTTTACCGAAGTATAGTCTGTTTCGAATT
 AGAAAAATGGGCGTCGTCGATTACTTGCTAGTGCTGGTGAATTACAAAAAGGTAAACGAGCTA
 GCCTTACCAACACAATTTATGAAGTTCTTATACCTTGCAAGTCGTTATAATGAGTCAAAAGG
 TAAACCAGAGGAGATTGAGAAGAAACAAGAATTTGTAAATCAACATGTCTCTTATTTTGAT
 GACATCCTTCAATTAATTAATGATTTTTCAAAACGAGTTATTCTAGCAGATGCTAATTTAGA
 GAAAAATCAATAAGCTTTACCAGGATAATAAGGAAAAATATACCAGTAGATGAACCTTGCTAAT
 AATATTATCAATCTATTTACTTTTACCAGTCTAGGAGCTCCAGCAGCTTTTAAATTTTTTGAT
 AAAATAGTTGATAGAAAACGCTATACATCAACTAAAGAAGTACTTAATTCTACTCTAATCCA
 TCAATCTATTACTGGACTTTATGAAACACGTATTGATTTGGGTAAATTAGGAGAAGATTGAT
 ATGGCAGGTTGGCGAACTGTTGTTGTAAATACACATTCTAAGCTCTCTTATAAAAAATAATCA
 TCTGATTTTTAAAGATTCTTATCAGACGGAAATGATTCATCTTTCAGAGATTGATATTCTAAT
 CATGGAACGACAGATATTGTTTTGTCGACTATGCTGATTAAACGTTTGGTTGATGAAAATA
 TTTTAGTCATATTTTGATGATAAACGCTTGCCAACAGCTATGTTAATGCCGTACTATGCTA
 GACATGATTGAGTTTACAATTATCTAGGCAGATGTCATGGATTGAGGATGTCAAAGCGGAT
 GTTTGGACATCAATTATTGCACAAAAAATTTGAATCAGTCCTTTTATCTCGGTGAGTGTTCT
 TTCTTTGAAAAATCCCAGTCTATTATGAATCTCTATCATGATTTAGAATCTTTTGACCCTTCC
 AATCGTGAAGGTCATGCAGCTAGGATTTATTTCAATACACTTTTTGGAAATGATTTTTCAAG
 AGAGCAGGATAATCCAATAAATGCTGGTTTAGACTATGGATATTCTCTGATTTTGAGTATGT
 TTGCGCGTGAAGTTGTTAAGTGTGGTTGCATGACACAATTTGGCTTAAAGCATGCTAATCAA
 TTTAATCAGTTCAACCTAGCAAGCGATATTATGGAACCATTTCCGCCAATCGTTGATAGGAT
 TATTTATGAAAATAGGCAGAGTGATTTTGTCAAAATGAAAAGAGAACTCTTTTCTATGTTTT
 CAGAGACATACAGCTACAACGGTAAAGAAATGTATCTTTCAAATATTGTCAGCGATTACAC
 CAAAAAAGTTATTAAGTCGCTAAATAGTGATGGGAATGGAATTCGGGAGTTTAGGATATGA
 GTTATCGGTATATGAGAATGATTTTAATGTTTGATATGCCTACTGAAACAGTAGAAGAACGT
 AAGGCGTATCGTAAGTTTAGAAAGTTTCTGTTGAGCGAAGGTTTTATTATGCATCAGTTCTC
 TGTTTATAGTAAATTATTGCTCAATAATACAGCTAATAATGCCATGATAGGTCGGCTTAAAG
 TGAATAATCCTAAGAAAGGGAGTATAACTCTTTTGACAGTTACCGAGAAGCAGTTTGCAAG
 GATGGTTTATCTACATGGTGAACATAATATGAGTGTGCCAACCTCTGATAGTCGCTTGTTTT
 CCTAGGAGATTCTTATGATCAAGATTAATTTTCCAATTTTAGATGAACCATTAGTGTTAAGT
 AATGCTACGATTTTAACGATAGAAGATGTTTCAGTTTATTCTTCATTGGTGAAACATTTTTAT
 CAATATGACGTAGATGAACATTTGAAATTATTTGATGATAAGCAGAAAAGTCTGAAGGCAA
 CGGAGTTAATGTTAGTTACAGATATCTTAGGATACGATGTCAACTCAGCACCTATTCTAAAG
 TTGATACATGGTGACTTAGAAAAATCAATTCACGAAAAGCCAGAAAGTGAATCAATGGTAG
 AAAAATTAGCAGCTACTATTACAGAACTTATCGCATTTGAGTGTCTAGAGAATGAGCTTGAT
 TTAGAATACGATGAAATTACGATTTTAGAATCATTAAAGGCACTGGGAGTCAAAATTGAGA
 CACAGAGCGACACTATCTTTGAAAAATGTTTTGAAATTATACAAGTTTACCATTATTTAACG
 AAAAAGAATCTCTTAGTTTTTGTAAATAGCGGAGCTTATCTTACCAAGATGAAGTTATAAA
 ATTATGTGAATACATCAATTTAATGCAAAAGTCAGTACTCTTTCTAGAACCTAGAAGACTCT
 ATGATTTACCGCAATATGTTATTGATAAGGATTATTTCTTGATAGGCGAAAAATATGGTATAA
 (SEQ ID NO:493)

SEQ ID NO:494:

ATGAATAAGCCATATTCAATAGGCCTTGACATCGGTACTAATTCCGTCGGATGGAGCATTAT
 TACAGATGATTATAAAGTACCTGCTAAGAAGATGAGAGTTTTAGGGAACACTGATAAAGAA
 TATATTAAGAAGAATCTCATAGGTGCTCTGCTTTTTGATGGCGGAATACTGCTGCAGATAG
 ACGCTTGAAGCGAACTGCTCGTCGTCGTTATACACGTCGTAGAAATCGTATTCTATATTTAC
 AAGAAATTTTTGCAGAGGAAATGAGTAAAGTTGATGATAGTTTCTTTCATCGATTAGAGGAT
 TCTTTTCTAGTTGAGGAAGATAAGAGAGGTAGCAAGTATCCTATCTTTGCAACAATGCAGGA
 GGAGAAATATTATCATGAAAAATTTCCGACAATCTATCATTTGAGAAAAGAATTGGCTGAC
 AAGAAAGAAAAAGCAGACCTTCGTCTTGTTTATCTGGCTCTAGCTCATATCATTAAATTCAG
 AGGGCATTTCCTAATTGAGGATGATAGATTTGATGTGAGGAATACCGATATTCAAAAACAA
 TATCAAGCCTTTTTAGAAATTTTTGATACTACCTTTGAAAATAATCATTGTATCTCAAAAT
 GTAGATGTAGAAGCAATTCTAACAGATAAGATTAGCAAGTCTGCGAAGAAGGATCGCATCT
 TAGCGCAGTATCCTAACCAAAAATCTACTGGTATTTTTGCAGAATTTTTGAAATTGATTGTC
 GGAAATCAAGCTGACTTCAAGAAACATTTCAATTTGGAGGATAAAACACCGCTTCAATTCTG
 CTAAGGATAGCTACGATGAAGATTTAGAAAAATCTTCTTGGACAGATTGGTGATGAATTTGCA
 GACTTATTCTCAGTAGCGAAAAAGCTATATGATAGTGTTCTTTTATCTGGCATTCTTACAGTA
 ACTGATCTCAGTACCAAGGCGCCACTTTCTGCCTCTATGATTACGCGTTATGATGAACATCA
 TGAGGACTTAAAGCATCTAAAACAATTCGTAAGGCTTCATTACCTGAAAAATTATCGGGAA
 GTATTTGCTGATTCATCAAAAGATGGCTACGCTGGCTATATTGAAGGCAAAACTAATCAAGA
 AGCTTTTATAAATATCTGTTAAAATTGTTGACCAACAAGAAGGTAGCGAGTATTTTCTTG
 AGAAAATTAAGAAATGAAGATTTTTTGAGAAAACAGAGAACCTTTGATAATGGCTCAATCCC
 GCATCAAGTCCATTTGACAGAATTGAGGGCTATTATTCGACGTCAATCAGAATACTATCCAT
 TCTTGAAAGAGAATCAAGATAGGATTGAAAAAATCCTTACCTTTAGAATTCCTTATTATGTC
 GGGCCACTAGCAGCTGAGAGAAGAGTGATTTTGTCATGGATGACTCGCAAAACAGATGACAGTA
 TTCGACCTTGAATTTTGAAGACTTGTTGATAAAGAAAAATCTGCGGAAGCTTTTATCCAT
 CGCATGACCAACAATGACCTCTATCTCCAGAAGAAAAAGTTTTACCAAAGCATAGTCTTAT
 TTATGAAAAATTTACTGTTTACAATGAATTAACGAAGGTTAGATTTTTTGGCAGAAGGCTTTA
 AAGATTTTCAATTTTAAATAGGAAGCAAAAAGAACTATCTTTAACAGCTTGTTTAAAGGAA
 AAACGTAAAGTAACTGAAAAGGATATTATTAGTTTTTTGAATAAAGTTGATGGATATGAAG
 GAATTGCAATCAAAGGAATTGAGAAACAGTTTAAACGCTAGCCTTTCAACCTATCATGATCTT
 AAAAAAATACTTGGCAAGGATTTCTTGATAATACAGATAACGAGCTTATTTTGGAGATAT
 CGTCCAACTCTAACCTTATTTGAAGATAGAGAAATGATTAAGAAGTGTCTTGACATCTATA
 AAGATTTTTTTACAGAGTCACAGCTTAAAAAGCTCTATCGCCGTCACTATACTGGCTGGGA
 CGATTGTCTGCTAAGCTAATAAATGGCATCCGAAATAAAGAGAATCAAAAACAATCTTGG
 ACTATCTTATTGATGATGGAAGTGCAACCGAACTTCATGCAGTTGATAAATGATGATGAT
 CTATCATTTAAACCAATTATTGACAAGGCACGAACTGGTAGTCATTCCGATAAATCTGAAAGA
 AGTTGTAGGTGAACCTTGCTGGTAGCCCTGCTATTAAGGAGGATTCTACAAAGTTTGAAAA
 TAGTTGATGAGCTGGTTAAAGTCATGGGCTATGAACCTGAACAAATCGTGGTTGAAATGGC
 ACGTGAGAACCAAACGACAGCAAAAGGATTAAGTCGTTACGACAACGCTTGACAACCTTG
 AGAGAATCTCTTGCTAATTTGAAGAGTAATATTTTGGGAAGAGAAAAAGCCTAAGTATGTGA
 AAGATCAAGTTGAAAATCATCATTTATCTGATGACCGTCTTTTCTTTACTACTTACAAAACG
 GAAGAGATATGTATACAAAAAGGCTCTGGATATTGATAATTTAAGTCAATATGATATTGA
 CCACATTATTCCTCAAGCTTTCATAAAAGATGATTCTATTGATAATCGTGTTTTGGTATCATC
 TGCTAAAAATCGTGGAATAATCAGATGATGTTCTTAGCATTGAAATTGTAAAAGCTCGCAAA
 ATGTTCTGGAATAATTTACTGGATGCTAAGTTAATGAGTCAGCGTAAGTATGATAATTTGAC
 TAAGGCAGAGCGCGGAGGCCTAACCTCCGATGATAAGGCAAGATTTATCCAACGTCAGTTG
 GTTGAGACTCGACAAATTACCAAGCATGTAGCTCGTATCTTGGATGAACGCTTCAATAATGA
 AGTTGATAATGGTAAAAAGATTTGCAAGGTTAAAATTGTAACCTTGAAGTCAAAATTTGGTTT
 CAAATTTCCGAAAAGAATTTGGATTCTATAAAATTCGTGAAGTTAATGATTATCACCATGCA
 CACGATGCTTATCTTAATGCAGTAGTTGCCAAAGCTATTCTAACCAAAATATCCACAGTTAGA
 GCCAGAGTTTGTCTACGGAATGTATAGACAGAAAAAACTTTGAAAATCGTTTATGAGGAT
 AAGGAAGAAAAATATAGTGAAGCAACCAGGAAAAATGTTTTCTACTCCAACCTTGATGAATA
 TGTTCAAAAGAGTTGTGAGGTTAGCAGATGGTTCTATTGTTGTAAGACCAGTAATAGAACT
 GGTAGATATATGAGAAAACTGCATGGGATAAAAAGAAACACTTTGCGACAGTTAGAAAA
 GTCTTGTACATACCCTCAGAACAATATCGTGAAGAAGACAGAGATTCAGACAGGTGGTTTCTC

TAAGGAATCAATCTTGGCGCATGGTAACTCAGATAAGTTGATTCCAAGAAAAACGAAGGAT
 ATTTATTTAGATCCTAAGAAATATGGAGGTTTTGATAGTCCGATAGTAGCTTACTCTGTTTTA
 GTTGTAGCTGATATCAAAAAAGGTAAAGCACAAAACTAAAAACAGTTACGGAACTTTTAG
 GAATTACCATCATGGAGAGGTCCAGATTTGAGAAAAATCCATCAGCTTTCCTTGAATCAAAA
 GGTTATTTAAATATTAGGGACGATAAATTAATGATTTTACCGAAGTATAGTCTGTTTCAATT
 AGAAAAATGGGCGTCGTCGATTACTTGCTAGTGCTGGTGAATTACAAAAAGGTAACGAGCTA
 GCCTTACCAACACAATTTATGAAGTTCTTATACCTTGCAAGTCGTTATAATGAGTCAAAAGG
 TAAACCAGAGGAGATTGAGAAGAAACAAGAATTTGTAATCAACATGTCTCTTATTTTGAT
 GACATCCTTCAATTAATTAATGATTTTCAAAACGAGTTATTCTAGCAGATGCTAATTTAGA
 GAAAAATCAATAAGCTTTACCAGGATAATAAGGAAAAATATACCAGTAGATGAACCTTGCTAAT
 AATATTATCAATCTATTTACTTTTACCAGTCTAGGAGCTCCAGCAGCTTTTAAATTTTTTGAT
 AAAATAGTTGATAGAAAACGCTATACATCAACTAAAGAAGTACTTAATTCTACTCTAATCCA
 TCAATCTATTACTGGACTTTATGAAACACGTATTGATTTGGGTAAATTAGGAGAAGATTGA
 (SEQ ID NO:494)

SEQ ID NO:495:

ATGGCAGGTTGGCGAACTGTTGTTGTAAATACACATTCTAAGCTCTCTTATAAAAAATAATCA
 TCTGATTTTTTAAAGATTCTTATCAGACGGAAATGATTCATCTTTCAGAGATTGATATTCTAAT
 CATGGAAACGACAGATATTGTTTTGTGCGACTATGCTGATTAAACGTTTGGTTGATGAAAATA
 TTTTAGTCATATTTTGTGATGATAAACGCTTGCCAACAGCTATGTTAATGCCGTACTATGCTA
 GACATGATTCGAGTTTACAATTATCTAGGCAGATGTCATGGATTGAGGATGTCAAAGCGGAT
 GTTTGGACATCAATTATTGCACAAAAAATTTTGAATCAGTCCTTTTATCTCGGTGAGTGTTCT
 TTCTTTGAAAAATCCCAGTCTATTATGAATCTCTATCATGATTTAGAATCTTTTGACCCTTCC
 AATCGTGAAGGTCATGCAGCTAGGATTTATTTCAATACACTTTTTTGAAATGATTTTTCAAG
 AGAGCAGGATAATCCAATAAATGCTGGTTTAGACTATGGATATTCTCTGATTTTGAGTATGT
 TTGCGCGTGAAGTTGTTAAGTGTGGTTGCATGACACAATTTGGCTTAAAGCATGCTAATCAA
 TTTAATCAGTTCAACCTAGCAAGCGATATTATGGAACCATTTGCCCCAATCGTTGATAGGAT
 TATTTATGAAAATAGGCAGAGTGATTTTGTCAAAATGAAAAGAGAAGTCTTTTCTATGTTTT
 CAGAGACATACAGCTACAACGGTAAAGAAATGTATCTTTCAAATATTGTGACGCGATTACAC
 CAAAAAGTTATTAAGTCGCTAAATAGTGATGGGAATGGAATTCCGGAGTTTAGGATATGA
 (SEQ ID NO:495)

SEQ ID NO:496:

ATGAGTTATCGGTATATGAGAATGATTTTAATGTTTGATATGCCTACTGAAACAGTAGAAGA
 ACGTAAGGCGTATCGTAAGTTTAGAAAGTTTCTGTTGAGCGAAGGTTTTATTATGCATCAGT
 TCTCTGTTTATAGTAAATTATTGCTCAATAATACAGCTAATAATGCCATGATAGGTGCGCTT
 AAAGTGAATAATCCTAAGAAAGGGAGTATAACTCTTTTGACAGTTACCGAGAAGCAGTTTG
 CAAGGATGGTTTATCTACATGGTGAACATAATATGAGTGTTGCCAACTCTGATAGTCGCTTG
 GTTTTCCTAGGAGATTCTTATGATCAAGATTAA (SEQ ID NO:496)

SEQ ID NO:497:

ATGATCAAGATTAATTTTTCCAATTTTAGATGAACCATTAGTGTTAAGTAATGCTACGATTTTA
 ACGATAGAAGATGTTTCAGTTTATTCTTCATTGGTGAAACATTTTTATCAATATGACGTAGAT
 GAACATTTGAAATTATTTGATGATAAGCAGAAAAAGTCTGAAGGCAACGGAGTTAATGTTAG
 TTACAGATATCTTAGGATACGATGTCAACTCAGCACCTATTCTAAAGTTGATACATGGTGAC
 TTAGAAAAATCAATTCAACGAAAAGCCAGAAGTGAAATCAATGGTAGAAAAATTAGCAGCTA
 CTATTACAGAACTTATCGCATTGAGTGCTAGAGAATGAGCTTGATTTAGAATACGATGAA
 ATTACGATTTTAGAACTCATTAAGGCACTGGGAGTCAAAAATTGAGACACAGAGCGACACTA
 TCTTTGAAAAATGTTTTGAAATTATACAAGTTTACCATTATTTAACGAAAAAGAATCTCTTA
 GTTTTTGTTAATAGCGGAGCTTATCTTACCAAAGATGAAGTTATAAAATTATGTGAATACAT
 CAATTTAATGCAAAAGTCAGTACTCTTTCTAGAACCTAGAAGACTCTATGATTTACCGCAAT
 ATGTTATTGATAAGGATTATTTCTTGATAGGCGAAAAATATGGTATAA (SEQ ID NO:497)

SEQ ID NO:498:

ATGAAAAAACCTTACTCTATTGGACTTGATATTGGAACCAATTCTGTTGGTTGGGCTGTTGT
 GACAGATGACTACAAAGTTCCTGCTAAGAAGATGAAGGTTCTGGGAAATACAGATAAAAGT
 CATATCGAGAAAAAATTTGCTTGGCGCTTTATTATTTGATAGCGGGAATACTGCAGAAGACAG
 ACGGTAAAGAGAACTGCTCGCCGTCGTTACACACGTCGCAGAAATCGTATTTTATATTTGC
 AAGAGATTTTTTCAGAAGAAATGGGCAAGGTAGATGATAGTTTCTTTCATCGTTTAGAGGAT
 TCTTTTCTTGTTACTGAGGATAAACGAGGAGAGCGCCATCCCATTTTTGGGAATCTTGAAGA
 AGAAGTTAAGTATCATGAAAATTTTCCAACCATTTATCATTTGCGGCAATATCTTGCGGATA
 ATCCAGAAAAAGTTGATTTGCGTTTAGTTTATTTGGCTTTGGCACATATAATTAAGTTTAGA
 GGTCAATTTTTAATTGAAGGAAAGTTTGATACACGCAATAATGATGTACAAAGACTGTTTCA
 AGAATTTTTAGCAGTCTATGATAATACTTTTGAGAATAGTTTCGCTTCAGGAGCAAAATGTTT
 AAGTTGAAGAAATTCTGACTGATAAAATCAGTAAATCTGCTAAGAAAGATAGAGTTTTGAA
 ACTTTTTCTAATGAAAAGTCTAATGGCCGCTTTGCAGAAATTTCTAAAACATAATTGTTGGTA
 ATCAAGCTGATTTTAAAAAGCATTTTGAATTAGAAGAGAAAGCACCATTGCAATTTTCTAAA
 GATACTTATGAAGAAGAGTTAGAAGTACTATTAGCTCAAATTGGAGATAATTACGCAGAGC
 TCTTTTATCAGCAAAGAACTGTATGATAGTATCCTTTTATCAGGGATTTTAACAGTTACTG
 ATGTTGGTACCAAAGCGCCTTTATCTGCTTCGATGATTCAGCGATATAATGAACATCAGATG
 GATTTAGCTCAGCTTAAACAATTCATTCGTCAGAAATTATCAGATAAATAACGAAGTTTT
 TTCTGATGTTTCAAAGACGGCTATGCGGGTTATATTGATGGGAAAACAAATCAAGAAGCTT
 TTTATAAATACCTTAAAGGTCTATTAATAAGATTGAGGGAAGTGGCTATTTCTTGATAAA
 ATTGAGCGTGAAGATTTTCTAAGAAAGCAACGTACCTTTGACAATGGCTCTATTCCACATCA
 GATTCATCTTCAAGAAATGCGTGCTATCATTGCTAGACAGGCTGAATTTATCCGTTTTAGC
 AGACAATCAAGATAGGATTGAGAAATTATTGACTTTCCGTATTCCCTACTATGTTGGTCCAT
 TAGCGCGCGGAAAAAGTGATTTTGCTTGGTTAAGTCGGAATCGGCTGATAAAATTACACC
 ATGGAATTTTGATGAAATCGTTGATAAAGAATCCTCTGCAGAAGCTTTTATCAATCGTATGA
 CAAATTATGATTTGTAATGTTGCAAAATCAAAAAGTTCTTCTAAACATAGTTTATTATACGAA
 AAATTTACTGTTTACAATGAATTAACAAAGGTTAAATATAAAACAGAGCAAGGAAAAACAG
 CATTTTTTGATGCCAATATGAAGCAAGAAATCTTTGATGGCGTATTTAAGGTTTATCGAAAA
 GTAATAAAGATAAATTAATGGATTTCTTGAAAAAGAAATTTGATGAATTTCTGATTGTTGA
 TTTAACAGGTCTGGATAAAGAAAAATAAAGTATTTAACGCTTCTTATGGAACCTTATCATGATT
 TGTGTAATAATTTAGATAAAGATTTTCTCGATAATTCAAAGAATGAAAAGATTTTAGAAGAT
 ATTGTGTTGACCTTAACGTTATTTGAAGATAGAGAAATGATTAGAAAACGTCTAGAAAATTA
 CAGTGATTTATTGACCAAAGAACAAGTGAAAAAGCTGGAAAGACGTCATTATACTGGTTGG
 GGAAGATTATCAGCTGAGTTAATTCATGGTATTCGCAATAAAGAAAGCAGAAAAACAATTC
 TTGATTATCTCATTGATGATGGCAATAGCAATCGGAACCTTTATGCAACTGATTAACGATGAT
 GCTCTTTCTTCAAAGAAGAGATTGCTAAGGCACAAGTTATTGGAGAAACAGACAATCTAA
 ATCAAGTTGTTAGTGATATTGCTGGCAGCCCTGCTATTAATAAAAGGAATTTTACAAAGCTTG
 AAGATTGTTGATGAGCTTGTCAAAATTATGGGACATCAACCTGAAAATATCGTCGTGGAGAT
 GGCGCGTGAAAACAGTTTACCAATCAGGGACGACGAAATTCACAGCAACGTTTGAAAGGT
 TTGACAGATTCTATTAAGAATTTGGAAGTCAAATCTTAAAGAACATCCGGTTGAGAATTC
 ACAGTTACAAAATGATAGATTGTTTCTATATTATTTACAAAACGGCAGAGATATGTATACTG
 GAGAAGAATTGGATATTGATTATCTAAGCCAGTATGATATAGACCATATTATCCCGCAAGCT
 TTTATAAAGGATAATTCTATTGATAATAGAGTATTGACTAGCTCAAAGGAAAAATCGTGGA
 AATCGGATGATGTACCAAGTAAAGATGTTGTTTCGTAATAATGAAATCCTATTGGAGTAAGCT
 ACTTTTCGGCAAAGCTTATTACACAACGTAATTTTGATAATTTGACAAAAGCTGAACGAGGTG
 GATTGACCGACGATGATAAAGCTGGATTCATCAAGCGTCAATTAGTAGAAACACGACAAAT
 TACCAAACATGTAGCACGTATTCTGGACGAACGATTTAATACAGAAACAGATGAAAACAAC
 AAGAAAATTCGTCAAGTAAAAATTTGTGACCTTGAAATCAAATCTTGTTCCTAATTTCCGTAA
 AGAGTTTGAACCTCTACAAAGTTCGTGAAATTAATGACTATCATCATGCACATGATGCCTATC
 TCAATGCTGTAATTGGAAAGGCTTTACTAGGTGTTTACCCACAATTGGAACCTGAATTTGTT
 TATGGTGATTATCCTCATTTTTCATGGACATAAAGAAAAATAAAGCAACTGCTAAGAAATTTT
 CTATTCAAATATTATGAACCTCTTTAAAAAAGATGATGTCCGTACTGATAAAAATGGTGAAA
 TTATCTGAAAAAAGATGAGCATATTTCTAATATTAATAAAGTGCTTTCTTATCCACAAGTT
 AATATTGTTAAGAAAGTAGAGGAGCAAACGGGAGGATTTTCTAAGAATCTATCTTGCCGA
 AAGGTAATTCTGACAAGCTTATTCTCGAAAAACGAAGAAATTTTATTGGGATACCAAGAA

ATATGGAGGATTTGATAGCCCGATTGTTGCTTATTCTATTTTAGTTATTGCTGATATTGAAAA
 AGGTAAATCTAAAAAATTGAAAACAGTCAAAGCCTTAGTTGGTGTCACTATTATGGAAAAG
 ATGACTTTTGAAAGGGATCCAGTTGCTTTTCTTGAGCGAAAAAGGCTATCGAAATGTTCAAGA
 AGAAAAATATTATAAAGTTACCAAAATATAGTTTATTTAACTAGAAAAACGGACGAAAAAAGG
 CTATTGGCAAGTGCTAGGGAACTTCAAAGGGGAAATGAAATCGTTTTGCCAAATCATTTAG
 GAACCTTGCTTTATCACGCTAAAAATATTCATAAAGTTGATGAACCAAAGCATTTGGACTAT
 GTTGATAAACATAAAGATGAATTTAAGGAGTTGCTAGATGTTGTGTCAAACCTTTTCTAAAAA
 ATATACTTTAGCAGAAGGAAATTTAGAAAAAATCAAAGAATTATATGCACAAAATAATGGT
 GAAGATCTTAAAGAATTAGCAAGTTCATTTATCAACTTATTAACATTTACTGCTATAGGAGC
 ACCGGCTACTTTTAAATTCTTTGATAAAAAATATTGATCGAAAACGATATACTTCAACTACTG
 AAATTCTCAACGCTACCCCTCATCCACCAATCCATCACCGGTCTTTATGAAACGCGGATTGAT
 CTCAATAAGTTAGGAGGAGACTAATGGGCTGGCGGACAGTGGTTGTTAATACGCATTCCAA
 GTTGTCTTATAAGAACAACCACTTGATTTTTAAAGATGCTTATCAGACAGAGATGATTTCATC
 TGTCTGAGATTGACATCTTATTACTTGAGACAACAGATATTGTTTTGTCAAACCTATGCTAATCA
 AACGCTTGGTTGATGAGAATATTTTGGTCATTTTTTGTGATGACAAACGTCTGCCAACAGCC
 ATGCTCATGCCTTACTATGCGCGTCACGATTCCAGCTTGACGCTGAGTCATCAGATTTCTTGG
 ACAGAAGAAGTGAAATGCGATGTCTGGACAACAATCATCGCTCAAAGATTTTGAATCAGT
 CATGTTATTTGGGAGAATGTTTTTATTTTGA AAAATCTCAGTCAATTATGGATTTATATCATG
 ACTTAGAGCCTTTTGACCCTAGTAATCGAGAAGGACATTCTGCGCGGATTTATTTCAATACC
 TTATTTGGAAATGTTTTTCCAGAGAACAAGATAATGATATTAATGCAGGTCTTGACTATGG
 TTATACGCTGCTGTTAAGTATGTTTGCGCGTGAAGTGGTTGTATCTGGCTGTATGACACAATT
 TGGTCTCAAGCATGCCAACCAATTCAATCAGTTTAACTTTGCCAGTGATATTATGGAGCCTT
 TTCGTCCAATTGTTGACCGTATTGTTTATGAAAATCGAAATAACTCTTTTATTA AAAATAAAAC
 GTGAGCTATTCAGCATGTTTTTCAGACACCTATCTTTATAATAATAAGGAGATGTAATTTGACA
 AATATTGTCAGCGATTATACCAAAAAGGTAATCAAGGCGCTGAATAATGATGGGAAAGGAG
 TTCCTGAGTTTAGGATATGAGTTACCGATATATGCGAATGATTTTAATGTTTGATATGCCAA
 CAGATACTGCTGAGGAACGCAAAGCTTATCGTAAATTTTCGGAAATTTTACTGAGCGAAGGT
 TTCATCATGCATCAGTTTTTCAGTATACAGCAAGCTGCTTTTGAATAACTCTGCCAATACAGC
 CATGATTGCCCCGCTTGAAGGAGAATAATCCAAAGAAGGGCAATATCACCTTGTTGACCGTG
 ACTGAAAAGCAGTTTGCCCGTATGATTTACCTGAATGGTGAGCGTGATACTAGCATTGCTAA
 TTCGGATTACAGACTGGTCTTTCTAGGGGAGGCTTTTCCTGATGAAACTTAATTTTCTTATAT
 TGGATGAACCAATAACTCTTGAAAAATCTACGATTTTGGTATTAGAAGATGTGCAAGTTTTT
 GCTCAAATGGTGAGAAATCTTTATCAATATGATGAAGATAGTGAACTTAAATTTTTTAATAG
 AAAATTTAAGAGTCTGAAACCATCTGAGTTAATGCTTGACAGATATTTTAGGTTATGATG
 TCAATGCCCGTCCCTTGCTGAAGTTGGTTCACGCTGATTTAGAAAATCAGTTTAATGAAAAA
 CCAGAGGTTAAGTCTATGGTTGAAAAACTGGCAAATACCATTACGGAATTAATTGCTTATGA
 ATGTTTAGAAAATGAATTGGACTTAGAATATGATGAGATTACTATTTTAGAGTTAATCAAAG
 CTTTAGGCGTCAAAATTGAAACACAAAGTGATACCATTTTTGAAAAAATGTTTGAAGTCCTT
 CAAGTTTATAAGTATCTAAATAAAAAGAAGCTTCTCGTTTTTATCAATACTTTATCCTATTTT
 AAAAGAGAAGAAATCGCGCAAATTCTAGAATATATTCACCTATCCGATATGGTTGTTTTATT
 TCTTGAACCCCGTAAATTTGATGGTTTTGCTCAATATATTTTAGATGAAGATTATTTCTTGAT
 AACAGAAAGCAACAATAACGAATAATAAGATAGTTTCTAAATCAGGGGCTGTCTTTT
 ATTATGGATTGACAAATGCGTATAATGCGTATAAAAATAAAAAGAGAAATGTTATTTGCCATT
 AACAGGGAAAGAATTAGCTAAATTAGCGATAAACAATGGATGGGAAGAAGTTCGGGTGAG
 AGGAAGTCATCATCATTTCAAGAAAGATGGAGTATCTTATATTGTGACGATTCCTATTTCATG
 GAAATAAAGTGCTTAAATTTGGTCTTGAAAAGAACTCTTAAGGGATTTAAATTTATTATGA
 TAGAGGAGGAAGTCGTCATGTTAAAATCATATCCTGTAATTTTTCATAAGGAAGAGGAAGG
 GTATTGGGTTGAATTTCTGAATTTGGCGGTGGTACGCAAGGGGAAGATTTGGAAGAAGCC
 ATGAAGAACGCTCGTCAGATGTTAGAAAGTGTGTTGGCATCTTATCTTGATGAAGGGTTGGT
 TCTACCCATTTCAAGCGATATTCAGAAAATATCTGTTGAAGATGGTTTTGCGACCATGATTC
 AAGCTGATCCTAGTCCTTATCTCAAAAATAACAAAGCTATTCGGAAAAATGTTACCGTGCCT
 GAGTGGTTGATACGATTAGCAGACCGTGACCGAGTAAATTATTCTGAAGTATTAACAAAGG
 CTTTGAAAAGAACTACAATTATAA (SEQ ID NO:498)

SEQ ID NO:499:

ATGAAAAAACCTTACTCTATTGGACTTGATATTGGAACCAATTCTGTTGGTTGGGCTGTTGT
 GACAGATGACTACAAAGTTCCTGCTAAGAAGATGAAGGTTCTGGGAAATACAGATAAAAGT
 CATATCGAGAAAAATTTGCTTGGCGCTTTATTATTTGATAGCGGGAATACTGCAGAAGACAG
 ACGGTAAAGAGAACTGCTCGCCGTCGTTACACACGTCGCAGAAATCGTATTTTATATTGTC
 AAGAGATTTTTTCAGAAGAAATGGGCAAGGTAGATGATAGTTTCTTTTCATCGTTTAGAGGAT
 TCTTTTCTTGTTACTGAGGATAAACGAGGAGAGCGCCATCCCATTTTTGGGAATCTTGAAGA
 AGAAGTTAAGTATCATGAAAATTTTCCAACCATTTATCATTTGCGGCAATATCTTGCGGATA
 ATCCAGAAAAAGTTGATTTGCGTTTAGTTTATTTGGCTTTGGCACATATAATTAAGTTTAGA
 GGTCATTTTTTAATTGAAGGAAAGTTTGATACACGCAATAATGATGTACAAAGACTGTTTCA
 AGAATTTTTAGCAGTCTATGATAATACTTTTGAGAATAGTTTCGCTTCAGGAGCAAAATGTTT
 AAGTTGAAGAAATTCTGACTGATAAAATCAGTAAATCTGCTAAGAAAGATAGAGTTTTGAA
 ACTTTTTCTAATGAAAAGTCTAATGGCCGCTTTGCAGAAATTTCTAAAACATAATTGTTGGTA
 ATCAAGCTGATTTTAAAAAGCATTTTGAATTAGAAGAGAAAGCACCATTGCAATTTTCTAAA
 GATACTTATGAAGAAGAGTTAGAAGTACTATTAGCTCAAATTGGAGATAATTACGCAGAGC
 TCTTTTTATCAGCAAAGAACTGTATGATAGTATCCTTTTATCAGGGATTTTAAACAGTTACTG
 ATGTTGGTACCAAAGCGCCTTTATCTGCTTCGATGATTCAGCGATATAATGAACATCAGATG
 GATTTAGCTCAGCTTAAACAATTCATTCGTCAGAAATTATCAGATAAATAACGAAGTTTT
 TTCTGATGTTTCAAAGACGGCTATGCGGGTTATATTGATGGGAAAACAAATCAAGAAGCTT
 TTTATAAATACCTTAAAGGTCTATTAATAAGATTGAGGGAAAGTGGCTATTTCTTGATAAA
 ATTGAGCGTGAAGATTTTCTAAGAAAGCAACGTACCTTTGACAATGGCTCTATTCCACATCA
 GATTCATCTTCAAGAAATGCGTGCTATCATTCGTAGACAGGCTGAATTTTATCCGTTTTTAGC
 AGACAATCAAGATAGGATTGAGAAATTATTGACTTTCCGTATTCCCTACTATGTTGGTCCAT
 TAGCGCGCGGAAAAAGTGATTTTGCTTGGTTAAGTCGGAAATCGGCTGATAAAATTACACC
 ATGGAATTTTGATGAAATCGTTGATAAAGAATCCTCTGCAGAAGCTTTTATCAATCGTATGA
 CAAATTATGATTTGTAATGAAATCAAAAAGTTCTTCTAAACATAGTTTATTATACGAA
 AAATTTACTGTTTACAATGAATTAACAAAGGTTAAATATAAAACAGAGCAAGGAAAAACAG
 CATTTTTTGATGCCAATATGAAGCAAGAAATCTTTGATGGCGTATTTAAGGTTTATCGAAAA
 GTAACATAAGATAAATTAATGGATTTCTTGAAAAAGAATTTGATGAATTTCTGATTGTTGA
 TTTAACAGGTCTGGATAAAGAAAAATAAAGTATTTAACGCTTCTTATGGAACCTTATCATGATT
 TGTGTAATTTTAGATAAAGATTTTCTCGATAATTCAAAGAATGAAAAGATTTTAGAAGAT
 ATTGTGTTGACCTTAACGTTATTTGAAGATAGAGAAATGATTAGAAAACGTCTAGAAAATTA
 CAGTGATTTATTGACCAAAGAACAAGTGAAAAAGCTGGAAAGACGTCATTATACTGGTTGG
 GGAAGATTATCAGCTGAGTTAATTCATGGTATTCGCAATAAAGAAAGCAGAAAAACAATTC
 TTGATTATCTCATTGATGATGGCAATAGCAATCGGAACCTTTATGCAACTGATTAACGATGAT
 GCTCTTTCTTCAAAGAAGAGATTGCTAAGGCACAAGTTATTGGAGAAACAGACAATCTAA
 ATCAAGTTGTTAGTGATATTGCTGGCAGCCCTGCTATTA AAAAAGGAATTTTACAAAGCTTG
 AAGATTGTTGATGAGCTTGTCAAATTAATGGGACATCAACCTGAAAATATCGTCGTGGAGAT
 GGCGCGTGAAAACCAGTTTACCAATCAGGGACGACGAAATTCACAGCAACGTTTGAAAGGT
 TTGACAGATTCTATTAAGAATTTGGAAGTCAAATCTTAAAGAACATCCGGTTGAGAATTC
 ACAGTTACAAAATGATAGATTGTTTCTATATTATTTACAAAACGGCAGAGATATGTATACTG
 GAGAAGAATTGGATATTGATTATCTAAGCCAGTATGATATAGACCATATTATCCCGCAAGCT
 TTTATAAAGGATAATTCTATTGATAATAGAGTATTGACTAGCTCAAAGGAAAATCGTGGA
 AATCGGATGATGTACCAAGTAAAGATGTTGTTTCGTA AAAATGAAATCCTATTGGAGTAAGCT
 ACTTTTCGGCAAAGCTTATTACACAACGTAAATTTGATAATTTGACAAAAGCTGAACGAGGTG
 GATTGACCGACGATGATAAAGCTGGATTTCATCAAGCGTCAATTAGTAGAAACACGACAAAT
 TACCAAACATGTAGCACGTATTCTGGACGAACGATTTAATACAGAAACAGATGAAAACAAC
 AAGAAAATTCGTCAAGTAAAAATTGTGACCTTGAAATCAAATCTTGTTTCCAATTTCCGTAA
 AGAGTTTGAACCTCTACAAAGTTCGTGAAATTAATGACTATCATCATGCACATGATGCCTATC
 TCAATGCTGTAATTGGAAAGGCTTTACTAGGTGTTTACCCACAATTGGAACCTGAATTTGTT
 TATGGTGATTATCCTCATTTTTCATGGACATAAAGAAAAATAAAGCAACTGCTAAGAAATTTT
 CTATTCAAATATTATGAACTTCTTTAAAAAAGATGATGTCCGTA CTGATAAAAATGGTGAAA
 TTATCTGGA AAAAAGATGAGCATATTTCTAATATTA AAAAAGTGCTTTCTTATCCACAAGTT
 AATATTGTTAAGAAAGTAGAGGAGCAAACGGGAGGATTTTCTAAGAATCTATCTTGCCGA
 AAGGTAATTCTGACAAGCTTATTCCTCGAAAAACGAAGAAATTTTATTGGGATACCAAGAA

ATATGGAGGATTTGATAGCCCGATTGTTGCTTATTCTATTTTAGTTATTGCTGATATTGAAAA
 AGGTAAATCTAAAAAATTGAAAACAGTCAAAGCCTTAGTTGGTGTCACTATTATGGAAAAAG
 ATGACTTTTGAAAGGGATCCAGTTGCTTTTCTTGAGCGAAAAGGCTATCGAAATGTTCAAGA
 AGAAAATATTATAAAGTTACCAAAATATAGTTTATTTAACTAGAAAACGGACGAAAAAGG
 CTATTGGCAAGTGCTAGGGAACCTTCAAAAGGGAAATGAAATCGTTTTGCCAAATCATTAG
 GAACCTTGCTTTATCACGCTAAAAATATTCATAAAGTTGATGAACCAAAGCATTGGAAGTAT
 GTTGATAAACATAAAGATGAATTTAAGGAGTTGCTAGATGTTGTGTCAAACCTTTTCTAAAAA
 ATATACTTTAGCAGAAGGAAATTTAGAAAAAATCAAAGAATTATATGCACAAAAATAATGGT
 GAAGATCTTAAAGAATTAGCAAGTTCAATTTATCAACTTATTAACATTTACTGCTATAGGAGC
 ACCGGCTACTTTTAAATTCTTTGATAAAAAATATTGATCGAAAACGATATACTTCAACTACTG
 AAATTCTCAACGCTACCCTCATCCACCAATCCATCACCGGTCTTTATGAAACGCGGATTGAT
 CTCAATAAGTTAGGAGGAGACTAA (SEQ ID NO:499)

SEQ ID NO:500:

ATGGGCTGGCGGACAGTGGTTGTTAATACGCATTCCAAGTTGTCTTATAAGAACAACCACTT
 GATTTTAAAGATGCTTATCAGACAGAGATGATTCATCTGTCTGAGATTGACATCTTATTACT
 TGAGACAACAGATATTGTTTTGTCAACTATGCTAATCAAACGCTTGGTTGATGAGAATATTT
 TGGTCATTTTTTGTGATGACAAACGTCTGCCAACAGCCATGCTCATGCCTTACTATGCGCGTC
 ACGATTCCAGCTTGCAGCTGAGTCATCAGATTTCTTGGACAGAAGAAGTGAAATGCGATGTC
 TGGACAACAATCATCGCTCAAAAGATTTTGAATCAGTCATGTTATTTGGGAGAATGTTTTTA
 TTTTGA AAAATCTCAGTCAATTATGGATTTATATCATGACTTAGAGCCTTTTGACCCTAGTAA
 TCGAGAAGGACATTCTGCGCGGATTTATTTCAATACCTTATTTGGAAATGTTTTTCCAGAG
 AACAAGATAATGATATTAATGCAGGTCTTGACTATGGTTATACGCTGCTGTTAAGTATGTTT
 GCGCGTGAAGTGGTTGTATCTGGCTGTATGACACAATTTGGTCTCAAGCATGCCAACCAATT
 CAATCAGTTTAACTTTGCCAGTGATATTATGGAGCCTTTTCGTCCAATTGTTGACCGTATTGT
 TTATGAAAATCGAAATAACTCTTTTATTAATAAATAAACGTGAGCTATTCAGCATGTTTTAG
 ACACCTATCTTTATAATAATAAGGAGATGTATTTGACAAATATTGTCAGCGATTATACCAA
 AAGGTAATCAAGGCGCTGAATAATGATGGGAAAGGAGTTCCTGAGTTTAGGATATGA (SEQ
 ID NO:500)

SEQ ID NO:501:

ATGCGAATGATTTTAATGTTTGATATGCCAACAGATACTGCTGAGGAACGCAAAGCTTATCG
 TAAATTTGCGAAATTTTACTGAGCGAAGGTTTCATCATGCATCAGTTTTCAGTATACAGCA
 AGCTGCTTTTGAATAACTCTGCCAATACAGCCATGATTGCCCGCTTGAAGGAGAATAATCCA
 AAGAAGGGCAATATCACCTTGTTGACCGTGACTGAAAAGCAGTTTGCCCGTATGATTTACCT
 GAATGGTGAGCGTGATACTAGCATTGCTAATTCGGATTACGACTGGTCTTTCTAGGGGAGG
 CTTTTCTGATGAAACTTAA (SEQ ID NO:501)

SEQ ID NO:502:

ATGGTGAGAAATCTTTATCAATATGATGAAGATAGTGAACCTTAAATTTTTTAATAGAAAATT
 TAAGAGTCTGAAACCATCTGAGTTAATGCTTGTGACAGATATTTTAGGTTATGATGTCAATG
 CCCCCTCCTTGCTGAAGTTGGTTCACGCTGATTTAGAAAATCAGTTTAATGAAAAACCAGAG
 GTTAAGTCTATGGTTGAAAAACTGGCAAATACCATTACGGAATTAATTGCTTATGAATGTTT
 AGAAAATGAATTGGACTTAGAATATGATGAGATTACTATTTTAGAGTTAATCAAAGCTTTAG
 GCGTCAAAATTGAAACACAAAGTGATACCATTTTTGAAAAAATGTTTGAAGTCCTTCAAGTT
 TATAAGTATCTAAATAAAAAGAAGCTTCTCGTTTTTATCAATACTTTATCCTATTTTAAAGA
 GAAGAAATCGCGCAAATTCTAGAATATATTCATTATCCGATATGGTTGTTTTATTTCTTGAA
 CCCCCTAAAATTGATGGTTTTGCTCAATATATTTTAGATGAAGATTATTTCTTGATAACAGA
 AAGCAACAATAA (SEQ ID NO:502)

SEQ ID NO:503:

ATGTTAAATCATATCCTGTAATTTTTCATAAGGAAGAGGAAGGGTATTGGGTTGAATTTCC
 TGAATTTGGCGGTGGTACGCAAGGGGAAGATTTGGAAGAAGCCATGAAGAACGCTCGTCAG

ATGTTAGAAAGTGTGTTGGCATCTTATCTTGATGAAGGGTTGGTTCTACCCATTTCAAGCGA
TATTCAGAAAATATCTGTTGAAGATGGTTTTGCGACCATGATTCAAGCTGATCCTAGTCCTT
ATCTCAAAAATAACAAAGCTATTCGGAAAAATGTTACCGTGCCTGAGTGGTTGATACGATT
GCAGACCGTGACCGAGTAAATTATTCTGAAGTATTAACAAAGGCTTTGGAAAAGAACTAC
AATTATAA (SEQ ID NO:503)

SEQ ID NO:504:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCCGATGGGCGGTGA
TCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCA
CAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCG
ACTCGTCTCAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCT
ACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAG
AGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTA
GATGAAGTTGCTTATCATGAGAAATATCCAACCTATCTATCATCTGCGAAAAAAATTGGTGA
TTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCG
TGGTCATTTTTTGATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTA
TCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCCTATTAACGCAAGTGGAGTA
GATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGC
TCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCATTGGGT
TGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAA
GATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTT
GTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTCAGATATCCTAAGAGTAAATAC
TGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAACGCTACGATGAACATCATCAAG
ACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACCTTCCAGAAAAGTATAAAGAAATCTTT
TTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAAT
TTTATAAATTTATCAAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAA
CTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCA
AATTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCA
TTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCC
ATGGAATTTTGAAGAAGTTGTGATAAAGGTGCTTCAGCTCAATCATTATTTGAACGCATGA
CAAACCTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGAG
TATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAAACC
AGCATTCTTTTCAGGTGAACAGAAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAA
AAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTT
GAAATTTTCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAA
AATTATTAAGATAAAGATTTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATT
GTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGC
TCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGAC
GTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGAT
TTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTG
ACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAAC
ATATTGCAATTTAGCTGGTAGCCCTGCTATTAaaaaaAGGTATTTTACAGACTGTAAAAGTT
GTTGATGAATGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAAATATCGTTATTGAAATGG
CACGTGAAAATCAGACAACCTCAAAAGGGCCAGAAAAATTTCGCGAGAGCGTATGAAACGAA
TCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATAC
TCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAAATGGAAGAGACATGTATGTGG
ACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTGATCATTGTTCCACAAAGT
TTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAAACGCGTTCTGATAAAAAATCGTGGTAA
ATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCT
CTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAG
GTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAAACGCCAATTGGTTGAAACTCGCCAAATC
ACTAAGCATGTGGCACAAATTTTGGATAGTCGATGAATACTAAATACGATGAAAATGATA
AACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAA
GATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTATCT

AAATGCCGTCGTTGGAACGTGCTTTGATTAAAGAAATATCCAAAACCTGAATCGGAGTTTGTCT
 ATGGTGATTATAAAGTTTATGATGTTTCGTAAAAATGATTGCTAAGTCTGAGCAAGAAATAGGC
 AAAGCAACCGCAAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC
 ACTTGCAAATGGAGAGATTTCGCAACGCCCTCTAATCGAACTAATGGGGAAACCTGGAGAA
 ATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCAGCAAAGTATTGTCCATGCCCAAGT
 CAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCA
 AAAAGAAATTTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTG
 GTTTTGATAGTCCAACGGTAGCTTATTTCAGTCTAGTGGTTGCTAAGGTGGAAAAAGGGAAA
 TCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCCT
 TTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTT
 AATCATTAACCTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
 CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTT
 TTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAA
 CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATT
 TTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGCATATAACAAAC
 ATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGACGAA
 TCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTC
 TACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACAC
 GCATTGATTTGAGTCAGCTAGGAGGTGACTGATGGCTGGTTGGCGTACTGTTGTGGTAAATA
 CCCACTCGAAATTATCCTATAAGAATAATCATCTGATTTTTAAGGATGCCTATAAAACGGAG
 CTGATCCATTTATCAGAAATTGATATTTTGTATTAGAAAACGACCGATATTGTCTTGTCCACT
 ATGCTGGTAAAACGGCTAGTGGATGAGAATGTCCTTGTTCATATTCTGTGATGATAAACGATT
 ACCAACAGCTATGCTGATGCCTTTTTATGGTCGTTCATGATTTCGAGTTTACAGCTTGGGAAAC
 AAATGTCCTGGTCAGAAACAGTCAAATCGCAGGTTTGGACGACGATTATTGCTCAAAAGAT
 TTTGAATCAATCTTGCTATCTAGGAGCATGCTCCTATTTTGAAAAATCCCAATCTATTATGGA
 TTTATATCATGGTTTGGAAAAATTTGATCCGAGTAATCGAGAAGGGCATGCAGCGAGAATTT
 ATTTAATACACTTTTGGGAACGATTTCTCAAGAGATTGGAGCATCCAATCAATGCAGGT
 CTGGATTATGGTTATACCTTTATTATTGAGTATGTTTGCAGGTGAAGTGTTGTGTCTGGATGT
 ATGACTCAGTTTGGGCTTAAACACGCTAATCAGTTTAAATCAGTTCAATTTTGCTAGCGATATT
 ATGGAACCATTTAGGCCTTTAGTGGATAAGATTGTTTATGAAAATCGAAATCAGCCTTTTCC
 CAAAATAAAGAGAGAGTTATTTACTTTGTTTTAGATAACATTTTCATATAATGGTAAAGAGA
 TGTATCTCACGAATATTATTAGCGATTATACTAAAAAAGTTGTCAAAGCTCTGAATAATGAA
 GGGAAAGGAGTTTCTGAATTTAGGATATGAGTTATAGATATATGAGAATGATACTTATGTTT
 GATATGCCGACGGACACCGCTGAGGAACGAAAAGCCTATCGAAAATTTTCGGAAATTTTAC
 TTAGTGAAGGGTTTATCATGCATCAATTTTCTATTTATAGTAAGTTGCTGTTGAATAATACAG
 CTAACAATGCCATGATTGGTCGGCTGAGGGAGCATAATCCTAATAAAGGAAATATTACATT
 ACTAACGGTCACGGAAAAACAGTTTGCACGAATGATTTATTTACATGGTGAAAGAAATAAT
 TGTATTGCAAACCTCCGATGAAAGACTTGTATTTCTTGGGGAGGCTTTTGATGAATCTTAATTT
 TTCCTTACTAGATGAACCGATTCCATTAAGAGGCGGTACAATTCCTGTGCTCGAAGATGTCT
 GTGTATTTTCAAAAATAGTGCAATATTGTTACCAATATGAGGAAGATTCTGAACTTAAATTT
 TTTGATCACAAGATGAAAACAATCAAAGAATCAGAAATCATGCTTGTAACAGATATTTTAG
 GATTTGATGTTAACTCCTCAACCATTTTAAAAATTGATTCATGCAGATTTAGAATCTCAATTTA
 ATGAGAAACCCGAAGTGAAATCGATGATTGACAAATTGGTTGCTACGATTACAGAAGTGAAT
 TGTCTTTGAATGCTTAGAAAAATGAATTAGATTTAGAGTATGATGAAATCACAATCCTGGAAT
 TGATTAAGTCCTTAGGAGTAAAAGTAGAAACGCAAGTGATACTATTTTGA AAAATGTCTA
 GAGATACTTCAAAATTTTCAAAATATCTCAAGAAAAAGTTGCTTATTTTGTCAATAGCGG
 AGCTTTTCTAACAAAGGATGAAGTGGCTAGTTTACAAGAGTATATATCATTGACAAATTTAA
 CAGTTCTCTTTTAGAACCACGTGAACCTATATGATTTTCCGCAGTATATTTTAGATGAAGATT
 ATTTCTTAATAACTAAAAATATGGTATAA (SEQ ID NO:504)

SEQ ID NO:505:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGA
 TCACTGATGAATATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCA
 CAGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCG
 ACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCT

ACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAG
 AGTCTTTTTTGGTGGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTA
 GATGAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAAAAATTTGGTAGA
 TTCTACTGATAAAGCGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCG
 TGGTCATTTTTTGAATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTA
 TCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCCTATTAACGCAAGTGGAGTA
 GATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGC
 TCAGCTCCCCGGTGAGAAGAAAAATGGCTTATTTGGGAATCTCATTGCTTTGTCAATTGGGTT
 TGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAA
 GATACTTACGATGATGATTTAGATAATTTATTTGGCGCAAATTGGAGATCAATATGCTGATTT
 GTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTACTTTTTCAGATATCCTAAGAGTAAATAC
 TGAAATAACTAAGGCTCCCTATCAGCTTCAATGATTAAACGCTACGATGAACATCATCAAG
 ACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACCTCCAGAAAAGTATAAAGAAATCTTT
 TTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAAT
 TTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAA
 CTAATCGTGAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCCATCA
 AATTCATTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTAA
 AAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCCA
 TTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCC
 ATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGCATGA
 CAACTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGTCTTATGAG
 TATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATGTTACTGAAGGAATGCGAAAACC
 AGCATTTCTTTAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCGAA
 AAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTT
 GAAATTTAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCTAAA
 AATTATTAAGATAAAGATTTTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATT
 GTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATATGC
 TCACCTCTTTGATGATAAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGGGAC
 GTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGAT
 TTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTG
 ACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATGAAC
 ATATTGCAAATTTAGCTGGTAGCCCTGCTATTAATAAAGGTATTTTACAGACTGTAAAAGTT
 GTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAATATCGTTATTGAAATGG
 CACGTGAAAATCAGACAACCTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAA
 TCGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATAC
 TCAATTGCAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATGTGG
 ACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAAAGT
 TTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAAACGCGTTCTGATAAAAAATCGTGGTAA
 ATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAACCTT
 CTAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAG
 GTTTGAGTGAACCTTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATC
 ACTAAGCATGTGGCACAAATTTTGATAGTCGCATGAATACTAAATACGATGAAAAATGATA
 AACTTATTCGAGAGGTTAAAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAAAA
 GATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCATGATGCGTATCT
 AAATGCCGTCGTTGGAAGTCTTTGATTAAGAAATATCCAAAACCTTGAATCGGAGTTTGTCT
 ATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGC
 AAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTAC
 ACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAGAA
 ATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCAGCAAGTATTGTCCATGCCCCAAGT
 CAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCA
 AAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATGGTG
 GTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGAAAAAAGGGAAA
 TCGAAGAAGTTAAATCCGTTAAAGAGTTACTAGGGATCACAATTATGGAAAGAAGTTCTT
 TTGAAAAAATCCGATTGACTTTTTAGAAAGCTAAAGGATATAAGGAAGTTAAAAAAGACTT
 AATCATTAACCTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG

CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTT
 TTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAA
 CAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAATT
 TTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGTCATATAACAAAC
 ATAGAGACAAACCAATACGTGAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACGAA
 TCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTC
 TACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACAC
 GCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:505)

SEQ ID NO:506:

ATGGCTGGTTGGCGTACTGTTGTGGTAAATACCCACTCGAAATTATCCTATAAGAATAATCA
 TCTGATTTTTAAGGATGCCTATAAACCGGAGCTGATCCATTTATCAGAAATTGATATTTTGT
 ATTAGAAACGACCGATATTGTCTTGCTCACTATGCTGGTAAACGGCTAGTGGATGAGAATG
 TCCTTGTCATATTCTGTGATGATAAACGATTACCAACAGCTATGCTGATGCCTTTTTATGGTC
 GTCATGATTCGAGTTTACAGCTTGGGAAACAAATGTCCTGGTCAGAAACAGTCAAATCGCA
 GGTTTGGACGACGATTATTGCTCAAAAGATTTTGAATCAATCTTGCTATCTAGGAGCATGCT
 CCTATTTTGA AAAATCCCAATCTATTATGGATTTATATCATGGTTTGGAAAATTTTGATCCGA
 GTAATCGAGAAGGGCATGCAGCGAGAATTTATTTTAATACACTTTTTGGGAACGATTTCTCA
 AGAGATTTGGAGCATCCAATCAATGCAGGTCTGGATTATGGTTATACTTTATTATTGAGTAT
 GTTTGCGCGTGAAGTGGTTGTGTCTGGATGTATGACTCAGTTTGGGCTTAAACACGTAATC
 AGTTTAATCAGTTCAATTTTGCTAGCGATATTATGGAACCATTTAGGCCTTTAGTGGATAAG
 ATTGTTTATGAAAATCGAAATCAGCCTTTTCCCAAAATAAGAGAGAGTTATTTACTTTGTT
 TTCAGATACATTTTCATATAATGGTAAAGAGATGTATCTCACGAATATTATTAGCGATTATA
 CTA AAAAAGTTGTCAAAGCTCTGAATAATGAAGGGAAAGGAGTTCCTGAATTTAGGATATG
 A (SEQ ID NO:506)

SEQ ID NO:507:

ATGAGTTATAGATATATGAGAATGATACTTATGTTTGATATGCCGACGGACACCGCTGAGGA
 ACGAAAAGCCTATCGAAAATTTTCGGAAATTTTTACTTAGTGAAGGGTTTATCATGCATCAAT
 TTTCTATTTATAGTAAGTTGCTGTTGAATAATACAGCTAACAATGCCATTGTTGTCGGCTG
 AGGGAGCATAATCCTAATAAAGGAAATATTACATTACTAACGGTCACGGAAAAACAGTTTG
 CACGAATGATTTATTTACATGGTGAAGAAATAATTGTATTGCAAACCTCCGATGAAAGACTT
 GTATTTCTTGGGGAGGCTTTTGATGAATCTTAA (SEQ ID NO:507)

SEQ ID NO:508:

ATGAATCTTAATTTTTCTTACTAGATGAACCGATTCCATTAAGAGGCGGTACAATTCTTGTC
 CTCGAAGATGTCTGTGTATTTTCAAAAATAGTGCAATATTGTTACCAATATGAGGAAGATT
 TGAACCTTAAATTTTTTGATCACAAGATGAAAACAATCAAAGAATCAGAAATCATGCTTGTA
 CAGATATTTTAGGATTTGATGTTAACTCCTCAACCATTTTAAAATTGATTCATGCAGATTTAG
 AATCTCAATTTAATGAGAAACCCGAAGTGAAATCGATGATTGACAAATTGGTTGCTACGATT
 ACAGAACTGATTGTCTTTGAATGCTTAGAAAAATGAATTAGATTTAGAGTATGATGAAATCAC
 AATCCTGGAATTGATTAAGTCCTTAGGAGTAAAAGTAGAAACGCAAAGTGATACTATTTTTG
 AAAAATGTCTAGAGATACTTCAAATTTTCAAATATCTCCTAAGAAAAAGTTGCTTATTTTT
 GTCAATAGCGGAGCTTTTCTAACAAGGATGAAGTGGCTAGTTTACAAGAGTATATATCATT
 GACAAATTTAACAGTTCTCTTTTGAACACGTGAACCTATATGATTTTCCGCAGTATATTTT
 AGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAA (SEQ ID NO:508)

SEQ ID NO:517:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGA
 TCACTGATGATTATAAGGTTCCGTCTAAAAAGCTCAAGGGTCTGGGAAATACAGACCGCCA
 CGGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCG
 ACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCCGAAGAATCGTATTTGTTATCT

ACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAG
 AGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTA
 GATGAAGTTGCTTATCATGAGAAATATCCAACATCTATCATCTGCGAAAAAAATTGGCAGA
 TTCTACTGATAAAGTGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAAGTTTCG
 TGGTCATTTTTTGAATTGAGGGAGATTTAAATCCTGATAAATAGTGATGTGGACAAACTATTTA
 TCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCCTATTAACGCAAGTAGAGTA
 GATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAAATCTCATTGC
 TCAGCTCCCCGGTGAGAAGAAAAATGGATTGTTTGGGAATCTCATTGCTTTGTCATTGGGAT
 TGACCCCTAATTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAA
 GATACTTACGATGATGATTTAGATAATTTATTGGCGCAAATTGGAGATCAATATGCTGATTT
 GTTTTTGGCAGCTAAGAATTTATCAGATGCTACTTTACTTTTACAGATATCCTAAGAGTAAATA
 GTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAA
 GACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACTTCCAGAAAAAGTATAAAGAAATCTT
 TTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAA
 TTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGCGAA
 ACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCTATC
 AAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTAA
 AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCC
 ATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCC
 CATGGAATTTTGAAGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTATTGAACGCATG
 ACAAACCTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGCTTTATGA
 GTATTTTACGGTTTATAACGAATTGACAAAAGTCAAATATGTTACTGAGGGAATGCGAAAA
 CCAGCATTCTTTTCAAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAACAATCG
 AAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGT
 GTTGAAATTTCAAGGAGTTGAAGATAGATTTAATGCTTCATTAGGTACCTACCATGATTTGCT
 AAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAAAACGAAGATATCTTAGAGGAT
 ATTGTTTTAACATTGACCTTATTTGAAGATAGGAGATTGATTGAGGAAAGACTTAAACATA
 TGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGG
 GACGTTTGTCTCGAAAATTGATTAATGCTATTAGGGATAAGCAATCTGGCAAAACAATATTA
 GATTTTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGT
 TTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATG
 AACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAAAAAAGGTATTTTACAGACTGTAAAA
 GTTGTTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAAATATCGTTATTGAAA
 TGGCACGTGAAAATCAGACAACCTCAAAAGGGCCAGAAAAATTCGCGTGAGCGTATGAAAC
 GTATTGAAGAAGGAATAAAGAAGTGAAGTGAATTTCTAAAGGAGTATCCTGTTGAAAA
 CACTCAATTACAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGTATG
 TGGACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAA
 AGTTTCTTAAAGACGATTCAATAGACAATAAGGTCTTAACGCGTTCTGATAAAAAATCGTGG
 TAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
 CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACAAAAGCTGAACGTGG
 AGGTTTGAGTGAACCTGATAAAGTTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAA
 TCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGA
 TAAACTTATTCGAGAGGTTAGAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAA
 AAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATGCGTAT
 CTTAATGCCGTCGTTGGAACCTGCTTTGATTAAGAAATATCCAAAACCTGAATCGGAGTTTGT
 CTATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAGGAAATAG
 GCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATT
 ACACTTGCAAATGGAGAGATTTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAG
 AAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCA
 AGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTA
 CCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATG
 GTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGTTGCTAAGGTGGAAAAAGGG
 AAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACTAGGGATCACAATAATGGAAAGAAGCT
 CTTTTGAAAAAGATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAGAAAAGA
 CTTAATCATTAACCTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTTCGTAAACGGATGC

TGGCTAGTGCCGGAGAATTGCAAAAAGGAAATGAGCTAGCTCTGCCAAGCAAATATGTGAA
TTTTTATATTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAA
AACAAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAA
TTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAAGTTCTTAGTGCAATATAACAA
ACATAGAGACAAACCAATACGTGAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACG
AATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATAC
GTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAA
CACGCATTGATTTGAGTCAGCTAGGAGGTGACTGATGGCTGGTTGGCGTACTGTTGTGGTAA
ATACCCACTCGAAATTATCCTATAAGAATAATCATCTGATTTTTAAGGATGCCTATAAAACG
GAGCTGATCCATTTATCAGAAATTGATATTTTGTATTAGAAACGACCGATATTGTCTTGCC
ACTATGCTGGTAAAACGGCTAGTGGATGAGAATGTCCTTGTCAATTTCTGTGATGATAAACG
ATTACCAACAGCTATGCTGATGCCTTTTTATGGTCGTCAATCGAGTTTACAGCTTGGGA
AACAAATGTCCTGTGAGAACAGTCAAATCGCAGGTTTGGACGACGATTATTGCTCAAAA
GATTTTGAATCAATCTTGCTATCTAGGAGCATGCTCCTATTTTGA AAAATCCCAATCTATTAT
GGATTTATATCATGGTTTGGAAAATTTTGATCCGAGTAATCGAGAAGGGCATGCAGCGAGA
ATTTATTTTAATACACTTTTGGGAACGATTTCTCAAGAGATTTGGAGCATCCAATCAATGC
AGGTCTGGATTATGGTTATACTTTATTATTGAGTATGTTTGC GCGTGAAGTGGTTGTGTCTGG
ATGTATGACTCAATTTGGACTCAAACACGCCAATCAGTTTAATCAGTTCAATTTTGCTAGCG
ATATTATGGAACCATTTAGGCCTTTGGTGGATAAGATTGTTTATGAAAATCGAAATCAGCCT
TTTCCCAAAATAAAGAGAGAGATTATTTACTTTGTTTTCAGATACATTTTCATATAATGGTAAA
GAGATGTATCTCACGAATATTATTAGCGATTATACTAAAAAAGTTGTCAAAGCTCTGAATAA
TGAAGGGAAAGGAGTTCCTGAATTTAGGATATGAGTTATAGATATATGAGAATGATACTTA
TGTTTGATATGCCGACGGACACTGCTGAGGAACGAAAAGCTTATCGAAAATTTCCGAAATTT
TTACTTAGTGAAGGGTTTATCATGCATCAATTTTCTATTTATAGTAAGTTACTGTTGAATAAT
ACAGCTAACAACGCCATGATTGGTCCGCTGAGGGAGCATAATCCTCATAAAGGAAATATTA
CATTACTAACAGTCACAGAAAAACAGTTTGCACGAATGATTTATTTACATGGTGAAAGAAA
TAATTGTATTGCAAACCTCCGATGAGAGACTTGTATTTCTTGGGGAGGCTTTTGATGAATCTT
AATTTTCCCTTATTAGATGAACCGATTCCATTAAGAGGGCGGTACAATTCCTGTGCTCGAAGA
TGTCTGTGTATTTTCAAAAATAGTGCAATATTGTTACAAATATGAGGAAGATTCTGAACTTA
AATTTTGTGATCACAAGATGAAAACCATCAAAGAATCAGAAATCATGCTTGTAACAGATATT
TTAGGATTTGATGTTAACTCCTCAACCATTTTAAAAATTGATTCATGCAGATTTAGAATCTCAA
TTAATGAGAAACCCGAAGTGAAATCGATGATTGACAAATTTGGTTGCTACGATTACAGAAC
TGATTGTCTTTGAATGCTTAGAAAAATGAATTAGATTTAGAGTATGATGAAATCACAATCCTG
GAATTGATTAAGTCCTTAGGAGTAAAAGTAGAAACGCAAAGTGATACTATTTTGA AAAAT
GTCTAGAGATACTTCAAATTTTCAAATATCTCACTAAGAAAAAGTTGCTTATTTTGTCAATA
GCGGAGCTTTTCTAACAAAGGATGAAGTGGCTAGTTTACAAGAGTATATATCATTGACAAAT
TTAACAGTTCTCTTTTGAACCACGTGAACATATGATTTTCCGCAGTATATTTTAGATGAA
GATTATTTCTTAATAACTAAAAATATGGTATAA (SEQ ID NO:517)

SEQ ID NO:518:

ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGA
TCACTGATGATTATAAGGTTCCGTCTAAAAAGCTCAAGGGTCTGGGAAATACAGACCGCCA
CGGTATCAAAAAAATCTTATAGGGGCTCTTTTATTTGACAGTGGAGAGACAGCGGAAGCG
ACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACACGTCGGAAGAATCGTATTTGTTATCT
ACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCATCGACTTGAAG
AGTCTTTTTTGGTGGAAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTA
GATGAAGTTGCTTATCATGAGAAATATCCAATCTATCATCTGCGAAAAAATTTGGCAGA
TTCTACTGATAAAGTGGATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCG
TGGTCATTTTTTGAATTGAGGGAGATTTAAATCCTGATAATAGTGATGTGGACAAACTATTTA
TCCAGTTGGTACAAACCTACAATCAATTATTTGAAGAAAACCCTATTAACGCAAGTAGAGTA
GATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCATTGC
TCAGCTCCCCGGTGAGAAAGAAAAATGGATTGTTTGGGAATCTCATTGCTTTGTCAATTGGGAT
TGACCCCTAATTTTAAATCAAATTTTGATTGCGCAGAAAGATGCTAAATTACAGCTTTCAAAA
GATACTTACGATGATGATTAGATAATTTATTGGCGCAAATTTGGAGATCAATATGCTGATTT
GTTTTTGGCAGCTAAGAATTTATCAGATGCTACTTTACTTTCAAGATATCCTAAGAGTAAATA

GTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGAACATCATCAA
 GACTTGACTCTTTTAAAAGCTTTAGTTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTT
 TTTTGATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAA
 TTTTATAAATTTATCAAACCAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGCGAA
 ACTAAATCGTGAAGATTTGCTGCGCAAGCAACGGACCTTTGACAACGGCTCTATTCCCTATC
 AAATTCACCTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGACTTTTATCCATTTTTA
 AAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTCC
 ATTGGCGCGTGGCAATAGTCGTTTTGTCATGGATGACTCGGAAGTCTGAAGAAACAATTACCC
 CATGGAATTTTGAAGAAGTTGTGCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATG
 ACAAACCTTTGATAAAAAATCTTCCAAATGAAAAAGTACTACCAAAACATAGTTTGGCTTTATGA
 GTATTTTACGGTTTATAACGAATTGACAAAAGTCAAAATATGTTACTGAGGGGAATGCGAAAA
 CCAGCATTCTTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAACAAATCG
 AAAAGTAACCGTTAAGCAATTAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGT
 GTTGAAATTTTCAGGAGTTGAAGATAGATTAAATGCTTCATTAGGTACCTACCATGATTTGCT
 AAAAATTATTAAAGATAAAGATTTTTTGGATAATGAAGAAACGAAGATATCTTAGAGGAT
 ATTGTTTTAACATTGACCTTATTTGAAGATAGGGAGATGATTGAGGAAAGACTTAAACATA
 TGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTATACTGGTTGGG
 GACGTTTGTCTCGAAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTA
 GATTTTTTGAATCAGATGGTTTTTGCCTTCGCAATTTTATGCAGCTGATCCATGATGATAGT
 TTGACATTTAAAGAAGACATTCAAAAAGCACAAGTGTCTGGACAAGGCGATAGTTTACATG
 AACATATTGCAAATTTAGCTGGTAGCCCTGCTATTAATAAAGGTATTTTACAGACTGTAAAA
 GTTGTGATGAATTGGTCAAAGTAATGGGGCGGCATAAGCCAGAAAAATATCGTTATTGAAA
 TGGCACGTGAAAAATCAGACAACCTCAAAAGGGCCAGAAAAATTCGCGTGAGCGTATGAAAC
 GTATTGAAGAAGGAATAAAGAAGCTAGGAAGTGATATTCTAAAGGAGTATCCTGTTGAAAA
 CACTCAATTACAAAATGAAAAGCTCTATCTCTATTATCTCCAAAATGGAAGAGACATGATG
 TGGACCAAGAATTAGATATTAATCGTTTAAAGTGATTATGATGTCGATCACATTGTTCCACAA
 AGTTTCCTTAAAGACGATTCAATAGACAATAAGGTCTTAAACGCGTTCTGATAAAAAATCGTGG
 TAAATCGGATAACGTTCCAAAGTGAAGAAGTGAAGTCAAAAAGATGAAAACTATTGGAGACAA
 CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACAAAAGCTGAACGTGG
 AGGTTTGAGTGAACCTTGATAAAGTTGGTTTTATCAAAACGCCAATTGGTTGAAACTCGCCAAA
 TCACTAAGCATGTGGCACAAATTTTGGATAGTCGCATGAATACTAAATACGATGAAAATGA
 TAACTTATTTCGAGAGGTTAGAGTGATTACCTTAAATCTAAATTAGTTTCTGACTTCCGAA
 AAGATTTCCAATTCTATAAAGTACGTGAGATTAAACAATTACCATCATGCCCATGATGCGTAT
 CTTAATGCCGTCGTTGGAACCTGCTTTGATTAAAGAAATATCCAAAACCTTGAATCGGAGTTTGT
 CTATGGTGATTATAAAGTTTATGATGTTTCGTAAAATGATTGCTAAGTCTGAGCAGGAAATAG
 GCAAAGCAACCGCAAAATATTTCTTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATT
 AACTTGCATAATGGAGAGATTTCGCAAACGCCCTCTAATCGAAACTAATGGGGAAACTGGAG
 AAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTCCATGCCCCA
 AGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTA
 CCAAAAAGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAATATG
 GTGGTTTTGATAGTCCAACGGTAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGG
 AAATCGAAGAAGTTAAAAATCCGTTAAAGAGTTACTAGGGATCACAATAATGGAAAGAAGCT
 CTTTTGAAAAAGATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAAGTTAGAAAAAG
 CTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGC
 TGGCTAGTGCCGGAGAATTGCAAAAAGGAAATGAGCTAGCTCTGCCAAGCAAATATGTGAA
 TTTTTTATTTTAGCTAGTCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAA
 AACAATTGTTTGTGGAGCAGCATAAGCATTATTTAGATGAGATTATTGAGCAAATCAGTGAA
 TTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTCTTAGTGATATAACAA
 ACATAGAGACAAACCAATACGTGAACAAGCAGAAAAATATTATTCATTTATTTACGTTGACG
 AATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATAC
 GTCTACAAAAGAAGTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAA
 CACGCATTGATTTGAGTCAGCTAGGAGGTGACTGA (SEQ ID NO:518)

SEQ ID NO:519:

ATGGCTGGTTGGCGTACTGTTGTGGTAAATACCCACTCGAAATTATCCTATAAGAATAATCA
TCTGATTTTTAAGGATGCCTATAAAACGGAGCTGATCCATTTATCAGAAATTGATATTTTGT
ATTAGAAACGACCGATATTGTCTTGTCCACTATGCTGGTAAACGGCTAGTGGATGAGAATG
TCCTTGTGATATTCTGTGATGATAAACGATTACCAACAGCTATGCTGATGCCTTTTTATGGTC
GTCATGATTCGAGTTTACAGCTTGGGAAACAAATGTCCTGGTCAGAAACAGTCAAATCGCA
GGTTTGGACGACGATTATTGCTCAAAAGATTTTGAATCAATCTTGCTATCTAGGAGCATGCT
CCTATTTTGAAAAATCCCAATCTATTATGGATTTATATCATGGTTTGGAAAAATTTTGATCCGA
GTAATCGAGAAGGGCATGCAGCGAGAATTTATTTAATACACTTTTTGGGAACGATTTCTCA
AGAGATTTGGAGCATCCAATCAATGCAGGTCTGGATTATGGTTATACTTTATTATTGAGTAT
GTTTGC CGTGAAGTGGTTGTGTCTGGATGTATGACTCAATTTGGACTCAAACACGCCAATC
AGTTTAATCAGTTCAATTTTGTAGCGATATTATGGAACCATTTAGGCCTTTGGTGGATAAG
ATTGTTTATGAAAATCGAAATCAGCCTTTTCCCAAAATAAAGAGAGAGTTATTTACTTTGTT
TTCAGATACATTTTCATATAATGGTAAAGAGATGTATCTCACGAATATTATTAGCGATTATA
CTAAAAAAGTTGTCAAAGCTCTGAATAATGAAGGGAAAGGAGTTCCTGAATTTAGGATATG
A (SEQ ID NO:519)

SEQ ID NO:520:

ATGAGAATGATACTTATGTTTGATATGCCGACGGACACTGCTGAGGAACGAAAAGCTTATC
GAAAATTTTCGGAATTTTACTTAGTGAAGGGTTTATCATGCATCAATTTTCTATTTATAGTA
AGTTACTGTTGAATAATACAGCTAACACGCCATGATTGGTCGCTGAGGGAGCATAATCCT
CATAAAGGAAATATTACATTACTAACAGTCACAGAAAAACAGTTTGCACGAATGATTTATTT
ACATGGTGAAAGAAATAATTGTATTGCAAACCTCCGATGAGAGACTTGTATTTCTTGGGGAG
GCTTTTGATGAATCTTAA (SEQ ID NO:520)

SEQ ID NO:521:

ATGAATCTTAATTTTCCCTTATTAGATGAACCGATTCCATTAAGAGGGCGGTACAATTCTTGTG
CTCGAAGATGTCTGTGTATTTTCAAAAAATAGTGCAATATTGTTACAAATATGAGGAAGATTC
TGAACCTAAATTTTGTATCACAAGATGAAAACCATCAAAGAATCAGAAATCATGCTTGTA
CAGATATTTTAGGATTTGATGTTAACTCCTCAACCATTTTAAATTTGATTTCATGCAGATTTAG
AATCTCAATTTAATGAGAAACCCGAAGTGAATCGATGATTGACAAATTTGGTTGCTACGATT
ACAGAACTGATTGTCTTTGAATGCTTAGAAAAATGAATTAGATTTAGAGTATGATGAAATCAC
AATCCTGGAATTGATTAAGTCCTTAGGAGTAAAAGTAGAAACGCAAAGTGATACTATTTTGT
AAAAATGTCTAGAGATACTTCAAATTTTCAAATATCTCACTAAGAAAAAGTTGCTTATTTT
GTCAATAGCGGAGCTTTTCTAACAAAGGATGAAGTGGCTAGTTTACAAGAGTATATATCATT
GACAAATTTAACAGTTCTCTTTTGAACCACGTGAACCTATATGATTTTCCGCAGTATATTT
AGATGAAGATTATTTCTTAATAACTAAAAATATGGTATAA (SEQ ID NO:521)

con secuencias repetidas: SEQ ID NO: 13 hasta SEQ ID NO: 19

Combinación funcional #4:

5

secuencias cas: SEQ ID NO:509 hasta SEQ ID NO:516 (todas las cuales son de *S. pyogenes*), como se muestran a continuación:

SEQ ID NO:509:

ATGAGAATGATTTTAGCACACTATGACTGTAAAAAAGATAAAAAAGCAATCTTTAGATGAGC
ATTTATGGCATGTGGCCTGTTCTAGTCGACAGGAAGCATCTATAATTGGTCAAGGAGATGTG
CTTTTTTTAATTGGTCTTTACCACGACCTGGGCAAAGCTGATCGAACCTTTCAAGATAAATTA
TTAAATAATCCAAATCGGCATGTTGATCACTCTTATGCAGGGGGCAAAATACTTATGTTCTAT
TATTGGGCCTCATCTAAAAAACCAGGGGGTTGATAAAAAATGAGAGAATGACATTCAACGAA
ATGGTGGGGTATGTCATCTCTGCTCATCATGGGATGTATGATTTATGCTACTATTTTGACGAT

GCTGAATATTATGGCTTTAATAAGTTTAAAAATCGTATCAATAGAGACTTAGATGGTTATCA
 CTATCATGAAGATATTAAAGGGTACGCTCTAAAATTAGAAAAAAATTATGTGATTATGGCT
 ACAAAGATTAAAGGGAGCTTATTGATAAAGCTTTTGATAATTACCAACAAGCCATGTCTTCC
 TTAAACTGGCAAGATAAGAGTGAGTGGGATTATTATCAGTCTTGTATGGTGAGACTTTACTT
 GTCACTCTTAAAAAACGCTGATATTTTGGACACAGTAAATGCCTATGGCCTTAAGATAAGTC
 CTATGGATAAAACAGAGCGATCCTTTCTAAAACACTCCTATTTAGCGGCCATTGAACAAAAA
 TATGCTAGCTTTGGACAGCCAAACAATCAGTTGAACACTATTCGGACAGAAATCGCTGAGC
 GTGTTAAAGAAAGAGGTAAACGAGATTCCAAGGGGATTATCGCTTAGATTTACCGACAGG
 AGCTGGCAAGACTAATCTTAGTATGCGTTATGCGTTTCACCAATTAGTTCATCACGACAAAT
 CAAGGTTTTTTTACATAACTCCCTTTCTTTTCGGTTCTTGAGCAAAATGCTTCGGAAATTAGAA
 AAGTTACAGGTGACCTTGGCGTTCTAGAACACCATTCCAATGTGGTGAAACAGGCTAATGA
 AGATGATGATGATAAGGACAGTTTATTGTCAGCTTATCTTAGTGATAGCTGGGACAGTCAAG
 TAGTCTTGACTTCTATGGTTCAATTTTTCCAAACACTTTTCAAAACAAAATCAGCTAATCTGA
 GACGTTTTTCAAGTTTGATTAATAGTGTTGTGATTCTAGATGAAGTTCAATCCCTGCCTATTG
 AAGTCACCCTTTGTTTTAATTTAACGATGAATTTTTTAAATAAAGTTATGGATACAACCATC
 GTTCTTTGCACAGCGACACAACCTGCTTATGATTCTTCAGAGATTGACCATCGTATCTGTTAT
 GGAGGGAACCTTGGGAGAATTAGCTGAAATAGTTGAGTTAACGATTGAAGAAAAACAGATTT
 TTTCAAGGACAGAGCTTAGAAAAATTTGATGATAGTGATCAGAAAGTTCACTTGACTGATGTT
 ATTAACCTTATTCTAGGTGAGGAAAACCTCAGTTCTTGCTATTTTTAATACGAAAAAACGGT
 TCATAACTGCTATACTATGCTAAAAGACATGACTGATAGACCGGTCTATCAGCTTTCGACAA
 ATATGTGTGCGCAGCATAGACTTGACTTGATTGCTAAGATCAAAACGGAGTTACAAAATAA
 TATCCCTATTATTTGTATTAGCACGCAATTAATTGAAGCAGGTGTAGATGTTGATTTTCATCG
 CGTCATTCGTTCCCTACTCAGGGATTGATTCTATTGTTACAGGCTGCTGGACGGTGTAACCGAG
 AAGGCAAACGAGATAAAGGGCAAGTCACTCTTGTCATCTGACCAATGAAGAGGAAAATAT
 TTCTAGGCTGACAGAAATAAAAACTAAAAAAGAAGCCACAGAATCTATTCTTCATAAGATT
 GGGTCTCCAATTGATATCTCAACTTTAAACCGTGACTTTTTTGAGTATTATTATGCCAATAAT
 CAGGGACTGATGGATTATCCTTTGGAAGACAACCTATCAATCTACGACTATTTAAGCCTTAA
 TATTTATCAGACGGCAAATAAAAAAGTTCAAAGGTAAGTTAAAACAAGCTTTTAAAACAGCA
 GGAGCCAAAATGAACCTCATCAATAATGATATGATAGGAATTCTCGTACCTTATGGCGAAG
 CTGAGAAAAAATTGGCTTATTTAGAAGAATTAGGTGTGTCACATTTTTTATCAGCAAAAGAT
 TATCAAACGATAAAATCATTACTAAAAGAGTTACAACCTTTTACGGTTAATGTCCGCGAGAA
 CGATCCTCTCTTTGAGACAACAAAATCTTATCTAAATGGTCAGATTCTGGTTTTGACGTCGG
 AGTATTATGACACGGAAAGAGGAGTTAAATACGATTACAGCTAGCTTTTACTTCTAACTCAAA
 ACGAAAGAAGATTAAACAAAAGGTTGTTAGAGGACCTTGTTAACCTGCCAATCATCATTAGT
 AATTATTATCAATTTAGACTATTTAATAAAATTAGATTACAAAAAACAGAAGGAGGAAAG
 TAGCTTGTACAGATCTAGAGACTTCTACGTGAGAGTAAGTGGTCAGCGAGCTCTTTTTACAA
 ATCCAGCCACAAAAGGGGGATCGGAACGCTCATCCTATTCGGTTCCGACTAGACAGGCACT
 GAATGGTATCGTTGATGCCATCTATTATAAGCCGACCTTTACTAATATCGTACACAGAGGTTA
 AGGTTATTAACCAGATTCAAACCGAATTACAGGGGTGTCAGGGGCTCTGTTACATGATTATAGT
 GCAGATTTAAGTTATGTATCCTATTTGAGTGATGTTGTTTTATCTGATCAAGTTTCATTTTGTTT
 GGAATGAAGATAGAAAAGATTTGAACTCAGATAGACTTCCAGCTAAACATGAAGCCATTAT
 GGAGCGTTCTATTCGTAAAGGGGGACGTCGAGATGTGTTTTTGGGTACAAGAGAATGTTTAG
 GGCTTGATAGATGATATCAGCCAAGAAGAGTATGAGACTACTGTGTCGTATTATAATGGTGTC
 AATATCGACTTGGGAATCATGTTCCATTCCCTTTGCCTATCCGAAGGACAAAAAGACACCATT
 AAAATCATACTTTACAAAGACTGTGATGAAAAATGGAGTCATTACGTTTAAAGCACAGTCT
 GAATGCGATATTGTTAACACGCTTTCTAGTTATGCTTTTTAAAGCACCCAGAGGAGATAAAATC
 GGTAAACGATGAATGCATGGAGTATGATGCCATGGAGAAAGGAGAAAACCTGATGGATTTTT
 TTAATTCTCTCTTGAAGACTTATGAAAAAGCAGAGCTAGCAGACTTGGTTGATCATCAAAAA
 AGAAATAATGAGCCGGTTTTACTGCCGATTTATCATACGAGTTTAAAGTCAAATGGTAAAAA
 TATCATTTTCAGTGAAACTTGACAAAGATGGCCAGTTTCACAAGGCAGAATTTATGGCAGATA
 AGCAAATGATTATTTTTCTGTAAACGGCTGATTCTGTTGCTAGGTACGGTAGTCATCCTGCAC
 CGCATCCCTAGTCGATAAATTTGCTTATTATAGTGCTGAAATGGGGCAGATTTCAGTATGAT
 TCTTTTCATAAGCAACTGAATAACTGGATTGATTATTGTGAGGAGGGTGATGTCAAGAAATT
 TTTAACCTTTGTTTCAGCAGTTCAATTTGAAGCCAGAATTTCTAACATTGATTCTTGATTCTTT
 AATTGGTCCTGATTATCAACATAATCAATTAAGTCAATTTTGTGATGCCACTGGAAAAG

AAAAATTAATTGATTTATCAGCTTGCTTTTTAGAATTTTCAATTGATCAGTTCCAGGGCTTTA
 AAAATGAATCGGTTTCGACATTTAAAGCCTTACACCAATCCTATATTTCTTTTGTGAAGCCA
 ATCGTGAAAATCTCGGTATTTGTAATATTAGTGGACGAGAGGAACAGCTTACCGATAAGCA
 TAGAGGTTTGATGGGGAATGCTAAAATCATCTCTGTTAGTAATAAAAGAGAAGCTTATAAA
 GGACGTTTTAGAGAACGCGAAGACGTTTTTAGTGTTGGCTATGAAACTTCCGAAAAGATTCA
 TTTAATGCTCAAGTACCTTTTAGAAAAATAAAAAATACCAGTACTTGGTTAGGGTCTTCTCAAT
 ATTTAATCAACTGGTTCAGCGATGATTTAACAAATGATAGTCGGTTGGATATTGTATCACCA
 ATCTTTGATGATGGACTTGAAGAAGATGATGATGACGATACGCCTCCTGTTATAACATTAGC
 AACTGAAGACAATAAAAGAATTGGTAAATCATTATCAAGGGACAAAAATTATTTGCTAAT
 GATGCCACTTACTACGTTGCTATTTTGAATAAAACCAGCAATGGGCGGATTGCTTTAAAAATA
 TTTTCGTCAGCTTCAAGCGTCCCAATTACTACCAATCTTAACAAGTGGCAGGAAACATACA
 GTTGGGAGTCGCGATCTAAGTTTGGGAAAAGTCGCTTAAGAACCCCTACTTTTCATGACATC
 CTTAATGTGTCCTACGGGGTTGATAGGGATCGCTTCCTTGAATTAGATAATGATAACTTCAA
 AAGTGATCAAATTCAAAAGTTAGTGGCAAGTTTGATTGATGGTAAACCGATGCCACAGTCC
 ATTTGCAAAAAGTTAGGTAAACAATGTTAAAGAACGACATCGTTACCGTAAGCACTGGTATC
 AAGTTGAGCAGGTCTGCTTAGCAATTTTACACAAACAAAATGGGGAGGAATTTTCACCGAT
 GCTAGATCATACCAATCAAAATCGTTCCTATCTTTTGGACGATTATTAGCAATTTTGAATT
 AATCGAGACCTTGCGTTATGGCTTGGATGGAAACAATAACGACCGTATTACCAATGCTGAA
 CGTTATTGGACAGCCTATACTGGACAACCAACAAAATTGATGATGTTATTGGAAAATAAAA
 TTAAGCCTTACGAAGAACCATTGAAATTAATCGTCGTGGCAGTTGGATGAAATTAGAAAA
 AGAAAAAGAAGAGATTTTAGAACTGTTAAATCCTCTGTTAGAAACAGAAACAATGGAAAAA
 CCCTTAGATTACCGCTTTATTTTGGGTATTATGCTGAGAAAACTATTACTATACAAAACA
 AAACACGGAAGTAACAGAAAGTGAGGAGTAAAAAGATGTTGGAACACAAAAATTGATTTTAT
 GGTAACTCTTGAAGTGAAAGAAGCAAAATGCAAAATGGTGATCCCTTAAATGGAAACATGCCT
 CGTACAGATGCCAAAGGATATGGTGTGATGAGTGATGCTCCATTAAACGTAAGATTCTGTA
 ATCGTTTGCAAGATATGGGGAAGTCTATTTTGTGCAAGCTAATGAGCGTATTGAAGATGAT
 TTTGTTCACTGGAAAAACGCTTTTCGCAACATTTTACAGCTAAGACACCTGACAAAGAAAT
 TGAAGAAAAAGCAAAATGCATTATGGTTTGATGTTCTGTGCTTTTGGACAAGTTTTTACTTATCT
 GAAAAAATCAATTGGGGTGCCTGGACAGTTTCCATCAGTATGGCTAAGTCCTTGGAGCCA
 ATTTGTCATTTCCAGCCTTCAAATTACGCGTAGTACCAATGGTATGGAAGCTAAGAATAATAG
 TGGCCGCTCTTCTGATACGATGGGGACAAAACATTTTGTAGATTATGGTGTGTATGTACTTA
 AAGTTCTATCAATGCTTATTTTGTGAAAAGACTGGTTTTTCTCAGGAAGATGCTGAGGCT
 ATTAAGAAGTTTTGGTTAGCTTGTGTGAAAATGATGCGTCGTCTGCACGTCCGGAAGGCTC
 TATGCGAGTTTGTGAAGTCTTTTGGTTTACGCATTCAAGCAAATGGGAAATGTTTCAAGTG
 CGCGTGTCTTTGACTTGTAGAGTATCATCAATCAATAGAAGAAAAAGCACTTATGACGCT
 TATCAGATTCATCTAAATCAAGAAAAATTGGCTAAATATGAAGCGAAAGGGTTAACGCTTG
 AAATCCTAGAAGGACTCTAGTATGGTCTATGCCGAAGATGATTATTTAATGCTGTCAGGTAT
 TCAGCATTTCCAATTTTGTAAACGTCAATGGGCGTTGATCCATATTGAGCAACAATGGCTTG
 ATAATGAAGCGACAGCGCATGGACAGGTTTACATACTAAAGCAGATAACCCTTACATTAA
 AGAAAAACGAAAAGAGCTTTTGGTCTCACGTGCTATGCCCATTTCTTCTGCAGAACTTGGAC
 TTTACAGGAATTATGGATGTTGTGGAATTTTATAAAGATGATCAAGGTGTGTCTTTGAGGGGA
 AAACGTGGGAAATGGTTACCAAAAGTTGTGGAATACAAGCGCGGAAAACCTAAAAAAGAT
 ACCAGAGATATTGTCCAGTTGGTGGCTCAGACCATGTGTTTAGAAGAAACGCTAGACTGCG
 ACATTAACGAAGGTTGTCTTTATTACCATAGTGTCAATCAAAGAGTGATTGTTCTTATGACA
 TCAGCTTTGCGTCAAGAAGTGAAGGAATTAGCCGCAGAGATGCATGAGGTTTATCAGAGTC
 AAATGCTACCTAAAGCAGCTTATTTTAAAAAAGTGTGAGCTTTGTTCTTTAGTCGATATTTGTA
 AGCCCAGGTTGAGTAAAAAAACAAGGAGTGTGTGCGGTTACATCAATGAGGCTATGACCAG
 TGAGGAGATGGACCTATGAAGAAGTTGCTAAATACCTTGTATTTGACGCAAGAAGATTTTAA
 TGTCACTAAAGAGGGGCGATAACATTGTTATCAAGCAAGAAGGTAAGGTTCTCAAACGGTTT
 CCGTTTCGGATTATTGACGGTATTGTCTGTTTTTCTTATTTGGGTGTGTGCTGCTTTGGTGA
 AGTTATGTACGGAGAATCAGATTAATTTATCGTTTCATACACCACAAGGGCGTTTTTGTGGT
 CGCTATATTGGTTCAACCAATGGGAATGTGTTGTTGCGTAGAGAACATTATCGTTTATCTGA
 TCGTGAGGAATCTTTGGAATACGCAAAGCGGTTTATTTTGGCTAAAATTTCCAACCTCAAGGA
 AATACTTGCTACGCTTTAAACGAGATCATCGTCAACAGATTGATACCAAGCTTTTGGAGGCT
 GTTAATGACGAATTGATATGGGCTTTAGAGATGGTTCAGGCAGCAGATAATAAAGACTCTTT

AAGAGGGATTGAAGGCCAAGCTGCTAATCAGTATTTTCGCATATTTAATGACCTGGTGTGA
 CGGACAAAAAACGTTTTACTTCCAAGGTCCGAGTAAACGACCACCCTTAGATTGTGTTAAT
 GCCCTCTTGTCTTTTGGTTACAGTTTACTGACCTTTGAATGTCAATCTGCCTTGGAAAGCTGTC
 GGATTAGACAGTTACGTTGGTTTCTTTCACACGGATCGTCCTGGGCGTGCTAGTTTAGCGCTT
 GATTTAGTTGAAGAGTCCGCTCATATATTGTAGATCGTTTTGTCTTTTCATTAATTAATAAA
 GGACAACTTCAGAAAAAACACTTTGAGGTTAAAGAAAATGGTAGTATTTTATTGACGGAAA
 ATGGCAGAGCTATTTTTATTGATTTGTGGCAGAAGCGTAAGCATACTGAGGTAGAACATCCT
 TTTACAAAAGAGAAAGTAAACTTATGTTATTACCCTATGTACAAGCGCAGCTTTTAGCTAA
 GGCTATACGAGGAGATTTAGAAAGCTATCCACCTTTTATGGTTTAGGAGATGTTATATGATG
 GTTTTAGTCACCTTATGATGTAAATACGGAAACACCTGCTGGTAGAAAAAGATTGCGTCATGT
 TGCCAAACTCTGTGTGGACTATGGGCAACGTGTTCAAAATTCTGTTTTGAATGTTCTGTGAC
 ACCCGCAGAATTTGTGGATATAAAGCACCGCTTAACACAAATCATTGATGAGAAAACTGAT
 AGTATTCGCTTTTATTTATTGGGGAAAAATTGGCAGAGGCGTGTGAAACACTTGGTCGCTC
 AGACAGCTATGACCCAGATAAAGGTGTCTTATTATTGTAA (SEQ ID NO:509)

SEQ ID NO:510:

ATGAGAATGATTTTAGCACACTATGACTGTAAAAAGATAAAAAAGCAATCTTTAGATGAGC
 ATTTATGGCATGTGGCCTGTTCTAGTCGACAGGAAGCATCTATAATTGGTCAAGGAGATGTG
 CTTTTTTTAAATTGGTCTTTACCACGACCTGGGCAAAGCTGATCGAACCTTTCAAGATAAATTA
 TTAAATAATCCAAATCGGCATGTTGATCACTCTTATGCAGGGGCAAAATACTTATGTTCTAT
 TATTGGGCCTCATCTAAAAAACCGAGGGGTTGATAAAAAATGAGAGAATGACATTCAACGAA
 ATGGTGGGGTATGTCATCTCTGCTCATCATGGGATGTATGATTTATGCTACTATTTTGACGAT
 GCTGAATATTATGGCTTTAATAAGTTTAAAAATCGTATCAATAGAGACTTAGATGGTTATCA
 CTATCATGAAGATATTAAGGGTACGCTCTAAAATTAGAAAAAAAATTATGTGATTATGGCT
 ACAAAGATTTAAGGGAGCTTATTGATAAAGCTTTTGATAATTACCAACAAGCCATGTCTTCC
 TTAAACTGGCAAGATAAGAGTGAGTGGGATTATTATCAGTCTTGATGGTGAGACTTTACTT
 GTCACTCTTAAAAAACGCTGATATTTTGGACACAGTAAATGCCTATGGCCTTAAGATAAGTC
 CTATGGATAAAACAGAGCGATCCTTTCTAAACACTCCTATTTAGCGGCCATTGAACAAAAA
 TATGCTAGCTTTGGACAGCCAAACAATCAGTTGAACACTATTCGGACAGAAATCGCTGAGC
 GTGTTAAAGAAAGAGGTAAACGAGATTCCAAGGGGATTTATCGCTTAGATTTACCGACAGG
 AGCTGGCAAGACTAATCTTAGTATGCGTTATGCGTTTCACCAATTAGTTCATCACGACAAAT
 CAAGGTTTTTTTACATAACTCCCTTTCTTTCGGTTCTTGAGCAAAATGCTTCGGAATTAGAA
 AAGTTACAGGTGACCTTGGCGTTCTAGAACACCATTCCAATGTGGTGAAACAGGCTAATGA
 AGATGATGATGATAAGGACAGTTTATTGTGAGCTTATCTTAGTGATAGCTGGGACAGTCAAG
 TAGTCTTGACTTCTATGGTTCAATTTTTCCAACACTTTTCAAAACAAAATCAGCTAATCTGA
 GACGTTTTTCAAGTTTGATTAATAGTGTTGTGATTCTAGATGAAGTTCAATCCCTGCCTATTG
 AAGTCACCACTTTGTTTAAATTAACGATGAATTTTTTAAATAAAGTTATGGATACAACCATC
 GTTCTTTGCACAGCGACACAACCTGCTTATGATTCTTCAGAGATTGACCATCGTATCTGTTAT
 GGAGGGAACCTTGGGAGAATTAGCTGAAATAGTTGAGTTAACGATTGAAGAAAAACAGATTT
 TTTCAAGGACAGAGCTTAGAAAAATTTGATGATAGTGATCAGAAAGTTCACTTGACTGATGTT
 ATTAACCTTATTCTAGGTGAGGAAAACTCAGTTCTTGCTATTTTTAATACGAAAAAACGGT
 TCATAACTGCTATACTATGCTAAAAGACATGACTGATAGACCGGTCTATCAGCTTTTCGACAA
 ATATGTGTGCGCAGCATAGACTTGACTTGATTGCTAAGATCAAAACGGAGTTACAAAATAA
 TATCCCTATTATTTGTATTAGCACGCAATTAATTGAAGCAGGTGTAGATGTTGATTTTCATCG
 CGTCATTCGTTCCCTACTCAGGGATTGATTCTATTGTTCAAGGCTGCTGGACGGTGTAACCGAG
 AAGGCAAACGAGATAAAGGGCAAGTCACTCTTGTCATCTGACCAATGAAGAGGAAAATAT
 TTCTAGGCTGACAGAAATAAAAACTAAAAAAGAAGCCACAGAATCTATTCTTCATAAGATT
 GGGTCTCCAATTGATATCTCACTTTAAACCGTGACTTTTTTGAGTATTATTATGCCAATAAT
 CAGGGACTGATGGATTATCCTTTGGAAGACAACCTATCAATCTACGACTATTTAAGCCTTAA
 TATTTATCAGACGGCAAAATAAAAAAGTTCAAAGGTAAGTTAAACAAGCTTTTAAAAACAGCA
 GGAGCCAAAATGAACCTCATCAATAATGATATGATAGGAATTCTCGTACCTTATGGCGAAG
 CTGAGAAAAAATTGGCTTATTTAGAAGAATTAGGTGTGTCACATTTTTTATCAGCAAAAGAT
 TATCAAACGATAAAATCATTACTAAAAGAGTTACAACCTTTTACGGTTAATGTCCGCGAGAA
 CGATCCTCTCTTTGAGACAACAAAATCTTATCTAAATGGTCAGATTCTGGTTTTGACGTCGG

AGTATTATGACACGGAAAGAGGAGTTAAATACGATTCAGCTAGCTTTTACTTCTAA (SEQ ID NO:510)

SEQ ID NO:511

TTGTACAGATCTAGAGACTTCTACGTGAGAGTAAGTGGTCAGCGAGCTCTTTTACAAATCC
AGCCACAAAAGGGGGATCGGAACGCTCATCCTATTCGGTTCCGACTAGACAGGCACTGAAT
GGTATCGTTGATGCCATCTATTATAAGCCGACCTTTACTAATATCGTCACAGAGGTAAAGGT
TATTAACCAGATTCAAACCGAATTACAGGGGTGTCAGGGCTCTGTTACATGATTATAGTGCAG
ATTTAAGTTATGTATCCTATTTGAGTGATGTTGTTTATCTGATCAAGTTTCATTTTGTTGGA
ATGAAGATAGAAAAGATTTGAACTCAGATAGACTTCCAGCTAAACATGAAGCCATTATGGA
GCGTTCTATTCGTAAAGGGGGACGTGAGATGTGTTTTTGGGTACAAGAGAATGTTTAGGGC
TTGTAGATGATATCAGCCAAGAAGAGTATGAGACTACTGTGTCGTATTATAATGGTGTCAAT
ATCGACTTGGGAATCATGTTCCATTCCTTTGCCTATCCGAAGGACAAAAAGACACCATTA
ATCATACTTTACAAAGACTGTGATGAAAAATGGAGTCATTACGTTTAAAGCACAGTCTGAAT
GCGATATTGTTAACACGCTTTCTAGTTATGCTTTTAAAGCACCAGAGGAGATAAAATCGGTT
AACGATGAATGCATGGAGTATGATGCCATGGAGAAAGGAGAAAACCTGA (SEQ ID NO:511)

SEQ ID NO:512:

ATGGATTTTTTTACTTCTCTTGAAGACTTATGAAAAAGCAGAGCTAGCAGACTTGGTTGA
TCATCAAAAAAGAAATAATGAGCCGGTTTTACTGCCGATTTATCATACGAGTTTAAAGTCAA
ATGGTAAAAATATCATTTTCAGTGAAACTTGACAAAGATGGCCAGTTTCACAAGGCAGAATT
TATGGCAGATAAGCAAAATGATTATTTTCTGTAAACGGCTGATTCTGTTGCTAGGTCAGGTA
GTCATCCTGACCCGATCCCCTAGTCGATAAAATTTGCTTATTATAGTGCTGAAATGGGGCAG
ATTCAGTATGATTCTTTTCATAAGCAACTGAATAACTGGATTGATTATTGTGAGGAGGGTGA
TGTCAGAAATTTTTAACCTTTGTTTCAGCAGTTTCATTTTGAAGCCAGAATTTCTAACATTGAT
TCTTGATTCTTTAATTGGTCCTGATTATCAACATAATCAATTAAGAGTCACATTTTGTGATGC
CACTGGAAAAGAAAAATTAATTGATTTATCAGCTTGCTTTTTAGAAATTTCAATTGATCAGTT
CCAGGGCTTTAAAAATGAATCGGTTTCGACATTTAAAGCCTTACACCAATCCTATATTTCTTT
TGTTGAAGCCAATCGTGAAAATCTCGGTATTTGTAATATTAGTGGACGAGAGGAACAGCTTA
CCGATAAGCATAGAGGTTTGATGGGGAATGCTAAAATCATCTCTGTTAGTAATAAAAGAGA
AGCTTATAAAGGACGTTTTAGAGAACGCGAAGACGTTTTTAGTGTTGGCTATGAACTTCCG
AAAAGATTCATTTAATGCTCAAGTACCTTTTAGAAAAATAAAATACCAGTACTTGGTTAGGG
TCTTCTCAATATTTAATCAACTGGTTTCAGCGATGATTTAACAAATGATAGTCGGTTGGATATT
GTATCACCAATCTTTGATGATGGACTTGAAGAAGATGATGATGACGATACGCCTCCTGTTAT
AACATTAGCAACTGAAGACAATAAAAGAATTGGTAAATCATTATCAAGGGACAAAAATTA
TTTGCTAATGATGCCACTTACTACGTTGCTATTTTGAATAAAACCAGCAATGGGCGGATGC
TTTAAAAATATTTTCGTCAGCTTCAAGCGTCCCAATTACTCACCAATCTTAACAAGTGGCAGG
AAACATACAGTTGGGAGTCGCGATCTAAGTTTGGAAAAAGTCGCTTAAGAACCCTACTTTT
CATGACATCCTTAATGTGTCCTACGGGGTTGATAGGGATCGCTTCCTTGAATTAGATAATGA
TAACTTCAAAAGTGATCAAATTCAAAGTTAGTGGCAAGTTTGATTGATGGTAAACCGATGC
CACAGTCCATTGTCAAAAAGTTAGGTAACAATGTAAAGAACGACATCGTTACCGTAAGCA
CTGGTATCAAGTTGAGCAGGTCTGCTTAGCAATTTTACACAAACAAAATGGGGAGGAATTTT
CACCGATGETAGATCATACCAATCAAAATCGTTCCTATCTTTTGGACGATTATTAGCAATTT
TTGAATTAATCGAGACCTTGCGTTATGGCTTGGATGGAAACAATAACGACCGTATTACCAAT
GCTGAACGTTATTGGACAGCCTATACTGGACAACCAACAAAATTGATGATGTTATTGGAAA
ATAAAATTAAGCCTTACGAAGAACCATTGAAATTAATCGTTCGTGGCAGTTGGATGAAATT
AGAAAAAGAAAAAGAAGAGATTTTAGAACTGTTAAATCCTCTGTTAGAAACAGAAACAATG
GAAAAACCCTTAGATTACCGCTTTATTTTGGGTATTATGCTGAGAAAAACTATTACTATAC
AAAAACAAACACGGAAGTAACAGAAAGTGAGGAGTAA (SEQ ID NO:512)

SEQ ID NO:513:

ATGTTGGAACACAAAATTGATTTTATGGTAACTCTTGAAGTGAAAGAAGCAAATGCAAATG
GTGATCCCTTAAATGGAAACATGCCTCGTACAGATGCCAAAGGATATGGTGTGATGAGTGA
TGTCTCCATTAAACGTAAGATTTCGTAATCGTTTGAAGATATGGGGAAGTCTATTTTGTGC

AAGCTAATGAGCGTATTGAAGATGATTTTCGTTCACTGGAAAAACGCTTTTCGCAACATTTT
 ACAGCTAAGACACCTGACAAAGAAATTGAAGAAAAAGCAAATGCATTATGGTTTGATGTTT
 GTGCTTTTGGACAAGTTTTTACTTATCTGAAAAAATCAATTGGGGTGCGTGGACCAGTTTCC
 ATCAGTATGGCTAAGTCCTTGGAGCCAATTGTCAATTTCCAGCCTTCAAATTACGCGTAGTAC
 CAATGGTATGGAAGCTAAGAATAATAGTGGCCGCTCTTCTGATACGATGGGGACAAAACAT
 TTTGTAGATTATGGTGTGTATGTACTTAAAGGTTCTATCAATGCTTATTTTGCTGAAAAGACT
 GGTTTTCTCAGGAAGATGCTGAGGCTATTAAAGAAGTTTTGGTTAGCTTGTGTTGAAAATGA
 TCGTCTGCTGCACGTCCGGAAGGCTCTATGCGAGTTTGTGAAGTCTTTTGGTTTACGCATT
 AAGCAAATTGGGAAATGTTTCAAGTGCAGCTGTCTTTGACTTGTAGAGTATCATCAATCAA
 TAGAAGAAAAAAGCACTTATGACGCTTATCAGATTATCTAAATCAAGAAAAATTGGCTAA
 ATATGAAGCGAAAGGGTTAACGCTTGAAATCCTAGAAGGACTCTAG (SEQ ID NO:513)

SEQ ID NO:514:

ATGGTCTATGCCGAAGATGATTATTTAATGCTGTCAGGTATTCAGCATTTCCAATTTTGTA
 CGTCAATGGGCGTTGATCCATATTGAGCAACAATGGCTTGATAATGAAGCGACAGCGCATG
 GACAGGTTTACATACTAAAGCAGATAACCCTTACATTAAGAAAAACGAAAAGAGCTTTT
 GGTCTCACGTGCTATGCCATTTCTTCTGCAGAACTTGGACTTTCAGGAATTATGGATGTTGT
 GGAATTTTATAAAGATGATCAAGGTGTGTCTTTGAGGGGAAAACGTGGGAAATGGTTACCA
 AAAGTTGTGGAATACAAGCGCGGAAAACCTAAAAAAGATACCAGAGATATTGTCCAGTTGG
 TGGCTCAGACCATGTGTTTGAAGAAACGCTAGACTGCGACATTAACGAAGGTTGTCTTTAT
 TACCATAGTGTCAATCAAAGAGTGATTGTTCTATGACATCAGCTTTGCGTCAAGAAGTGAA
 GGAATTAGCCGCAGAGATGCATGAGGTTTATCAGAGTCAAATGCTACCTAAAGCAGCTTAT
 TTTAAAACTGTCAGCTTTGTTCTTTAGTCGATATTTGTAAGCCCAGGTTGAGTAAAAAAC
 AAGGAGTGTGTCGCGTTACATCAATGAGGCTATGACCAGTGAGGAGATGGACCTATGA
 (SEQ ID NO:514)

SEQ ID NO:515:

ATGAAGAAGTTGCTAAATACCTTGTATTTGACGCAAGAAGATTTTTATGTCCTAAAGAGGG
 CGATAACATTGTTATCAAGCAAGAAGGTAAGGTTCTCAAACGGTTTCCGTTTCGGATTATTG
 ACGGTATTGTCTGTTTTTCTTATTTGGGTGTGTCGTCTGCTTTGGTGAAGTTATGTACGGAGA
 ATCAGATTAATTTATCGTTTCATACACCACAAGGGCGTTTTTGTGGTTCGCTATATTGGTTCAA
 CCAATGGGAATGTGTTGTTGCGTAGAGAACATTATCGTTTATCTGATCGTGAGGAATCTTTG
 GAATACGCAAAGCGGTTTTATTTTGGCTAAAATTTCCAACCTCAAGGAAATACTTGCTACGCTT
 TAAACGAGATCATCGTCAACAGATTGATACCAAGCTTTTTGAGGCTGTTAATGACGAATTGA
 TATGGGCTTTAGAGATGGTTCAGGCAGCAGATAATAAAGACTCTTTAAGAGGGATTGAAGG
 CCAAGCTGCTAATCAGTATTTTCGCATATTTAATGACCTGGTGTGACGGACAAAAAACGT
 TTTACTTCCAAGGTCGGAGTAAACGACCACCCTTAGATTGTGTTAATGCCCTCTTGCTTTTG
 GTTACAGTTTACTGACCTTTGAATGTCAATCTGCCTTGGAAAGCTGTCGGATTAGACAGTTAC
 GTTGGTTTCTTTCACACGGATCGTCCTGGGCGTGCTAGTTTAGCGCTTGATTTAGTTGAAGAG
 TTCCGCTCATATATTGTAGATCGTTTTGTCTTTTCATTAATTAATAAAGGACAACCTCAGAAA
 AAACACTTTGAGGTTAAAGAAAAATGGTAGTATTTTATTGACGGAAAATGGCAGAGCTATTTT
 TATTGATTTGTGGCAGAAGCGTAAGCATACTGAGGTAGAATCCTTTTACAAAAGAGAAA
 GTAAACTTATGTTATTACCCTATGTACAAGCGCAGCTTTTAGCTAAGGCTATACGAGGAGA
 TTTAGAAAGCTATCCACCTTTTATGGTTTAG (SEQ ID NO:515)

SEQ ID NO:516:

ATGATGGTTTTAGTCACTTATGATGTAAATACGGAAACACCTGCTGGTAGAAAAAGATTGCG
 TCATGTTGCCAACTCTGTGTGGACTATGGGCAACGTGTTCAAATTTCTGTTTTGAATGTTT
 TGTGACACCCGCAGAAATTTGTGGATATAAAGCACCGCTTAACACAAATCATTGATGAGAAA
 ACTGATAGTATTCGCTTTTATTTATTGGGGAAAAATTGGCAGAGGCGTGTGGAAACACTTGG
 TCGTCTCAGACAGCTATGACCCAGATAAAGGTGTCTTATTATTGTAA (SEQ ID NO:516)

con secuencias repetidas: SEQ ID NO: 20 y SEQ ID NO: 22.

Todas las patentes y publicaciones mencionadas en la especificación son indicativas de los niveles de aquellos experimentos en el arte a los cuales la invención pertenece. Todas las patentes y publicaciones se incorporan en la presente por referencia en el mismo grado que si se indicara específica e individualmente que cada publicación individual se incorpora por referencia.

5 Aquellos de experiencia en el arte fácilmente aprecian que la presente invención está bien adaptada para llevar a cabo los objetivos y obtener los fines y ventajas mencionados, así como aquellos inherentes en ellos. Las composiciones y métodos descritos en la presente son representativos de modalidades preferidas, son ejemplares, y no se pretenden como limitaciones en el alcance de la invención.

10 La invención descrita ilustrativamente de forma adecuada en la presente puede practicarse en la ausencia de cualquier elemento o elementos, limitación o limitaciones que no se describen específicamente en la presente. Los términos y expresiones que se han empleado se usan como términos de la descripción y no de limitación, y no existe la intención de que el uso de tales términos y expresiones excluya cualquiera de los equivalentes de las características mostradas y descritas o porciones de las mismas, pero se reconoce que varias modificaciones son posibles dentro del alcance de la invención. De esta manera, deberá entenderse que aunque la presente invención se ha descrito específicamente por modalidades preferidas y características, modificación y variación opcionales de los conceptos descritos en la presente pueden recurrirse por aquellos experimentados en el arte, y que tales modificaciones y variantes se consideran dentro del alcance de esta invención.

20 La invención se ha descrito ampliamente y genéricamente en la presente. Cada una de las especies más limitadas y agrupaciones subgenéricas caen dentro de la descripción genética también forman parte de la invención. Esto incluye la descripción genérica de la invención con la condición de que o la limitación negativa remueva cualquier materia objeto del género, sin tener en cuenta si la materia removida se recita específicamente en la presente o no.

25

REIVINDICACIONES

1. Un método para generar un cultivo de inicio que comprende al menos dos variantes de una cepa resistente a bacteriófagos, que comprende los pasos de:

(a) la exposición de una cepa bacteriana precursora que comprende al menos una porción de una ubicación CRISPR a un bacteriófago para producir una mezcla de bacterias que comprende una variante de la cepa resistente a bacteriófagos que comprende una ubicación CRISPR modificada que comprende al menos un espaciador adicional en dicha ubicación CRISPR modificada;

(b) la exposición independiente de la misma cepa bacteriana precursora que en el paso (a) que comprende al menos una porción de una ubicación CRISPR al mismo bacteriófago que en el paso (a) para producir una mezcla de bacterias que comprende otra variante de la cepa resistente a bacteriófagos que comprende una ubicación CRISPR modificada que comprende al menos un espaciador adicional en dicha ubicación CRISPR modificada;

(c) la selección de dichas variantes de la cepa resistente a bacteriófagos de dichas mezclas de bacterias;

(d) la selección de dichas variantes de la cepa resistente a bacteriófagos que comprenden un espaciador adicional en dicha ubicación CRISPR modificada de dichas cepas resistentes a bacteriófagos seleccionadas en el paso (c); y

(e) el aislamiento de dichas variantes de la cepa resistente a bacteriófagos, donde dichas cepas comprenden un espaciador adicional en dicha ubicación CRISPR modificada;

donde la secuencia de al menos un espaciador adicional en la variante de la cepa resistente a bacteriófagos es diferente de la secuencia de al menos un espaciador adicional en la otra cepa resistente a bacteriófagos.

2. El método de la Reivindicación 1, donde dicho método comprende además la etapa de comparar dicha ubicación CRISPR o una porción de la misma de dicha cepa bacteriana precursora y dicha ubicación CRISPR modificada de las variantes de la cepa resistente a bacteriófagos para identificar las variantes de la cepa resistente a bacteriófagos que comprenden al menos un espaciador adicional en dicha ubicación CRISPR modificada que está ausente de dicha ubicación CRISPR de dicha cepa bacteriana precursora.

3. El método de la Reivindicación 1, donde dicho bacteriófago se selecciona del grupo de familias de virus que consiste en: Corticoviridae, Cystoviridae, Inoviridae, Leviviridae, Microviridae, Myoviridae, Podoviridae, Siphoviridae, y Tectiviridae.

4. El método de la Reivindicación 1, donde dicho bacteriófago es un bacteriófago de origen natural o es un bacteriófago mutado obtenido mediante presión selectiva utilizando una bacteria resistente a bacteriófagos.

5. El método de la Reivindicación 1, donde dichas cepas resistentes a bacteriófagos o dicha cepa bacteriana precursora es una mutante insensible a bacteriófagos.

6. El método de la Reivindicación 1, donde las porciones terminales 5' y/o 3' de dicha ubicación CRISPR de la cepa bacteriana precursora se comparan con dicha ubicación CRISPR modificada de dichas variantes de la cepa resistente a bacteriófagos.

7. El método de la Reivindicación 1, donde las porciones terminales 5' y/o 3' de al menos el primer espaciador CRISPR de dicha ubicación CRISPR de dicha cepa bacteriana precursora se comparan con dicha ubicación CRISPR modificada de dichas variantes de la cepa resistente a bacteriófagos.

8. El método de la Reivindicación 1, donde dicha al menos una porción de dicha ubicación CRISPR de dicha cepa bacteriana precursora y al menos una porción de dicha ubicación CRISPR modificada de dichas variantes de la cepa resistente a bacteriófagos se comparan al amplificar, preferiblemente utilizando la reacción en cadena de la polimerasa, al menos una porción de dicha ubicación CRISPR y al menos una porción de dicha ubicación CRISPR modificada, para producir una secuencia de ubicación CRISPR amplificada y una secuencia de ubicación CRISPR modificada amplificada.

9. El método de la Reivindicación 1, donde dicha al menos una porción de dicha ubicación CRISPR de dicha cepa bacteriana precursora y al menos una porción de dicha ubicación CRISPR modificada de dichas variantes de la cepa resistente a bacteriófagos se comparan al procesar por secuencia al menos una porción de dicha ubicación CRISPR y al menos una porción de dicha ubicación CRISPR modificada.

10. El método de la Reivindicación 8, que además comprende la etapa de procesar por secuencia dicha secuencia de ubicación CRISPR amplificada y dicha secuencia de ubicación CRISPR modificada amplificada.

11. El método de la Reivindicación 1, donde dicho espaciador adicional en dicha ubicación CRISPR modificada forma parte de una unidad espaciadora de repetición.
- 5 12. El método de la Reivindicación 1, donde al menos uno de dichos espaciadores adicionales forma parte de una unidad espaciadora de repetición que comprende al menos 44 nucleótidos o donde dicha unidad espaciadora de repetición adicional comprende entre 44 nucleótidos y 119 nucleótidos.
- 10 13. El método de la Reivindicación 1, donde al menos uno de dichos espaciadores adicionales forma parte de una unidad espaciadora de repetición que comprende al menos una secuencia de nucleótido que tiene al menos 95% de identidad con una repetición CRISPR en dicha ubicación CRISPR de dicha cepa bacteriana precursora.
- 15 14. El método de la Reivindicación 1, donde al menos uno de dichos espaciadores adicionales forma parte de una unidad espaciadora de repetición que comprende al menos una secuencia de nucleótido que tiene al menos 95% de identidad con una secuencia de nucleótido en el genoma de dicho bacteriófago.
- 20 15. El método de la Reivindicación 1, donde dicha cepa bacteriana precursora es una cepa útil en la industria.
- 25 16. El método de la Reivindicación 15, donde dicha cepa bacteriana precursora es susceptible a infección por al menos un bacteriófago.
- 30 17. El método de la Reivindicación 15, donde dicha cepa bacteriana precursora es una cepa obtenida de un cultivo seleccionado de cultivos de inicio, cultivos probióticos, y cultivos de suplementos dietéticos.
- 35 18. El método de la Reivindicación 1, donde dicha cepa bacteriana precursora se selecciona de *Escherichia*, *Shigella*, *Salmonella*, *Erwinia*, *Yersinia*, *Bacillus*, *Vibrio*, *Legionella*, *Pseudomonas*, *Neisseria*, *Bordetella*, *Helicobacter*, *Listeria*, *Agrobacterium*, *Staphylococcus*, *Streptococcus*, *Enterococcus*, *Clostridium*, *Corynebacterium*, *Mycobacterium*, *Treponema*, *Borrelia*, *Francisella*, *Brucella*, *Campylobacter*, *Klebsiella*, *Frankia*, *Bartonella*, *Rickettsia*, *Shewanella*, *Serratia*, *Enterobacter*, *Proteus*, *Providencia*, *Brochothrix*, *Bifidobacterium*, *Brevibacterium*, *Propionibacterium*, *Lactococcus*, *Lactobacillus*, *Pediococcus*, *Leuconostoc* y *Oenococcus*.
- 40 19. Un cultivo de inicio obtenido mediante el método de la Reivindicación 1.
- 45 20. Un método de fermentación que comprende agregar el cultivo de inicio de la Reivindicación 19 a un medio de fermentación, bajo condiciones tales que ocurra la fermentación de los componentes de dicho medio de fermentación.
- 50 21. El método de la Reivindicación 20, donde dicha fermentación no se ve afectada por la presencia de bacteriófagos.
- 55 22. El método de la Reivindicación 20, donde dicho medio de fermentación es un producto alimenticio tal como un producto lácteo, por ejemplo, leche.
- 60 23. El método de la Reivindicación 20, donde dicho medio de fermentación se expone secuencialmente a al menos dos cultivos de inicio diferentes.

FIGURA 1

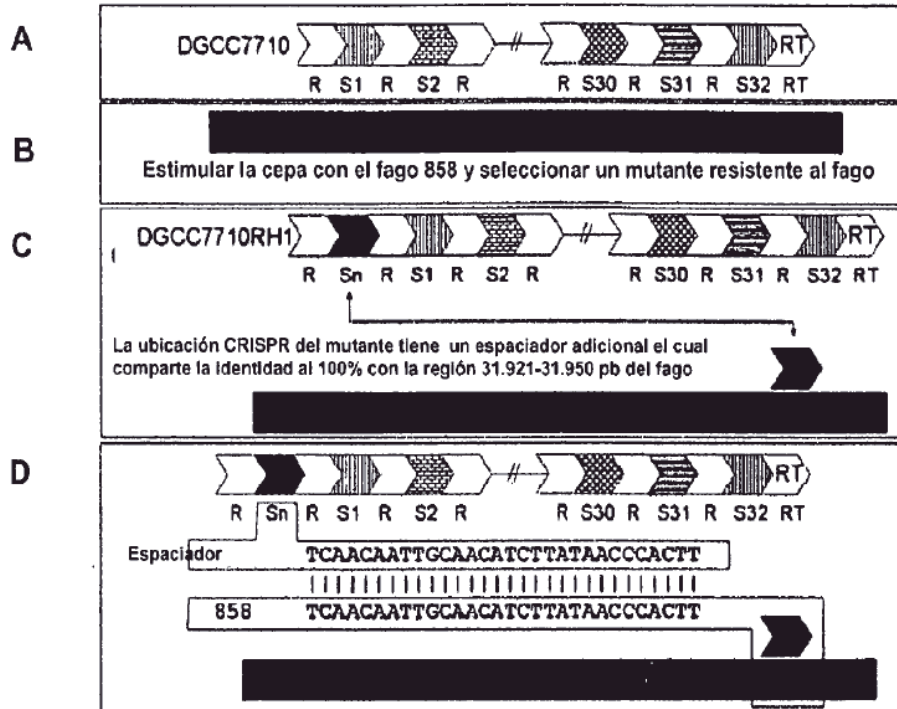


FIGURA 2

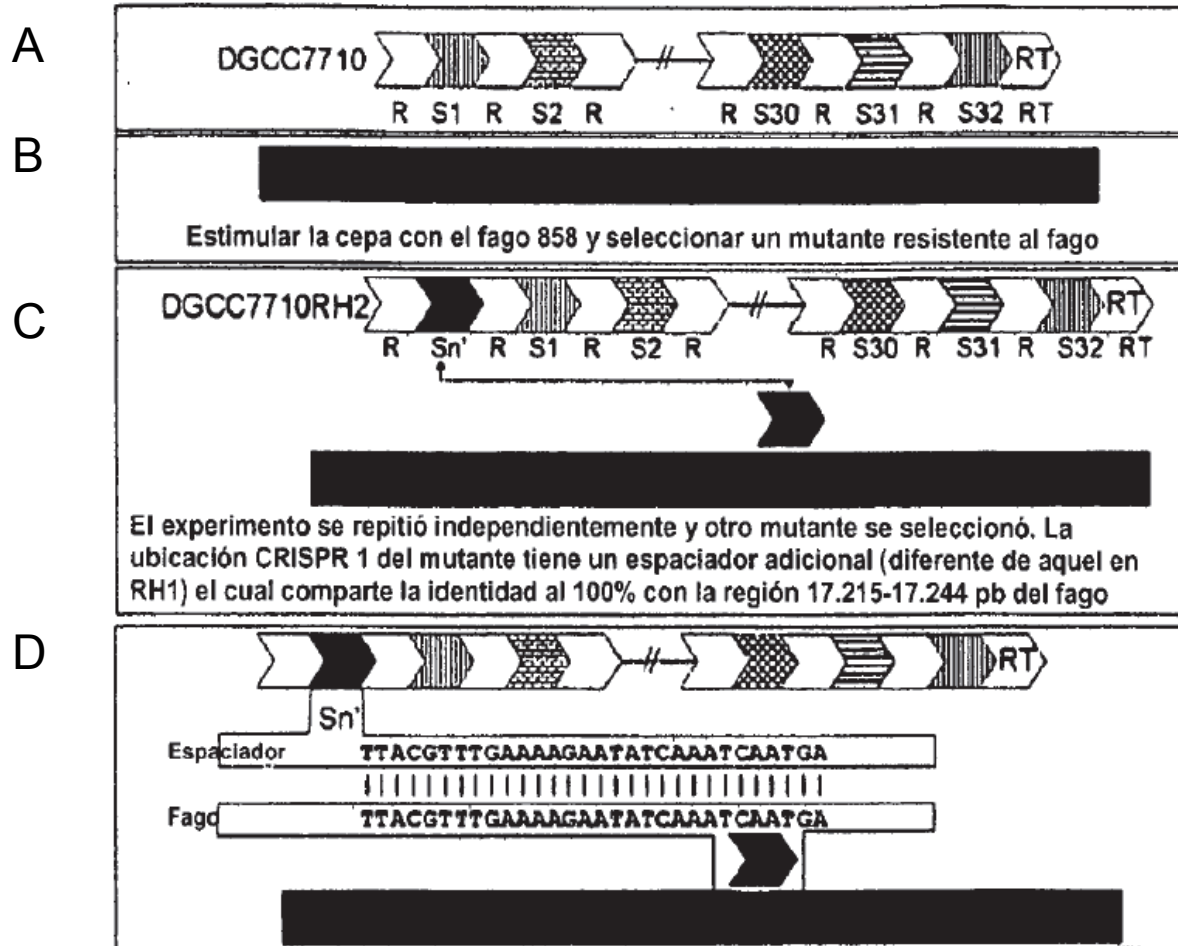


FIGURA 3

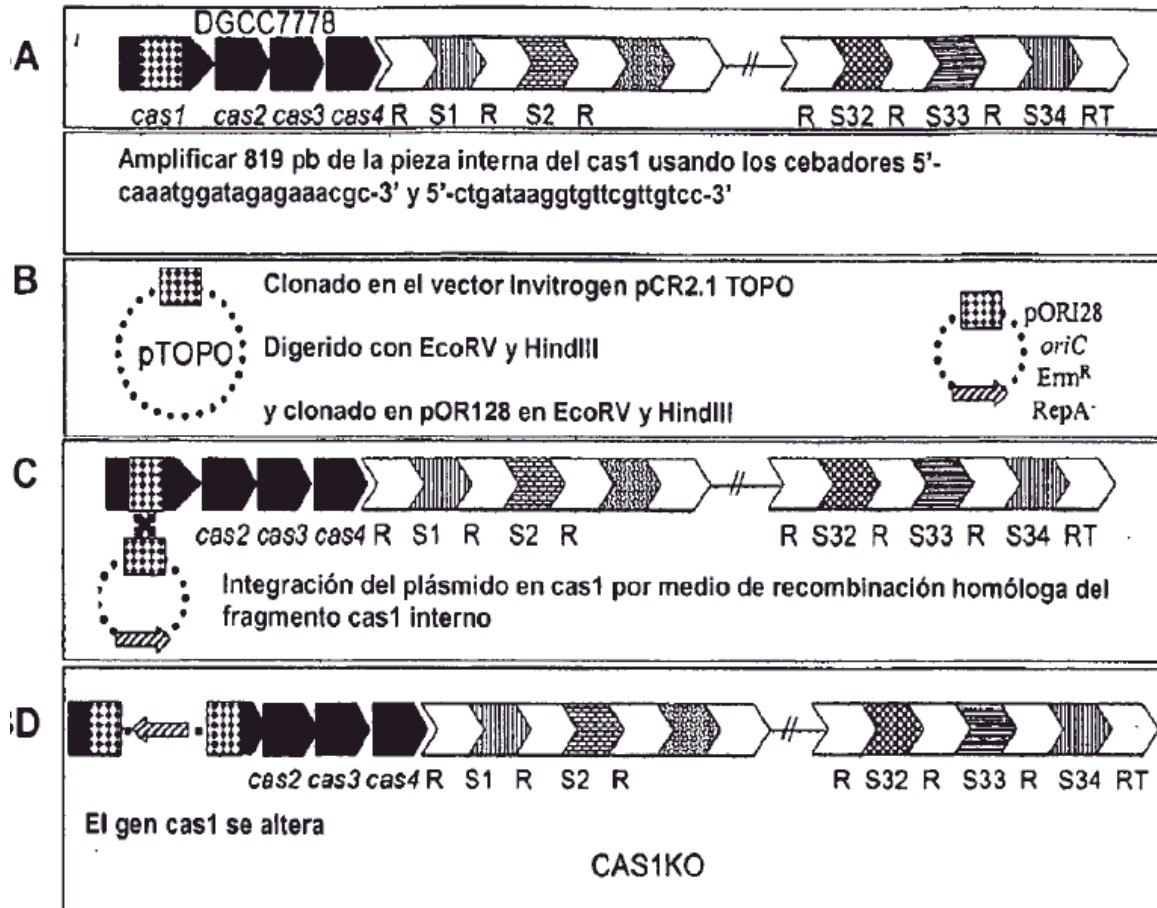


FIGURA 4

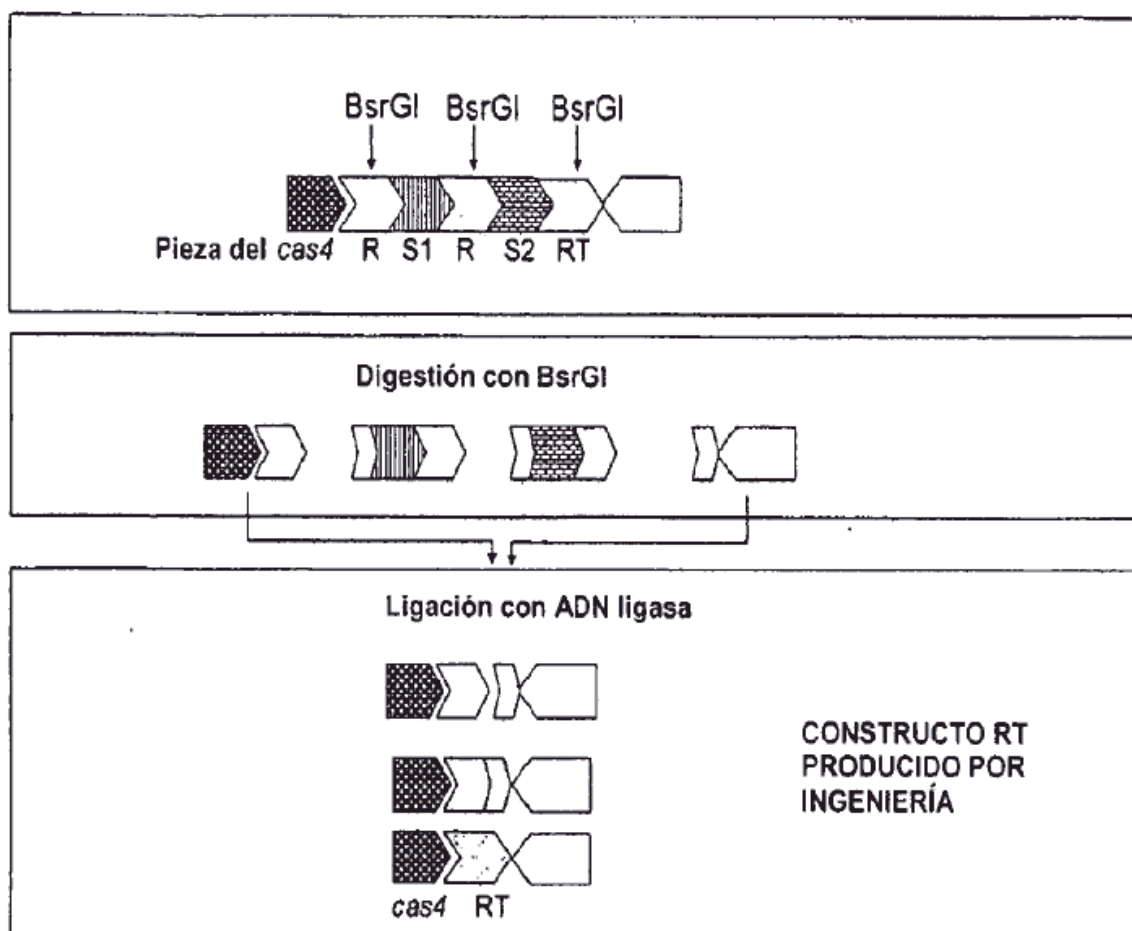


FIGURA 5

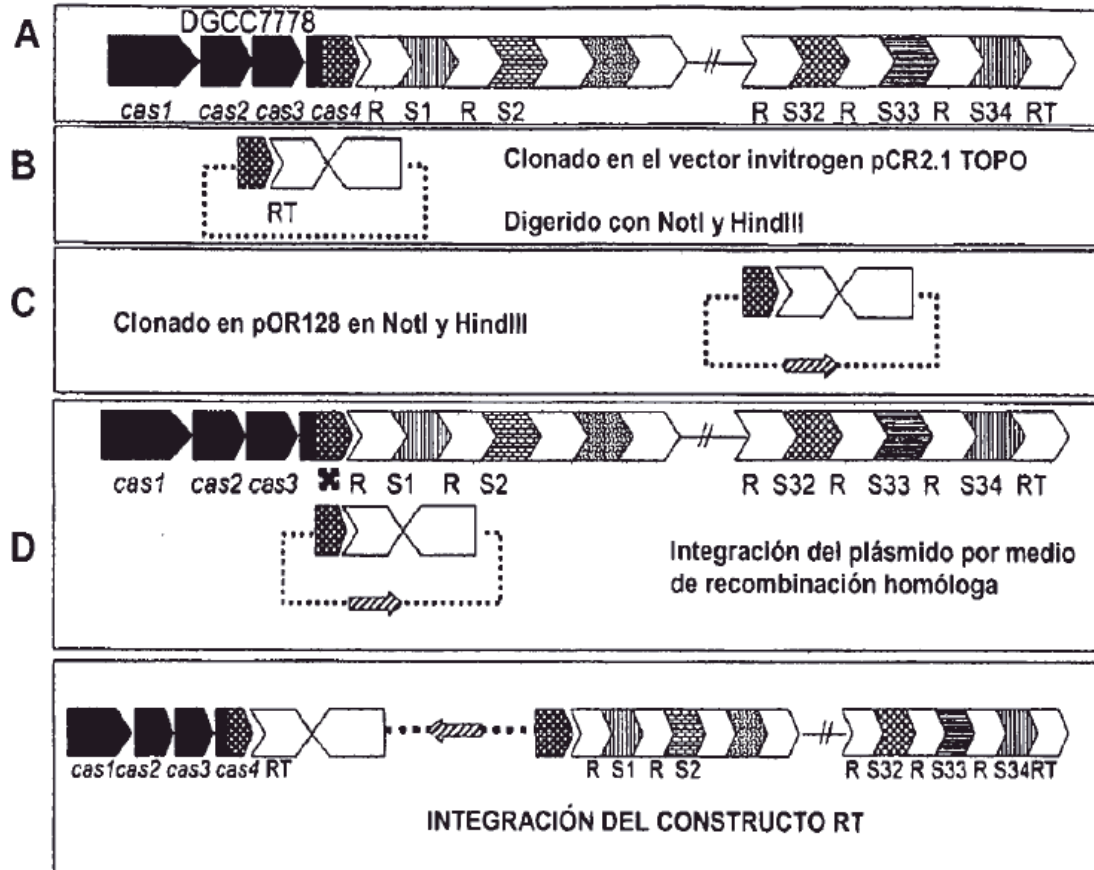


FIGURA 6

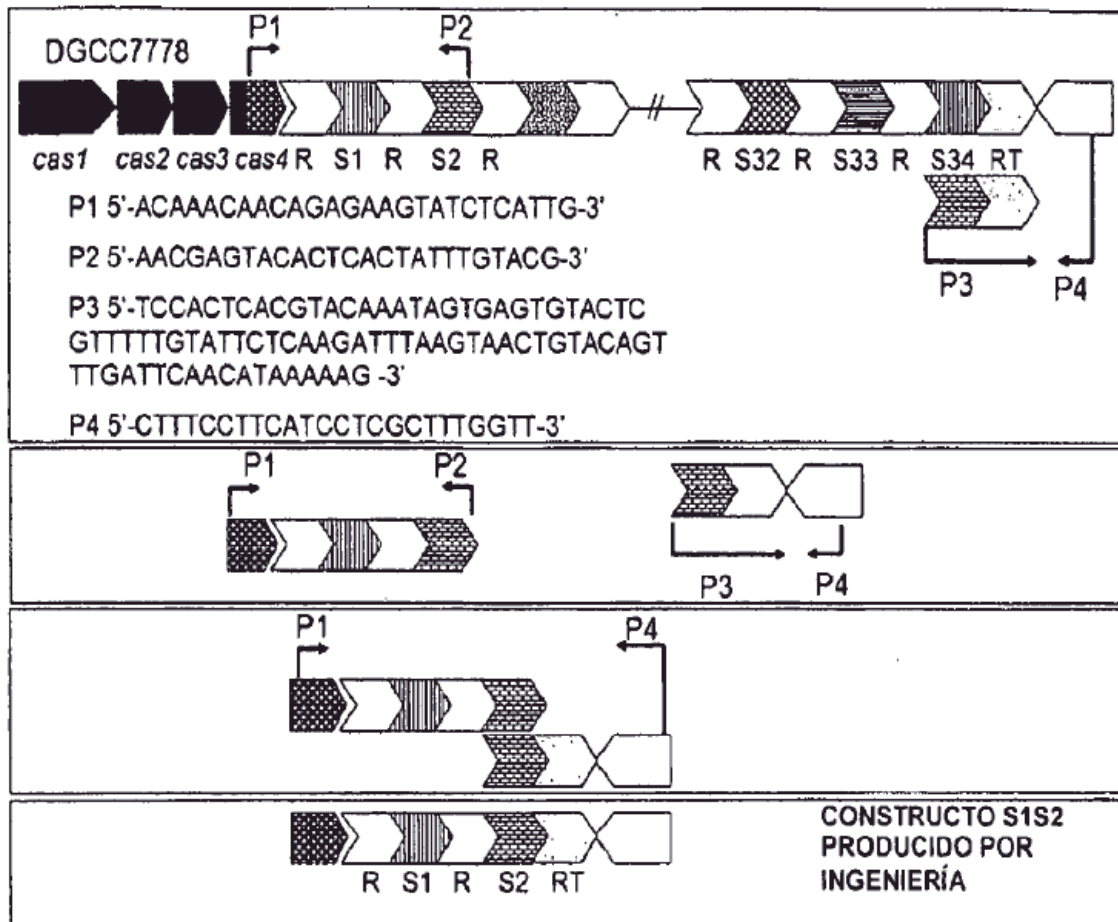


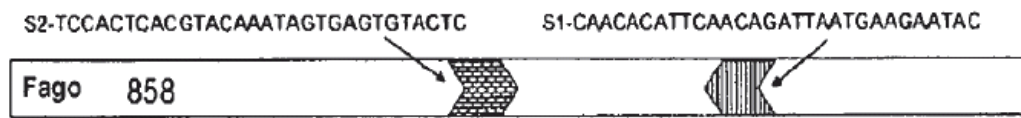
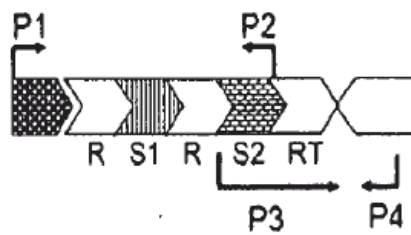
FIGURA 7

P2 5'-AACGAGTACACTCACTATTTGTACG-3' este es el complemento inverso de la parte de S2

P3 5'-TCCACTCACGTACAAATAGTGAGTGTACTC-
GTTTTGTATTCTCAAGATTAAAGTAACTGTACAGT TTGATTCAACATAAAAAG -3'

P3 se diseñó para contener:

- S2 TCCACTCACGTACAAATAGTGAGTGTACTC
- RT GTTTTGTATTCTCAAGATTAAAGTAACTGTACAGT



S1 y S2 ambos se originan del fago 858, S1 está en ORF
38 a 31,381 pb – S2 está en ORF 27 a 25,440 pb

FIGURA 8

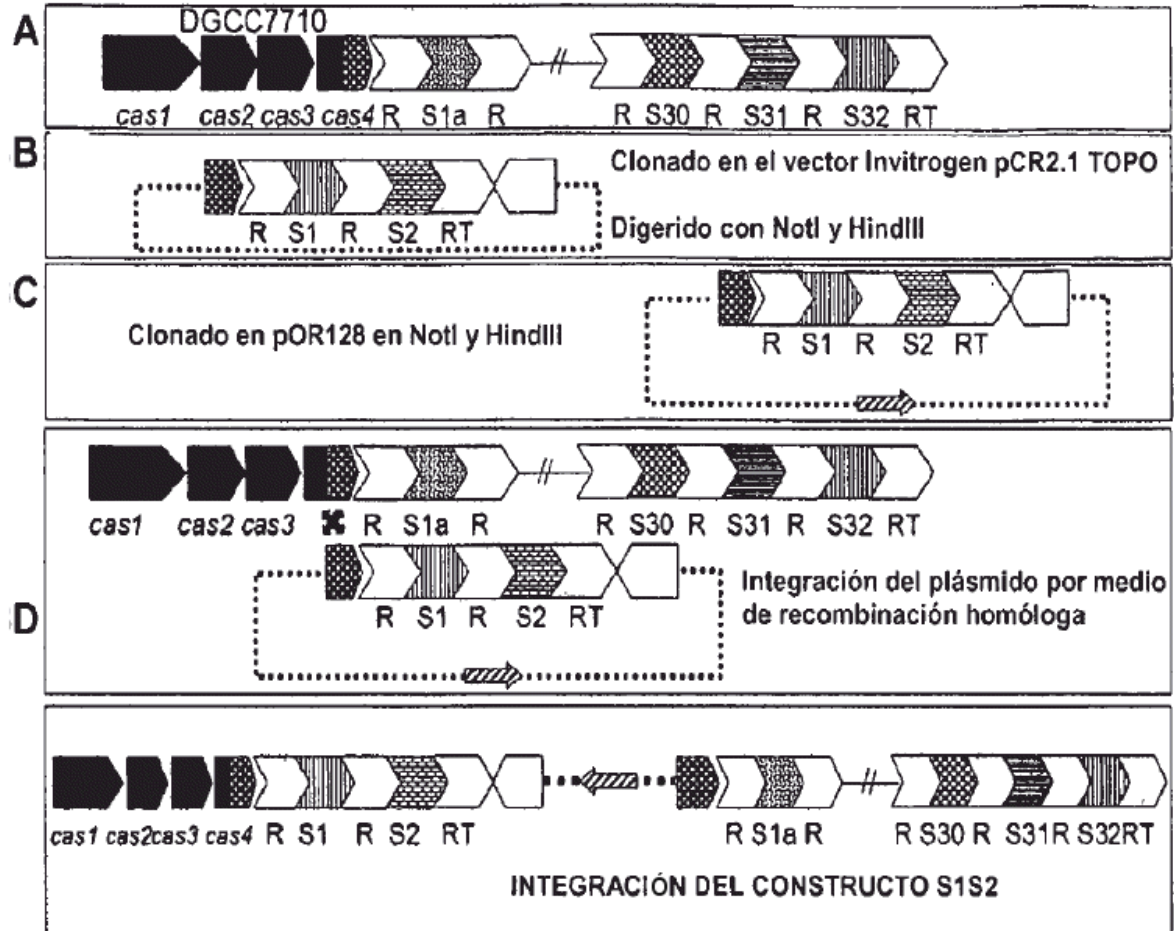


FIGURA 9

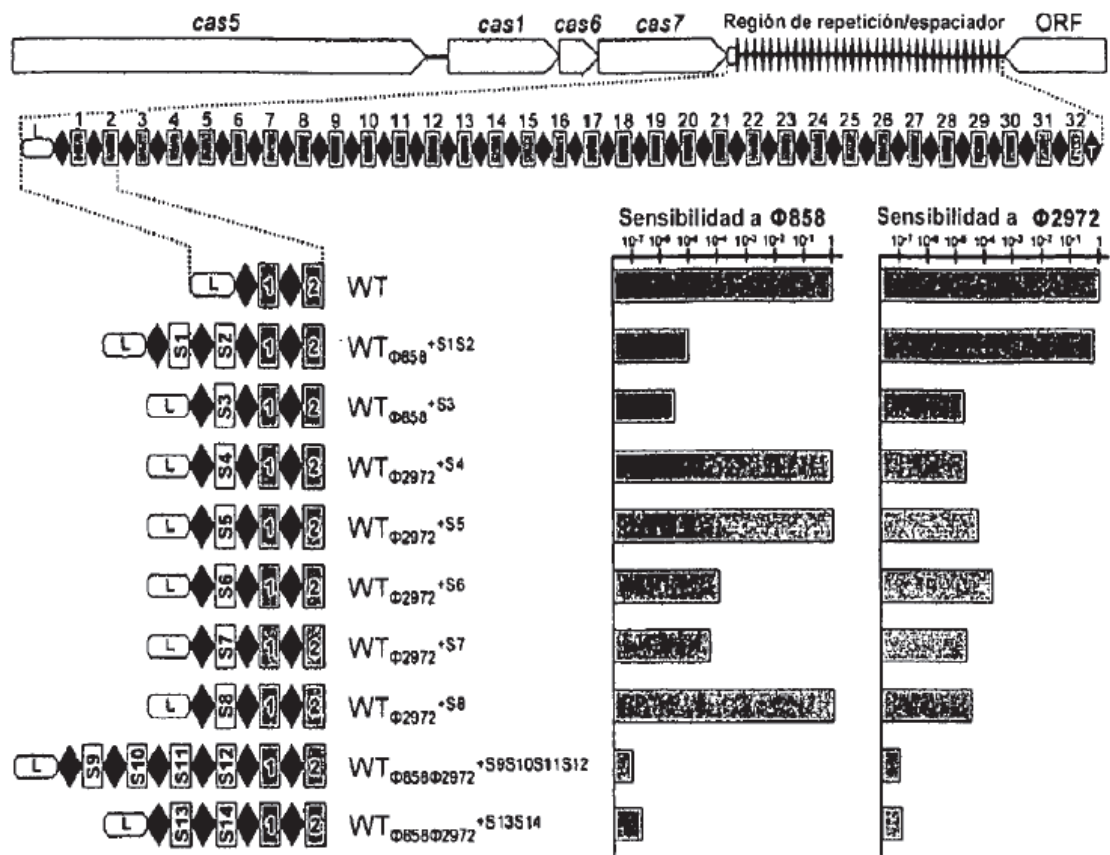


FIGURA 10

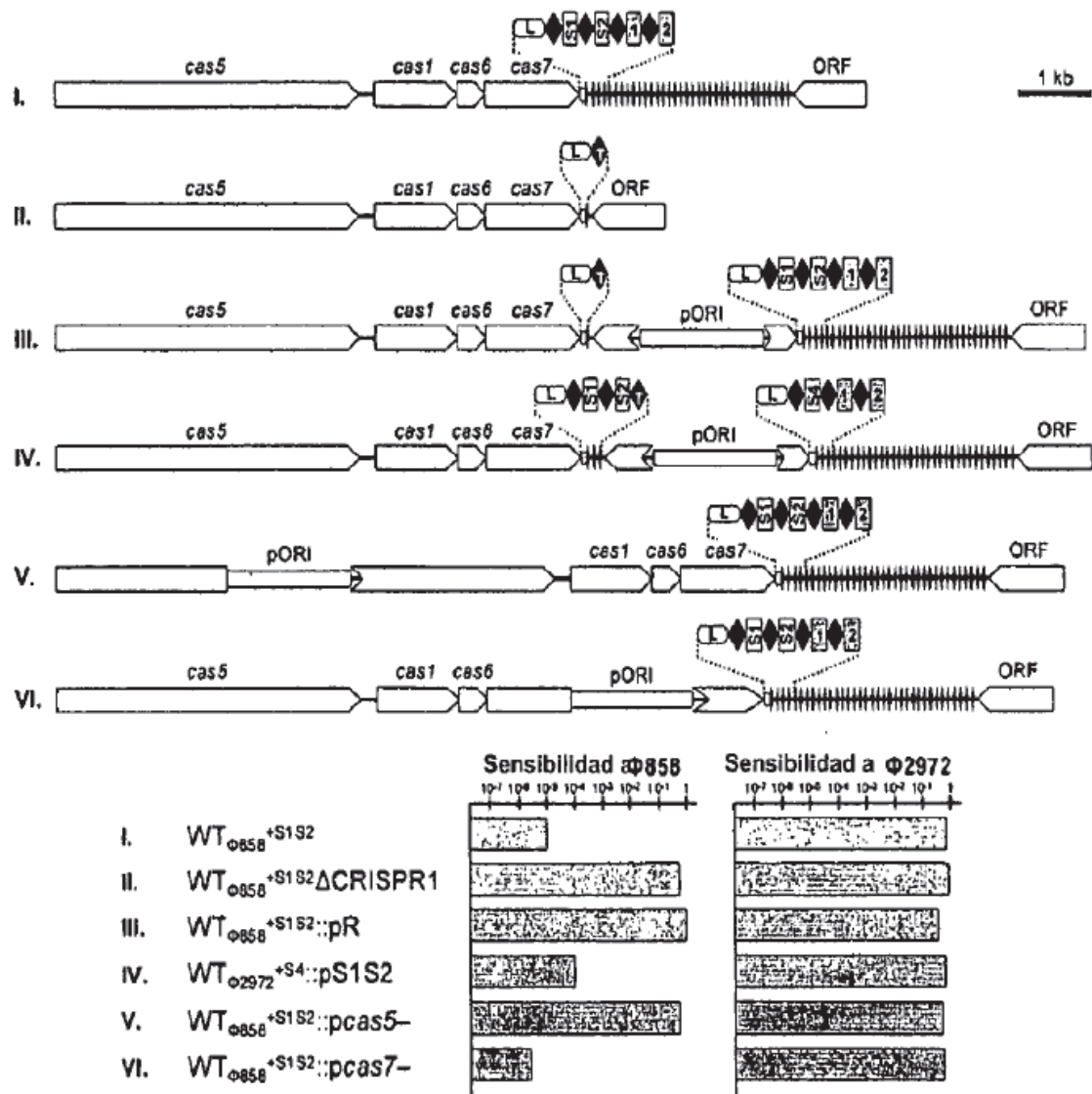


FIGURA 11

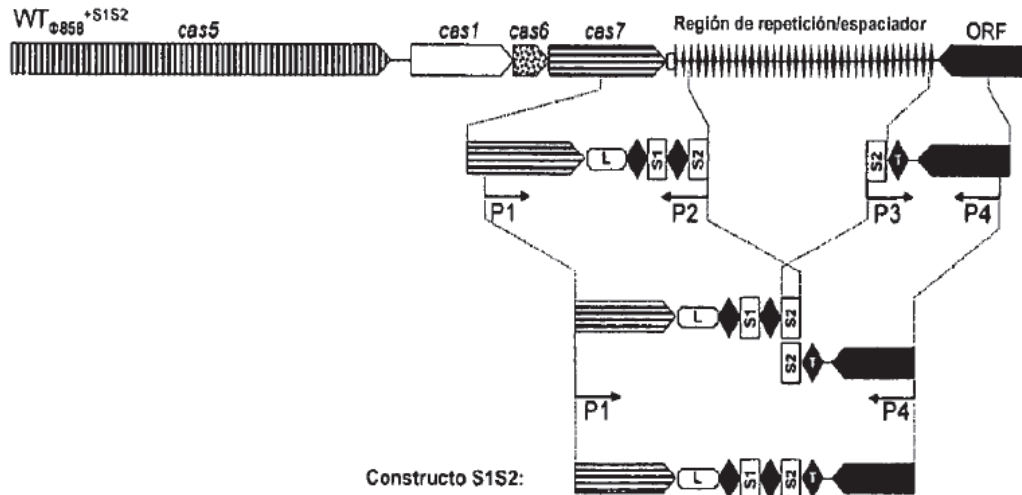


FIGURA 12

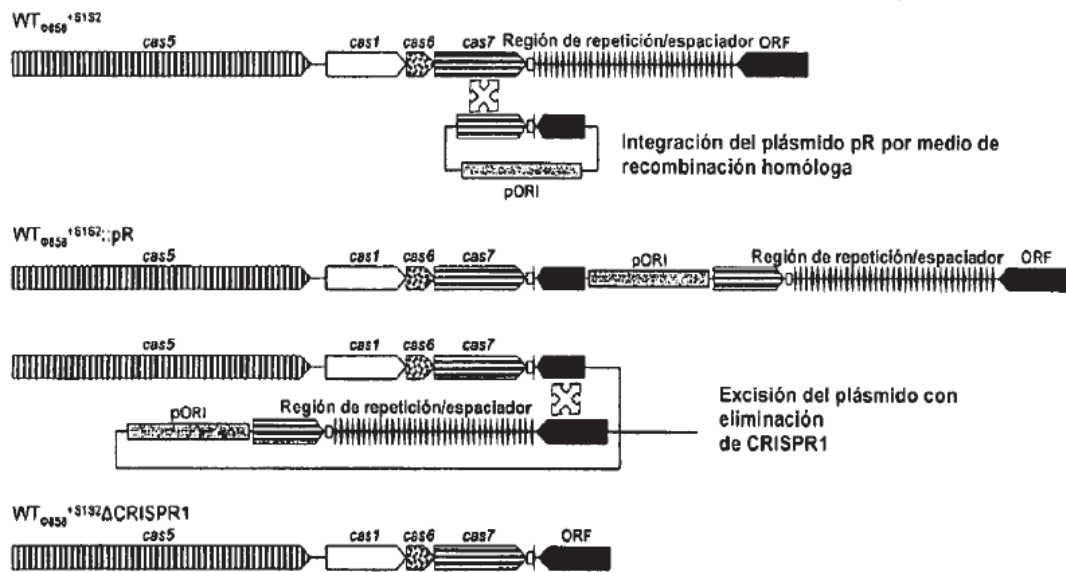


FIGURA 13

S1	CAACACATTCAACAGATTAATGAAGAATAC
Φ858	
Φ858 - A	A.....
Φ858 - B	C.....

FIGURA 14

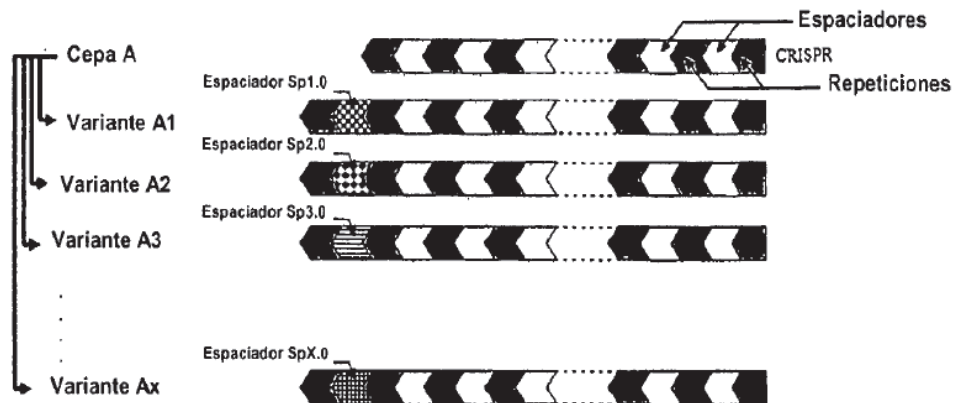


FIGURA 15

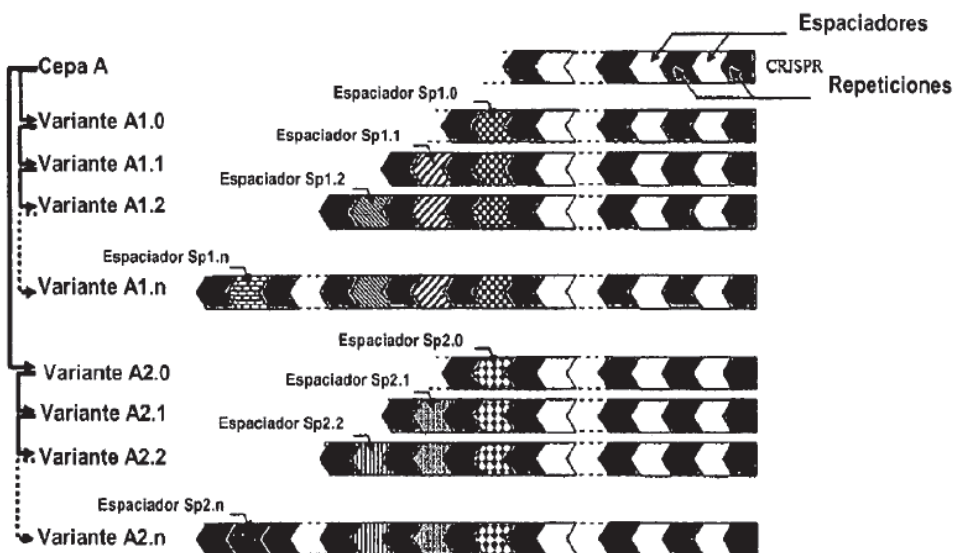


FIGURA 16



FIGURA 17

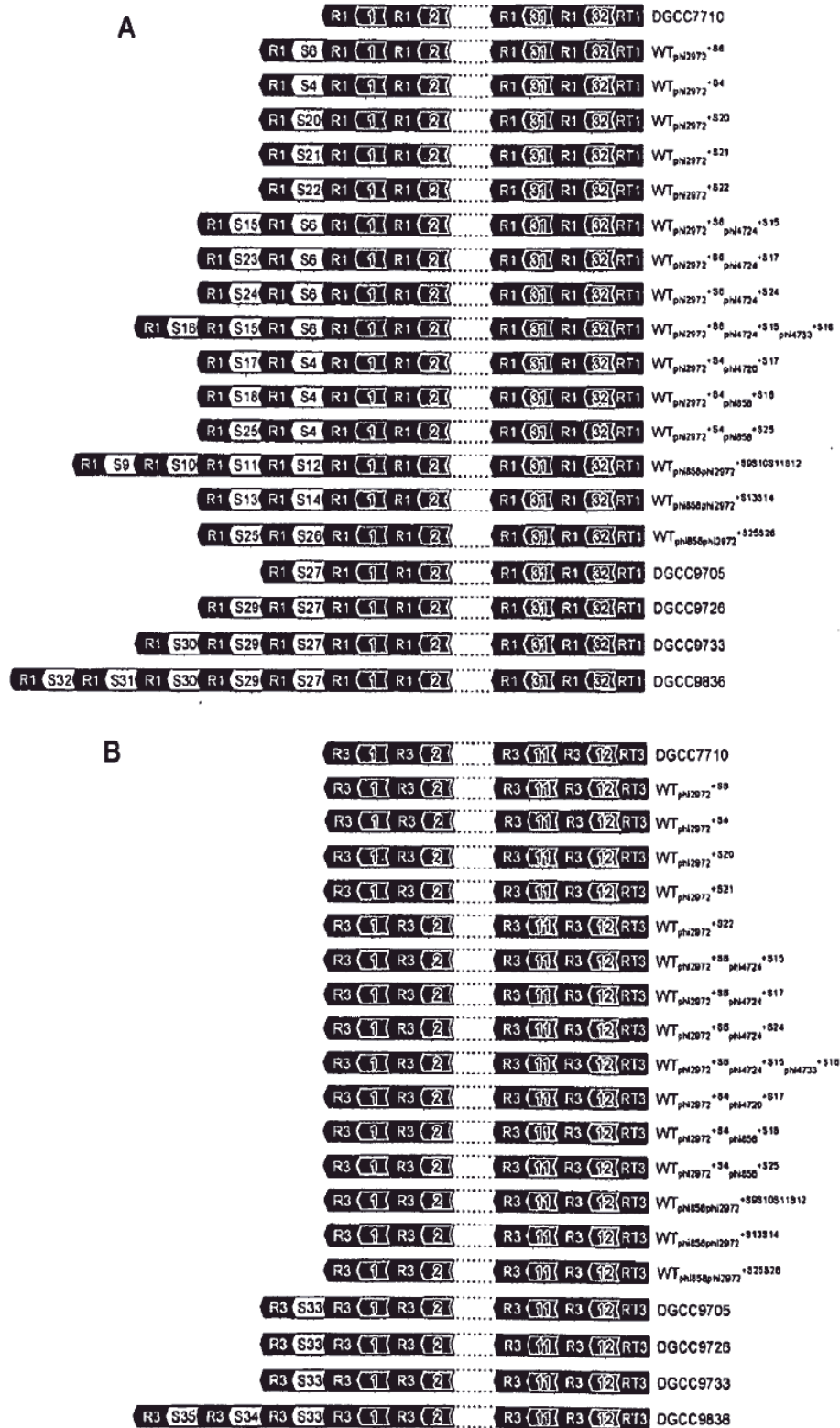


FIGURA 18

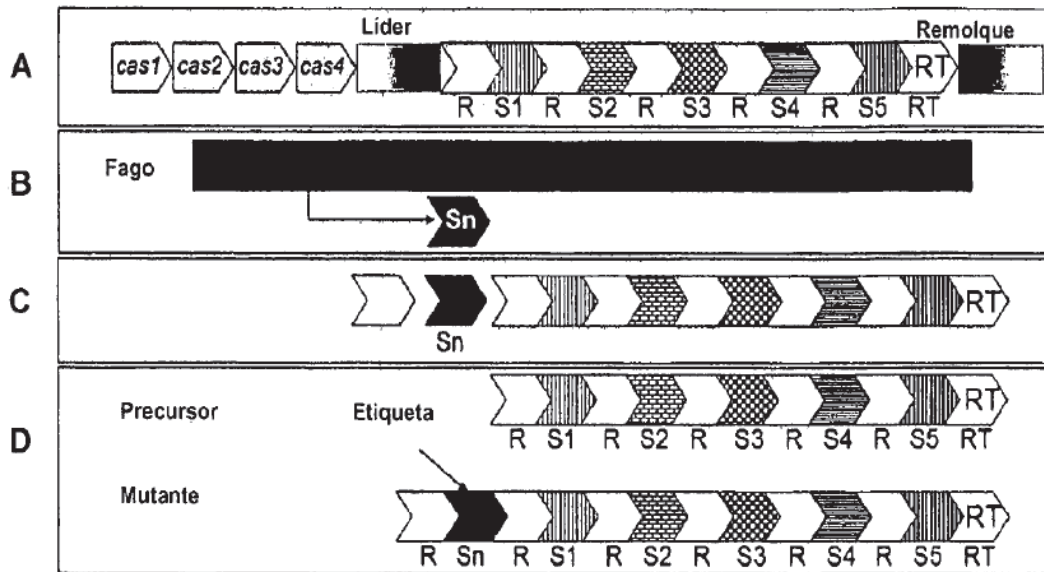


FIGURA 19

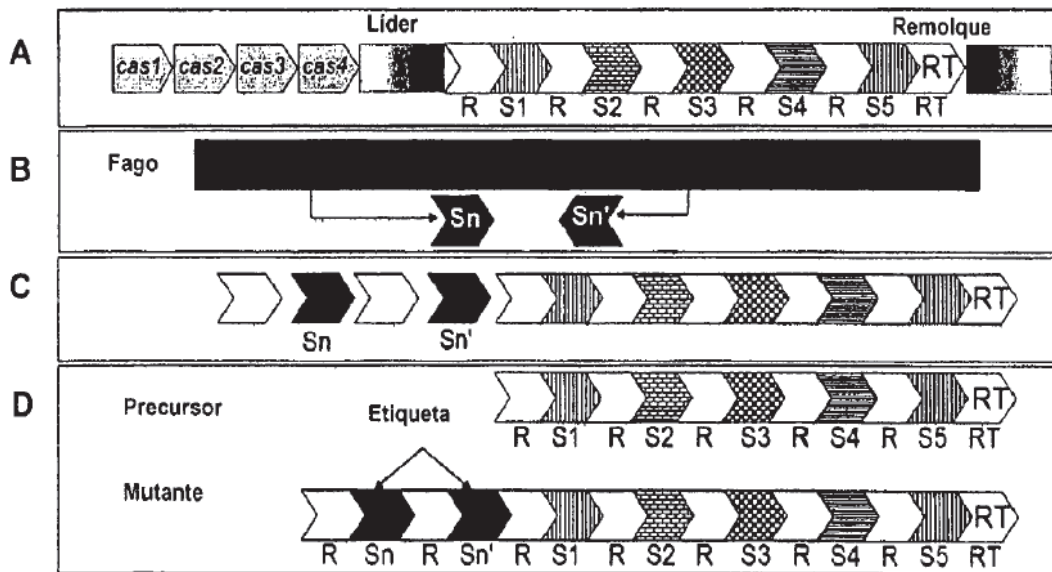


FIGURA 20

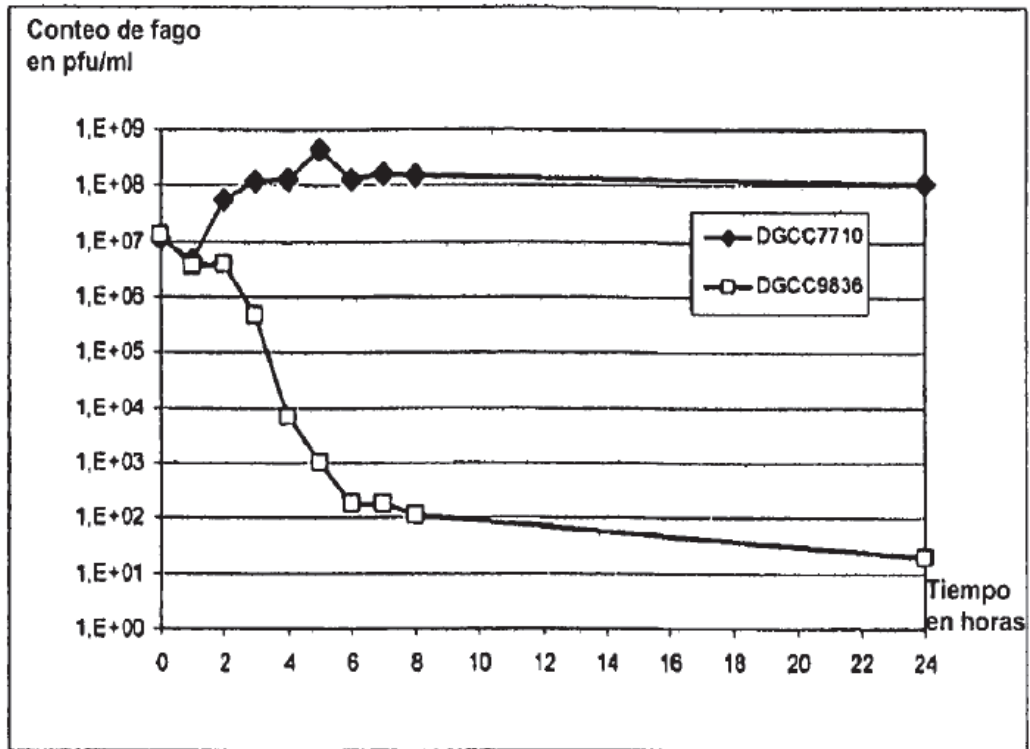


FIGURA 21

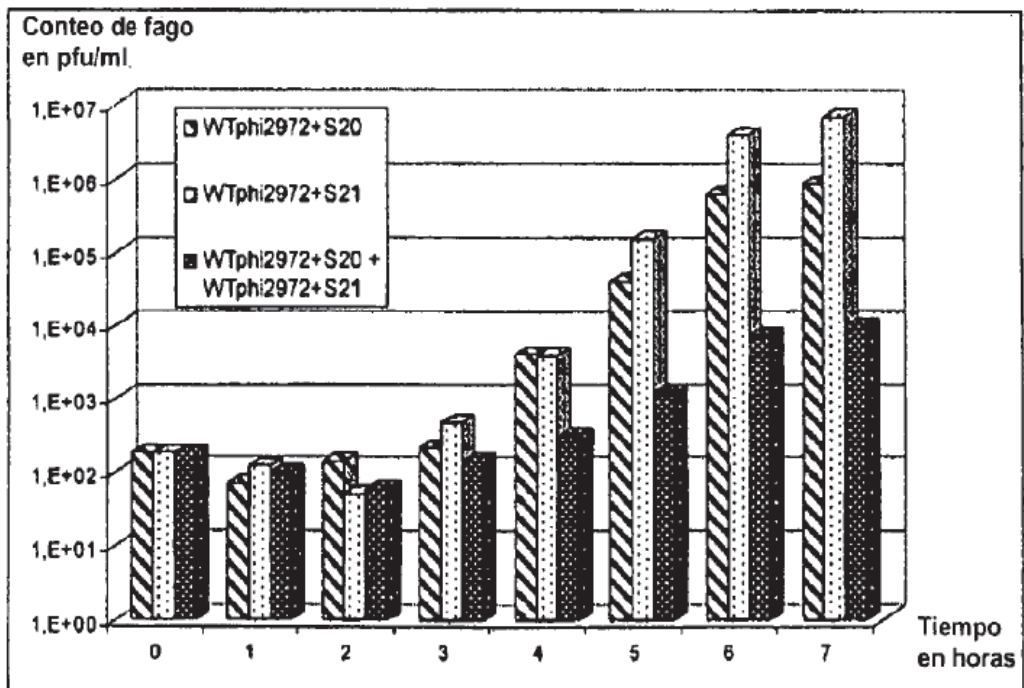


FIGURA 22

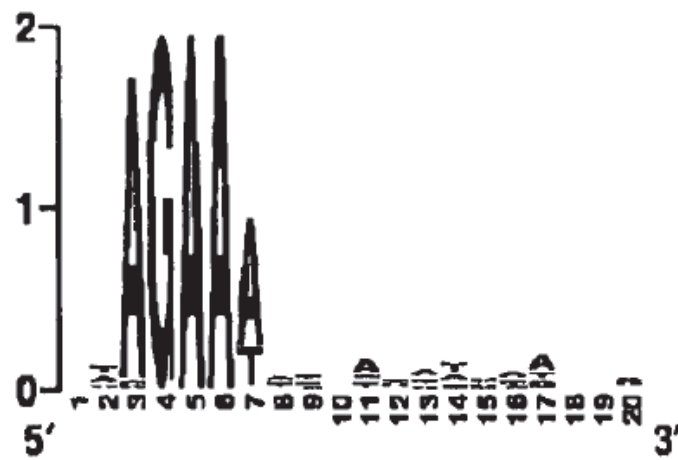


FIGURA 23

(S42) DGCC7710	^{+S40} _{Q2972}	^{+S41S42} _{Q3821}	TGCTCGACTTGTAAAAAACTACTGAAGA	TGGCG
(S41) DGCC7710	^{+S40} _{Q2972}	^{+S41S42} _{Q3821}	TAGAGGTAATGACGGCTTACCGGGTAAAGA	CGGGG
(S43) DGCC7710	^{+S1S2aCRISPR1} _{Q658}	^{+S43} _{Q658}	TACGCCAGAACAACTAGCGAAGAACATAGT	AGGAG
(S78) LMD-9	^{+S78} _{Q4241}		TGCAATTTCCATTAGTTCTTGACGCCCTT	AGGGG

