

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 738 589**

51 Int. Cl.:

<b>C12N 15/52</b>	(2006.01)
<b>C40B 30/02</b>	(2006.01)
<b>G06F 19/16</b>	(2011.01)
<b>C12P 19/18</b>	(2006.01)
<b>C12N 9/10</b>	(2006.01)
<b>C12N 15/72</b>	(2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **24.07.2013 PCT/US2013/051777**

87 Fecha y número de publicación internacional: **30.01.2014 WO14018596**

96 Fecha de presentación y número de la solicitud europea: **24.07.2013 E 13823740 (9)**

97 Fecha y número de publicación de la concesión europea: **15.05.2019 EP 2877574**

54 Título: **Alfa (1,2) fucosiltransferasas adecuadas para uso en la producción de oligosacáridos fucosilados**

30 Prioridad:

**25.07.2012 US 201213557655**

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

**23.01.2020**

73 Titular/es:

**GLYCOSYN LLC (100.0%)  
890 Winter Street Suite 208  
Waltham, MA 02451, US**

72 Inventor/es:

**HEIDTMAN, MATTHEW, IAN;  
MERIGHI, MASSIMO y  
MCCOY, JOHN, M.**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 738 589 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

Alfa (1,2) fucosiltransferasas adecuadas para uso en la producción de oligosacáridos fucosilados

## 5 Campo de la invención

La invención proporciona el uso de un constructo de ácido nucleico que comprende un ácido nucleico aislado que codifica una enzima  $\alpha(1,2)$ fucosiltransferasa que utiliza lactosa que tiene la secuencia de SEQ ID NO: 8 para producir un oligosacárido fucosilado en una bacteria, en particular ciertos oligosacáridos fucosilados que por lo general se encuentran en la leche humana.

## Antecedentes de la invención

La leche humana contiene un conjunto diverso y abundante de oligosacáridos neutros y ácidos. En la leche humana se han identificado más de 130 oligosacáridos complejos diferentes, y su diversidad y abundancia estructural es exclusiva de los seres humanos. Aunque estas moléculas pueden no ser utilizadas directamente por los bebés para la nutrición, sin embargo, tienen funciones críticas en el establecimiento de un microbioma intestinal saludable, en la prevención de enfermedades y en la función inmunológica. Antes de la invención que se describe en el presente documento, la capacidad de producir oligosacáridos de la leche humana (HMOS) era económicamente problemática. Por ejemplo, su producción a través de síntesis química se vio limitada por problemas de estereoespecificidad, disponibilidad de precursores, impurezas del producto y alto coste general. Como tal, existe una necesidad apremiante de nuevas estrategias para fabricar grandes cantidades de HMOS a bajo coste.

## Sumario de la invención

La invención presenta el uso de un constructo de ácido nucleico que comprende un ácido nucleico aislado que codifica una enzima  $\alpha(1,2)$ fucosiltransferasa que utiliza lactosa, estando dicho ácido nucleico unido de forma operativa a una o más secuencias de control heterólogas que dirigen la producción de la enzima en una cepa de producción de bacterias hospedadoras, en el que la secuencia de aminoácidos de dicha enzima codificada por dicho ácido nucleico comprende la secuencia de SEQ ID NO: 8 para producir un oligosacárido fucosilado en una bacteria. Por lo tanto, la invención presenta un uso o método eficaz y económico para producir oligosacáridos fucosilados. Una producción de ese tipo de un oligosacárido fucosilado se consigue usando un ácido nucleico aislado que comprende una secuencia que codifica un producto génico de fucosiltransferasa  $\alpha(1,2)$  que utiliza lactosa (por ejemplo, polipéptido o proteína), que se une de forma operativa a una o más secuencias de control heterólogas que dirigen la producción del producto génico de la fucosiltransferasa recombinante en una bacteria tal como *Escherichia coli* (*E. coli*). En un ejemplo, la bacteria es una bacteria entérica. La secuencia de aminoácidos del producto génico de la  $\alpha(1,2)$  fucosiltransferasa que acepta lactosa es la SEQ ID NO: 8.

Se desvela un constructo de ácido nucleico que comprende un ácido nucleico aislado que codifica una enzima  $\alpha(1,2)$  fucosiltransferasa que acepta lactosa, estando dicho ácido nucleico unido de forma operativa a una o más secuencias de control heterólogas que dirigen la producción de la enzima en una cepa de producción de bacterias hospedadoras, en el que la secuencia de aminoácidos del producto génico (enzima) codificada por el ácido nucleico comprende la identidad de aproximadamente un 70 % con la SEQ ID NO: 2. Por ejemplo, el constructo comprende la SEQ ID NO: 7, que codifica una proteína FutL. Por "heterólogo" se hace referencia a que la secuencia de control y la secuencia que codifica la proteína se originan a partir de diferentes cepas bacterianas. Una cepa bacteriana hospedadora de producción adecuada es una que no es la misma cepa bacteriana que la cepa bacteriana fuente a partir de la cual se identificó la secuencia de ácidos nucleicos que codifica fucosiltransferasa.

Un uso o método para producir un oligosacárido fucosilado, por ejemplo, un HMOS, en una bacteria se lleva a cabo proporcionando una bacteria tal como una cepa hospedadora de producción, *Escherichia coli* (*E. coli*), que se caracteriza por un nivel reducido de actividad de  $\beta$ -galactosidasa, una ruta de síntesis del ácido colónico defectuosa, una mutación en una proteasa intracelular dependiente de ATP, una mutación en un gen *lacA* y un gen de  $\alpha(1,2)$  fucosiltransferasa exógeno. Preferentemente, una mutación en un gen *thyA* en la bacteria hospedadora permite el mantenimiento de plásmidos que portan *thyA* como un gen marcador seleccionable. Los marcadores seleccionables alternativos a modo de ejemplo incluyen genes de resistencia a antibióticos tal como BLA (beta-lactamasa), o genes *proBA* (para complementar una auxotropía de prolina de cepa hospedadora de *proAB*) o *purC* (para complementar una auxotrofia de adenina de cepa hospedadora *purC*). La bacteria que comprende estas características se cultiva en presencia de lactosa, y un oligosacárido fucosilado se recupera de la bacteria o de un sobrenadante de cultivo de la bacteria. En algunos casos, el uso o método comprende además cultivar la bacteria en presencia de triptófano y en ausencia de timidina. En realizaciones preferentes, la cepa hospedadora de producción comprende K12 de *E. coli*. Otros organismos hospedadores de producción se enumeran a continuación.

Se desvela un oligosacárido fucosilado purificado producido por los métodos que se describen en el presente documento. El oligosacárido fucosilado se purifica para su uso en productos terapéuticos o nutricionales, o la bacteria se usa directamente en dichos productos. El oligosacárido fucosilado producido por la bacteria de modificada por ingeniería es 2'-fucosillactosa (2'-FL) o lactodifucotetraosa (LDFT). Las nuevas alfa 1,2-

fucosiltransferasas también son útiles para sintetizar HMOS de restos de fucosa alfa 1,2 de mayor peso molecular, por ejemplo, lacto-N-fucopentaosa (LNF I) y lacto-N-difucohexaosa (LDFH I).

5 La bacteria usada para producir los oligosacáridos se modificó mediante ingeniería genética para que comprendiera un aumento de la mezcla intracelular de difosfato de guanosina (GDP)-fucosa (en comparación con el tipo silvestre), un aumento en el grupo de lactosa intracelular (en comparación con el tipo natural) y para que comprendiera la actividad de fucosiltransferasa. Por consiguiente, un gen *lacZ* endógeno y un gen *lacI* endógeno de *E. coli* se eliminan o se inactivan funcionalmente para reducir el nivel de actividad de  $\beta$ -galactosidasa. La bacteria también puede comprender una mutación en el gen *lacA*. La bacteria aislada de *E. coli* también comprende un promotor del gen *lacIq* inmediatamente cadena arriba de un gen *lacY*. En algunos casos, la bacteria *E. coli* aislada comprende una ruta de síntesis de ácido colónico defectuosa debido a que un gen *wcaJ* endógeno de *E. coli* se elimina o se desactiva funcionalmente. La bacteria comprende una mutación en una proteasa intracelular dependiente de adenosina-5'-trifosfato (ATP). Por ejemplo, la bacteria comprende una mutación nula en un gen *lon*. La bacteria también comprende una mutación en un gen *thyA*. Preferentemente, la bacteria acumula un aumento de la acumulación de lactosa intracelular y un aumento de la acumulación intracelular de GDP-fucosa. En un aspecto, la bacteria *E. coli* comprende el genotipo  $\Delta ampC::P_{trp}^B cl$ ,  $\Delta(lacI-lacZ)::FRT$ ,  $P_{lacI} lacY^*$ ,  $\Delta wcaJ::FRT$ , *thyA::n10*,  $\Delta lon::(npt3, lacZ^*)$ ,  $\Delta lacA$ .

20 La bacteria posee una actividad (1,2) fucosil transferasa. Por ejemplo, la bacteria comprende un gen exógeno de  $\alpha$  (1,2) fucosiltransferasa. La actividad de  $\alpha$  (1,2) fucosiltransferasa codificada por el constructo de ácido nucleico usado de acuerdo con la invención tiene la secuencia de SEQ ID NO: 8. Se desvela una bacteria con un gen exógeno de  $\alpha$  (1,2) fucosiltransferasa que comprende al menos un 10 % de homología/identidad y menos de un 40 % al nivel de aminoácidos con alfa-(1,2) fucosiltransferasa (*futC*) de 26695 de *Helicobacter pylori*, por ejemplo, una identidad de al menos un 15 %, al menos un 20 %, al menos un 25 %, al menos un 30 %. Las secuencias pueden ser al menos un 50 %, al menos un 55 %, al menos un 60 %, al menos un 65 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 95 % de homología/identidad con alfa (1,2) fucosiltransferasa (*futC*) de 26695 de *Helicobacter pylori*. Por lo tanto, FutL es idéntico en un 70 % a FutC a nivel de aminoácidos.

30 El término "% de identidad", en el contexto de dos o más secuencias de ácidos nucleicos o polipéptidos, se refiere a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje específico de restos de aminoácido o nucleótido que son iguales, cuando se comparan y se alinean para una correspondencia máxima, de acuerdo con lo medido usando uno de los siguientes algoritmos de comparación de secuencias o mediante inspección visual. Por ejemplo, el % de identidad es relativo a la longitud total de las regiones de codificación de las secuencias que se comparan.

35 Para la comparación de secuencias, normalmente una secuencia actúa como una secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de ensayo y referencia se introducen un ordenador, las coordenadas de la subsecuencia se designan, si es necesario, y se designan los parámetros del programa de algoritmos de secuencia. El algoritmo de comparación de secuencias a continuación calcula el porcentaje de identidad de secuencia para la secuencia de ensayo en relación con la secuencia de referencia, en función de los parámetros de programa designados.

45 El porcentaje de identidad se determina usando algoritmos de búsqueda tales como BLAST y PSI-BLAST (Altschul *et al.*, 1990, J Mol Biol 215: 3, 403-410; Altschul *et al.*, 1997, Nucleic Acids Res 25: 17, 3389-402). Para la búsqueda de PSI-BLAST, se emplean los siguientes parámetros a modo de ejemplo: (1) El umbral esperado fue 10; (2) el coste del hueco fue Existencia: 11 y Extensión: 1; (3) La Matriz empleada fue BLOSUM62; (4) El filtro para regiones de baja complejidad estaba "activado". La bacteria expresa un producto génico de fucosiltransferasa codificado por una secuencia que no es idéntica a *futC*.

50 Los genes de  $\alpha$  (1,2) fucosiltransferasa como se desvela en el presente documento incluye *Escherichia coli* 0126 *wbgL*, *Helicobacter mustelae* 12198 (ATCC 43772) alfa-1,2-fucosiltransferasa (*futL*), y proteína de la familia de glicosil transferasa de *Bacteroides vulgatus* ATCC 8482 (*futN*). Un gen exógeno de  $\alpha$  (1,2) fucosiltransferasa se selecciona entre el grupo que consiste en *Escherichia coli* 0126 *wbgL*, *Helicobacter mustelae* 12198 (ATCC 43772) alfa-1,2-fucosiltransferasa (*futL*), *Bacteroides fragilis* (NCTC) 9343 fucosil transferasa (*bft3/wcfB*), *Escherichia coli* O55:H7 (cepa CB9615) fucosiltransferasa (*wbgN*), *Helicobacter bilis* ATCC 437879 *futD*, *Vibrio cholera* O22 *wbIA*, *Bacteroides fragilis* (NCTC) 9343 alfa-1,2-fucosiltransferasa (*bft1*), *Bacteroides ovatus* ATCC 8483 *futO*, y *Helicobacter cinaedi* CCUG 18818 alfa-1,2-fucosiltransferasa (*futE*). La presente invención usa proteína de la familia de glicosil transferasa de *Bacteroides vulgatus* ATCC 8482 (*futN*) para producir un oligosacárido fucosilado en una bacteria.

65 Se desvela un vector, por ejemplo, un vector que contiene ácido nucleico. El vector puede incluir además uno o más elementos reguladores, por ejemplo, un promotor heterólogo. Los elementos reguladores se pueden unir de forma operativa a un gen que codifica una proteína, un constructo genético que codifica un gen de proteína de fusión, o una serie de genes unidos en un operón con el fin de expresar la proteína de fusión. También se desvela una célula recombinante aislada, por ejemplo, una célula bacteriana que contiene una molécula o vector de ácido nucleico

mencionados anteriormente. El ácido nucleico se integra opcionalmente en el genoma.

El constructo de ácido nucleico puede comprender al menos un promotor de bacteriófago λ, un gen *rcaA* de *E. coli*, un gen *bla*, y un gen *thyA* nativo.

5 Esto se muestra a modo de ejemplo para el constructo plasmídico pG171, no comprendido en la presente invención. El mapa plasmídico de pG171 se muestra en la Figura 5.

La secuencia de pG171 se presenta a continuación con anotaciones de GenBank con respecto a características específicas (SEQ ID NO: 1):

```

LOCUS          pEC2-futC-MYC-rcaA-thyA_(pG171)          6244 pb          ADN
SYN circular  24-MAYO-1995
DEFINICIÓN    Vector de clonación de fusión pTRXFUS,  secuencia completa.
ACCESO       U16857
VERSIÓN      U16857.1  GI: 57 54 47
PALABRAS CLAVE  vector de fusión genética de tiorredoxina.
FUENTE       Vector de clonación pTRXFUS (desconocido)
              ORGANISMO  Vector de clonación pTRXFUS
              otras secuencias; secuencias artificiales; vectores.
REFERENCE    1 (bases 1 a 3585)
              AUTORES   LaVallie ,E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F.
              y McCoy, J.M.
              TÍTULO    Un sistema de expresión de fusión genética de tiorredoxina que elude
              la formación de cuerpos de inclusión en el citoplasma de E. coli
REVISTA      Biotechnology (N.Y.) 11 (2), 187-193 (1993)
PUBMED      7763371
REFERENCIA  2 (bases 1 a 3585)
              AUTORES   LaVallie, E.R.
              TÍTULO    Direct Submission
              REVISTA   Presentado (03-NOV-1994) Edward R. LaVallie, Genetics Institute,
77
              CambridgePark Drive, Cambridge, MA 02140, USA
CARACTERÍSTICAS  Localización/Calificadores
              Cebador    163..183
              /etiqueta=lacZR6
              CDS      243..481
              /gen="lgt"
              /nota="ECK2824:JW2796:b2828"
              /codón_inicio=1
              /trad_tabla=11
              /producto="fosfatidilglicerol-
              prolipoproteínadiacilglicerilo
    
```

## ES 2 738 589 T3

```

transferasa"
/ proteína_id="BAE76897.1"
/db_xref="GI:85675644"
/ traducción="MTSSYLHFPEFDPVIFSIGPVALHWYGLMYLVGFIFAMWLATRRA
NRPGSGWTKNEVENLLYAGFLGVFLGGRIGYVLFYNFPQFMADPLYLFRVWDSFHG
GLIGVIVVMIIFARRTKRSFFQVSDFIAPLIPFGLGAGRLGNFINGELWGRVDPNFPFA
MLFPGSRTEIDILLQTNPQWQSIFDTYGVLP RHPSQLYELLLEGVVLFIILNLYIRKPR
PMGAVSGLFLIGYGAFRIIVEFFRQPDAQFTGAWVQYISMGQILSIPMIVAGVIMMVA
YRRSPQQHVS"

```

```

Fuente      complemento (243..1365)
            /organismo="Escherichia coli W3110"
            /mol_tipo="ADN genómico"
            /cepa="K-12"
            /sub_cepa="W3110"
            /db_xref="taxón:316407"
            /nota="sinónimo: Escherichia coli cepa K12 subcepa
W3110"

```

```

Fuente      complemento (242^243)
            /organismo="Escherichia coli W3110"
            /mol_tipo="ADN genómico"
            /cepa="K-12"
            /sub_cepa="W3110"
            /db_xref="taxón:316407"
            /nota="sinónimo: Escherichia coli cepa K12 subcepa
W3110"

```

```

Cebador      243..266
            /nota=cagtcagtcaggcgccTCCTCAACCTGTATATTCGTAAAC
            /etiqueta =THYA-F
Promotor     359..364
            /etiqueta ="thyA -35"
Promotor     380..385
            /etiqueta ="thyA WEAK -10"
Unión_sitio  479..484
            /etiqueta ="thyA RBS"
Gen          488..1282
            /gen ="thyA"
CDS          488..1282
            /gen ="thyA"
            /nota="ECK2823:JW2795:b2827"
            /codón_inicio=1
            /trad_tabla=11

```

```

        / producto= "timidilato sintetasa"
        / proteína_id="BAE76896.1"
        / db_xref="GI:85675643"
        / traducción ="MKQYLELMQKVLDEGTQKNDRTGTGTLSIFGHQMRFNLDGPFPLV
TTKRCHLRSIIHELLWFLQGDTNIAYLHENNVTIWDEWADENGDLGPVYGKQWRTP
DGRHIDQITTVLNQLKNPDSRRIIVSAWNVGELDKMALAPCHAFFQFYVADGKLSL
YQRSCDVFLGLPFNIASYALLVHMAQQCDLEVGDFVWTGGDTHLYSNHMDQLSR
EPRPLPKLI IKRKPESIFDYRFEDFEIEGYDPHPGIKAPVAI"
    Horquilla_bucle      1304..1310
                          /etiqueta=Terminador
    Horquilla_bucle      1317..1323
                          /etiqueta=Terminador
    Cebador               complemento (1345..1365)
                          / nota=cagtcagtcaggcgccTCGGGAAGGCGTCTCGAAGA
                          /etiqueta=THYA-R
    Cebador               complemento (1468..1489)
                          /etiqueta=lacZF5
    Cebador               1508..1524
                          /etiqueta=aspAseq
    Gen                   1536..1588
                          /gen="dsrB"
    Cebador               1536..1558
                          / nota=cagtcagtcaaagcttTCTTTAATGAAGCAGGGCATCAG
                          /etiqueta=rcsA-R
    Horquilla_bucle      complemento (1600..1610)
                          /etiqueta= Horquilla
    Horquilla_bucle      complemento (1615..1625)
                          /etiqueta= Horquilla
    CDS                   complemento (1632..2255)
                          /gen="rcsA"
                          / nota="ECK1949:JW1935:b1951"
                          /codón_inicio=1
                          / trad_tabla=11
                          / producto= (co-regulador transcriptional de unión a ADN
                          con RcsB"
                          /proteína_id="BAA15776.1"
                          /db_xref="GI:1736617"
        / traducción ="MSTIIMDLCSYTRLGLTGYLLSRGVKKREINDIETVDDLAIACDS
QRPSVVFINECDFIHDASNSQRIKLIINQHPNTLFIVFMAIANVHFDEYLLVRKNLLIS
SKSIKPESLDDILGDILKKETTITSFLNMPTLSLSRTESSMLRMWMAGOGTIQISDQMN
IKAKTVSSHKGNIKRKIKTHNKQVIYHVVRRLTDNVTNGIFVNMR"

```



## ES 2 738 589 T3

	<code>/etiqueta="cepa 26695 de futC (fijada)"</code>
Fuente	<code>2582..3517</code>
	<code>/organismo="Helicobacter pylori 26695"</code>
	<code>/mol_tipo="ADN genómico"</code>
	<code>/cepa="26695"</code>
	<code>/db_xref="taxón:85962"</code>
MYC-etiqueta	<code>complemento (2585..2617)</code>
Cebador	<code>2968..2990</code>
	<code>/nota=GATAGTCAATACCAAGCTGACAG</code>
	<code>/etiqueta=3-6-R</code>
Cebador	<code>2968..2990</code>
	<code>/etiqueta="42 (3-6-R)"</code>
Gen	<code>complemento(3052..3517)</code>
	<code>/locus_etiqueta="HP0094"</code>
	<code>/db_xref="GenID:899021"</code>
Cebador	<code>complemento (3495..3517)</code>
	<code>/nota=GAAttcaagaaggagatataCATATGGCTTTTAAGGTGGTGCAAAT</code>
	<code>/etiqueta=pLfutC-F</code>
Fuente	<code>3517^3518</code>
	<code>/organismo="Escherichia coli W3110"</code>
	<code>/mol_tipo="ADN genómico"</code>
	<code>/cepa="K-12"</code>
	<code>/sub_cepa="W3110"</code>
	<code>/db_xref="taxón:316407"</code>
	<code>/nota="sinónimo: Escherichia coli cepa K12 cepa</code>
W3110"	
RBS	<code>complemento (3521..3535)</code>
	<code>/etiqueta="T7 gen 10 RBS"</code>
Fuente	<code>complemento (3541..3715)</code>
	<code>/nota="se origina a partir de LAMCG"</code>
	<code>/etiqueta="lambda ADN"</code>
Cebador	<code>3541..3567</code>
	<code>/nota=cagtcagtcagaattcTAACAATTGATTGAATGTATGCAAATA</code>
	<code>/etiqueta=pLnut-R</code>
Región	<code>complemento(3544)</code>
	<code>/etiqueta="sitio de terminación TR1"</code>
nota	<code>3544</code>
	<code>/nota="el plásmido pEC2-BfT2-MYC secuenciado original es</code>
	<code>una mezcla de C y A aquí. Este plásmido es A"</code>
Misc. _feature	<code>complemento (3547..3553)</code>
	<code>/etiqueta="caja consenso de rho-dep de TR1"</code>

## ES 2 738 589 T3

Región	complemento (3566..3571) /etiqueta="sitio de terminación de TR1"
Misc._feature	complemento(3580..3599) /nota="sitio B de utilización de rho (rutB)"
Misc._unión	complemento (3600..3616) /nota="utilización de N hacia la derecha; supuesto" /resto_unido="Nutr"
Región	complemento (3601..3632) /etiqueta=nutR
Misc._feature	complemento (3615..3632) /nota=" sitio A de utilización de rho (rutA)"
Variación	complemento (3636..3715) /nota="región imm434"
CDS	complemento (3640..3715) /codón_inicio=1 /trad_tabla=11 /producto="cro (antirrepresor; también tof; 66)" /proteína_id="AAA96582.1" /db_xref="GI:215148"
/traducción="MEQRITLKDYAMRFGQTKTAKDLGVYQSAINKAIHAGRKIFLTIN ADGSVYAEVVKPFPSNKKTTA"	
Región	complemento (3729..3730) /etiqueta="sitio 1/2 HaellI que usó Rosenberg"
CDS	complemento (3731..3907) /codón_inicio=1 /trad_tabla=11 /producto="N (regulador de gen temprano; 133)" /proteína_id="AAA96578.1" /db_xref="GI:508997"
/traducción="MCQSRGVFVQDYNCHTPPKLTDRRIQMDAQTRRRERRAEKQAQWK AANPLLVGVS AKPVNRPILSLNRKPKSRVESALNPIDLTVLAEYHKQIESNLQRIERKNQRTWYSKPGERGITCS GRQKIKGKSIPLI"	
Fuente	complemento (3731..4023) /nota="se origina a partir de LAMCG rev" /etiqueta="lambda ADN"
Región	complemento (3731..3733) /etiqueta=" sitio 1/2 HpaI que usó Rosenberg"
Variación	3731..4184 /nota="región imm21"
Misc._unión	complemento (3987..4003)

## ES 2 738 589 T3

	/nota="utilización de N hacia la izquierda; supuesto"
	/unido_resto="NutI"
Región	complemento (3 988 ..4018) /etiqueta=nutL
Fuente	complemento (4 024 ..4184) /nota="se origina a partir de LAMCG rev" /etiqueta="lambda ADN"
Cebador	4 024 .. 4044 /nota=cagtcagtcagaaTTCATGGTGGTCAGTGCGTCC /etiqueta=PLvect3
Región	complemento (4 024 ..4369) /nota="originates from EC1 pL region" /etiqueta="región pL de ECI"
ARNm	complemento (4051) /nota="ARNm-pl (alt.; mediante terminador t2a)" /etiqueta="inicio de ARNm de pLI"
Variación	4 052 .. 4184 /nota="región imm434"
Promotor	complemento(4 058 ..4 063) /etiqueta="pLI -10"
operator	complemento (4 060 ..4076) /nota="operador-11 (primera base en la hebra comp)" /etiqueta=OL1
Cebador	complemento (4 075 ..4092) /etiqueta=pLseq
Promotor	complemento (4081 ..4086) /etiqueta="pL1 -35"
operator	complemento (4 084 ..4100) /nota="operador-12 (primera base en la hebra comp)" /etiqueta=OL2
ARNm	complemento (4 093) /etiqueta="inicio de ARNm de pL2"
Promotor	complemento (4100..4105) /etiqueta="pL2 -10"
operator	complemento (4104..4120) /nota="operador-13 (primera base en la hebra comp)" /etiqueta=OL3
Promotor	complemento (4122..4127) /etiqueta="pL2 -35"
Fuente	complemento (4185..6244) /etiqueta="ADN de pUC18"

## ES 2 738 589 T3

Cebador	complemento (4 350..4 369) /nota=cagtcagtcaACATGTTCTTTCCTGCGTTA /etiqueta=pLnut-F
Cebador	complemento (4 414..4 4 30) /etiqueta=pLnutseq-F
Replicación_ori	complemento (4 4 25..5013) /etiqueta="origen de replicación"
ARN_transcrito	complemento (4 4 25..4 9 77) /etiqueta=ARNII
Promotor	4 8 32 .. 4 8 36 /etiqueta="ARNI -35"
Promotor	4 8 53 .. 4 8 58 /etiqueta="ARNI -10"
ARN_transcrito	4 8 6 7 .. 4 9 7 4 /etiqueta=RNAI
Promotor	complemento (4988..4992) /etiqueta="ARNII -10"
Promotor	complemento (5 008..5013) /etiqueta="ARNII -35"
CDS	complemento (5184..6 0 44) /etiqueta=beta-lactamasa
Restricción_sit	5 9 1 6 .. 5 9 2 7 /etiqueta="sitio EcoK"
Señal_péptido	complemento (5 9 7 6..6044) /etiqueta=beta-lactamasa
Promotor	complemento (6088..6093) /etiqueta="beta-lactamasa -10"
Promotor	complemento (6109..6114) /etiqueta="beta-lactamasa -35"

### ORIGEN

```

1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
61 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
121 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
181 ACCATATATG CGGTGTGAAA TACCGCACAG ATGCGTAAGG AGAAAATACC GCATCAGGCG
241 CCTCCTCAAC CTGTATATTC GTAAACCACG CCCAATGGGA GCTGTCTCAG GTTTGTTCCT
301 GATTGGTTAC GGCGCGTTTC GCATCATTGT TGAGTTTTTC CGCCAGCCCG ACGCGCAGTT
361 TACCGGTGCC TGGGTGCAGT ACATCAGCAT GGGGCAAATT CTTTCCATCC CGATGATTGT
421 CGCGGGTGTG ATCATGATGG TCTGGGCATA TCGTCGCAGC CCACAGCAAC ACGTTTCCTG
481 AGGAACCATG AAACAGTATT TAGAACTGAT GCAAAAAGTG CTCGACGAAG GCACACAGAA
541 AAACGACCGT ACCGGAACCG GAACGCTTTC CATTTTTGGT CATCAGATGC GTTTTAACCT
601 GCAAGATGGA TTCCCCTGG TGACAATAA ACGTTGCCAC CTGCGTTCCA TCATCCATGA

```

661 ACTGCTGTGG TTTCTGCAGG GCGACACTAA CATTGCTTAT CTACACGAAA ACAATGTCAC  
721 CATCTGGGAC GAATGGGCCG ATGAAAACGG CGACCTCGGG CCAGTGTATG GTAAACAGTG  
781 GCGCGCCTGG CCAACGCCAG ATGGTCGTCA TATTGACCAG ATCACTACGG TACTGAACCA  
841 GCTGAAAAAC GACCCGGATT CGCGCCGCAT TATTGTTTCA GCGTGGAACG TAGGCGAACT  
901 GGATAAAATG GCGCTGGCAC CGTGCCATGC ATTCTTCCAG TTCTATGTGG CAGACGGCAA  
961 ACTCTCTTGC CAGCTTTATC AGCGCTCCTG TGACGTCTTC CTCGGCCTGC CGTTCAACAT  
1021 TGCCAGCTAC GCGTTATTGG TGCATATGAT GGCAGCAGCAG TGCGATCTGG AAGTGGGTGA  
1081 TTTTGTCTGG ACCGGTGGCG ACACGCATCT GTACAGCAAC CATATGGATC AAACCTCATCT  
1141 GCAATTAAGC CGCGAACCGC GTCCGCTGCC GAAGTTGATT ATCAAACGTA AACCCGAATC  
1201 CATCTTCGAC TACCGTTTCG AAGACTTTGA GATTGAAGGC TACGATCCGC ATCCGGGCAT  
1261 TAAAGCGCCG GTGGCTATCT AATTACGAAA CATCCTGCCA GAGCCGACGC CAGTGTGCGT  
1321 CGGTTTTTTT ACCCTCCGTT AAATTCCTCG AGACGCCTTC CCGAAGGCCG CATTCCGCAT  
1381 TCAGGCTGCG CAACTGTTGG GAAGGGCGAT CGGTGCGGGC CTCTTCGCTA TTACGCCAGC  
1441 TGCGGAAAGG GGGATGTGCT GCAAGGCGAT TAAGTTGGGT AACGCCAGGG TTTTCCCAGT  
1501 CACGACGTTG TAAAACGACG GCCAGTGCCA AGCTTTCTTT AATGAAGCAG GGCATCAGGA  
1561 CGGTATCTTT GTGGAGAAAG CAGAGTAATC TTATTCAGCC TGACTGGTGG GAAACCACCA  
1621 GTCAGAATGT GTTAGCGCAT GTTGACAAAA ATACCATTAG TCACATTATC CGTCAGTCGG  
1681 ACGACATGGT AGATAACCTG TTTATTATGC GTTTTGATCT TACGTTTAAAT ATTACCTTTA  
1741 TGCGATGAAA CGGTCTTGCC TTTGATATTC ATTTGGTCAG AGATTTGAAT GGTTCCCTGA  
1801 CCTGCCATCC ACATTCGCAA CATACTCGAT TCGGTTCCGC TCAATGATAA CGTCGGCATA  
1861 TTTAAAAACG AGGTTATCGT TGCTCTTTTT TTCAGAATAT CGCCAAGGAT ATCGTCGAGA  
1921 GATTCCGGTT TAATCGATTT AGAACTGATC AATAAATTTT TTCTGACCAA TAGATATTCA  
1981 TCAAAATGAA CATTGGCAAT TGCCATAAAA ACGATAAATA ACGTATTGGG ATGTTGATTA  
2041 ATGATGAGCT TGATACGCTG ACTGTTAGAA GCATCGTGGA TGAAACAGTC CTCATTAATA  
2101 AACACCACTG AAGGGCGCTG TGAATCACAA GCTATGGCAA GGTCATCAAC GTTTTCAATG  
2161 TCGTTGATTT CTCTTTTTTT AACCCCTCTA CTCAACAGAT ACCCGGTTAA ACCTAGTCGG  
2221 GTGTAACTAC ATAAATCCAT AATAATCGTT GACATGGCAT ACCCTCACTC AATGCGTAAC  
2281 GATAATTCCC CTTACCTGAA TATTTATCA TGAATAACG GAACAACATG GGTCACCTAA  
2341 TGCGCCACTC TCGCGATTTT TCAGGCGGAC TTACTATCCC GTAAAGTGTT GTATAATTTG  
2401 CCTGGAATTG TCTTAAAGTA AAGTAAATGT TGCGATATGT GAGTGAGCTT AAAACAAATA  
2461 TTTGCTGCA GGAGTATCCT GGAAGATGTT CGTAGaagct tACTGCTCAC AAGAAAAAAG  
2521 GCACGTCATC TGACGTGCCT TTTTATTTG TACTACCCTG TACGATTACT GCAGCTCGAG  
2581 TTTAattcaa atcttcttca gaaatcaatt tttgttcAGC GTTATACTTT TGGGATTTTA  
2641 CCTCAAAATG GGATTCATTT TTCACCCACT CCTTACAAAG GATATTCTCA TGCCCAAAAA  
2701 GCCAGTGTTC GGGGCCAATA ATGATTTTTT CTGGATTTTC TATCAAATAG GCCGCCACC  
2761 AGCTATAAGT GCTATTAGCG ATAATGCCAT GCTGACAAGA TTGCATGAGC AGCATGTCCC  
2821 AATACGCCTC TTCTTCTTTA TCCCTAGTGG TCATGTCCAT AAAAGGGTAG CCAAGATCAA  
2881 GATTTTGCCT GAATTCCTAAG TCTTCGAAA ACACAAAAAG CTCCATGTTT GGCACGCGCT  
2941 TTGCCATATA CTCAAGCGCC TTTTTTTGAT AGTCAATACC AAGCTGACAG CCAATCCCCA  
3001 CATAATCCCC TCTTCTTATA TGCACAAACA CGCTGTTTTT AGCGGCTAAA ATCAAAGAAA  
3061 GCTTGCCTG ATATTCCTCC TCTTTTTTAT TATTATTCTT ATTATTTTCG GGTGGtGGtG

3121 GTAGAGTGAA GGTTTGCTTG ATTAAAGGGG ATATAGCATC AAAGTATCGT GGATCTTGGA  
3181 AATAGCCAAA AAAATAAGTC AAGCGGCTTG GCTTTAGCAA TTTAGGCTCG TATTCAAAAA  
3241 CGATTTCTTG ACTCACCTA TCAAATCCCA TGCATTTGAG CGCGTCTCTT ACTAGCTTGG  
3301 GGAGGTGTTG CATTTTAGCT ATAGCGATTT CTTTCGCGCT CGCATAGGGC AAATCAATAG  
3361 GGAAAAGTTC TAATTGCATT TTCCTATCGC TCCAATCAAA AGAAGTGATA TCTAACAGCA  
3421 CAGGCGTATT AGAGTGTTTT TGCAAACTTT TAGCGAAAAGC GTATTGAAAC ATTTGATTCC  
3481 CAAGCCCTCC GCAAATTTGC ACCACCTTAA AAGCCATATG tataatctect tcttgaaTTC  
3541 TAAaAATTGA TTGAATGTAT GCAAATAAAT GCATACACCA TAGGTGTGGT TTAATTTGAT  
3601 GCCCTTTTTT AGGGCTGGAA TGTGTAAGAG CGGGGTATT TATGCTGTTG TTTTTTGT  
3661 ACTCGGGAAG GGCTTTACCT CTTCCGCATA AACGCTTCCA TCAGCGTTTA TAGTTAAAAA  
3721 AATCTTTCGG AACTGGTTTT GCGCTTACCC CAACCAACAG GGGATTTGCT GCTTTCATT  
3781 GAGCTGTTT CTCTGCGCGA CGTTCGCGGC GCGGTGTTT TGCATCCATC TGGATTCTCC  
3841 TGTCAGTTAG CTTTGGTGGT GTGTGGCAGT TGTTAGTCTG AACGAAAACC CCCC GCGATT  
3901 GGCACATTGG CAGCTAATCC GGAATCGCAC TTACGGCCAA TGCTTCGTTT CGTATCACAC  
3961 ACCCAAAGC CTTCTGCTTT GAATGCTGCC CTTCTCAGG GCTTAATTTT TAAGAGCGTC  
4021 ACCTTCATGG TGGTCAGTGC GTCCTGCTGA TGTGCTCAGT ATCACCGCCA GTGGTATTTA  
4081 TGTC AACACC GCCAGAGATA ATTTATCACC GCAGATGGTT ATCTGTATGT TTTTATATG  
4141 AATTTATTTT TTGCAGGGGG GCATTGTTTG GTAGGTGAGA GATCAATTCT GCATTAATGA  
4201 ATCGGCAAAC GCGCGGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCTCGCTC  
4261 ACTGACTCGC TGCCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG  
4321 GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC  
4381 CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC  
4441 CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA  
4501 CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCCTCTCC TGTTCCGACC  
4561 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTCTCAT  
4621 AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG  
4681 CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATCG TCTTGAGTCC  
4741 AACCCGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA  
4801 GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT  
4861 AGAAGGACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT  
4921 GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGAAG  
4981 CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG  
5041 TCTGACGCTC AGTGGAACGA AAATCAGCT TAAGGGATT TGGTCATGAG ATTATCAAAA  
5101 AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA  
5161 TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG  
5221 ATCTGTCTAT TTCGTTTATC CATAGTTGCC TGAATCCCG TCGTGTAGAT AACTACGATA  
5281 CGGGAGGGCT TACCATCTGG CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACCG  
5341 GCTCCAGATT TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCTT  
5401 GCAACTTTAT CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGGAGCTAG AGTAAGTAGT  
5461 TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC  
5521 TCGTCGTTTGTATGTATGCTTC ATTACAGCTCC GGTTCCTAAC GATCAAGGCG AGTTACATGA

```

5581 TCCCCCATGT TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT
5641 AAGTTGGCCG CAGTGTATC ACTCATGGTT ATGGCAGCAC TGCATAATTC TCTTACTGTC
5701 ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA
5761 TAGTGTATGC GCGGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA
5821 CATAGCAGAA CTTTAAAAGT GTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA
5881 AGGATCTTAC CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT
5941 TCAGCATCTT TTACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC
6001 GCAAAAAAGG GAATAAGGGC GACACGGAAA TGTTGAATAC TCATACTCTT CCTTTTTCAA
6061 TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT
6121 TAGAAAAATA AACAAATAGG GGTTCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC
6181 TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAAATA GCGGTATCAC GAGGCCCTTT
6241 CGTC (SEQ ID NO:1; plásmido G171)

```

Se desvela un constructo de ácido nucleico que comprende un gen de  $\alpha$  (1,2) fucosiltransferasa que comprende, por ejemplo, una identidad de al menos 10 % y de menos de un 40 % a nivel del aminoácido con respecto a alfa-(1,2) fucosiltransferasa 26695 de *Helicobacter pylori* (*futC*). El gen exógeno de  $\alpha$  (1,2) fucosiltransferasa se puede seleccionar entre el grupo que consiste en alfa-(1,2) fucosiltransferasa 26695 de *Helicobacter pylori* (*futC*), *Vibrio cholera* O22 *wbIA*, *Escherichia coli* 0126 *wbgL*, *Helicobacter bilis* ATCC 437879 *futD*, *Helicobacter cinaedi* CCUG 18818 alfa-1,2-fucosiltransferasa (*futE*), *Helicobacter mustelae* 12198 (ATCC 43772) alfa-1,2-fucosiltransferasa (*futL*), *Bacteroides ovatus* ATCC 8483 *futO*, *Escherichia coli* O55:H7 (cepa CB9615) fucosiltransferasa (*wbgN*), *Bacteroides fragilis* (NCTC) 9343 alfa-1,2-fucosiltransferasa (*bft1*), y *Bacteroides fragilis* (NCTC) 9343 fucosiltransferasa (*bft3/wcfB*). De acuerdo con la presente invención, el gen de  $\alpha$  (1,2) fucosiltransferasa es la proteína de la familia de glicosil transferasa de *Bacteroides vulgatus* ATCC 8482 (*futN*).

Se desvela una bacteria de *E. coli* aislada como se ha descrito anteriormente y caracterizada como comprendiendo un nivel reducido de actividad de  $\beta$ -galactosidasa, una ruta de síntesis del ácido colónico defectuosa, una mutación en el gen *lacA*, una mutación en una proteasa intracelular dependiente de ATP, y una mutación en un gen *thyA*. También se desvelan métodos para identificar un gen de  $\alpha$  (1,2) fucosiltransferasa capaz de sintetizar 2'-fucosillactosa (2'-FL) en *E. coli*. El método de identificación de enzima  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa y no es FutC, comprende las siguientes etapas:

- 1) realizar una búsqueda informática de bases de datos de secuencias para definir un amplio grupo de homólogos de secuencia sencilla de cualquier  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa, conocida;
- 2) usar el listado de la etapa (1), obteniendo un perfil de búsqueda que contiene secuencias comunes y/o motivos estructurales compartidos por los miembros del listado;
- 3) buscar bases de datos de secuencias, usando un perfil de búsqueda derivado basándose en la secuencia o motivo estructural común de la etapa (2) como búsqueda, e identificar secuencias candidatas, en las que una homología de secuencia con respecto a una  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa es de un 40 % o inferior;
- 4) compilar un listado de organismos candidatos, siendo dichos organismos caracterizados por la expresión de  $\alpha$ (1,2)fucosil-glicanos en un estado de origen natural;
- 5) seleccionar secuencias candidatas que se obtienen a partir de organismos candidatos para generar un listado de enzimas que utilizan lactosa;
- 6) expresar la enzima que utiliza lactosa candidata en un organismo hospedador; y
- 7) someter a ensayo la actividad de  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa, en el que la detección de 2'-FL en dicho organismo indica que la secuencia candidata comprende una  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa y que no es FutC. Por ejemplo, la homología de secuencia con respecto a una  $\alpha$ (1,2)fucosiltransferasa que utiliza lactosa de referencia es de un 40 % o inferior.

Se desvela un oligosacárido fucosilado purificado producido de acuerdo con el uso o con los métodos que se han descrito anteriormente. En oligosacárido purificado (2'-FL) obtenido al final del proceso es un polvo dulce, cristalino, de color blanco/ligeramente blanquecino. A diferencia de los métodos de producción de oligosacárido que usan FutC, los métodos que utilizan ciertas enzimas que no son FutC (por ejemplo FutL) no tienen actividad de  $\alpha$  (1,3) fucosiltransferasa que conducía reacciones secundarias. La falta de actividad de  $\alpha$  (1,3) fucosiltransferasa asociada a FutL contribuye a su eficacia para producir 2'FL y es una ventaja en comparación con FutC. FutL no tiene actividad de alfa 1,3 fucosiltransferasa. Por ejemplo, una célula de modificada por ingeniería *E. coli*, sobrenadante de cultivo de *E. coli*, o lisado celular de *E. coli* comprende 2'-FL recombinante y básicamente no comprende una lactosa consignada en 1,3 antes de la purificación de 2'-FL a partir de la célula, sobrenadante de cultivo, o lisado. Parece

que tanto FutN como WbgL tienen actividad de alfa 1,3 fucosiltransferasa (similar a FutC). Sin embargo como una materia general, el oligosacárido fucosilado producido con los métodos contienen una cantidad insignificante de 3-FL en una célula, lisado o cultivo o sobrenadante celular, que contienen 2'-FL, por ejemplo, menos de un 1 % del nivel de 2'-FL o un 0,5 % del nivel de 2'-FL.

5 Un oligosacárido purificado, por ejemplo, 2'-FL o LDFT, es uno que tiene al menos un 90 %, un 95 %, un 98 %, un 99 %, o un 100 % (p/p) del oligosacárido deseado en peso. La pureza se evalúa con cualquier método conocido, por ejemplo, cromatografía en capa fina u otras técnicas de cromatografía conocidas en la técnica. Se desvela un método para purificar un oligosacárido fucosilado producido con la bacteria modificada mediante ingeniería genética que se ha descrito anteriormente, método que comprende la separación del oligosacárido fucosilado deseado (por ejemplo, 2'-FL) de contaminantes en un lisado celular bacteriano o cultivo o sobrenadante celular bacteriano de la bacteria.

15 Los oligosacáridos se purifican y se usan en una serie de productos para el consumo humano así como por animales, tal como animales de compañía (perros, gatos) así como ganado (animales bovino, equino, ovino, caprino o porcino, así como aves). Por ejemplo, una composición farmacéutica comprende 2'-FL purificada y un excipiente farmacéuticamente aceptable que es adecuado para la administración oral. Grandes cantidades de 2'-FL se producen en hospedador es bacterianos, por ejemplo, una bacteria *E. coli* que comprende un gen de  $\alpha$  (1,2) fucosiltransferasa exógeno.

20 Una bacteria de *E. coli* que comprende una combinación citoplásmica mejorada de lactosa y GDP-fucosa es útil en sistemas de producción de ese tipo. Las rutas metabólicas y los genes endógenos de *E. coli* se manipulan en formas que dan como resultado la generación de aumentos de concentraciones citoplásmicas de lactosa y/o GDP-fucosa, en comparación con los niveles encontrados en *E. coli* de tipo silvestre. Por ejemplo, las bacterias contienen al menos un 10 %, un 20 %, un 50 %, o 2X, 5X, 10X o más de los niveles en comparación con una bacteria de tipo silvestre correspondiente que carece de las modificaciones genéticas que se han descrito anteriormente.

30 Un método para producir una composición farmacéutica que comprende un oligosacárido purificado de leche humana (HMOS) se lleva a cabo cultivando la bacteria que se ha descrito anteriormente, purificando el HMOS producido por la bacteria y combinando el HMOS con un excipiente o vehículo para obtener un suplemento dietético para la administración oral. Estas composiciones son útiles en métodos para prevenir o tratar enfermedades entéricas y/o respiratorias en bebés y adultos. Por consiguiente, las composiciones se administran a un sujeto que padece o está en riesgo de desarrollar una enfermedad de este tipo.

35 También se desvelan métodos para aumentar la concentración intracelular de lactosa en *E. coli*, para células cultivadas en presencia de lactosa, mediante el uso de manipulaciones de genes de *E. coli* endógenos involucrados en la importación, exportación y catabolismo de lactosa. En particular, en el presente documento se describen métodos para aumentar los niveles de lactosa intracelular en *E. coli* modificada por ingeniería genética para producir un oligosacárido de la leche humana mediante la eliminación simultánea del gen endógeno de la  $\beta$ -galactosidasa (*lacZ*) y el gen represor del operón lactosa (*lacI*). Durante la construcción de esta delección, el promotor *lacIq* se coloca inmediatamente cadena arriba de (contiguo con) el gen de la lactosa permeasa, *lacY*, es decir, la secuencia del promotor *lacIq* está directamente cadena arriba y adyacente al inicio de la secuencia que codifica el gen *lacY*, de modo que el gen *lacY* está bajo regulación transcripcional por el promotor *lacIq*. La cepa modificada mantiene su capacidad para transportar lactosa desde el medio de cultivo (a través de *LacY*), pero sufre delección por la copia cromosómica de tipo silvestre del gen *lacZ* (que codifica la  $\beta$ -galactosidasa) responsable del catabolismo de la lactosa. Por lo tanto, se crea un grupo de lactosa intracelular cuando la cepa modificada se cultiva en presencia de lactosa exógena. Otro método para aumentar la concentración intracelular de lactosa en *E. coli* implica la delección del gen *lacA*. La mutación *lacA* previene la formación de acetil-lactosa intracelular, que no solo elimina esta molécula como contaminante de las purificaciones posteriores, sino que también elimina la capacidad de *E. coli* para exportar el exceso de lactosa de su citoplasma (Danchin A. Cells need safety valves. Bioessays 2009, Julio 31 (7): 769-73). Por lo tanto facilita en gran medida las manipulaciones intencionadas de la combinación de lactosa intracelular de *E. coli*.

55 También se desvelan métodos para aumentar los niveles intracelulares de GDP-fucosa en *Escherichia coli* mediante la manipulación de la ruta de biosíntesis del ácido colánico endógeno del organismo. Este aumento se logra a través de una serie de modificaciones genéticas de los genes endógenos de *E. coli* implicados directamente en la biosíntesis del precursor del ácido colánico, o en el control general del regulón sintético del ácido colánico. En el presente documento se describen métodos para aumentar los niveles intracelulares de GDP-fucosa en *E. coli* genéticamente modificada para producir un oligosacárido de la leche humana mediante la eliminación del gen *wcaJ*, que codifica la transferasa vehículo lipídico de UDP-glucosa. En un fondo nulo de *wcaJ*, la GDP-fucosa se acumula en el citoplasma de *E. coli*.

65 En un aspecto, el oligosacárido de la leche humana producido por bacterias modificadas genéticamente que comprenden una molécula de ácido nucleico exógena que codifica una  $\alpha$ (1,2) fucosiltransferasa es 2'-FL (2'-fucosillactosa). De acuerdo con la invención, la  $\alpha$ (1,2) fucosiltransferasa tiene la secuencia de SEQ ID NO: 8. La  $\alpha$ (1,2) fucosiltransferasa usada es una  $\alpha$ (1,2) fucosiltransferasa capaz de usar lactosa como sustrato aceptor de

azúcar para la síntesis de 2'-FL. Se desvela un gen de  $\alpha(1,2)$  fucosiltransferasa exógena que comprende al menos un 10 % de identidad a nivel de aminoácidos y menos de aproximadamente un 40 % con la alfa-(1,2) fucosiltransferasa de *Helicobacter pylori* 26695 (FutC).

5 También se desvelan composiciones que comprenden *E. coli* genéticamente modificada para producir el tetrasacárido de la leche humana lactodifucotetraosa (LDFT). La *E. coli* en este caso comprende una molécula de ácido nucleico exógena que codifica una  $\alpha(1,2)$  fucosiltransferasa que también tiene actividad de  $\alpha(1,3)$  fucosiltransferasa.

10 Se desvela un método para tratar, prevenir o reducir el riesgo de infección en un sujeto que comprende administrar a dicho sujeto una composición que comprende un oligosacárido de leche humana recombinante purificada, en la que el HMOS se une a un patógeno y en la que el sujeto está infectado o en riesgo de infección con el patógeno. La infección puede ser causada por un virus similar al de Norwalk o *Campylobacter jejuni*. El sujeto puede ser un mamífero que necesite un tratamiento de ese tipo. El mamífero es, por ejemplo, cualquier mamífero, por ejemplo, un ser humano, un primate, un ratón, una rata, un perro, un gato, una vaca, un caballo o un cerdo. El mamífero puede ser un ser humano. Por ejemplo, las composiciones se formulan en piensos para animales (por ejemplo, gránulos, croquetas, puré) o complementos alimenticios para animales de compañía, por ejemplo, perros o gatos, así como ganado o animales criados para consumo alimentario, por ejemplo, ganado, ovejas, cerdos, gallinas y cabras. El HMOS purificado se puede formular en un polvo (por ejemplo, polvo de fórmula infantil o polvo de suplemento nutricional para adultos, cada uno de los cuales se mezcla con un líquido como agua o uno antes del consumo) o en forma de comprimidos, cápsulas o pastas o se incorpora como componente en productos lácteos como leche, crema, queso, yogur o kéfir, o como componente en cualquier bebida, o se combina en una preparación que contiene cultivos microbianos vivos destinados a servir como probióticos, o en preparaciones prebióticas para mejorar el crecimiento de microorganismos benéficos ya sea *in vitro* o *in vivo*.

25 Los polinucleótidos, polipéptidos y oligosacáridos se purifican y/o se aíslan. Purificado define un grado de esterilidad que es seguro para la administración a un sujeto humano, por ejemplo, que carece de agentes infecciosos o tóxicos. De forma específica, como se usa en el presente documento, una molécula de ácido nucleico, polinucleótido, polipéptido, proteína u oligosacárido "aislada" o "purificada", está sustancialmente libre de otro material celular o medio de cultivo cuando se produce mediante técnicas recombinantes, o precursores químicos u otros productos químicos cuando se sintetizan por vía química. Por ejemplo, las composiciones de HMOS purificadas tienen al menos un 60 % en peso (peso seco) del compuesto de interés. Preferentemente, la preparación tiene al menos un 75 %, más preferentemente al menos un 90 %, y lo más preferentemente al menos un 99 %, en peso del compuesto de interés. La pureza se mide mediante cualquier método convencional apropiado, por ejemplo, mediante cromatografía en columna, cromatografía en capa fina o análisis de cromatografía líquida de alto rendimiento (HPLC). Por ejemplo, una "proteína purificada" se refiere a una proteína que se ha separado de otras proteínas, lípidos y ácidos nucleicos con los que está naturalmente asociada. Preferentemente, la proteína constituye al menos un 10, 20, 50, 70, 80, 90, 95, 99-100 % en peso seco de la preparación purificada.

40 Del mismo modo, por "sustancialmente puro" se hace referencia a un oligosacárido que se ha separado de los componentes que lo acompañan de manera natural. Por lo general, el oligosacárido es sustancialmente puro cuando está libre en al menos un 60 %, un 70 %, un 80 %, un 90 %, un 95 %, o incluso un 99 %, en peso libre de proteínas y moléculas orgánicas de origen natural con las que está naturalmente asociado.

45 Por "ácido nucleico aislado" se hace referencia a un ácido nucleico que está libre de los genes que, en el genoma de origen natural del organismo a partir del que se obtiene el ADN, flanquean al gen. La expresión abarca, por ejemplo: (a) un ADN que forma parte de una molécula de ADN genómico natural, pero que no está flanqueado por ambas secuencias de ácido nucleico que flanquean esa parte de la molécula en el genoma del organismo en el que se produce de manera natural (b) un ácido nucleico incorporado en un vector o en el ADN genómico de un procarionte o eucariota de manera tal que la molécula resultante no sea idéntica a ningún vector natural o ADN genómico; (c) una molécula separada tal como un ADNc, un fragmento genómico, un fragmento producido por reacción en cadena de la polimerasa (PCR), o un fragmento de restricción; y (d) una secuencia de nucleótidos recombinante que forma parte de un gen híbrido, es decir, un gen que codifica una proteína de fusión. Las moléculas de ácido nucleico aisladas de acuerdo con la presente invención incluyen además moléculas producidas sintéticamente, así como cualquier ácido nucleico que se haya alterado químicamente y/o que tenga cadenas principales modificadas.

Un "promotor heterólogo" es un promotor que es diferente del promotor al que un gen o secuencia de ácidos nucleicos se une de forma operativa en la naturaleza.

60 El término "sobrexpresar" o "sobrexpresión" se refiere a una situación en la que una célula alterada mediante ingeniería genética expresa más factores de lo que lo haría, en las mismas condiciones, una célula de tipo silvestre. De manera similar, si una célula no alterada no expresa un factor que se modifica mediante ingeniería genética para producir, se usa el término "expresar" (a diferencia de "sobrexpresar") que indica que la célula de tipo silvestre no expresó el factor en absoluto antes de la manipulación genética.

65 Los términos "tratar" y "tratamiento", como se usan en el presente documento se refieren a la administración de un

agente o formulación a un individuo clínicamente sintomático afectado por una afección, trastorno o enfermedad adversos, con el fin de conseguir una reducción en la gravedad y/o frecuencia de los síntomas, una eliminación de los síntomas y/o su causa subyacente, y/o facilitar la mejora o la reparación del daño. Los términos "prevención" y "prevenir" se refieren a la administración de un agente o composición a un individuo clínicamente asintomático que es susceptible a una afección, trastorno o enfermedad adversa particular, y por lo tanto se relaciona con la prevención de la aparición de síntomas y/o su causa subyacente.

Con los términos "cantidad eficaz" y "cantidad terapéuticamente eficaz" de una formulación o componente de formulación se hace referencia a una cantidad no tóxica pero suficiente de la formulación o componente para proporcionar el efecto deseado.

El término de transición "que comprende", que es sinónimo de "que incluye", "que contiene" o "caracterizado por", es inclusivo o de extremos abiertos y no excluye elementos o etapas de método adicionales sin mencionar. Por el contrario, la expresión de transición "que consiste en" excluye cualquier elemento, etapa o ingrediente no especificado en la reivindicación. La expresión de transición "que consiste esencialmente en" limita el alcance de una reivindicación a los materiales o etapas especificados" y aquellos que no lo hacen influyen materialmente en las características básicas y novedosas" de la invención que se reivindica.

El organismo hospedador usado para expresar el producto génico de fucosiltransferasa que acepta lactosa y no es FutC es generalmente la enterobacteria K-12 de *Escherichia coli* (*E. coli*). La cepa K-12 de *E. coli* no se considera un patógeno humano o animal ni es toxicogénica. *E. coli* K-12 es una cepa de producción de bacterias convencional y se indica por su seguridad debido a su escasa habilidad para colonizar el colon y establecer infecciones (véase, por ejemplo, [epa.gov/oppt/biotech/pubs/fra/fra004.htm](http://epa.gov/oppt/biotech/pubs/fra/fra004.htm)). Sin embargo, una diversidad de especies bacterianas se puede usar en los métodos de biosíntesis de oligosacáridos, por ejemplo, *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, o *Xanthomonas campestris*. También se pueden usar bacterias del género *Bacillus*, incluyendo *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, y *Bacillus circulans*. Del mismo modo, las bacterias de los géneros *Lactobacillus* y *Lactococcus* se pueden modificar usando los métodos de la presente invención, que incluyen, pero no se limitan a, *Lactobacillus acidophilus*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, y *Lactococcus lactis*. *Streptococcus thermophilus* y *Propionibacterium freudenreichii* también son especies bacterianas adecuadas para la invención que se describe en el presente documento. Como parte de la presente invención también se incluyen cepas, modificadas como se describe en el presente documento, de los géneros *Enterococcus* (por ejemplo, *Enterococcus faecium* y *Enterococcus thermophilus*), *Bifidobacterium* (por ejemplo, *Bifidobacterium longum*, *Bifidobacterium infantis*, y *Bifidobacterium bifidum*), *Sporolactobacillus* spp., *Micromonospora* spp., *Micrococcus* spp., *Rhodococcus* spp., y *Pseudomonas* (por ejemplo, *Pseudomonas fluorescens* y *Pseudomonas aeruginosa*). Las bacterias que comprenden las características que se describe en el presente documento se cultivan en presencia de lactosa, y un oligosacárido fucosilado se recupera, ya sea de la propia bacteria o de un sobrenadante de tipo de la bacteria. El oligosacárido fucosilado se purifica para su uso en productos terapéuticos o nutricionales, o las bacterias se usan directamente en productos de ese tipo.

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que entiende comúnmente un experto en la materia a la que pertenece la presente invención. Aunque los métodos y materiales similares o equivalentes a los que se describen en el presente documento se pueden usar en la práctica o en el ensayo de la presente invención, a continuación se describen métodos y materiales adecuados. En caso de conflicto con patentes extranjeras publicadas y solicitudes de patentes citadas en el presente documento, las solicitudes de Genbank y NCBI indicadas por el número de acceso mencionado en el presente documento, y todas las demás referencias, documentos, manuscritos y bibliografía científica publicados en el presente documento, la presente memoria descriptiva, incluyendo las definiciones, prevalecerá. Además, los materiales, métodos y ejemplos son solo ilustrativos y no pretenden ser limitantes.

#### Breve descripción de las figuras

La Figura 1 es una ilustración esquemática que muestra la ruta de síntesis de los fucosil oligosacáridos de leche humana. Se y Le indican síntesis con fucosiltransferasas de los genes secretores y de Lewis, respectivamente. Se proporciona al nombre bioquímico abreviado [con la estructura bioquímica alternativa entre corchetes] (análogo antigénico del grupo histo-sanguíneo entre paréntesis).

La Figura 2 es un esquema que demuestra las rutas metabólicas y los cambios introducidos en las mismas para modificar mediante ingeniería la síntesis de 2'-fucosillactosa (2'-FL) en *Escherichia coli* (*E. coli*).

La Figura 3 es una serie de fotografías que muestran análisis de cromatografía en capa fina de 2'-FL producida en cepas de *E. coli* con  $\alpha$  (1,2) fucosiltransferasas candidatas. La Figura 3A muestra la producción significativa de 2'-FL con WbgL. La Figura 3B muestra la producción significativa de 2'-FL con FutL. La Figura 3C muestra la producción significativa de 2'-FL con FutN.

La Figura 4 es un diagrama y una fotografía de análisis de cromatografía en capa fina que muestra que la

digestión con fucosidasa confirma la síntesis de 2'-FL auténtica con WbgL. Los oligosacáridos producidos con una cepa de *E. coli* expresan *wbgL* se aislaron y se sometieron a digestión durante la noche con diferentes fucosidasas. Los productos de reacción se analizaron por TLC. La producción de fucosa y lactosa mediante tratamiento con  $\alpha(1,2)$  fucosidasa se ilustra en la calle 2.

5 La Figura 5 es un diagrama del plásmido pG171.

La Figura 6 es un diagrama de un constructor cromosómico de *Placq lacY<sup>+</sup>*.

La Figura 7 es un diagrama de la delección cromosómica de *wcaJ*.

La Figura 8 es un diagrama de la inserción de *kan*, *lacZ<sup>+</sup>* en el locus *lon*.

La Figura 9 es un diagrama del plásmido pG204.

10 La Figura 10 es un diagrama del plásmido pG216.

La Figura 11 es un diagrama del plásmido pG217.

#### Descripción detallada de la invención

15 Aunque algunos estudios sugieren que los glicanos de la leche humana se podrían usar como agentes antiadhesión antimicrobiana, la dificultad y el gasto para producir cantidades adecuadas de éstos agentes de una calidad adecuada para consumo humano ha limitado su ensayo a escala completa y utilidad percibida. Lo que se necesita es un método adecuado para producir los glicanos apropiados en cantidades suficientes con un coste razonable. Antes de la invención que se describe en el presente documento, se realizaron intentos de uso de varios enfoques de síntesis distintos para síntesis de glicanos. Algunos enfoques químicos pueden sintetizar oligosacáridos (Flowers, H. M. *Methods Enzymol* 50, 93-121 (1978); Seeberger, P. H. *Chem Commun (Camb)* 1115-1121 (2003)), pero los reactivos por estos métodos son caros y potencialmente tóxicos (Koeller, K. M. y Wong, C. H. *Chem Rev* 100, 4465-4494 (2000)). Las enzimas expresadas a partir de organismos modificados por ingeniería (Albermann, C., Piepersberg, W. y Wehmeier, U. F. *Carbohydr Res* 334, 97-103 (2001); Bettler, E., Samain, E., Chazalet, V., Bosso, C., *et al.*, *Glycoconj J* 16, 205-212 (1999); Johnson, K. F. *Glycoconj J* 16, 141-146 (1999); Palcic, M. M. *Curr Opin Biotechnol* 10, 616-624 (1999); Wymer, N. y Toone, E. J. *Curr Opin Chem Biol* 4, 110-119 (2000)) proporcionan una síntesis precisa y eficaz (Palcic, M. M. *Curr Opin Biotechnol* 10, 616-624 (1999)); Crout, D. H. y Vic, G. *Curr Opin Chem Biol* 2, 98-111 (1998)), pero el coste elevado de los reactivos, especialmente los nucleótidos de azúcar, limita su utilidad para producción a gran escala con bajo coste. Algunos microbios se han modificado mediante ingeniería genética para expresar las glicosiltransferasas necesarias para sintetizar oligosacáridos a partir de la combinación innata de azúcares de nucleótidos de las bacterias (Endo, T., Koizumi, S., Tabata, K., Kakita, S. y Ozaki, A. *Carbohydr Res* 330, 439-443 (2001); Endo, T., Koizumi, S., Tabata, K. y Ozaki, A. *Appl Microbiol Biotechnol* 53, 257-261 (2000); Endo, T. y Koizumi, S. *Curr Opin Struct Biol* 10, 536-541 (2000); Endo, T., Koizumi, S., Tabata, K., Kakita, S. y Ozaki, A. *Carbohydr Res* 316, 179-183 (1999); Koizumi, S., Endo, T., Tabata, K. y Ozaki, A. *Nat Biotechnol* 16, 847-850 (1998)). Sin embargo, antes de la invención que se describe en el presente documento, había una necesidad creciente de identificar y caracterizar glicosiltransferasas adicionales que son útiles para la síntesis de HMOS en los operadores bacterianos modificados mediante ingeniería metabólicamente.

40 No todas las  $\alpha(1,2)$  fucosiltransferasas pueden utilizar la lactosa como un azúcar aceptor. Una enzima deseada utiliza GDP-fucosa como donante, y la lactosa es el aceptor para ese donante. Se llevó a cabo un método para identificar nuevas enzimas  $\alpha(1,2)$  fucosiltransferasa capaces de utilizar lactosa como aceptor utilizando las siguientes etapas: 1) realizar una búsqueda informática de bases de datos de secuencias para definir un amplio grupo de homólogos de secuencia sencilla de cualquier  $\alpha(1,2)$  fucosiltransferasa que utiliza lactosa, conocida; 2) usar el listado de homólogos de la etapa 1 para obtener un perfil de búsqueda que contiene una secuencia común y/o motivos estructurales compartidos por los miembros del grupo amplio, por ejemplo, mediante el uso de programas informáticos tales como MEME (Multiple Em for Motif Elicitation disponible en <http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) o PSI-BLAST (BLAST con iteración de posición específica disponible en [ncbi.nlm.nih.gov/blast](http://ncbi.nlm.nih.gov/blast) con información adicional en [cnx.org/cont-tent/m11040/latest/](http://cnx.org/cont-tent/m11040/latest/)); 3) búsquedas en bases de datos de secuencias (por ejemplo, usando programas informáticos tales como PSI-BLAST, o MAST (Motif Alignment Search Tool disponible en <http://meme.sdsc.edu/meme/cgi-bin/mast.cgi>), usando este perfil de búsqueda derivado como una búsqueda e identificando "secuencias candidatas" cuya simple homología de secuencia con la  $\alpha(1,2)$  fucosiltransferasa que acepta lactosa original es de un 40 % o inferior; 4) escaneando la bibliografía científica y desarrollando un listado de "organismos candidatos" conocidos por expresar  $\alpha(1,2)$  fucosil-glicanos; 5) seleccionando solo aquellas "secuencias candidatas" que se obtienen a partir de "organismos candidatos" para generar un listado de "enzimas candidatas que utilizan lactosa"; y 6) expresando cada "enzima candidata que utiliza lactosa" y analizando la actividad de  $\alpha(1,2)$  fucosiltransferasa que utiliza lactosa.

60 El conjunto de herramientas de análisis de secuencia MEME ([meme.sdsc.edu/meme/cgi-bin/meme.cgi](http://meme.sdsc.edu/meme/cgi-bin/meme.cgi)) también se puede usar como una alternativa a PSI-BLAST. Los motivos de secuencia se descubren usando el programa "MEME". A continuación estos motivos se pueden usar para buscar bases de datos de secuencias usando el programa "MAST". Los algoritmos de búsqueda BLAST y PHI-BLAST son otras alternativas bien conocidas.

65 Para someter al ensayo la actividad de utilización de lactosa, la producción de 2'-FL se evalúa en un organismo hospedador que expresa la enzima candidata y que contiene tanto GDP-fucosa citoplasmática como combinaciones de lactosa. La producción de 2'-FL indica que la secuencia codificante de la enzima candidata funciona como una  $\alpha(1,2)$  fucosiltransferasa que utiliza lactosa.

Para encontrar enzimas similares a FutC, se usó el aminoácido completo de FutC como una consulta en PSI-BLAST. Los resultados del método de identificación de  $\alpha(1,2)$  fucosiltransferasa que utiliza lactosa de la presente invención son sorprendentes, ya que el % de identidad de varias de los  $\alpha(1,2)$  fucosiltransferasas que utilizan lactosa son menos de un 40% de la secuencia de FutC de referencia. Otro aspecto más sorprendente del estudio es que 8 de los 10 candidatos analizados pudieron utilizar la lactosa como aceptor, 3 de los cuales lo hicieron a niveles muy cercanos a la enzima FutC usada como "criterio de referencia". Esta fue una "tasa de éxito" mayor que se anticipó. Aunque 6 de cada 10 de las enzimas candidatas se encuentran en bacterias que incorporan  $\alpha(1,2)$  fucosa en su estructura de LPS, los oligosacáridos a los que se une la fucosa son muy diferentes de la lactosa que se le pide a cada enzima candidata que utilice en la consulta. Además, fue sorprendente que tanto WbIA como WbgN pudieran utilizar la lactosa como un aceptor, porque estas dos enzimas se encuentran en bacterias que no incorporan fucosa en su estructura de LPS. En su lugar, utilizan un azúcar relacionado llamado colitosa.

#### Glicanos de leche humana

La leche humana contiene un conjunto diverso y abundante de oligosacáridos neutros y ácidos (Kunz, C., Rudloff, S., Baier, W., Klein, N., y Strobel, S. (2000). *Annu Rev Nutr* 20, 699-722; Bode, L. (2006). *J Nutr* 136, 2127-130). Más de 130 oligosacáridos complejos diferentes se han identificado en la leche humana, y su diversidad y abundancia estructural es exclusiva de los seres humanos. Aunque estas moléculas pueden no ser utilizadas directamente por los bebés para la nutrición, sin embargo, cumplen funciones críticas en el establecimiento de un microbioma intestinal saludable (Marcobal, A., Barboza, M., Froehlich, J. W., Block, D. E., *et al.*, *J Agric Food Chem* 58, 5334-5340 (2010)), en la prevención de enfermedades (Newburg, D. S., Ruiz-Palacios, G. M. y Morrow, A. L. *Annu Rev Nutr* 25, 37-58 (2005)), y en la función inmunológica (Newburg, D. S. y Walker, W. A. *Pediatr Res* 61, 2-8 (2007)). A pesar de los millones de años de exposición a los oligosacáridos de la leche humana (HMOS), los patógenos aún tienen que desarrollar formas de eludir la capacidad del HMOS para prevenir la adhesión a las células diana e inhibir la infección. La capacidad de utilizar HMOS como inhibidores de la adherencia de patógenos promete abordar la crisis actual de la creciente resistencia a los antibióticos. Los oligosacáridos de la leche humana producidos por biosíntesis representan los compuestos directores de una nueva clase de terapias contra algunos de los azotes más difíciles de la sociedad.

Una estrategia alternativa para la síntesis eficaz de HMOS a escala industrial es la ingeniería metabólica de las bacterias. Este enfoque implica la construcción de cepas microbianas que sobreexpresan glicosiltransferasas heterólogas, transportadores de membrana para la importación de azúcares precursores al citosol bacteriano, y que poseen conjuntos mejorados de azúcares de nucleótidos regeneradores para su uso como precursores biosintéticos (Dumon, C., Samain, E., y Priem, B. (2004). *Biotechnol Prog* 20, 412-19; Ruffing, A., y Chen, R.R. (2006). *Microb Cell Fact* 5, 25). Un aspecto clave de este enfoque es la glicosiltransferasa heteróloga seleccionada para la sobreexpresión en el hospedador microbiano. La elección de la glicosiltransferasa puede influir de forma significativa en el rendimiento final del oligosacárido sintetizado deseado, dado que las enzimas pueden variar mucho en términos de cinética, especificidad de sustrato, afinidad por las moléculas donantes y aceptoras, estabilidad y solubilidad. Se identificaron y caracterizaron algunas glicosiltransferasas obtenidas a partir de diferentes especies bacterianas en términos de su capacidad para catalizar la biosíntesis de HMOS en cepas hospedadoras de *E. coli* (Dumon, C., Bosso, C., Utille, J.P., Heyraud, A., y Samain, E. (2006). *Chembiochem* 7, 359-365; Dumon, C., Samain, E., y Priem, B. (2004). *Biotechnol Prog* 20, 412-19; Li, M., Liu, X.W., Shao, J., Shen, J., Jia, Q., Yi, W., Song, J.K., Woodward, R., Chow, C.S., y Wang, P.G. (2008). *Biochemistry* 47, 378-387). La identificación de glicosiltransferasas adicionales con una cinética más rápida, una mayor afinidad por los donantes de nucleótidos de azúcar y/o moléculas aceptoras, o una mayor estabilidad dentro del hospedador bacteriano mejora de forma significativa los rendimientos de HMOS terapéuticamente útiles. Antes de la invención que se describe en el presente documento, las síntesis químicas de HMOS eran posibles, pero estaban limitadas por problemas de estereoespecificidad, disponibilidad de precursores, impurezas del producto y alto coste general (Flowers, H. M. *Methods Enzymol* 50, 93-121 (1978); Seeberger, P. H. *Chem Commun (Camb)* 1115-1121 (2003); Koeller, K. M. y Wong, C. H. *Chem Rev* 100, 4465-4494 (2000)). La invención supera los inconvenientes de estos intentos previos al proporcionar nuevas estrategias para fabricar grandes cantidades de oligosacáridos de la leche humana (HMOS) para su uso como suplementos dietéticos. Las ventajas incluyen la expresión eficaz de la enzima, la estabilidad mejorada y/o la solubilidad del producto génico (2'-FL) y una toxicidad reducida para el organismo hospedador. Por ejemplo, las  $\alpha(1,2)$  fucosiltransferasas obtenidas a partir de cepas de *E. coli* (por ejemplo WbgL) son más estables y se expresan en niveles más altos dentro de las cepas hospedadoras de producción de *E. coli* en comparación con FutC. En otro ejemplo, la fucosiltransferasa altamente activa (futN) se obtiene a partir de un microbio comensal (*Bacteroides*) en lugar de un patógeno. Dado que muchas cepas de producción obtenidas por ingeniería usan genes de fucosiltransferasa obtenidos a partir de patógenos, la seguridad y/o la mayor aceptación por parte del consumidor son ventajas añadidas de esta secuencia/enzima.

Como se describe con detalle a continuación, *E. coli* (u otras bacterias) está modificada por ingeniería para producir 2'-FL a niveles comercialmente viables. Por ejemplo, los rendimientos son > 5 gramos/litro en un proceso de fermentación bacteriana.

#### Papel de los glicanos de leche humana en enfermedades infecciosas

Los glicanos de leche humana, que comprenden oligosacáridos no unidos y sus glicoconjugados, desempeñan un papel importante en la protección y el desarrollo del tracto gastrointestinal (GI) infantil. Los oligosacáridos fucosilados neutros, incluyendo 2'-fucosillactosa (2'-FL), protegen a los bebés frente a varios patógenos importantes. Los oligosacáridos de la leche que se encuentran en varios mamíferos difieren mucho, y la composición en seres humanos es única (Hamosh M., 2001 *Pediatr Clin North Am*, 48: 69-86; Newburg D.S., 2001 *Adv Exp Med Biol*, 501: 3-10). Además, los niveles de glicano en la leche materna cambian a lo largo de la lactancia y también varían ampliamente entre individuos (Morrow A.L. *et al.*, 2004 *J Pediatr*, 145: 297-303; Chaturvedi P *et al.*, 2001 *Glycobiology*, 11: 365-372). Se han identificado aproximadamente 200 oligosacáridos distintos de la leche humana y las combinaciones de epítomos simples son responsables de esta diversidad (Newburg D.S., 1999 *Curr Med Chem*, 6:117-127; Ninonuevo M. *et al.*, 2006 *J Agric Food Chem*, 54: 7471-74801).

Los oligosacáridos de la leche humana están compuestos por 5 monosacáridos: D-glucosa (Glc), D-galactosa (Gal), N-acetilglucosamina (GlcNAc), L-fucosa (Fuc) y ácido siálico (N-acetil neuramínico, Neu5Ac, NANA). Los oligosacáridos de la leche humana generalmente se dividen en dos grupos según sus estructuras químicas: compuestos neutros que contienen Glc, Gal, GlcNAc y Fuc, unidos a un núcleo de lactosa (Gal $\beta$ 1-4Glc), y compuestos ácidos que incluyen los mismos azúcares y, a menudo, el mismo estructuras centrales, más NANA (Charlwood J. *et al.*, 1999 *Anal Biochem*, 273:261-277; Martin-Sosa *et al.*, 2003 *J Dairy Sci*, 86: 52-59; Parkkinen J. y Finne J., 1987 *Methods Enzymol*, 138: 289-300; Shen Z. *et al.*, 2001 *J Chromatogr A*, 921: 315-321).

Aproximadamente un 70-80 % de los oligosacáridos en la leche materna están fucosilados, y se cree que sus rutas de síntesis proceden como se muestra en la Figura 1 (las rutas de Tipo I y de Tipo II comienzan con diferentes moléculas precursoras). Una proporción menor de los oligosacáridos está sialilada o fucosilada y sialilada, pero sus rutas de síntesis no están completamente definidas. La comprensión de los oligosacáridos ácidos (sialilados) está limitada en parte por la capacidad de medir estos compuestos. Se han diseñado métodos sensibles y reproducibles para el análisis de oligosacáridos neutros y ácidos. Los oligosacáridos de la leche humana como clase sobreviven al tránsito a través del intestino de los bebés de manera muy eficaz, siendo esencialmente no digeribles (Chaturvedi, P., Warren, C. D., Buescher, C. R., Pickering, L. K. y Newburg, D. S. *Adv Exp Med Biol* 501, 315-323 (2001)).

### Los glicanos de la leche humana inhiben la unión de enteropatógenos a sus receptores

Los glicanos de la leche humana tienen homología estructural con los receptores celulares para los enteropatógenos y funcionan como señuelos de los receptores. Por ejemplo, las cepas patógenas de *Campylobacter* se unen específicamente a glicanos que contienen H-2, es decir, 2'-fucosil-N-acetilglucosamina o 2'-fucosillactosa (2'-FL); La unión de *Campylobacter* y la capacidad de infección son inhibidas por 2'-FL y otros glicanos que contienen este epítomo H-2. De forma similar, algunos patógenos diarreogénicos de *E. coli* son fuertemente inhibidos *in vivo* por los oligosacáridos de la leche humana que contienen restos de fucosa unidos en la posición 2. Varias cepas principales de calicivirus humanos, especialmente los norovirus, también se unen a los glicanos fucosilados en la posición 2, y esta unión es inhibida por los glicanos fucosilados en la posición 2 unidos a la leche humana. El consumo de leche humana que tiene niveles altos de estos fucosiloligosacáridos unidos a la posición 2 se asoció con un menor riesgo de norovirus, *Campylobacter*, ST de diarrea asociada a *E. coli* y diarrea moderada a grave de todas las causas en una cohorte mejicana de niños en periodo de lactancia (Newburg D.S. *et al.*, 2004 *Glycobiology*, 14: 253-263; Newburg D.S. *et al.*, 1998 *Lancet*, 351: 1160-1164). Varios agentes patógenos utilizan glicanos sialilados como sus receptores anfitriones, tales como la gripe (Couceiro, J. N., Paulson, J. C. y Baum, L. G. *Virus Res* 29, 155-165 (1993)), parainfluenza (Amonsén, M., Smith, D. F., Cummings, R. D. y Air, G. M. *J Virol* 81, 8341-8345 (2007)), y rotovirus (Kuhlenschmidt, T. B., Hanafin, W. P., Gelberg, H. B. y Kuhlenschmidt, M. S. *Adv Exp Med Biol* 473, 309-317 (1999)). El epítomo sialil-Lewis X es usado por *Helicobacter pylori* (Mahdavi, J., Sondén, B., Hurtig, M., Olfat, F. O., *et al.*, *Science* 297, 573-578 (2002)), *Pseudomonas aeruginosa* (Scharfman, A., Delmotte, P., Beau, J., Lamblin, G., *et al.*, *Glycoconj J* 17, 735-740 (2000)), y algunas cepas de norovirus (Rydell, G. E., Nilsson, J., Rodríguez-Díaz, J., Ruvoën-Clouet, N., *et al.*, *Glycobiology* 19, 309-320 (2009)).

### Modificación por ingeniería de *E. coli* para producir oligosacárido 2'-FL de leche humana

En el presente documento se describe un enfoque de identificación sistemática de genes, que se usó para identificar nuevas  $\alpha$  (1,2) fucosiltransferasas ( $\alpha$  (1,2) FT) para la síntesis de oligosacáridos unidos a fucosilo en *E. coli* manipulada metabólicamente. De particular interés son las  $\alpha$  (1,2) FT que son capaces de la síntesis de la 2'-fucosillactosa de HMOS (2'-FL). 2'-FL es el oligosacárido fucosilado más abundante presente en la leche humana, y este oligosacárido ofrece protección a los recién nacidos contra la diarrea infecciosa causada por patógenos bacterianos tales como *Campylobacter jejuni* (Ruiz-Palacios, G.M., *et al.*, (2003). *J Biol Chem* 278, 14112-120; Morrow, A.L. *et al.*, (2004). *J Pediatr* 145, 297-303; Newburg, D.S. *et al.*, (2004). *Glycobiology* 14, 253-263).

La vía de síntesis de los fucosil oligosacáridos de la leche humana se ilustra en la Figura 1. Estructuralmente, 2'-FL consiste en una molécula de fucosa unida en  $\alpha$  1,2 a la parte de galactosa de la lactosa (Fuc $\alpha$ 1-2Gal $\beta$ 1-4Glc). Se ha utilizado una  $\alpha$  (1,2) FT de la cepa 26695 de *H. pylori* denominada FutC para catalizar la síntesis de 2'-FL en *E. coli* mediante ingeniería metabólica (Drouillard, S. *et al.*, (2006). *Angew Chem Int Ed Engl* 45, 1778-780). Por lo tanto, la secuencia de aminoácidos de FutC se usó como una consulta en el algoritmo de búsqueda PSI-BLAST (Position

Specific Iterated Basic Local Alignment Search Tool) para identificar nuevas  $\alpha$  (1,2) FT candidatas para la producción de 2'-FL en *E. coli*. Usando PSI-BLAST, se crea un listado de secuencias de proteínas estrechamente relacionadas en función de la secuencia de consulta. El algoritmo a continuación genera una secuencia de perfil, que resume los motivos significativos presentes en estas secuencias. Este perfil a continuación se usa como una nueva consulta para identificar un grupo más grande de secuencias candidatas, y el proceso se itera para generar un grupo aún mayor de candidatos.

La secuencia de aminoácidos de FutC se usó como una consulta para 2 iteraciones del algoritmo de búsqueda PSI-BLAST. Esta búsqueda dio como resultado un grupo de 277 candidatos con similitud con FutC, algunos de los cuales estaban más estrechamente relacionados (identidad de aminoácidos compartida mayor que un 25 %), así como un grupo que estaba más relacionado con FutC (identidad de aminoácidos compartida menor de un 25 %). Del grupo más estrechamente relacionado, se analizaron las  $\alpha$ (1,2) FT predichas de especies bacterianas que incorporan fucosa en el antígeno O de su lipopolisacárido (LPS) o en las subunidades de polisacárido que componen la cápsula de la superficie celular. Las  $\alpha$ (1,2) FT de estos tipos de organismos tienen más probabilidades de utilizar fucosa como sustrato, dada la presencia de fucosa en sus estructuras de carbohidratos de superficie. También se analizaron  $\alpha$ (1,2) FT de especies bacterianas entéricas conocidas, comensales o patógenas. Dichos organismos a veces exhiben estructuras de carbohidratos en su superficie celular que contienen fucosa e imitan varias estructuras de antígeno de Lewis que contienen 2'-fucosilo encontradas en organismos superiores (Appelmelk, B.J. *et al.*, (1998). *Infect Immun* 66, 70-76; Coyne, M.J. *et al.*, (2005). *Science* 307, 1778-781). Es más probable que las  $\alpha$ (1,2) FT candidatas de estos tipos de organismos usen fucosa como sustrato y también que catalicen el enlace de la fucosa a oligosacáridos aceptables útiles.

Se analizaron diez  $\alpha$ (1,2) FT con una homología superior a un 25 % a nivel del aminoácido con respecto a FutC identificada a partir de la identificación sistemática (Tabla 1).

Tabla 1. Sumario de  $\alpha$  (1,2) fucosiltransferasas candidatas sometidas a ensayo para su capacidad para estimular 2'-FL en cepas de *E. coli* modificadas mediante ingeniería. La actividad de cada candidato se comparó con FutC y se describió de forma semicuantitativa usando el símbolo "+" en la última columna, en la que se evalúa la actividad más elevada de FutC con 4 símbolos "+".

Nombre del Gen	N.º de Acceso (NCBI)	Organismo	% de Identidad c/ FutC	¿Fucosa en LPS o cápsula?	Síntesis de 2'-FL
<i>futC</i>	NP_206893 NP_206894	<i>H. pylori</i> 26695	-	Sí	++++
<i>wblA</i>	BAA33632	<i>V. cholerae</i> O22	28 %	No	+
<i>wbgL</i>	ADN43847	<i>E. coli</i> 0126	25 %	Sí	+++
<i>futD</i>	ZP_04580654	<i>H. bilis</i> ATCC 437879	39 %	Sí	+
<i>futE</i>	ZP_07805473	<i>H. cinaedi</i> CCUG 18818	44 %	Desconocido	-
<i>futL</i>	YP_003517185	<i>H. mustelae</i> ATCC 43772	70 %	Sí	+++
<i>futN</i>	YP_001300461	<i>B. vulgatus</i> ATCC 8482	27 %	Desconocido	++
<i>futO</i>	ZP_02065239	<i>8. ovatus</i> ATCC 8483	27 %	Desconocido	-
<i>wbgN</i>	YP_003500093	<i>E. coli</i> O55:H7	28 %	No	+
<i>bft1</i>	CAH09369	<i>B. fragilis</i> 9343	34 %	Sí	-
<i>bft3/wcfB</i>	CAH06753	<i>B. fragilis</i> 9343	28 %	Sí	+

La secuencia de aminoácidos de de *Helicobacter pylori* 26695 alfa-(1,2) fucosiltransferasa (FutC) se presenta a continuación (SEQ ID NO: 2; Número de Acceso en GenBank NP\_206893 y NP\_206894 (GI: 15644723 y 15644724)).

## ES 2 738 589 T3

1 mafkvvqicg glgnqmfqya fakslqkhl n tpvllditsf dwsnrkmqle lfpidlpyas  
61 akeiaiakmq hlpklvrtdl kcmgfdrvsq eivfeyepgl lkpsrltyfy gyfqdpryfd  
121 aisplikqtf tlpppengnn kkkeeyhrk lalilaakns vfvhvrrgdy vgigcqlgid  
181 yqkkaleyia krvpnmelfv fcedlkftqn ldlgypfmdm ttrdkeeeay wdmlmqsk  
241 hgiianstys wwaaylinnp ekiiigpkhw lfghenilck ewvkieshfe vkskky na

La secuencia de aminoácidos de *Vibrio cholera* O22 WbIA se presenta a continuación (SEQ ID NO: 3; Número de Acceso en GenBank BAA33632 (GI: 3721682)).

5

1 mivmkisggl gnqlfqyavg raiaiqygvv lkldvsaykn yklhngyrld qfninadian  
61 edeifhllkgs snrlsrilrr lgwlkkntyy aekqrtiydv svfmqapryl dgywqneqyf  
121 sqiravllqe lwpnqplsln aqahqikiqq thavsihvrr gdylnhpeig vldidyykra  
181 vdyikekiea pvffvfsndv awckdnfnfi dspvfiedtq teiddlmlmc qcqhnivans  
241 sfswwaawln snvdki viap ktwmaenpkg ykwvpswre i

La secuencia de aminoácidos de *Escherichia coli* 0126 WbgL se presenta a continuación (SEQ ID NO: 4; Número de Acceso en GenBank ADN43847 (GI: 307340785)).

10

1 msiirlqggl gnqlfqfsfg yalskingtp lyfdishyae nddhggyrln nlqipeeylq  
61 yytpkinniy kflvrgsrly peiflflgfc nefhaygydf eyiaqkwksk kyigywqseh  
121 ffhkhildlk effipknvse qanllaakil esqsslsihi rrgdyiknt atlthgvcsl  
181 eyykkalnki rdlamirdvf ifsddifwck enietllskk yniyysedls qeedlwlmsl  
241 anhhianss fswwgaylgt sasqiviypt pwyditpknt yipivnhwin vdkhssc

La secuencia de aminoácidos de *Helicobacter bilis* ATCC 437879 FutD se presenta a continuación (SEQ ID NO: 5; Número de Acceso en GenBank ZP\_04580654 (GI: 237750174)).

15

1 mgdykivelt cglgnqmfqy afakalqkhl qvpvlldktw ydtqdnstqf sldifnvdle  
61 yatntqieka karvsklppl lrkmfglkkh niaysqsfdf hdeyllpndf tyfsgffqna  
121 kylkgleqel ksifyydsnn fsnfgkqrle lilqaknsif ihirrgdyck igwelgmdyy  
181 kraiyyimdr veepkffifg atdmsfteqf qknlglenn sanlsektit qdnqhedmfl  
241 mcyckhaila nssysfwsay lnndannivi aptpwllndd niicddwiki ssk

La secuencia de aminoácidos de *Helicobacter cinaedi* CCUG 18818 alfa-1,2-fucosiltransferasa (FutE) se presenta a continuación (SEQ ID NO: 6; Número de Acceso en GenBank ZP\_07805473 (GI: 313143280)).

20

1 mlfpfkfiyn rlrykairli rrrasyrpfy efyahivwge egvndrink hyressfkpy  
61 afpyginmsf vysndvyall kddfrlkipl rydnamlkkq iqntdksvfl hirrgdylqs  
121 eglyvvlgt yyqkaleilk skitnphifv fsndmcwcke ylmryvdfsg ctidfiegent  
181 egnaveemel mrscqhaila nstfswwaay lienpdkivi mpkeylndss rflpkqflal  
241 knwflvdhiw gsvelan

La secuencia de aminoácidos de *Helicobacter mustelae* 12198 (ATCC 43772) alfa-1,2-fucosiltransferasa (FutL) se presenta a continuación (SEQ ID NO: 7; Número de Acceso en GenBank YP\_003517185 (GI: 291277413)).

25

1 mdfkivqvhg glgnqmfqya fakslqthln ipvllddttwf dygnrelglh lfpidlqcas  
 61 aqgiaaahmq nlprlvrgal rrmglgrvsk eivfeympel fepsriayfh gyfqdpryfe  
 121 displikqtf tlphpteae qysrklsql aaknsvfvi rrgdymrlgw qldisyqlra  
 181 iaymakrvqn leflfcedl efvqnlldlgy pfvdmtrdg aahwdmmlmq sckhgiitns  
 241 tyswaaayli knpekiiigp shwiygneni lckdvwkies qfetks

5 Se identificó una  $\alpha(1,2)$ fucosiltransferasa mediante la identificación sistemática que tenía actividad enzimática comparable con respecto a FutC que se denominó FutL. Se encontró que FutL dirigía la síntesis de 2'-FL a ~75 % el nivel de FutC en la cepa de producción de *E. coli* modificada mediante ingeniería metabólicamente (Tabla 1 y Figura 3). Además, los datos indicaron que FutL es significativamente menos eficaz para estimular la síntesis de LDFT, un producto secundario que se observó con otras  $\alpha(1,2)$ FT. Por lo tanto, FutL ofrece ventajas con respecto a las otras, por ejemplo, la capacidad de producir 2'-FL de manera resistente sin la preocupación de producir de forma simultánea otros oligosacáridos contaminantes no deseados. FutL se obtiene a partir de *Helicobacter mustelae* y es  
 10 idéntica en un 70 % a FutC a nivel del aminoácido.

15 La secuencia de aminoácidos de la proteína de la familia de glicosil transferasa *Bacteroides vulgatus* ATCC 8482 (FutN) se presenta a continuación (SEQ ID NO: 8; Número de Acceso en GenBank YP\_001300461 (GI: 150005717)). Esta enzima está codificada por el constructo de ácido nucleico usado para producir un oligosacárido fucosilado en una bacteria de acuerdo con la invención.

1 mrlkvtggl gnqmfyafy lrmkkyypkv ridlsdmmhy kvhygyemhr vfnlphtefc  
 61 inqplkkvie flffkkiyer kqapnslraf ekkyfwplly fkgfyqserf fadikdevre  
 121 sftfdknkan srslnmleil dkdenavslh irrgdylqpk hwattgsvqc lpyyqnaiae  
 181 msrrvaspsy yifsdiaaw kenlplqnav yidwntdeds wqdmmlmshc khhiicnstf  
 241 swwgawlhn mdktvivpsr wfqhseapdi yptgwikvpv s

20 La secuencia de aminoácidos de *Bacteroides ovatus* ATCC 8483 FutO se presenta a continuación (SEQ ID NO: 9; Número de Acceso en GenBank ZP\_02065239 (GI: 160884236)).

1 mkivnilggl gnqmfvyamy lalkeahpee eillcrrsyk gyplhngyel erifgveape  
 61 aalsqlarva ypffnyksq lmrhflplr ksmasgttqip fdysevtrnd nvyydgywqn  
 121 eknflsirdk vikaftfpef rdeknkalsd klksvktasc hirrgdylkd piygvnsdy  
 181 ytraitelnq svnpdmycif sddigwcken fkfligdkev vfvdnkgqe sfydmqlmsl  
 241 chyniianss fswwgawlhn nddkvvvape rwmnktlend picdnwkrik ve

25 La secuencia de aminoácidos de *Escherichia coli* O55:H7 (cepa CB9615) fucosiltransferasa (WbgN) se presenta a continuación (SEQ ID NO: 10; Número de Acceso en GenBank YP\_003500093 (GI: 291283275)).

1 msivvarlag glgnqmfqya kgyaesvern sslkldlrgy knytlhggfr ldklnidntf  
 61 vmskkemcif pnfivraink fpklslcskr feseqyskki ngsmkgsvef igfwqneryf  
 121 lehkeklrei ftpininlda kelsdvirct nsvsvhirrg dyvsnvealk ihglcteryy  
 181 idsiryker fnnlvffvfs ddiewckkyk neifsrddv kfiegtqev dmwlmsnaky  
 241 hiianssfsw wgawlknndl gitiaptpwf ereelnsfdp cpekvwriek

30 La secuencia de aminoácidos de *Bacteroides fragilis* (NCTC) 9343 alfa-1,2-fucosiltransferasa (Bft1) se presenta a continuación (SEQ ID NO: 11; Número de Acceso en GenBank CAH09369 (GI: 60494568)).

1 mffrcmkiv qiigglgnqm fqfafylalk ekyvvnkldt ssgaythng feldkvfhve  
 61 ylkasireri klsyqgseiw irvlrklkkr kkteyvepyl cfdenaaisls cdkyyigywq  
 121 sykyftniea airgqfhfsk vlsdknefik kqmqsnsvs lhvrlgdyn npaysnict  
 181 aaynkainii qskvsepkkf vfsddtvwck dhkipnchi idwnnkeesy wdmclmtyck  
 241 hniianssfs wwgawlntnp eriviapgkw inddrvqvsd iipsdwicv

La secuencia de aminoácidos de *Bacteroides fragilis* (NCTC) 9343 fucosil transferasa (Bft3/WcfB) se presenta a continuación (SEQ ID NO: 12; Número de Acceso en GenBank CAH06753 (GI: 60491992)).

1 mlyvilrgrl gnnlfqiata asltqnfifc tvnkqerqv llykdsffkn ikvmkgvdpd  
 61 ipyykepfhe fsripyeegk dliidgyfqs ekyfkrsvvl dlyritdelr kkiwnicgni  
 121 lekgetvsih vrrgdylklp halpfcgksy yknaiqyige dkifiicsdd idwckknfig  
 181 kryyfientt plldlyiqsl cthniisnss fswwgawlne nsnkiviapq mwfgisvklg  
 5 241 vsdllpvsww rlpnnytlgr ycfalykvve dyllnilrli wkrkknm

Matriz de comparación de homología de fucosiltransferasas examinada en este estudio:

	FutC	WbsJ	WbgL	WblA	WbgN	Bft1	Bft3	FutD	FutE	FutL	FutN	FutO
FutC	-	30 %	28 %	28 %	28 %	34 %	28 %	39 %	44 %	70 %	27 %	27 %
WbsJ	30 %	-	33 %	33 %	36 %	36 %	40 %	30 %	35 %	30 %	33 %	36 %
WbgL	28 %	33 %	-	33 %	37 %	32 %	39 %	31 %	32 %	25 %	32 %	33 %
WblA	28 %	33 %	33 %	-	36 %	37 %	33 %	31 %	38 %	29 %	31 %	35 %
WbgN	28 %	36 %	37 %	36 %	-	32 %	37 %	30 %	38 %	30 %	32 %	35 %
Bft1	34 %	36 %	32 %	37 %	32 %	-	30 %	32 %	37 %	33 %	35 %	38 %
Bft3	28 %	40 %	39 %	33 %	37 %	30 %	-	30 %	33 %	29 %	34 %	35 %
FutD	39 %	30 %	31 %	31 %	30 %	32 %	30 %	-	34 %	40 %	28 %	31 %
FutE	44 %	35 %	32 %	38 %	38 %	37 %	33 %	34 %	-	33 %	33 %	36 %
FutL	70 %	30 %	25 %	29 %	30 %	33 %	29 %	40 %	33 %	-	30 %	28 %
FutN	27 %	33 %	32 %	31 %	32 %	35 %	34 %	28 %	34 %	30 %	-	37 %
FutO	27 %	36 %	33 %	35 %	35 %	38 %	35 %	31 %	36 %	28 %	37 %	-

10 Todas estas proteínas se encuentran en bacterias que interactúan con el sistema gastrointestinal de organismos superiores. Además, 6 de las 10 seleccionadas incorporan fucosa en sus glicanos de la superficie celular. Se predijo que dichos genes tendrían la actividad más fuerte en términos de la síntesis de fucosil-oligosacáridos. En este grupo de 10 candidatos, también se incluyeron 2 enzimas encontradas en cepas bacterianas que no incorporan fucosa en glicanos de la superficie celular (WblA y WbgN). Se predijo que estos candidatos tendrían poca o ninguna actividad de síntesis de fucosil-oligosacáridos y, por lo tanto, podrían servir como un control negativo útil para validar el enfoque de selección.

20 Las  $\alpha(1,2)$  FT candidatas se clonaron mediante técnicas convencionales de biología molecular en un plásmido de expresión. Este plásmido utiliza el promotor fuerte hacia la izquierda del bacteriófago  $\lambda$  (denominado P<sub>L</sub>) para dirigir la expresión de los genes candidatos (Sanger, F. *et al.*, (1982). *J Mol Biol* 162, 729-773). El promotor es controlable, por ejemplo, un constructo trp-cl se integra de manera estable en el genoma del hospedador de *E. coli* (en el locus ampC), y el control se implementa agregando triptófano a los medios de crecimiento. La inducción gradual de la expresión de proteínas se lleva a cabo utilizando un represor cl sensible a la temperatura. Se ha descrito otra estrategia de control similar (sistema de expresión independiente de la temperatura) (Mieschendahl *et al.*, 1986, *Bio/Technology* 4: 802-808). El plásmido también transporta el gen *rcaA* de *E. coli* para regular de forma positiva la síntesis de GDP-fucosa, un precursor crítico para la síntesis de oligosacáridos unidos a fucosilo. Además, el plásmido lleva un gen de  $\beta$ -lactamasa (*bla*) para mantener el plásmido en las cepas hospedadoras mediante la selección de ampicilina (por conveniencia en el laboratorio) y un gen nativo de *thyA* (timidilato sintetasa) como un medio alternativo de selección en los hospedadores de *thyA*. Los marcadores seleccionables alternativos incluyen los genes proBA para complementar la auxotrofia de prolina (Stein *et al.*, (1984), *J Bacteriol* 158: 2, 696-700 (1984) o purC para complementar la auxotrofia de adenina (Parker, J., (1984), *J Bacteriol* 157: 3, 712-7). Para actuar como marcadores seleccionables de plásmidos, cada uno de estos genes se inactiva primero en el cromosoma de la célula hospedadora, a continuación se proporcionan copias de tipo silvestre de los genes en el plásmido. Como alternativa, se puede usar un gen de resistencia a fármacos en el plásmido, por ejemplo, beta-lactamasa (este gen ya se encuentra en el plásmido de expresión que se ha descrito anteriormente, lo que permite la selección con ampicilina). La selección con ampicilina se conoce bien en la técnica y se describe en manuales convencionales tales como Maniatis *et al.*, (1982) *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring, NY.

40 Los constructos de expresión se transformaron en una cepa hospedadora útil para la producción de 2'-FL. La biosíntesis de 2'-FL requiere la generación de un conjunto celular mejorado tanto de lactosa como de GDP-fucosa (Figura 2). La cepa prototrófica W3210 de *Eschericia coli* K12 de tipo silvestre se seleccionó como fondo principal

para someter a ensayo la capacidad de los candidatos para catalizar la producción de 2'-FL (Bachmann, B.J. (1972). *Bacteriol Rev* 36, 525-557). El derivado W3110 particular empleado fue uno que previamente se había modificado por la introducción (en el locus *ampC*) de un casete represor *P<sub>trpB</sub><sup>cl+</sup>* inducible por triptófano, generando una cepa de *E. coli* conocida como GI724 (LaVallie, E.R. *et al.*, (2000). *Methods Enzymol* 326, 322-340). Otras características de GI724 incluyen mutaciones de los promotores *lacIq* y *lacPL8*. La cepa GI724 de *E. coli* ofrece una producción económica de proteínas recombinantes del promotor del fago  $\lambda$  P<sub>L</sub> después de la inducción con bajos niveles de triptófano exógeno (LaVallie, E.R. *et al.*, (1993). *Biotechnology (N Y)* 11, 187-193; Mieschendahl, *et al.*, (1986). *Bio/Technology* 4, 802-08). Se hicieron alteraciones genéticas adicionales a esta cepa para estimular la biosíntesis de 2'-FL. Esto se consiguió en la cepa GI724 mediante varias manipulaciones del cromosoma usando la recombinación de  $\lambda$  Red (Court, D.L. *et al.*, (2002). *Annu Rev Genet* 36, 361-388) y la transducción de fagos P1 generalizada.

En primer lugar, la capacidad de la cepa hospedadora de *E. coli* para acumular lactosa intracelular se diseñó mediante la eliminación simultánea del gen endógeno de  $\beta$ -galactosidasa (*lacZ*) y el gen represor del operón de lactosa (*lacI*). Durante la construcción de esta deleción, el promotor *lacIq* se colocó inmediatamente cadena arriba del gen de la lactosa permeasa, *lacY*. La cepa modificada mantiene su capacidad para transportar lactosa desde el medio de cultivo (a través de LacY), pero se produce una deleción para la copia de tipo silvestre del gen *lacZ* ( $\beta$ -galactosidasa) responsable del catabolismo de la lactosa. Por lo tanto, se crea un grupo de lactosa intracelular cuando la cepa modificada se cultiva en presencia de lactosa exógena. Un esquema del constructo cromosómico *P<sub>lacIq</sub>lacY<sup>+</sup>* se muestra en la Figura 6.

La secuencia de ADN genómico del constructo cromosómico *P<sub>lacIq</sub>lacY<sup>+</sup>* se presenta a continuación (SEQ ID NO: 13) (el codón de inicio ATG está subrayado):

```

CACCATCGAATGGCGCAAACCTTTTCGCGGTATGGCATGATAGCGCCCGGAAGA
GAGTCAAGTGTAGGCTGGAGCTGCTTCGAAGTTCCTATACTTTCTAGAGAATAGG
AACTTCGGAATAGGAACTTCGGAATAGGAACTAAGGAGGATATTCATATGTACT
ATTTAAAAAACACAACTTTTGGATGTTTCGGTTTATTCTTTTTCTTTACTTTTTTA
TCATGGGAGCCTACTTCCCGTTTTTCCCGATTTGGCTACATGACATCAACCATATC
AGCAAAAGTGATACGGGTATTATTTTTGCCGCTATTTCTCTGTTCTCGCTATTATT
CCAACCGCTGTTTGGTCTGCTTTCTGACAACTCGGGCTGCGCAAATACCTGCTG
TGGATTATTACCGGCATGTTAGTGATGTTTGCGCCGTTCTTTATTTTTATCTTCGG
GCCACTGTTACAATAACAATTTTAGTAGGATCGATTGTTGGTGGTATTTATCTA
GGCTTTTGTTTTAAACGCCGGTGCGCCAGCAGTAGAGGCATTTATTGAGAAAGTCA
GCCGTCGCAGTAATTTCGAATTTGGTCGCGCGCGGATGTTTGGCTGTGTTGGCTG
GGCGCTGTGTGCCCTCGATTGTCCGCATCATGTTACCCATCAATAATCAGTTTGT
TCTGGCTGGGCTCTGGCTGTGCACTCATCCTCGCCGTTTTACTCTTTTTCGCCAAA
ACGGATGCGCCCTCTTCTGCCACGGTTGCCAATGCGGTAGGTGCCAACCATTTCGG
CATTTAGCCTTAAGCTGGCACTGGAAGTTCAGACAGCCAAAAGTGTGGTTTTT
GTCACTGTATGTTATTGGCGTTTCTTGCACCTACGATGTTTTTGACCAACAGTTTG
CTAATTTCTTTACTTCGTTCTTTGCTACCGGTGAACAGGGTACGCGGGTATTTGGC
TACGTAACGACAATGGGCGAATTACTTAACGCCTCGATTATGTTCTTTGCGCCAC
TGATCATTAAATCGCATCGGTGGGAAAACGCCCTGCTGCTGGCTGGCACTATTAT
GTCTGTACGTATTATTGGCTCATCGTTCGCCACCTCAGCGCTGGAAGTGGTTATTC
TGAAAACGCTGCATATGTTTGAAGTACCGTTCTGCTGGTGGGCTGCTTTAAATA
TATTACCAGCCAGTTTGAAGTGCCTTTTTTCAGCGACGATTTATCTGGTCTGTTTCT
GCTTCTTTAAGCAACTGGCGATGATTTTTATGTCTGTACTGGCGGGCAATATGTAT
GAAAGCATCGGTTTCCAGGGCGCTTATCTGGTGTGTTGCTGGTGGCGCTGGGCT
TCACCTTAATTTCCGTGTTACACGCTTAGCGGCCCGGCCCGCTTTCCCTGCTGCGT
CGTCAGGTGAATGAAGTCGCTTAAGCAATCAATGTCGGATGCGGCGCGAGCGCC
TTATCCGACCAACATATCATAACGGAGTGATCGCATTGTAAATTATAAAAATTGC
CTGATACGCTGCGCTTATCAGGCCTACAAGTTCAGCGATCTACATTAGCCGCATC
CGGCATGAACAAAGCGCAGGAACAAGCGTCGCA

```

En segundo lugar, la capacidad de la cepa hospedadora de *E. coli* para sintetizar el ácido colánico, un polisacárido

capsular extracelular, se eliminó mediante la delección del gen *wcaJ*, que codifica la transferasa vehículo lipídico de UDP-glucosa (Stevenson, G. *et al.*, (1996). J Bacteriol 178, 4885-893). En un fondo nulo de *wcaJ* la GDP-fucosa se acumula en el citoplasma de *E. coli* (Dumon, C. *et al.*, (2001). Glycoconj J 18, 465-474). Un esquema de la delección cromosómica de *wcaJ* se muestra en la Figura 7.

5 La secuencia de la región cromosómica que porta la mutación  $\Delta wcaJ::FRT$  se presenta a continuación (SEQ ID NO: 14) (la mutación está subrayada):

GTTCGGTTATATCAATGTCAAAAACCTCACGCCGCTCAAGCTGGTGATCAACTCC  
 GGAACGGCGCAGCGGGTCCGGTGGTGACGCCATTGAAGCCCGCTTTAAAGCC  
 CTCGGCGCGCCCGTGAATTAATCAAAGTGCACAACACGCCGGACGGCAATTC  
 CCCAACGGTATTCCTAACCCACTACTGCCGGAATGCCGCGACGACACCCGCAATG  
 CGGTCATCAAACACGGCGCGGATATGGGCATTGCTTTTGATGGCGATTTTGACCG  
 CTGTTTCCTGTTTGACGAAAAGGGCAGTTTATTGAGGGGCTACTACATTGTCCGGC  
 CTGTTGGCAGAAGCATTCCCTCGAAAAAATCCCGGCGCGAAGATCATCCACGAT  
 CCACGTCTCTCCTGGAACACCGTTGATGTGGTGACTGCCGCGAGGTGGCACGCCGG  
 TAATGTGCGAAAACCGGACACGCCTTTATTAAGAAGCGTATGCGCAAGGAAGACG  
 CCATCTATGGTGGCGAAATGAGCGCCACCATTACTTCCGTGATTTGCTTACTG  
 CGACAGCGGCATGATCCCGTGGCTGCTGGTCGCCGAACCTGGTGTGCCTGAAAGA  
 TAAAACGCTGGGCGAACTGGTACGCGACCGGATGGCGGGCCTTCCGGCAAGCGG  
 TGAGATCAACAGCAAACCTGGCGCAACCCGTTGAGGCGATTAACCGCGTGGAACA  
 GCATTTTAGCCGTGAGGCGCTGGCGGTGGATCGCACCGATGGCATCAGCATGAC  
 CTTTGCCGACTGGCGCTTTAACCTGCGCACCTCCAATACCGAACCGGTGGTGC  
 CTGAATGTGGAATCGCGCGGTGATGTGCCGCTGATGGAAGCGCGAACGCGAACT  
 CTGCTGACGTTGCTGAACGAGTAATGTCGGATCTTCCCTTACCCACTGCGGGTA  
 AGGGGCTAATAACAGGAACAACGATGATCCGGGGATCCGTCGACCTGCAGTTC  
GAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCGAAGCAGCTCCAGCCTACAGT  
 TAACAAAGCGGCATATTGATATGAGCTTACGTGAAAAAACCATCAGCGGCGCGA  
 AGTGGTCCGGCGATTGCCACGGTGATCATCATCGGCTTATTTTTGACACCAGACCA  
 ACTGGTAATTTATTTTTGACACCAGACCAACTGGTAATTTATTTTTGACACCAGAC  
 CAACTGGTTATTTTTGACACCAGACCAACTGGTTATTTTTGACACCAGACCAACT  
 GGCTCGGGCTGGTGCAGATGACCGTGCTGGCGCGGATTATCGACAACCACCAGT  
 TCGGCCTGCTTACCGTGTGCTGGTGATTATCGCGCTGGCAGATACGCTTTCTGA  
 CTTCGGTATCGCTAACTCGATTATTCAGCGAAAAGAAATCAGTCACCTTGA  
 ACTCACCAGTTGTACTGGCTGAACGTCCGGGCTGGGGATCGTGGTGTGCGTGGCGGTGT  
 TTTTGTGAGTGATCTCATCGGCGACGTGCTGAATAACCCGGACCTGGCACCAGT  
 GATTAACACATTATCGCTGGCGTTTGTGGTAATCCCCACGGGCAACAGTTCGCG  
 GCGTTGATGCAAAAAGAGCTGGAGTTCAACAAAATCGGCATGATCGAAACCAGC  
 GCGGTGCTGGCGGGCTTCACTTGTACGGTGGTTAGCGCCATTTCTGGCCGCTGG  
 CGATGACCGCGATCCTCGGTTATCTGGTCAATAGTGCGGTGAGAACGCTGCTGTT  
 TGGCTACTTTGGCCGCAAAATTTATCGCCCCGGTCTGCATTTCTCGCTGGCGTCCG  
 TGGCACCGAACTTACGCTTTGGTGCCTGGCTGACGGCGGACAGCATCATCAACTA  
 TCTCAATACCAACCTTTCAACGCTCGTGCTGGCGCGTATTCTCGGCGCGGGCGTG  
 GCAGGGGGATACAACCTGGCGTACAACGTGGCCGTTGTGCCACCGATGAAGCTG  
 AACCAATCATCACCCGCGTGTGTTTCCGGCATTCCGCAAAATTCAGGACGATA  
 CCGAAAAGCTGCGTGTTAACTTCTACAAGCTGCTGTCGGTAGTGGGGATTATCAA  
 CTTCCGGCGCTGCTCGGGCTAATGGTGGTGTGCAATAACTTTGTACCGCTGGTC  
 TTTGGTGAGAAGTGGAACAGCATTATCCGGTGTGCAATTGCTGTGTGGTGG  
 GTCTGCTGCGCTCCG

10 En tercer lugar, la magnitud de la combinación de GDP-fucosa citoplasmática aumentó mediante la introducción de una mutación nula en el gen *lon*. Lon es una proteasa intracelular dependiente de ATP que es responsable de la

degradación de RcsA, que es un regulador transcripcional positivo de la biosíntesis del ácido colánico en *E. coli* (Gottesman, S. y Stout, V. Mol Microbiol 5, 1599-1606 (1991)). En un fondo nulo de *lon*, RcsA se estabiliza, los niveles de RcsA aumentan, los genes responsables de la síntesis de GDP-fucosa en *E. coli* se regulan de forma positiva, y las concentraciones de GDP-fucosa intracelular aumentan. El gen *lon* sufrió una delección casi completa y se substituyó mediante gen *lacZ*<sup>+</sup> de *E. coli* de tipo silvestre funcional insertado, pero sin promotor ( $\Delta lon::(kan, lacZ^+)$ ). La recombinación de  $\lambda$  Red se usó para llevar a cabo la construcción. Un esquema de la inserción *kan, lacZ*<sup>+</sup> en el locus *lon* se muestra en la Figura 8.

5 La secuencia de ADN genómico que rodea la inserción de *lacZ*<sup>+</sup> en la región *lon* en la cepa de *E. coli* se presenta a continuación (SEQ ID NO: 15):

10  
 GTGGATGGAAGAGGTGGAAAAAGTGGTTATGGAGGAGTGGGTAATTGATGGTGA  
 AAGGAAAGGGTTGGTGATTTATGGGAAGGGGGAAGGGGAAGAGGGATGTGGTG  
 AATAATTAAGGATTGGGATAGAATTAGTTAAGGAAAAAGGGGGGATTTTATGTG  
 GGGTTTAATTTTTGGTGTATTGTGGGGGTTGAATGTGGGGGAAAGATGGGGATAT  
 AGTGAGGTAGATGTTAATAGATGGGGTGAAGGAGAGTGGTGTGATGTGATTAGG  
 TGGGGGAAATTAAGTAAGAGAGAGGTGTATGATTGGGGGGATGGGTGGAGGT  
 GGAGTTGGAAGTTGGTATTGTGTAGAAAGTATAGGAAGTTGAGAGGGGTTTTGA  
 AGGTGAGGGTGGGGGAAGGAGTGAGGGGGGAAGGGGTGGTAAAGGAAGGGGA  
 AGAGGTAGAAAGGGAGTGGGGAGAAAGGGTGGTGAGGGGGGATGAATGTGA  
 GGTAGTGGGGTATGTGGAGAAGGGAAAAGGGAAAGGGGAAAGAGAAAGGAGGTA  
 GGTGGAGTGGGGTTAGATGGGGATAGGTAGAGTGGGGGGTTTTATGGAGAGGA  
 AGGGAAGGGGAATTGGGAGGTGGGGGGGGGTGTGGTAAGGTTGGGAAGGGGTG  
 GAAAGTAAAGTGGATGGGTTTTGTTGGGGGGAAAGGATGTGATGGGGGAGGGGAT  
 GAAGATGTGATGAAGAGAGAGGATGAGGATGGTTTGGGATGATTGAAGAAGAT  
 GGATTGGAGGGAGGTTGTGGGGGGGGTGGGTGGAGAGGGTATTGGGGTATGAG  
 TGGGGAGAAGAGAGAATGGGGTGGTGTGATGGGGGGGTGTTGGGGGTGTGAGG  
 GGAGGGGGGGGGGGTGTTTTTGTGAAGAGGGAGGTGTGGGGTGGGGTGAATGA  
 AGTGGAGGAGGAGGGAGGGGGGGTATGGTGGGTGGGGAGGAGGGGGGTTGGTT  
 GGGGAGGTGTGGTGGAGGTTGTGAGTGAAGGGGGAAAGGGAGTGGGTGGTATTG  
 GGGGAAGTGGGGGGGGAGGATGTGGTGTGATGTGAGGTTGGTGGTGGGGAGAA  
 AGTATGGATGATGGGTGATGGAATGGGGGGGGTGGATAGGGTTGATGGGGGTAG  
 GTGGGGATTGGAGGAGGAAGGGAAAGATGGGATGGAGGGAGGAGGTAGTGGGA  
 TGGAAGGGGGTGTGTGGATGAGGATGATGTGGAGGAAGAGGATGAGGGGGTG  
 GGGGGAGGGGAAGTGTGGGGAGGGTGAAGGGGGGATGGGGGAGGGGGAGGAT  
 GTGGTGGTGAAGGATGGGGATGGGTGGTTGGGGAATATGATGGTGGAAAATGGG  
 GGGTTTTGTGGATTGATGGAGTGTGGGGGGGTGGGTGTGGGGGAGGGGTATGAG  
 GAGATAGGGTTGGGTAGGGGTGATATTGGTGAAGAGGTTGGGGGGGAATGGG  
 GTGAGGGGTTGGTGGTGGTTTAGGGTATGGGGGGTGGGGATTGGGAGGGGATGG  
 GGTTGTATGGGGTTGTTGAGGAGTTGTTGTAATAAGGGGATGTTGAAGTTGGTAT  
 TGGAAGTTGGTATTGTGTAGAAAGTATAGGAAGTTGGAAGGAGGTGGAGGGTA  
 GATAAAGGGGGGGGTTATTTTTGAGAGGAGAGGAAGTGGTAATGGTAGGGAGG  
 GGGGGTGAAGTGAATTGGGGGGATAGTGAGGGGGTGGAGGAGTGGTGGGGAG  
 GAATGGGGATATGGAAAGGGTGGATATTGAGGGATGTGGGTTGTTGGGGGTGGA  
 GGAGATGGGGATGGGTGGTTTGGATGAGTTGGTGTGAGTGTAGGGGGTGTGAT  
 TGAAGTGAAGTGGGGGGGGGAGTGGTGTGGGGGATAATTGAATTGGGGGGTG  
 GGGGAGGGGAGAGGGTTTTGGGTGGGGAAAGAGGTAGGGGGTATAGATGTTGAG

AATGGGAGATGGGAGGGGTGAAAAGAGGGGGGAGTAAGGGGGTGGGGATAGTT  
 TTGTTGGGGGGGTAATGGGAGGGAGTTTAGGGGGTGTGGTAGGTGGGGGAGGTG  
 GGAGTTGAGGGGAATGGGGGGGGATGGGGTGTATGGGTGGGGAGTTGAAGAT  
 GAAGGGTAATGGGGATTTGAGGAGTAGGATGAATGGGGTAGGTTTTGGGGGTGA  
 TAAATAAGGTTTTGGGGTGTGGTGGGAGGGGTGAGGGGTGGTAATGAGGAGGG  
 GATGAGGAAGTGTATGTGGGGTGGAGTGGAGAAGGGTGGTTGGGGGTGGTAAT  
 GGGGGGGGGGGTTGGAGGGTTGGAGGGAGGGGTTAGGGTGAATGGGGGTGGGT  
 TGAGTTAGGGGAATGTGGTTATGGAGGGGTGGAGGGGTGAAGTGTGGGGGAG  
 GGGGGTGAAGGAGTTGTTTTTATGGGGAATGGAGATGTGTGAAAGAAAGGGTGA  
 GTGGGGGTAAATTGGGAAGGGTTATTAGGGAGGTGGATGGAAAAATGGATTTG  
 GGTGGTGGTGAAGTGGGGGATGGGGTGGGAGGGGGGGGGGAGGGTGAAGTGA  
 GGTTTTGGGGGAGAGGGGAGTGGTGGGAGGGGGTGTATGTGGGGGGGTGTGAG  
 GATGGGGTGGGGTGGGTGGAGTAGGGGTAGTGTGAGGGAGAGTTGGGGGGG  
 GGTGTGGGGGTGGGGTAGTTGAGGGAGTTGAATGAAGTGTTAGGTTGTGGAGG  
 GAGATGGAGAGGGAGTTGAGGGGTGGGAGGGGGTTAGGATGGAGGGGGAGGA  
 TGGAGTGGAGGAGGTGGTTATGGGTATGAGGGAAGAGGTATTGGGTGGTGAAGT  
 GGATGGTTTGGGGGGATAAAGGGAAAGTGGAAAAAGTGGTGGTGGTGTTTTGGT  
 GGGTGAAGGGTGGATGGGGGGTGGGGTGGGAAAGAGGAGAGGGTTGATAGAG  
 AAGTGGGGATGGTTGGGGGTATGGGGAATGAGGGGGGTAAGGGGAGGAGGG  
 GTTGGGGTTTTGATGATATTAATGAGGGAGTGTGGAGGGAGTGGGAGAGGAA  
 GGGGGGGTGTAAAGGGGGATAGTGAAGAAAGGGGTGGGAGTATTTAGGGAAAG  
 GGGGAAGAGTGTAGGGATGGGGTGGGGTATTGGGAAAGGATGAGGGGGGGG  
 GTGTGTGGAGGTAGGGAAAGGGATTTTTGATGGAGGATTTGGGGAGAGGGGGG  
 AAGGGGTGGTGTGATGGAGGGGGGGGTAGATGGGGGAAATAATATGGGTGGG  
 GGTGGTGTGGGGTGGGGGGGGTTGATAGTGGAGGGGGGGGGAAGGATGGAGAG  
 ATTTGATGGAGGGATAGAGGGGGTGGTGAATTAGGGGGGTGGGGTGAATTGG  
 GGAGGGAGGAGATGATGAGAGTGGGGTGAATTAGGATGGGGGTGGAGGATTGGG  
 GTTAGGGGTGGGTGATGGGGGGTAGGGAGGGGGGATGATGGGTGAGAGGATT  
 GATTGGGAGGATGGGGTGGGTTTGAATATTGGGTGATGGAGGAGATAGAGGGG  
 GTAGGGGTGGGAGAGGGTGTAGGAGAGGGGATGGTTGGGATAATGGGAAGAGG  
 GGAGGGGGTTAAAGTTGTTGTGGTTGATGAGGAGGATATGGTGGAGGATGGTGT  
 GGTGATGGATGAGGTGAGGATGGAGAGGATGATGGTGGTGAAGGGTTAAGGGGT  
 GGAATGAGGAAGGGGTGGGGTTGAGGAGGAGGAGAGGATTTTGAATGGGGAG  
 GTGGGGGAAAGGGAGATGGGAGGGTTGTGTTGAATGAGGGTGGGGTGGGGGG  
 TGTGGAGTTGAAGGAGGGGAGGATAGAGATTGGGGATTTGGGGGGTGGAGAGTT  
 TGGGGTTTTGGAGGTTGAGAGGTAGTGTGAGGGGATGGGGATAAGGAGGAGGGT  
 GATGGATAATTTGAGGGGGGAAAGGGGGGGTGGGGGTGGGGAGGTGGGTTTGA  
 GGGTGGGATAAAGAAAGTGTAGGGGTAGGTAGTGAAGGAAGTGGGGGGGAGAT  
 GTGAAGTTGAGGGTGGAGTAGAGGGGGGGTGAATGATGATTAAGGGAGTGG  
 GAAGATGGAAATGGGTGATTTGTGTAGTGGTTTTATGGAGGAAGGAGAGGTGAG  
 GGAAAAATGGGGGTGATGGGGGAGATATGGTGTGTTGGAGATAAGTGGGGTGA  
 GTGGAGGGGAGGAGGATGAGGGGGAGGGGGTTTTGTGGGGGGGGTAAAAATGG  
 GGTGAGGTGAAATTGAGAGGGGAAAGGAGTGTGGTGGGGGTAAAGGAGGGAGG  
 GGGGGTTGGAGGAGAGATGAAAGGGGGAGTTAAGGGGATGAAAAATAATTGGG  
 GTGTGGGGTTGGTGTAGGGAGGTTTGTGAAGATTAAATGTGAGGGAGTAAGAA  
 GGGGTGGGATTGTGGGTGGGAAGAAAGGGGGGATTGAGGGTAATGGGATAGGT  
 GAGGTTGGTGTAGATGGGGGGATGGTAAGGGTGGATGTGGGAGTTTGAAGGGGAG  
 GAGGAGAGTATGGGGGTGAGGAAGATGGGAGGGAGGGAGGTTTGGGGGAGGGG  
 TTGTGGTGGGGGAAAGGAGGGAAAGGGGGATTGGGGATTGAGGGTGGGGAAAGT  
 GTTGGGAAGGGGGATGGGTGGGGGGGTGTTGGGTATTAGGGGAGGTGGGGAAA  
 GGGGGATGTGGTGGAAAGGGGATTAAGTTGGGTAAAGGGAGGGTTTTGGGAGTGA

GGAGGTTGTAAAAGGAGGGGGAGTGAATGGGTAATGATGGTGATAGTAGGTTTG  
 GTGAGGTTGTGAGTGGAAAATAGTGAGGTGGGGGAAAATGGAGTAATAAAAAG  
 AGGGGTGGGAGGGTAATTGGGGGTTGGGAGGGTTTTTTTTGTGTGGGTAAGTTAG  
 ATGGGGGATGGGGGTTGGGGTTATTAAGGGGTGTTGTAAGGGGATGGGTGGGGT  
 GATATAAGTGGTGGGGGTTGGTAGGTTGAAGGATTGAAGTGGGATATAAATTAT  
 AAAGAGGAAGAGAAGAGTGAATAAATGTGAATTGATGGAGAAGATTGGTGGAG  
 GGGGTGATATGTGTAAGGTGGGGGTGGGGGTGGGTTAGATGGTATTATTGGTT  
 GGGTAAGTGAATGTGTGAAAGAAGG

- 5 Preferentemente, la secuencia de ADN genómico que rodea la inserción *lacZ*<sup>+</sup> en la región *lon* en la cepa de *E. coli* es la que se presenta a continuación (SEQ ID NO: 19) (el sitio de unión al ribosoma se designan en letra minúscula, el codón complementario al codón *lacZ* de inicio ATG en la posición 5', y el codón complementario al codón de PARADA está en letra negra y subrayado):

GTCCATGGAAGACGTCGAAAAAGTGGTTATCGACGAGTCGGTAATTGATGGTCA  
 AAGCAAACCGTTGCTGATTTATGGCAAGCCGGAAGCGCAACAGGCATCTGGTGA  
 ATAATTAACCATCCCATACAATTAGTTAACCAAAAAGGGGGGATTTTATCTCCC  
 CTTTAATTTTTCTCTATTCTCGGCGTTGAATGTGGGGGAAACATCCCCATATACT  
 GACGTACATGTTAATAGATGGCGTGAAGCACAGTCGTGTCATCTGATTACCTGGC  
 GAAATTAACCTAAGAGAGAGCTCTATGATCCGGGGATCCGTCGACCTGCAGT  
 TCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCAGAGCGCTTTTGAAGCTCA  
 CGCTGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAG  
 AAAGCCAGTCCGCAGAAACGGTGTGACCCCGGATGAATGTCAGCTACTGGGCT  
 ATCTGGACAAGGGAAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGG  
 GCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACAGCAAGCGAACC GGA  
 ATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTG  
 GATGGCTTTCTTGCCGCCAAGGATCTGATGGCGCAGGGGATCAAGATCTGATCAA  
 GAGACAGGATGAGGATCGTTTTCGCATGATTGAACAAGATGGATTGCACGCAGGT  
 TCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACA  
 ATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGGCGCCCGGTTT  
 TTTTTGTCAAGACCGACCTGTCCGGTGCCTGAATGAACTGCAGGACGAGGCAGC  
 GCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTT  
 GTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGAT  
 CTCCTGTCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAA  
 TGGCGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAA  
 ACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGA  
 TGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCCGCCAGGCT  
 CAAGGCGCGCATGCCCGACGGCGAGGATCTCGTTCGTGACCCATGGCGATGCCTG  
 CTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTTCTGGATTCATCGACTGTGGC  
 CGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTG  
 CTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCTCGTGCTTTACGGTATCGC  
 CGCTCCCGATTCCGACGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTAAT  
 AAGGGGATCTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAA  
 CTTTCGAAGCAGCTCCAGCCTACATAAGCGGGCCGCT**TTA**TTTTTTGACACCAGACCAA  
 CTGGTAATGGTAGCGACCGGCGCTCAGCTGGAATTCCGCCGATACTGACGGGCTC  
 CAGGAGTCGTGCCACCAATCCCATATGGAAACCGTCGATATTCAGCCATGTGC  
 CTTCTTCCGCGTGCAGCAGATGGCGATGGCTGGTTCCATCAGTTGCTGTTGACT  
 GTAGCGGCTGATGTTGAACTGGAAGTCGCCGCGCCACTGGTGTGGGCCATAATTC  
 AATTCGCGCGTCCCGCAGCGCAGACCGTTTTTCGCTCGGGAAGACGTACGGGGTAT

ACATGTCTGACAATGGCAGATCCCAGCGGTCAAACAGGCGGCAGTAAGGCGGT  
 CGGGATAGTTTTCTTGC GGCCCTAATCCGAGCCAGTTTACCCGCTCTGCTACCTGC  
 GCCAGCTGGCAGTTCAGGCCAATCCGCGCCGGATGCGGTGTATCGCTCGCCACTT  
 CAACATCAACGGTAATCGCCATTTGACCACTACCATCAATCCGGTAGGTTTTCCG  
 GCTGATAAATAAGGTTTTCCCTGATGCTGCCACGCGTGAGCGGTGTAATCAGC  
 ACCGCATCAGCAAGTGTATCTGCCGTGCACTGCAACAACGCTGCTTCGGCCTGGT  
 AATGGCCCCGCCGCTTCCAGCGTTCGACCCAGGCGTTAGGGTCAATGCGGGTTCG  
 TTCCTTACGCCAATGTCGTTATCCAGCGGTGCACGGGTGAACTGATCGCGCAGC  
 GGCGTCAGCAGTTGTTTTTATCGCCAATCCACATCTGTGAAAGAAAGCCTGACT  
 GGCGGTTAAATTGCCAACGCTTATTACCCAGCTCGATGCAAAAATCCATTTTCGCT  
 GGTGGTCAGATGCGGGATGGCGTGGGACGCGGGCGGGGAGCGTCACACTGAGGTT  
 TTCCGCCAGACGCCACTGCTGCCAGGCGCTGATGTGCCCGGCTTCTGACCATGCG  
 GTCGCGTTCGGTTGCACTACGCGTACTGTGAGCCAGAGTTGCCCGGCGCTCTCCG  
 GCTGCGGTAGTTCAGGCAGTTCAATCAACTGTTTACCTTGTGGAGCGACATCCAG  
 AGGCACTTCACCGCTTGCCAGCGGCTTACCATCCAGCGCCACCATCCAGTGCAGG  
 AGCTCGTTATCGCTATGACGGAACAGGTATTCGCTGGTCACTTCGATGGTTTGCC  
 CGGATAAACGGAAGTGGAAAACTGCTGCTGGTGTGTTTTGCTTCCGTCAGCGCTGG  
 ATGCGGCGTGC GGTCGGCAAAGACCCAGACCGTTCATACAGAAGTGGCGATCGTT  
 CGGCGTATCGCCAAAATCACCGCCGTAAGCCGACCACGGGTTGCCGTTTTCATCA  
 TATTTAATCAGCGACTGATCCACCCAGTCCCAGACGAAGCCGCCCTGTAAACGGG  
 GATACTGACGAAACGCCTGCCAGTATTTAGCGAAACCGCCAAGACTGTTACCCAT  
 CGCGTGGGCGTATTCGCAAAGGATCAGCGGGCGCGTCTCTCCAGGTAGCGAAAG  
 CCATTTTTTATGATGGACCATTTCCGGCACAGCCGGGAAGGGCTGGTCTTCATCCAG  
 CGCGCGTACATCGGGCAAATAATATCGGTGGCCGTGGTGTCCGGCTCCGCCGCTT  
 CATACTGCACCGGGCGGGAAGGATCGACAGATTTGATCCAGCGATAACAGCGCGT  
 CGTGATTAGCGCCGTGGCCTGATTCATTCCCCAGCGACCAGATGATCACACTCGG  
 GTGATTACGATCGCGCTGCACCATTCCGCGTTACGCGTTCGCTCATCGCCGGTAGC  
 CAGCGCGGATCATCGGTCAGACGATTCATTGGCACCATGCCGTGGGTTTTCAATAT  
 TGGCTTCATCCACCACATACAGGCCGTAGCGGTCGCACAGCGGTGTACCACAGCG  
 GATGGTTCCGATAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCAG  
 CAGGATATCCTGCACCATCGTCTGCTCATCCATGACCTGACCATGCAGAGGATGA  
 TGCTCGTGACGGTTAACGCCTCGAATCAGCAACGGCTTGCCGTTACGACAGCAGCA  
 GACCATTTTCAATCCGCACCTCGCGGAAACCGACATCGCAGGCTTCTGCTTCAAT  
 CAGCGTGCCGTCCGGCGGTGTGCAGTTCAACCACCGCACGATAGAGATTCGGGAT  
 TTCGGCGCTCCACAGTTTTCCGGTTTTTCGACGTTTCAGACGTAGTGTGACGCGATCG  
 GCATAACCACCACGCTCATCGATAATTTACCGCCGAAAGGCGCGGTGCCGCTG  
 GCGACCTGCGTTTTACCCTGCCATAAAGAAACTGTTACCCGTAGGTAGTCACGCA  
 ACTCGCCGCACATCTGAACTTCAGCCTCCAGTACAGCGCGGCTGAAATCATCATT  
 AAAGCGAGTGGCAACATGGAAATCGCTGATTTGTGTAGTCGGTTTTATGCAGCAA  
 CGAGACGTCACGGAAAATGCCGCTCATCCGCCACATATCCTGATCTTCCAGATAA  
 CTGCCGTCACCTCAGCGCAGCACCATCACCGCGAGGCGGTTTTCTCCGGCGCGTA  
 AAAATGCGCTCAGGTCAAATTCAGACGGCAAACGACTGTCTTGCCGTAACCGA  
 CCCAGCGCCCGTTGCACCACAGATGAAACGCCGAGTTAACGCCATCAAAAATAA  
 TTCGCGTCTGGCCTTCCCTGTAGCCAGCTTTCATCAACATTAATGTGAGCGAGTA  
 ACAACCCGTCGGATTCTCCGTGGGAACAAACGGCGGATTGACCGTAATGGGATA  
 GGTCACGTTGGTGTAGATGGGCGCATCGTAACCGTGCATCTGCCAGTTTGAGGGG  
 ACGACGACAGTATCGGCCTCAGGAAGATCGCACTCCAGCCAGCTTTCGGGCACC  
 GCTTCTGGTGCCGGAAACCAGGCAAAGCGCCATTCCGCCATTACGGCTGCGCAACT  
 GTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAG  
 GGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTAC  
 GACGTTGTAAAACGACGGCCAGTGAATCCGTAATCATGGTCAtagtagtttctCAGGT

TGTGACTGCAAAATAGTGACCTCGCGCAAAATGCACTAATAAAAACAGGGCTGG  
 CAGGCTAATTCGGGCTTGCCAGCCTTTTTTTGTCTCGCTAAGTTAGATGGCGGATC  
 GGGCTTGCCCTTATTAAGGGGTGTTGTAAGGGGATGGCTGGCCTGATATAACTGC  
 TGCGCGTTCGTACCTTGAAGGATTCAAGTGCATATAAATTATAAAGAGGAAGA  
 GAAGAGTGAATAAATCTCAATTGATCGACAAGATTGCTGCAGGGGCTGATATCT  
 CTAAGCTGCGGCTGGCCGTGCGTTAGATGCTATTATTGCTTCCGTAACCTGAATC  
 TCTGAAAGAAGG

La secuencia de ácidos nucleicos que codifica la proteína de *lacZ* se presenta a continuación (SEQ ID NO: 20) (el codón de parada ATG está subrayado y el codón de PARADA está en letra negra y subrayado):

5

ATGACCATGATTACGGATTCACTGGCCGTGCTTTTTACAACGTCGTGACTGGGAAA  
 ACCCTGGCGTTACCCAACCTAATCGCCTTGACAGCACATCCCCCTTCGCCAGCTG  
 GCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCT  
 GAATGGCGAATGGCGCTTTGCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAG  
 CTGGCTGGAGTGCATCTTCCCTGAGGCCGATACTGTCGTCGTCCTCCCTCAAACCTGG  
 CAGATGCACGGTTACGATGCGCCCATCTACACCAACGTGACCTATCCCATACGG  
 TCAATCCGCCGTTTGTTCACCGGAGAATCCGACGGGTTGTTACTCGCTCACATTT  
 AATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGAATTATTTTTGATGGC  
 GTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGTTACGGCCAGG  
 ACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGAGAAAA  
 CCGCCTCGCGGTGATGGTGCTGCGCTGGAGTGACGGCAGTTATCTGGAAGATCA  
 GGATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCG  
 ACTACACAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTACGCC  
 GCGCTGTACTGGAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACG  
 GGTAACAGTTTCTTTATGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCC  
 TTTCCGGCGGTGAAATTATCGATGAGCGTGGTGGTTATGCCGATCGCGTCACACTA  
 CGTCTGAACGTCGAAAACCCGAAACTGTGGAGCGCCGAAATCCCGAATCTCTAT  
 CGTGCGGTGGTTGAACTGCACACCGCCGACGGCACGCTGATTGAAGCAGAAGCC  
 TGCATGTTCGTTTCCGCGAGGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACG  
 GCAAGCCGTTGCTGATTCGAGGCGTTAACCCTCACGAGCATCATCTCTGCATGG  
 TCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAGCAGAA  
 CAACTTTAACGCCGTGCGCTGTTTCGCATTATCCGAACCATCCGCTGTGGTACACG  
 CTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACG  
 GCATGGTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGA  
 GCGAACGCGTAACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCA  
 TCTGGTCGCTGGGGAATGAATCAGGCCACGGCGCTAATCACGACGCGCTGTATC  
 GCTGGATCAAATCTGTCGATCCTTCCC GCCCGGTGCAGTATGAAGGCGGCGGAGC  
 CGACACCACGGCCACCGATATTATTTGCCCGATGTACGCGCGCGTGGATGAAGA  
 CCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATCAAAAAATGGCTTTCGCTACCT  
 GGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCACGCGATGGGTAACAGT  
 CTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAGG  
 GCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAATATGATGAAAACG  
 GCAACCCGTGGTTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCA  
 GTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACG  
 GAAGCAAAACACCAGCAGCAGTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCG  
 AAGTGACCAGCGAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGAT  
 GGTGGCGCTGGATGGTAAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGC  
 TCCACAAGGTAAACAGTTGATTGAACTGCCTGAACTACCGCAGCCGGAGAGCGC

CGGGCAACTCTGGCTCACAGTACGCGTAGTGCAACCGAACGCGACCCGCATGGTC  
 AGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGGCGTCTGGCGGAAAACCTCAG  
 TGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTGACCACCAGCGAAATG  
 GATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGCCAGTCAGGCT  
 TTCTTTCACAGATGTGGATTGGCGATAAAAAACAACCTGCTGACGCCGCTGCGCGA  
 TCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCG  
 CATTGACCCTAACGCCTGGGTGGAACGCTGGAAGGCGGCGGGCCATTACCAGGC  
 CGAAGCAGCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATT  
 ACGACCGCTCACGCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAA  
 ACCTACCGGATTGATGGTAGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGG  
 CGAGCGATACACCGCATCCGGCGCGGATTGGCCTGAACTGCCAGCTGGCGCAGG  
 TAGCAGAGCGGGTAAACTGGCTCGGATTAGGGCCGCAAGAAAACCTATCCCGACC  
 GCCTTACTGCCGCTGTTTTGACCGCTGGGATCTGCCATTGTCAGACATGTATAACC  
 CCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGGACGCGCGAATTGAATT  
 ATGGCCCACACCAGTGGCGCGGGGACTTCCAGTTC AACATCAGCCGCTACAGTCA  
 ACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAAGGCAC  
 ATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGC  
 CCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTTCGCTACCATTACCAGTTGG  
 TCTGGTGTCAAAAATTAA

En cuarto lugar, se introdujo una mutación *thyA* (timidilato sintetasa) en la cepa por transducción P1. En ausencia de timidina exógena, las cepas *thyA* son incapaces de producir ADN y mueren. El defecto se puede complementar en *trans* suministrando un gen *thyA* de tipo silvestre en un plásmido de múltiples copias (Belfort, M., Maley, G.F., y Maley, F. (1983). Proc Natl Acad Sci U S A 80, 1858-861). Esta complementación se usó aquí como un medio de mantenimiento del plásmido.

Una modificación adicional que es útil para aumentar el grupo citoplásmico de lactosa libre (y, por lo tanto, el rendimiento final de 2'-FL) es la incorporación de una mutación *lacA*. *LacA* es una lactosa acetiltransferasa que solo es activa cuando se acumulan altos niveles de lactosa en el citoplasma de *E. coli*. La alta osmolaridad intracelular (por ejemplo, causada por una alta acumulación de lactosa intracelular) puede inhibir el crecimiento bacteriano, y *E. coli* ha desarrollado un mecanismo para protegerse de la alta osmolaridad intracelular causada por la lactosa al "etiquetar" el exceso de lactosa intracelular con un grupo acetilo utilizando *LacA*, y luego expulsando activamente la acetil-lactosa de la célula (Danchin, A. Bioessays 31, 769-773 (2009)). La producción de acetil-lactosa en *E. coli* modificada por ingeniería para producir 2'-FL u otros oligosacáridos de leche humana es, por lo tanto, indeseable: reduce el rendimiento general. Además, la acetil-lactosa es un producto secundario que complica los esquemas de purificación de oligosacáridos. La incorporación de una mutación *lacA* resuelve estos problemas. La producción subóptima de oligosacáridos fucosilados se produce en cepas que carecen de una o ambas mutaciones en la vía del ácido colánico y la proteasa *lon*. El desvío de lactosa a un producto secundario (acetil-lactosa) se produce en cepas que no contienen la mutación *lacA*. Un esquema de la delección de *lacA* y la secuencia genómica correspondiente se ha proporcionado anteriormente (SEQ ID NO: 13).

La cepa usada para someter a ensayo los diferentes candidatos de  $\alpha(1,2)$  FT incorpora todas las modificaciones genéticas mencionadas anteriormente y tiene el siguiente genotipo:  
 $\Delta amp^C::P_{trp}^{Bcl}$ ,  $\Delta(lacI-lacZ)::FRT$ ,  $P_{lacIq}lacY^+$ ,  $\Delta wcaJ::FRT$ , *thyA*::Tn10,  $\Delta lon:(npt3, lacZ^+)$ ,  $\Delta lacA$ .

Las cepas de *E. coli* que albergan los diferentes plásmidos de expresión candidatos a  $\alpha(1,2)$  FT se analizaron. Las cepas se cultivaron en medios selectivos (sin timidina) hasta la fase exponencial temprana. A continuación se añadió lactosa a una concentración final de un 1 %, y se añadió triptófano (200  $\mu$ M) para inducir la expresión de cada candidata  $\alpha(1,2)$  FT del promotor  $P_L$ . Al final del periodo de inducción (~20 h) se recolectaron 600 unidades de DO equivalentes de cada cepa. Los lisados se prepararon y analizaron para detectar la presencia de 2'-FL mediante cromatografía en capa fina (TLC). Como se muestra en la Figura 3A, una cepa de control que produce FutC-Myc fue eficaz en la biosíntesis de 2'-FL y también produjo una menor cantidad de tetrasacárido lactodifucotetraosa (LDFT). El WbsJ de  $\alpha(1,2)$  FT de *E. coli* O128:B12 caracterizado previamente también fue capaz de catalizar la síntesis de 2'-FL, aunque solo al ~30 % del nivel producido por FutC-Myc (Figura 3A, calles 5 y 6). WbIA (obtenido a partir de *V. cholerae* O22) fue capaz de estimular la síntesis de 2'-FL, aunque a un nivel significativamente menor en comparación con FutC (Figura 3A, calles 7 y 8). Este resultado no fue inesperado, ya que *V. cholerae* O22 no incorpora fucosa en los glicanos de la superficie celular (Cox, A.D. et al., (1997). Carbohydr Res 304, 191-208). La cepa que produce WbgL (obtenido a partir de la cepa 0126 de *E. coli*) del plásmido pG204 sintetizó una cantidad

- significativa de 2'-FL, aproximadamente ~75 % de la cantidad producida por FutC-Myc (Figura 3A, calles 9 y 10). WbgL también fue capaz de sintetizar LDFT. La cepa que produce FutL (obtenido a partir de *H. mustelae* ATCC 43772) del plásmido pG216 fue capaz de dirigir la síntesis de cantidades robustas de 2'-FL, comparables a los niveles obtenidos utilizando FutC-Myc y WbgL (Figura 3B, calles 7 y 8). Además, una cepa que produce FutN (obtenido a partir de *B. vulgatus* ATCC 8482) del plásmido pG217 también produjo cantidades significativas de 2'-FL, aproximadamente ~50 % de la cantidad producida por FutC-Myc (Figura 3C, calles 5 y 6). FutN se obtiene a partir de la bacteria comensal *B. vulgatus*, y, por lo tanto, no puede estar sujeto a las mismas preocupaciones asociadas con la utilización de una  $\alpha(1,2)$  FT obtenida a partir de una bacteria patógena para la producción de un aditivo alimentario.
- 10 Un mapa del plásmido pG204 se muestra en la Figura 11. La secuencia del plásmido pG204 se presenta a continuación (SEQ ID NO: 16) (el codón complementario del codón de inicio ATG para *O126 WbgL* de *E. coli* está subrayado; y el codón complementario al codón de PARADA está subrayado y en negrita):

AATTCTAAAAATTGATTGAATGTATGCAAATAAATGCATACACCATAGGTGTGGT  
 TTAATTTGATGCCCTTTTTTCAGGGCTGGAATGTGTAAGAGCGGGGTTATTTATGCT  
 GTTGTTTTTTTGTTACTCGGGAAGGGCTTTACCTCTTCCGCATAAACGCTTCCATC  
 AGCGTTTATAGTTAAAAAAATCTTTCGGAAGTGGTTTTGCGCTTACCCCAACCAA  
 CAGGGGATTTGCTGCTTTCCATTGAGCCTGTTTCTCTGCGCGACGTTTCGCGGCGG  
 CGTGTGTGTGCATCCATCTGGATTCTCCTGTGAGTTAGCTTTGGTGGTGTGTGGCA  
 GTTGTAGTCCTGAACGAAAACCCCCGCGATTGGCACATTGGCAGCTAATCCGGA  
 ATCGCACTTACGGCCAATGCTTCGTTTCGTATCACACACCCCAAAGCCTTCTGCTT  
 TGAATGCTGCCCTTCTTCAGGGCTTAATTTTTAAGAGCGTCACCTTCATGGTGGTC  
 AGTGCCTCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTTATGTCAACACC  
 GCCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGTTTTTATATGAATTTA  
 TTTTTTGCAGGGGGGCATTGTTTGGTAGGTGAGAGATCAATTCTGCATTAATGAA  
 TCGGCCAACGCGCGGGGAGAGGGCGTTTGCATATTGGGCGCTCTTCCGCTTCCCTC  
 GCTCACTGACTCGCTGCGCTCGGTTCGGCTGCGGCGAGCGGTATCAGCTCAC  
 TCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC  
 ATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCT  
 GCGTTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCTCA  
 AGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCT  
 GGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTC  
 CGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATC  
 TCAGTTCGGTGTAGGTTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGT  
 TCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTA  
 AGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCG  
 AGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACA  
 CTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAA  
 AAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTT  
 TTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCT  
 TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTAAGGGA  
 TTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTTAAATTA  
 ATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTAC  
 CAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTCCATCCAT  
 AGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCT  
 GGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTAT  
 CAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTT  
 TATCCGCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTC  
 GCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCA  
 CGCTCGTCTGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAG  
 TTACATGATCCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGAT  
 CGTTGTGAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTG  
 CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGA  
 CTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCCG  
 GCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATC  
 ATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAAGGATCTTACCGCTGTTGAGAT  
 CCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTC  
 ACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAAATGCCGCAAAAAGGG  
 AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTAT  
 TGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT  
 AGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTG  
 ACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAGGCGTATCAC  
 GAGGCCCTTTCGTCTCGCGCGTTTCGGTGTGACGGTGAAAACCTCTGACACATG  
 CAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAA

GCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGCGGGGCTGGCTTAACTAT  
 GCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATATGCGGTGTGAAATAC  
 CGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCTCCTCAACCTGTATA  
 TTCGTA AACACGCCAATGGGAGCTGTCTCAGGTTTGTTCCTGATTGGTTACGG  
 CGCGTTTCGCATCATTGTTGAGTTTTTCCGCCAGCCCAGCGCAGTTTACCGGTG  
 CCTGGGTGCAGTACATCAGCATGGGGCAAATTCTTTCCATCCCGATGATTGTCGC  
 GGGTGTGATCATGATGGTCTGGGCATATCGTCGCAGCCCACAGCAACACGTTTCC  
 TGAGGAACCATGAAACAGTATTTAGA ACTGATGCAAAAAGTGCTCGACGAAGGC  
 ACACAGAAAACGACCGTACCGGAACCGGAACGCTTTCCATTTTTGGTCATCAG  
 ATGCGTTTTAACCTGCAAGATGGATTCCCCTGCTGGTGACA ACTAAACGTTGCCACC  
 TGCGTTCCATCATCCATGAACTGCTGTGGTTTTCTGCAGGGCGACACTAACATTGC  
 TTATCTACACGAAAACAATGTCACCATCTGGGACGAATGGGCCGATGAAAACGG  
 CGACCTCGGGCCAGTGTATGGTAAACAGTGGCGCGCCTGGCCAACGCCAGATGG  
 TCGTCATATTGACCAGATCACTACGGTACTGAACCAGCTGAAAAACGACCCGGA  
 TTCGCGCCGCATTATTGTTTCAGCGTGGAACGTAGGCGAACTGGATAAAATGGCG  
 CTGGCACCGTGCCATGCATTCTTCCAGTTCTATGTGGCAGACGGCAA ACTCTCTT  
 GCCAGCTTTATCAGCGCTCCTGTGACGTCTTCCCTCGGCCTGCCGTTCAACATTGCC  
 AGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCATCTGGAAGTGGGT  
 GATTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATATGGATCAA  
 CTCATCTGCAATTAAGCCGCGAACCGCGTCCGCTGCCGAAGTTGATTATCAAACG  
 TAAACCCGAATCCATCTTCGACTACCGTTTTCGAAGACTTTGAGATTGAAGGCTAC  
 GATCCGCATCCGGGCATTAAAGCGCCGGTGGCTATCTAATTACGAAACATCCTGC  
 CAGAGCCGACGCCAGTGTGCGTCGGTTTTTTTACCCTCCGTTAAATTCTTCGAGA  
 CGCCTTCCCGAAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCG  
 ATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCA  
 AGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTACGACGTTGTAAAACG  
 ACGGCCAGTGCCAAGCTTTCTTTAATGAAGCAGGGCATCAGGACGGTATCTTTGT  
 GGAGAAAGCAGAGTAATCTTATTCAGCCTGACTGGTGGGAAACCACCAGTCAGA  
 ATGTGTTAGCGCATGTTGACAAAAATACCATTAGTCACATTATCCGTCAGTCGGA  
 CGACATGGTAGATAACCTGTTTATTATGCGTTTTGATCTTACGTTAATATTACCT  
 TTATGCGATGAAACGGTCTTGGCTTTGATATTCATTTGGTCAGAGATTTGAATGG  
 TTCCCTGACCTGCCATCCACATTCGCAACATACTCGATTCCGGTTCGGCTCAATGAT  
 AACGTCGGCATAATTTAAAACGAGGTTATCGTTGTCTCTTTTTTTCAGAATATCGCC  
 AAGGATATCGTCGAGAGATTCCGGTTAATCGATTTAGA ACTGATCAATAAATTT  
 TTTCTGACCAATAGATATTCATCAAAATGAACATTGGCAATTGCCATAAAAACGA  
 TAAATAACGTATTGGGATGTTGATTAATGATGAGCTTGATACGCTGACTGTTAGA  
 AGCATCGTGGATGAAACAGTCCTCATTAATAAACACCACTGAAGGGCGCTGTGA  
 ATCACAAGCTATGGCAAGGTCATCAACGGTTTTCAATGTCGTTGATTTCTTTTTT  
 TAACCCCTCTACTCAACAGATACCCGGTTAAACCTAGTCGGGTGTA ACTACATAA  
 ATCCATAATAATCGTTGACATGGCATAACCTCACTCAATGCGTAACGATAATTCC  
 CCTTACCTGAATATTTTCATCATGACTAAACGGAACAACATGGGTACCTAATGCC  
 CCACTCTCGCATTTTTTTCAGGCGGACTTACTATCCCGTAAAGTGTGTATAATTTG  
 CCTGGAATTTGCTTAAAGTAAAGTAAATGTTGCGATATGTGAGTGAGCTTAAAAC  
 AAATATTTTCGCTGCAGGAGTATCCTGGAAGATGTTTCGTgAGAAGCTTACTGCTCA  
 CAAGAAAAAAGGCACGTCATCTGACGTGCCTTTTTTTATTTGACTACCCTGTACG  
 ATTACTGCAGCTCGAGCTAACACGAGCTATGTTTATCCACGTTTATCCAGTGATT  
 GACTATGGGGATATAAGTATTTTTTGGAGTTATATCGTACCAAGGAGTAGGATAA  
 ATAACAATCTGTGACGCTGATGTACCTAAATAAGCCCCCACC AACTAAA ACTAC  
 TATTCGCTATAATATGATGGTTAGCTAAGCTCATTAAACATAAATCTTCTTCTTGT  
 GATAAATCTTCTGAATAATATATATTATTTTTTACTGAGTAATGTTTCGATATT  
 TTCTTTACACCAAAAAATATCATCACTGAAAATAAACACGTCACGTATCATTGCC

AAATCGCGTATTTTATTTAAAGCTTTTTTGTAACTCTAACGAACAAACGCCAT  
 GAGTTAAAGTAGCTGTTTTGTTTTTATATAATCTCCTCTTCTTATATGAATAGAA  
 AGTGATGATTGAGATTCAAGAATTTTTGCTGCAAGTAAATTTGCTTGTTTCAGACA  
 CATTCTTTGGAATAAAAAATTCTTTTAGATCTAATATATGTTTTATGGAAAAAGTG  
 CTCAGATTGCCAATACCCTATATATTTTTTTGGATTTCCATTTTTGCGCTATATATTC  
 AAAATCATAACCATAGGCATGAAATTCATTGCAAAAACCTAAAAAAGAAAGAT  
 TTCAGGATATAATCTTGACCCACGAACCAAAAATTTATAAATATTATTAATTTTT  
 GGTGTGTAATACTGTAATATTCCTCTGGAATTTGTAGATTGTTTAGCCTGTAACC  
 ACCATGATCATCATTTTCAGCATAATGACTTATATCAAAATATAATGGTGTCCCA  
 TTAATTTTGGAAAGCGCATACCCAAATGAGAAGTGAAGAAAGTTGATTTCCAAGTC  
 CGCCTTGTAATCTTATAATAGACATTATATCTCCTTCTTG

5 La secuencia de ácidos nucleicos que codifica la proteína O126 WbgL de *E. coli* del plásmido G204 se presenta a continuación (SEQ ID NO: 21) (el codón de parada ATG está subrayado y el codón de PARADA está en letra negrita y subrayado):

ATGTCTATTATAAGATTACAAGGCGGACTTGGAATCAACTTTTTCAGTTCTCATT  
 TGGGTATGCGCTTTCAAAATTAATGGGACACCATTATATTTTGATATAAGTCAT  
 TATGCTGAAAATGATGATCATGGTGGTTACAGGCTAAACAATCTACAAATTCAG  
 AGGAATTTTACAGTATTACACACCAAAAATTAATAATTTATAAATTTTTGGT  
 TCGTGGGTCAAGATTATATCCTGAAATCTTTCTTTTTTTAGGTTTTTGAATGAAT  
 TTCATGCCTATGGTTATGATTTTGAATATATAGCGCAAAAATGGAAATCCAAAA  
 ATATATAGGGTATTGGCAATCTGAGCACTTTTCCATAAACATATATTAGATCTA  
 AAAGAATTTTTATTCCAAAGAATGTGTCTGAACAAGCAAATTTACTTGCAGCAA  
 AAATTCCTGAATCTCAATCATCACTTTCTATTCATATAAGAAGAGGAGATTATAT  
 AAAAAACAAAACAGCTACTTTAACTCATGGCGTTTGTTCGTTAGAGTATTACAAA  
 AAAGCTTTAAATAAAATACGCGATTTGGCAATGATACGTGACGTGTTATTTTCA  
 GTGATGATATTTTTTTGGTGTAAAGAAAATATCGAAACATTACTCAGTAAAAATA  
 TAATATATATTATTCAGAAGATTTATCACAGAAGAAGATTTATGGTTAATGAGC  
 TTAGCTAACCATCATATTATAGCGAATAGTAGTTTTAGTTGGTGGGGGGCTTATT  
 TAGGTACATCAGCGTCACAGATTGTTATTTATCCTACTCCTTGGTACGATATAACT  
 CCAAAAAATACTTATATCCCATAGTCAATCACTGGATAAACGTGGATAAACATA  
 GCTCGTGT**TAG**

10 Un mapa del plásmido pG216 se muestra en la Figura 9. La secuencia del plásmido pG216 se presenta a continuación (SEQ ID NO: 17) (el codón complementario al codón de parada ATG en la posición 5' para *FutL* de *H. mustelae* está subrayado, y el codón complementario al codón de PARADA está en letra negrita y subrayado):

TCTAGAATTCTAAAAATTGATTGAATGTATGCAAATAAATGCATACACCA  
 TAGGTGTGGTTTAATTTGATGCCCTTTTTCAGGGCTGGAATGTGTAAGAGCGGGG  
 TTATTTATGCTGTTGTTTTTTGTTACTCGGGAAGGGCTTACCTCTTCCGCATAA  
 ACGCTTCCATCAGCGTTTATAGTTAAAAAATCTTTCGGAACCTGGTTTTGCGCTTA  
 CCCCACCAACAGGGGATTTGCTGCTTTCCATTGAGCCTGTTTCTCTGCGCGACG  
 TTCGCGGCGGCGTGTGTTGTGCATCCATCTGGATTCTCCTGTCAGTTAGCTTTGGTG  
 GTGTGTGGCAGTTGTAGTCCTGAACGAAAACCCCCGCGATTGGCACATTGGCAG  
 CTAATCCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTATCACACACCCCAA  
 GCCTTCTGCTTTGAATGCTGCCCTTCTCAGGGCTTAATTTTTAAGAGCGTCACCT  
 TCATGGTGGTCAGTGCCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTT

ATGTCAACACCGCCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGTTTTT  
 ATATGAATTTATTTTTTGCAGGGGGGCATTGTTTGGTAGGTGAGAGATCAATTCT  
 GCATTAATGAATCGGCCAACGCGCGGGGAGAGGGCGTTTGCGTATTGGGCGCTC  
 TTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCCGGCTGCGGCGAGCG  
 GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAAC  
 GCAGGAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAA  
 GGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAA  
 AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACAG  
 GCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGCTTAC  
 CGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCAC  
 GCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGGGCTGTGTGCA  
 CGAACCCCGTTACGCCGACCCTGCGCCTTATCCGGTAACTATCGTCTTGAG  
 TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGG  
 ATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCCTGAAGTGGTGGCCT  
 AACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG  
 TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGG  
 TAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCT  
 CAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAAT  
 CACGTTAAGGGATTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCT  
 TTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGG  
 TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTAT  
 TTCGTTTATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGA  
 GGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCG  
 GCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGT  
 GGTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA  
 GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGG  
 CATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTACGCTCCGGTTCCCAAC  
 GATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTT  
 CGGTCCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAAGTGTATCACTCATGGTT  
 ATGGCAGCACTGCATAATTCTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGT  
 GACTGGTGAAGTCAACCAAGTCAATTCTGAGAATAGTGTATGCGGCGACCGAG  
 TTGCTCTTGCCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTA  
 AAAGTGCTCATCATTGAAAACGTTCTTCGGGGCGAAAATCTCAAGGATCTTAC  
 CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGC  
 ATCTTTACTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCC  
 GCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCTT  
 TTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATT  
 TGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAA  
 AGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAT  
 AGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCCGGTATGACGGTGAAAACC  
 TCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGG  
 GAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTCCGGGGCTG  
 GCTTAACTATGCGGCATCAGAGCAGATTGTAAGTACTGAGAGTGCACCATATATGCGGT  
 GTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCTCCTCA  
 ACCTGTATATTTCGTAACCCACGCCAATGGGAGCTGTCTCAGGTTTGTTCCTGAT  
 TGGTTACGGCGCGTTTCGCATCATTGTTGAGTTTTTCCGCCAGCCCAGCGCGAG  
 TTTACCGGTGCCTGGGTGCAGTACATCAGCATGGGGCAAATTTCCATCCCGA  
 TGATTGTCGCGGGTGTGATCATGATGGTCTGGGCATATCGTCCGAGCCCACAGCA  
 ACACGTTTCTGAGGAACCATGAAACAGTATTTAGAACTGATGCAAAAAGTGCT  
 CGACGAAGGCACACAGAAAAACGACCGTACCGGAACCGGAACGCTTTCCATTTT  
 TGTCATCAGATGCGTTTTAACCTGCAAGATGGATTCCCGCTGGTGACAATAA

CGTTGCCACCTGCGTTCCATCATCCATGAACTGCTGTGGTTTTCTGCAGGGCGACA  
 CTAACATTGCTTATCTACACGAAAACAATGTCACCATCTGGGACGAATGGGCCGA  
 TGA AACACGGCGACCTCGGGCCAGTGTATGGTAAACAGTGGCGCGCCTGGCCAAC  
 GCCAGATGGTCGTCATATTGACCAGATCACTACGGTACTGAACCAGCTGAAAAA  
 CGACCCGGATTTCGCGCCGCATTATTGTTTTAGCGTGGAAACGTAGGCGAACTGGAT  
 AAAATGGCGCTGGCACCGTGCCATGCATTCTTCCAGTTCTATGTGGCAGACGGCA  
 AACTCTCTTGCCAGCTTTATCAGCGCTCCTGTGACGTCTTCCCTCGGCCTGCCGTTT  
 AACATTGCCAGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCATCTGG  
 AAGTGGGTGATTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATAT  
 GGATCAAACCTCATCTGCAATTAAGCCGCGAACC GCGTCCGCTGCCGAAGTTGATT  
 ATCAAACGTAAACCCGAATCCATCTTCGACTACCGTTTTCGAAGACTTTGAGATTG  
 AAGGCTACGATCCGCATCCGGGCATTAAAGCGCCGGTGGCTATCTAATTACGAA  
 ACATCCTGCCAGAGCCGACGCCAGTGTGCGTCGGTTTTTTTTACCTCCGTTAAATT  
 CTTTCGAGACGCCTTCCCGAAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGG  
 GAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA  
 TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTT  
 GTAAAACGACGGCCAGTGCCAAGCTTTCTTTAATGAAGCAGGGCATCAGGACGG  
 TATCTTTGTGGAGAAAGCAGAGTAATCTTATTCAGCCTGACTGGTGGGAAACCAC  
 CAGTCAGAATGTGTTAGCGCATGTTGACAAAAATAACCATTAGTCACATTATCCGT  
 CAGTCGGACGACATGGTAGATAACCTGTTTATTATGCGTTTTGATCTTACGTTTAA  
 TATTACCTTTATGCGATGAAACGGTCTTGGCTTTGATATTCATTTGGTCAGAGATT  
 TGAATGGTTCCTGACCTGCCATCCACATTCGCAACATACTCGATTCCGTTCCGGC  
 TCAATGATAACGTCGGCATATTTAAAAACGAGGTTATCGTTGTCTCTTTTTTCAGA  
 ATATCGCCAAGGATATCGTCGAGAGATTCCGGTTTAATCGATTTAGA ACTGATCA  
 AATAAATTTTTCTGACCAATAGATATTCATCAAAATGAACATTGGCAATTGCCAT  
 AAAAACGATAAATAACGTATTGGGATGTTGATTAATGATGAGCTTGATACGCTG  
 ACTGTTAGAAGCATCGTGGATGAAACAGTCCTCATTATAA AACACCACTGAAGG  
 GCGCTGTGAATCACAAGCTATGGCAAGGTCATCAACGGTTTTCAATGTCGTTGATT  
 TCTCTTTTTTTAACCCCTCTACTCAACAGATACCCGGTTAAACCTAGTCGGGTGTA  
 ACTACATAAATCCATAATAATCGTTGACATGGCATAACCCTCACTCAATGCGTAAC  
 GATAATCCCCTTACCTGAATATTTTCATCATGACTAAACGGAACAACATGGGTCA  
 CCTAATGCGCCACTCTCGCGATTTTTCAGGCGGACTTACTATCCCGTAAAGTGTT  
 GTATAATTTGCCTGGAATTGTCTTAAAGTAAAGTAAATGTTGCGATATGTGAGTG  
 AGCTTAAAACAAATATTTTCGCTGCAGGAGTATCCTGGAAGATGTTTCGTAGAAGCT  
 TACTGCTCACAAGAAAAAAGGCACGTCATCTGACGTGCCTTTTTTTATTTGTA  
 CACTGACGATTACTGCAGCTCGAGTTAGGATTTTCGTTTCGAATTGGGATTCGAT  
 TTTAACCCAGTCTTTGCACAGGATGTTTTTCGTTACCGTAAATCCAGTGGGACGGA  
 CCAATGATAATTTTTTCCGGATTTTTGATCAGGTAGGCTGCCACCAGGAGTAAG  
 TGCTGTTAGTGATGATACCGTGTGCAAGACTGCATCAGCATCATGTCCCAGTG  
 GGCTGCACCATCACGCGTCGTCATGTCAACAAACGGGTAACCCAGATCCAGGTTT  
 TGACGAATTCAGATCCTCGCAGAACAGGAACAGTTCCAGATTTTGAACACGTT  
 TTGCCATATACGCAATGGCGCGCAGCTGGTAGGAGATGTCCAGCTGCCAGCCCA  
 GCGCATGTAATCGCCACGGCGGATGTGAACGAACACAGAGTTTTTTCGCAGCCA  
 GGATCTGGGACAGTTTACGAGAGTACTGTTCCGCGTGTTCGGTTCGGGTGAGGCAG  
 GGTGAAAGTTTGTGATCAGAGGGGAGATATCTTCGAAATAGCGCGGGTCTCTG  
 AAAGTAGCCATGGAAATACGCAATGCGGCTCGGTTCAAACAGTTCCGGCATGTA  
 CTCGAATACAATTTCTTTGCTAACGCGGCCAGACCCATACGACGCAGTGCACCA  
 CGCACCAGACGCGGCAGGTTCTGCATGTGTGCCGCGGCGATCTGCTGGGCGGAC  
 GCACACTGCAGGTCGATCGGGAACAGGTGCAGGCCAGTTACGGTTACCGTAA  
 TCGAACCAAGTGGTATCCAGCAGTACCGGAATGTTTCAGGTGAGTCTGCAGAGAT

TTAGCGAATGCGTACTGGAACATCTGGTTACCCAGGCCGCCGTGCACCTGAACGA  
TTTTGAAATCCATTATATCTCCTTCTTG

5 La secuencia de ácidos nucleicos que codifica la proteína FutL de *H. mustelae* del plásmido pG216 se presenta a continuación (SEQ ID NO: 22) (el codón de parada ATG está subrayado y el codón de PARADA está en letra negrita y subrayado):

ATGGATTTCAAATCGTTCAGGTGCACGGCCGGCCTGGGTAACCAGATGTTCCAGT  
ACGCATTTCGCTAAATCTCTGCAGACTCACCTGAACATTCCGGTACTGCTGGATAC  
CACTTGGTTCGATTACGGTAACCGTGAACCTGGGCCTGCACCTGTTCCCGATCGAC  
CTGCAGTGTGCGTCCGCCAGCAGATCGCCGCGGCACACATGCAGAACCTGCCG  
CGTCTGGTGCCTGGTGCCTGCGTCGTATGGGTCTGGGCCGCGTTAGCAAAGAAA  
TTGTATTCGAGTACATGCCGGAACCTGTTTGAACCGAGCCGCATTGCGTATTTCCA  
TGGCTACTTTTCAGGACCCGCGCTATTTTCGAAGATATCTCCCCTCTGATCAAACAA  
ACTTTCACCTGCCTCACCCGACCGAACACGCGGAACAGTACTCTCGTAAACTGT  
CCCAGATCCTGGCTGCGAAAACTCTGTGTTTCGTTACATCCGCCGTGGCGATTA  
CATGCGCCTGGGCTGGCAGCTGGACATCTCCTACCAGCTGCGCGCCATTGCGTAT  
ATGGCAAACGTTGTTCAAATCTGGAACCTGTTCTGTTCTGCGAGGATCTGGAAT  
TCGTACAGAACCTGGATCTGGGTTACCCGTTTGTGACATGACGACGCGTGATGG  
TGCAGCCCACTGGGACATGATGCTGATGCAGTCTTGCAAACACGGTATCATCACT  
AACAGCACTTACTCCTGGTGGGCGAGCTACCTGATCAAAAATCCGGAAAAAATT  
ATCATTGGTCCGTTCCCACTGGATTTACGGTAACGAAAACATCCTGTGCAAAGACT  
GGGTTAAAATCGAATCCCAATTCGAAACGAAATCCT**TAA**

10 Un mapa del plásmido pG217, usado de acuerdo con la invención para producir un oligosacárido fucosilado en una bacteria, se muestra en la Figura 10. La secuencia del plásmido pG217 se presenta a continuación (SEQ ID NO: 18) (el codón complementario al codón de parada ATG en la posición 5' para *FutN* de *B. vulgatus* está subrayado, y el codón complementario al codón de PARADA está en letra negrita y subrayado):

TCTAGAATTCTAAAAATTGATTGAATGTATGCAAATAAATGCATACACCATAGGT  
GTGGTTTAATTTGATGCCCTTTTTCAGGGCTGGAATGTGTAAGAGCGGGGTTATT  
TATGCTGTTGTTTTTTGTTACTCGGGAAGGGCTTTACCTCTTCCGCATAAACGCT  
TCCATCAGCGTTTATAGTTAAAAAAATCTTTCGGAACCTGGTTTTGCGCTTACCCCA  
ACCAACAGGGGATTTGCTGCTTTCCATTGAGCCTGTTTCTCTGCGCGACGTTTCGC  
GGCGGCGTGTGTTGTGCATCCATCTGGATTCTCCTGTCAGTTAGCTTTGGTGGTGTG  
TGGCAGTTGTAGTCTGAACGAAAACCCCCCGGATTGGCACATTGGCAGCTAAT  
CCGGAATCGCACTTACGGCCAATGCTTCGTTTCGTATCACACACCCCAAAGCCTT  
CTGCTTTGAATGCTGCCCTTCTTCAGGGCTTAATTTTTAAGAGCGTCACCTTCATG  
GTGGTCAGTGCCTGCTGATGTGCTCAGTATCACCGCCAGTGGTATTTATGTC  
AACACCGCCAGAGATAATTTATCACCGCAGATGGTTATCTGTATGTTTTTTTATAT  
GAATTTATTTTTTGCAGGGGGGCATTGTTTGGTAGGTGAGAGATCAATTCTGCAT  
TAATGAATCGGCCAACGCGCGGGGAGAGGGCGGTTTGCATATTGGGCGCTCTTCC  
GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGGGCAGCGGTAT  
CAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAG  
GAAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCC  
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATC  
GACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATAACCAGGCGT

15

TTCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGA  
 TACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTG  
 TAGGTATCTCAGTTCGGTGTAGGTCGTTTCGCTCCAAGCTGGGCTGTGTGCACGAA  
 CCCCCGTTTACGCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA  
 ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTA  
 GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACT  
 ACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC  
 CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAG  
 CGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAA  
 GAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAACGAAAACTCAC  
 GTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTT  
 AATTA AAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCT  
 GACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTT  
 GTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGG  
 CTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCT  
 CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT  
 CCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAG  
 TAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCAT  
 CGTGGTGTACGCTCGTTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT  
 CAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGG  
 TCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATG  
 GCAGCACTGCATAATTCTTACTGTGATGCCATCCGTAAGATGCTTTTTCTGTGAC  
 TGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGC  
 TCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGA ACTTTAAAA  
 GTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGC  
 TGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCA ACTGATCTTCAGCATC  
 TTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCA  
 AAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTT  
 AATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGA  
 ATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGT  
 GCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAAATAGG  
 CGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCT  
 GACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGA  
 GCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTCCGGGGCTGGC  
 TTA ACTATGCGGCATCAGAGCAGATTGTA CTGAGAGTGCACCATATATGCGGTGT  
 GAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGCGCCTCCTCAAC  
 CTGTATATTCGTAAACCACGCCCAATGGGAGCTGTCTCAGGTTTGTTCCTGATTG  
 GTTACGGCGCGTTTCGCATCATTGTTGAGTTTTTCCGCCAGCCCAGCGCAGTTT  
 ACCGGTGCCTGGGTGCAGTACATCAGCATGGGGCAAATCTTTCCATCCCGATGA  
 TTGTCGCGGGTGTGATCATGATGGTCTGGGCATATCGTCGCAGCCCACAGCAACA  
 CGTTTCCTGAGGAACCATGAAACAGTATTTAGAACTGATGCAAAAAGTGCTCGA  
 CGAAGGCACACAGAAAAACGACCGTACCGGAACCGGAACGCTTTCCATTTTTGG  
 TCATCAGATGCGTTTTAACCTGCAAGATGGATTCCCGCTGGTGACAACTAAACGT  
 TGCCACCTGCGTTCCATCATCCATGAACTGCTGTGGTTTCTGCAGGGCGACACTA  
 ACATTGCTTATCTACACGAAAACAATGTCACCATCTGGGACGAATGGGCCGATG  
 AAAACGGCGACCTCGGGCCAGTGTATGGTAAACAGTGGCGCGCCTGGCCAACGC  
 CAGATGGTTCGTCATATTGACCAGATCACTACGGTACTGAACCAGCTGAAAAACG  
 ACCCGGATTCGCGCCGCATTATTGTTTACGCGTGGAACGTAGGCGAACTGGATAA  
 AATGGCGCTGGCACCGTGCCATGCATCTTCCAGTTCTATGTGGCAGACGGCAA  
 CTCTCTTGCCAGCTTTATCAGCGCTCCTGTGACGTCTTCCTCGGCCTGCCGTTCAA  
 CATTGCCAGCTACGCGTTATTGGTGCATATGATGGCGCAGCAGTGCATCTGGAA

GTGGGTGATTTTGTCTGGACCGGTGGCGACACGCATCTGTACAGCAACCATATGG  
 ATCAAACCTCATCTGCAATTAAGCCGCGAACC GCGTCCGCTGCCGAAGTTGATTAT  
 CAAACGTAAACCCGAATCCATCTTCGACTACCGTTTCGAAGACTTTGAGATTGAA  
 GGCTACGATCCGCATCCGGGCATTAAGCGCCGGTGGCTATCTAATTACGAAAC  
 ATCCTGCCAGAGCCGACGCCAGTGTGCGTCCGTTTTTTTTACCCTCCGTTAAATTCT  
 TCGAGACGCCTTCCCGAAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGA  
 AGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATG  
 TGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGT  
 AAAACGACGCCAGTGCCAAGCTTTCTTTAATGAAGCAGGGCATCAGGACGGTA  
 TCTTTGTGGAGAAAGCAGAGTAATCTTATTCAGCCTGACTGGTGGGAAACCACCA  
 GTCAGAATGTGTTAGCGCATGTTGACAAAAATACCATTAGTCACATTATCCGTC  
 GTCGGACGACATGGTAGATAACCTGTTTATTATGCGTTTTGATCTTACGTTTAATA  
 TTACCTTTATGCGATGAAACGGTCTTGGCTTTGATATTCATTTGGTCAGAGATTTG  
 AATGGTTCCCTGACCTGCCATCCACATTCGCAACATACTCGATTCCGTTCCGGCTC  
 AATGATAACGTCGGCATATTTAAAAACGAGGTTATCGTTGTCTCTTTTTTCAGAA  
 TATCGCCAAGGATATCGTCGAGAGATTCCGGTTTAATCGATTTAGAACTGATCAA  
 TAAATTTTTCTGACCAATAGATATTCATCAAAATGAACATTGGCAATTGCCATA  
 AAAACGATAAATAACGTATTGGGATGTTGATTAATGATGAGCTTGATACGCTGAC  
 TGTTAGAAGCATCGTGGATGAAACAGTCCTCATTAAATAAACACCACTGAAGGGC  
 GCTGTGAATCACAAGCTATGGCAAGGTCATCAACGGTTTTCAATGTCGTTGATTTT  
 TCTTTTTTTAAACCCCTCTACTCAACAGATAACCCGGTTAAACCTAGTCGGGTGTAAC  
 TACATAAATCCATAATAATCGTTGACATGGCATAACCTCACTCAATGCGTAACGA  
 TAATCCCCTTACCTGAATATTTTCATCATGACTAAACGGAACAACATGGGTCACC  
 TAATGCGCCACTCTCGCGATTTTTTCAGGCGGACTTACTATCCCGTAAAGTGTGT  
 ATAATTTGCCTGGAATTGTCTTAAAGTAAAGTAAATGTTGCGATATGTGAGTGAG  
 CTAAAACAAATATTTTCGCTGCAGGAGTATCCTGGAAGATGTTTCGTAGAAGCTTA  
 CTGCTCACAAGAAAAAGGCACGTCATCTGACGTGCCTTTTTTATTTGTACTACC  
 CTGTACGATTACTGCAGCTCGAGTTAGGATACCGGCACCTTTGATCCAACCAGTCG  
 GGTAGATATCCGGTGCTTCGGAGTGCTGGAACCAACGGCTCGGCACAATAACAG  
 TCTTATCCATATTAGGGTTCAGCCAGGCACCCACCAAGAAAACGTGCTGTTACA  
 AATGATGTGATGTTTGCAATGAGACATCAGCATCATATCCTGCCAGGAGTCTTCA  
 TCAGTGTTCCAGTCAATATAAACCGCATTCTGCAGTGGCAGATTTTCTTTAACCC  
 ACGCGATATCGTCGGAGAAGATATAGTAAGATGGGCTAGCAACACGACGGGACA  
 TTTCCGCGATAGCATTCTGGTAATACGGCAGCTGGCACACGGAACCGGTAGTAGC  
 CCAGTGTTTCCGGCTGCAGATAGTCACCACGACGAATGTGCAGGGAAACCGCGTTT  
 TCATCTTTGTCCAGGATTTCCAGCATGTTCCAGGCTGCGGGAATTTGCTTTGTTCTT  
 ATCAAAGGTGAAGGATTCACGCACTTCGTCTTTGATATCAGCGAAGAAACGCTCG  
 CTCTGATAGAAACCTTTAAAGTACAGCAGCGGCCAGAAATACTTCTTCTCGAACG  
 CACGCAGAGAGTTCGGCGCCTGCTTGCCTTCGTAGATTTTTTTAAAAACAGGAA  
 TTCGATAACTTTTTTCAGCGGTTGGTTGATGCAGAATTCGGTGTGCGGCAGGTTG  
 AACACGCGGTGCATTTTCGTAACCGTAATGGACTTTGTAATGCATCATGTCGCTCA  
 GGTGATACGGACCTTCGGGTAATACTTTTTTCATACGCAGATAGAAAGCATAGAT  
 AAACATCTGGTTGCCAGACC GCCAGTCACTTTGATCAGACGCATTATATCTCCT  
 TCTTG

La secuencia de ácidos nucleicos que codifica la proteína *FutN* de *B. vulgatus* del plásmido pG217 se presenta a continuación (SEQ ID NO: 23) (el codón de parada ATG está subrayado y el codón de PARADA está en letra negrita y subrayado):

5

ATGCGTCTGATCAAAGTGACTGGCGGTCTGGGCAACCAGATGTTTATCTATGCTT  
TCTATCTGCGTATGAAAAAGTATTACCCGAAGGTCCGTATCGACCTGAGCGACAT  
GATGCATTACAAAGTCCATTACGGTTACGAAATGCACCCGCGTGTTC AACCTGCCG  
CACACCGAATTCTGCATCAACCAACCGCTGAAAAAAGTTATCGAATTCCTGTTTT  
TTAAAAAAATCTACGAACGCAAGCAGGCGCCGAACTCTCTGCGTGC GTTTCGAGA  
AGAAGTATTTCTGGCCGCTGCTGTACTTTAAAGGTTTCTATCAGAGCGAGCGTTT  
CTTCGCTGATATCAAAGACGAAGTGCGTGAATCCTTCACCTTTGATAAGAACAAA  
GCAAATTC CCGCAGCCTGAACATGCTGGAAATCCTGGACAAAGATGAAAACGCG  
GTTTCCCTGCACATTCGTCGTGGTGACTATCTGCAGCCGAAACACTGGGCTACTA  
CCGTTCCGTGTGCCAGCTGCCGTATTACCAGAATGCTATCGCGGAAATGTCCCG  
TCGTGTTGCTAGCCCATCTTACTATATCTTCTCCGACGATATCGCGTGGGTTAAAG  
AAAATCTGCCACTGCAGAATGCGGTTTTATATTGACTGGAACACTGATGAAGACTC  
CTGGCAGGATATGATGCTGATGTCTCATTGCAAACATCACATCATTTGTAACAGC  
ACGTTTTCTTGGTGGGGTGCCCTGGCTGAACCCTAATATGGATAAGACTGTTATTG  
TGCCGAGCCGTTGGTTCCAGCACTCCGAAGCACCGGATATCTACCCGACTGGTTG  
GATCAAAGTGCCGGTATCCTAA

Los oligosacáridos fucosilados producidos por células de *E. coli* metabólicamente modificadas por ingeniería se purifican a partir del caldo de cultivo posterior a la fermentación. Un procedimiento a modo de ejemplo comprende cinco pasos. (1) Clarificación: el caldo de cultivo de fermentación se recoge y las células se eliminan por sedimentación en una centrifugadora preparativa a 6000 x g durante 30 min. Cada ejecución de biorreactor produce aproximadamente 5-7 l de sobrenadante parcialmente clarificado. (2) Captura del producto en carbono grueso: una columna empaquetada con carbono grueso (Calgon 12 x 40 TR) de ~1000 ml de volumen (dimensión de 5 cm de diámetro x 60 cm de longitud) se equilibra con 1 columna de volumen (CV) de agua y se carga con agua clarificada sobrenadante de cultivo a un caudal de 40 ml/min. Esta columna tiene una capacidad total de alrededor de 120 g de azúcar. Después de la carga y la captura de azúcar, la columna se lava con 1,5 CV de agua, a continuación se eluye con 2,5 CV de etanol al 50 % o isopropanol al 25 % (las concentraciones más bajas de etanol en esta etapa (25-30 %) pueden ser suficientes para la elución del producto). Esta etapa de elución con solventes libera aproximadamente un 95 % del total de azúcares unidos en la columna y una pequeña parte de los cuerpos de color. En este primer paso, la captura de la cantidad máxima de azúcar es el objetivo principal. La resolución de contaminantes no es un objetivo. (3) Evaporación: un volumen de 2,5 l de eluato de etanol o isopropanol de la columna de captura se evapora rotativamente a 56 °C y se genera un jarabe de azúcar en agua. Los métodos alternativos que se podrían usar para esta etapa incluyen liofilización o secado por pulverización. (4) Cromatografía ultrarrápida en carbono fino y medios de intercambio iónico: Una columna (GE Healthcare HiScale50/40, 5 x 40 cm, presión máxima de 20 bares) conectada a un sistema de cromatografía Biotage Isolera One FLASH se empaqueta con 750 ml de un Carbón Activado Darco Carbon G60 (malla de 100): Celite 535 (grueso) mezcla a 1:1 (ambos empaquetamiento se columna se tuvieron en Sigma). La columna se equilibra con 5 CV de agua y se carga con azúcar de la etapa 3 (10-50 g, dependiendo de la proporción de 2'-FL con respecto a la lactosa contaminante), usando un cartucho de carga de celite o una inyección directa. La columna se conecta a un detector de dispersión de luz por evaporación (ELSD) para detectar picos de azúcares eluidos durante la cromatografía. Se ejecuta un gradiente de cuatro etapas de isopropanol, etanol o metanol para separar 2'-FL de los monosacáridos (si están presentes), la lactosa y los cuerpos de color. Las fracciones correspondientes a los picos de azúcar se recogen automáticamente en frascos de 120 ml, se agrupan y se dirigen a la etapa 5. En ciertas purificaciones, se realizan fermentaciones más largas de lo normal, el paso de la fracción que contiene 2'-FL a través de columnas de intercambio aniónico e intercambio catiónico puede eliminar el exceso de contaminantes proteicos/ADN/cuerpos de caramelo. Las resinas sometidas a ensayo de forma satisfactoria para esta finalidad son Dowex 22.

La identidad del oligosacárido principal sintetizado por WbgL se sometió a ensayo y se confirmó que era 2'-FL auténtica. Los oligosacáridos sintetizados en la cepa WbgL se inmovilizaron en una columna de carbono, se eluyeron y se resuspendieron en agua destilada. Este material se sometió a digestión durante la noche con fucosidasas de diferentes especificidades, y las reacciones se analizaron por TLC. Como se muestra en la Figura 4, el material no tratado consistió principalmente en un oligosacárido con la misma movilidad que el patrón de 2'-FL (calle 1). El tratamiento con  $\alpha$ 1,2 fucosidasa produjo tanto lactosa como fucosa, mientras que la presunta mancha de 2'-FL disminuyó significativamente la intensidad de la tinción (calle 2). El tratamiento de los oligosacáridos con una fucosidasa  $\alpha$ 1,3-4 no tuvo ningún efecto. Estos resultados demuestran que WbgL es capaz de la biosíntesis de 2'-FL auténtica en *E. coli* metabólicamente modificada mediante ingeniería genética.

El enfoque de identificación sistemática de genes se utilizó con éxito para identificar nuevas  $\alpha$ (1,2) FT para la biosíntesis eficaz de 2'-FL en cepas hospedadoras de *E. coli* metabólicamente modificadas mediante ingeniería genética. Los resultados de la identificación sistemática se resumen en la Tabla 1. De forma específica, tanto WbgL

como FutL dirigen directamente la síntesis de 2'-FL a aproximadamente un 75 % de los niveles alcanzados por el FutC de  $\alpha(1,2)$  FT previamente caracterizado. Además, WbgL también fue capaz de sintetizar LDFT, que es otra HMO terapéuticamente útil. Además, el FutN de la bacteria entérica comensal *B. vulgatus* se identificó como otra  $\alpha(1,2)$  FT útil para la síntesis de oligosacáridos fucosilados. El enfoque que se describe en el presente documento es útil en el análisis de las  $\alpha(1,2)$  FT candidatas adicionales e identifica enzimas adicionales que son útiles para la producción a gran escala de HMOS.

#### Cepas hospedadoras de producción

La cepa K-12 de *E. coli* es una bacteria bien estudiada que ha sido objeto de una amplia investigación en fisiología microbiana y genética y explotada comercialmente para una variedad de usos industriales. El hábitat natural de la especie precursora, *E. coli*, es el intestino grueso de los mamíferos. La cepa K-12 de *E. coli* tiene un historial de uso seguro, y sus derivados se utilizan en una gran cantidad de aplicaciones industriales, incluida la producción de productos químicos y fármacos para administración y consumo humanos. La cepa K-12 de *E. coli* se aisló originalmente a partir de un paciente con difteria convaleciente en 1922. Debido a que carece de características de virulencia, crece fácilmente en medios de laboratorio comunes y se ha utilizado ampliamente para la investigación de la fisiología microbiana y genética, se ha convertido en la cepa bacteriológica estándar utilizada en investigación microbiológica, enseñanza y producción de productos para la industria y la medicina. La cepa K-12 de *E. coli* ahora se considera un organismo debilitado como resultado de mantenerse en el entorno del laboratorio durante más de 70 años. Como resultado, las cepas K-12 son incapaces de colonizar los intestinos de los seres humanos y otros animales en condiciones normales. La información adicional sobre esta cepa conocida está disponible en <http://epa.gov/oppt/biotech/pubs/fra/fra004.htm>. Además de la cepa K12 de *E. coli*, se utilizan otras cepas bacterianas como cepas hospedadoras de producción, por ejemplo, una variedad de especies bacterianas se pueden usar en los métodos de biosíntesis de oligosacáridos, por ejemplo, *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, o *Xanthomonas campestris*. También se o bacterias del género *Bacillus*, incluyendo *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, y *Bacillus circulans*. Del mismo modo, las bacterias de los géneros *Lactobacillus* y *Lactococcus* se pueden modificar usando los métodos de la presente invención, que incluyen, pero no se limitan a, *Lactobacillus acidophilus*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, y *Lactococcus lactis*. *Streptococcus thermophilus* y *Propionibacterium freudenreichii* también son especies bacterianas adecuadas para la invención que se describe en el presente documento. Como parte de la presente invención también se incluyen cepas, modificadas como se describe aquí, de los géneros *Enterococcus* (por ejemplo, *Enterococcus faecium* y *Enterococcus thermophilus*), *Bifidobacterium* (por ejemplo, *Bifidobacterium longum*, *Bifidobacterium infantis*, y *Bifidobacterium bifidum*), *Sporolactobacillus* spp., *Micromonospora* spp., *Micrococcus* spp., *Rhodococcus* spp., y *Pseudomonas* (por ejemplo, *Pseudomonas fluorescens* y *Pseudomonas aeruginosa*).

Las cepas hospedadoras adecuadas son susceptibles de manipulación genética, por ejemplo, mantienen constructos de expresión, acumulan precursores del producto final deseado, por ejemplo, mantienen reservas de lactosa y GDP-fucosa, y acumulan producto final, por ejemplo, 2'-FL. Las cepas de ese tipo crecen bien en medios mínimos definidos que contienen sales simples y generalmente una única fuente de carbono. Las cepas diseñadas como se ha descrito anteriormente para producir el oligosacárido(s) fucosilado(s) deseado(s) se cultivan en un medio mínimo. Un medio mínimo a modo de ejemplo usado en un biorreactor, medio mínimo "FERM", se detalla a continuación.

Ferm (10 litros): Medio mínimo que comprende:

40 g de  $(\text{NH}_4)_2\text{HPO}_4$   
 100 g de  $\text{KH}_2\text{PO}_4$   
 10 g de  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$   
 40 g de NaOH

Elementos traza:

1,3 g de NTA (ácido nitrilotriacético)  
 0,5 g de  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$   
 0,09 g de  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$   
 0,09 g de  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$

0,01 g de  $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$

0,01 g de  $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$

5 0,02 g de  $\text{H}_3\text{BO}_3$

0,01 g de  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  (pH 6,8)

10 Agua hasta 10 l

Antiespumante DF204 (0,1 ml/l)

15 150 g de glicerol (crecimiento inicial discontinuo), seguido por alimentación en modo semicontinuo de glicerol al 90 %- $\text{MgSO}_4$  al 1 %-1X de elementos traza, a diversas tasas durante diversos periodos de tiempo.

20 Una cepa hospedadora de producción adecuada es una que no es la misma cepa bacteriana que la cepa bacteriana fuente a partir de la cual se identificó la secuencia de ácidos nucleicos que codifica fucosiltransferasa. Por ejemplo, la secuencia de ácidos nucleicos que codifica fucosiltransferasa FutL se identificó en *Helicobacter mustelae* y una cepa hospedadora adecuada es una bacteria distinta a *Helicobacter mustelae*, por ejemplo, FutL se produce en la cepa hospedadora K12 de producción de *E. coli* o cualquiera de las otras cepas que se han descrito anteriormente.

25 Las bacterias que comprenden las características que se describen en el presente documento se cultivan en presencia de lactosa, y se recupera un oligosacárido fucosilado, ya sea de la propia bacteria o de un sobrenadante de otra bacteria. El oligosacárido fucosilado se purifica para su uso en productos terapéuticos o nutricionales, a las bacterias se usan directamente en productos de ese tipo.

## REIVINDICACIONES

1. Uso de un constructo de ácido nucleico que comprende un ácido nucleico aislado que codifica una enzima  $\alpha(1,2)$ fucosiltransferasa que utiliza lactosa, dicho ácido nucleico estando unido de forma operativa a una o más secuencias de control heterólogas que dirigen la producción de la enzima en una cepa de producción de bacterias hospedadoras, en el que la secuencia de aminoácidos de dicha enzima codificada por dicho ácido nucleico comprende la secuencia de SEQ ID NO: 8 para producir un oligosacárido fucosilado en una bacteria en presencia de lactosa.
2. El uso de la reivindicación 1, en el que dicha cepa de producción comprende K12 de *Escherichia coli*.
3. El uso de la reivindicación 1, en el que dicho ácido nucleico codifica una proteína FutN.
4. El uso de la reivindicación 1, en el que dicha secuencia de control heteróloga comprende un promotor y un operador bacterianos, un sitio de unión al ribosoma bacteriano, un terminador de transcripción bacteriano, o un marcador seleccionable plasmídico.
5. El uso de la reivindicación 1, en el que dicha cepa de producción es un miembro del género *Bacillus*, *Pantoea*, *Lactobacillus*, *Lactococcus*, *Streptococcus*, *Propionibacterium*, *Enterococcus*, *Bifidobacterium*, *Sporolactobacillus*, *Micromonospora*, *Micrococcus*, *Rhodococcus*, o *Pseudomonas*.
6. El uso de la reivindicación 1, en el que dicha cepa de producción se selecciona entre el grupo que consiste en *Bacillus licheniformis*, *Bacillus subtilis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus*, *Bacillus circulans*, *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum*, *Xanthomonas campestris*, *Lactobacillus acidophilus*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus gasserii*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii*, *Lactococcus lactis*, *Streptococcus thermophiles*, *Propionibacterium freudenreichii*, *Enterococcus faecium*, *Enterococcus thermophiles*, *Bifidobacterium longum*, *Bifidobacterium infantis*, *Bifidobacterium bifidum*, *Pseudomonas fluorescens* y *Pseudomonas aeruginosa*.

FIG. 1

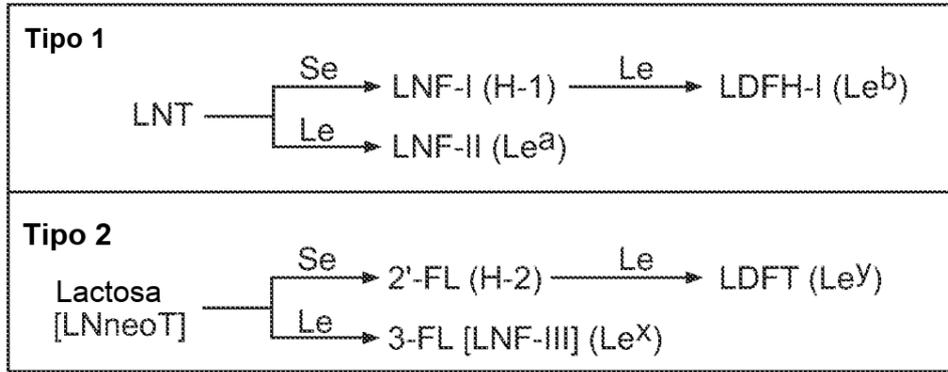
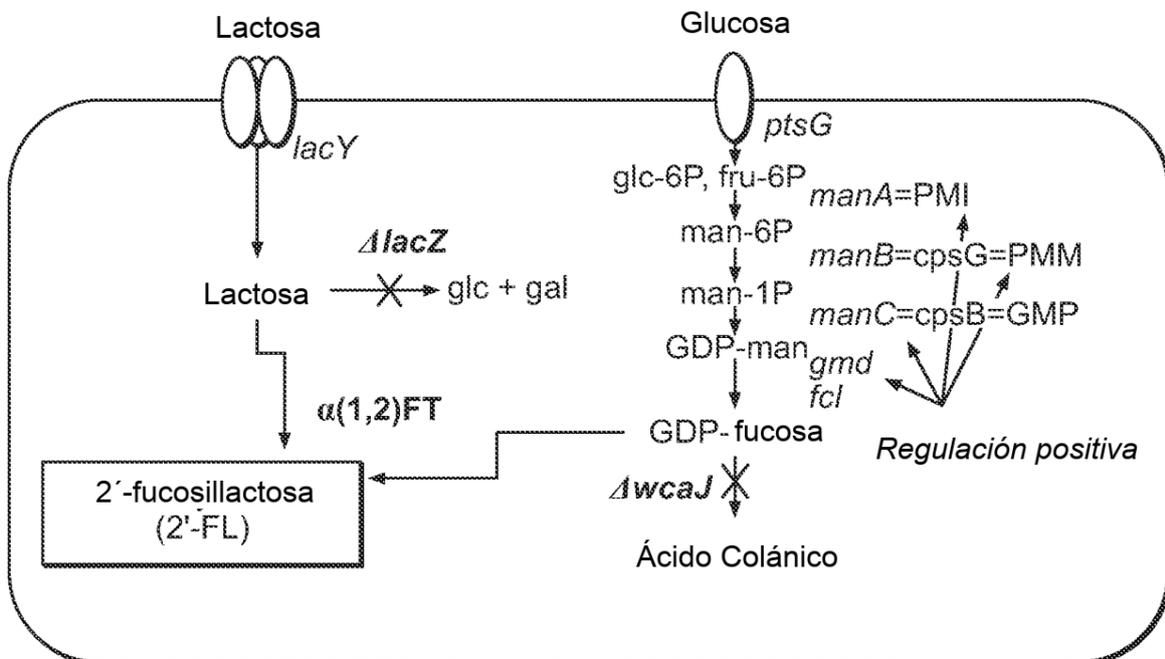


FIG. 2



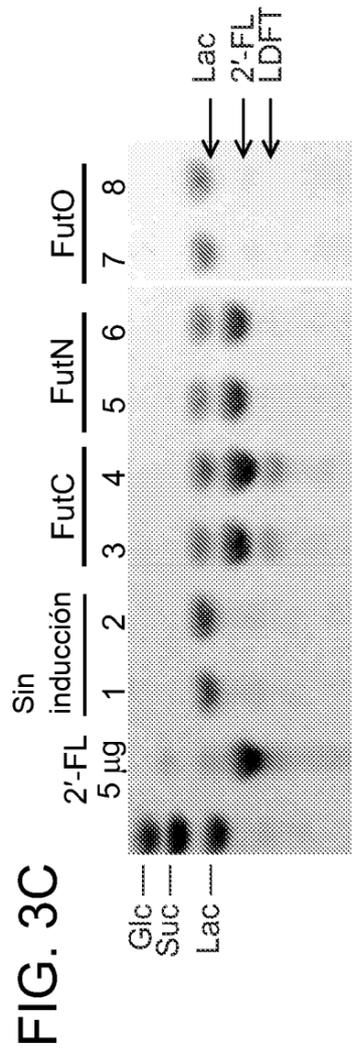
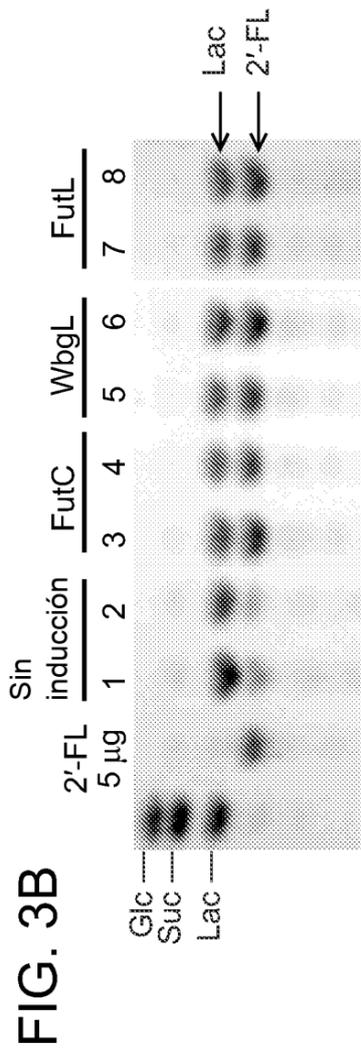
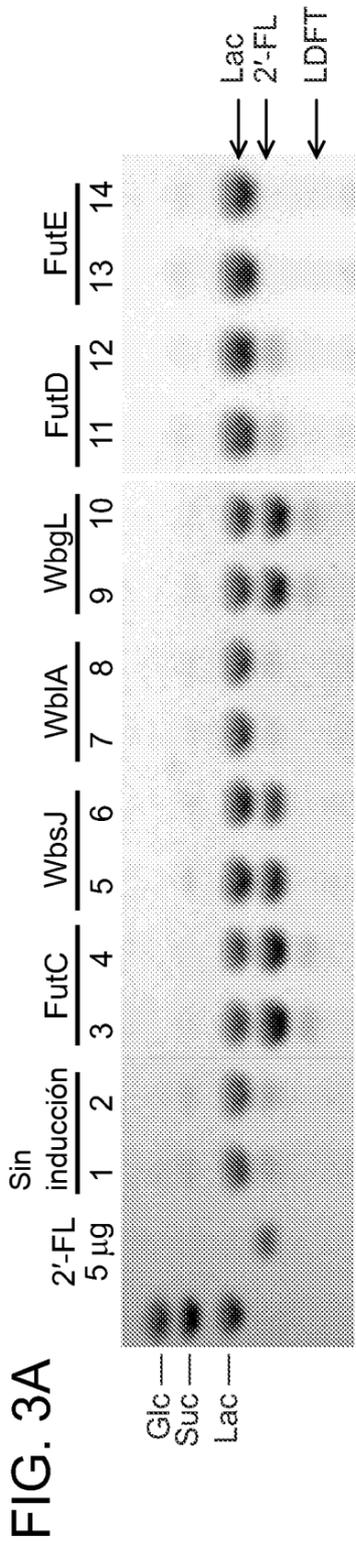
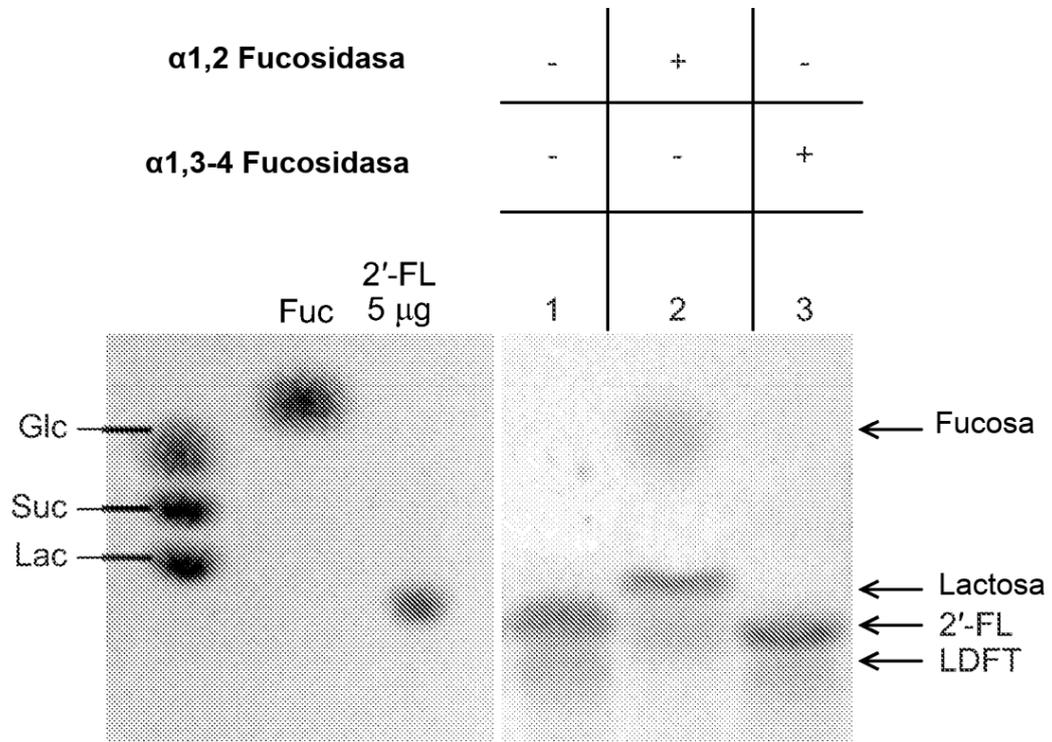


FIG. 4



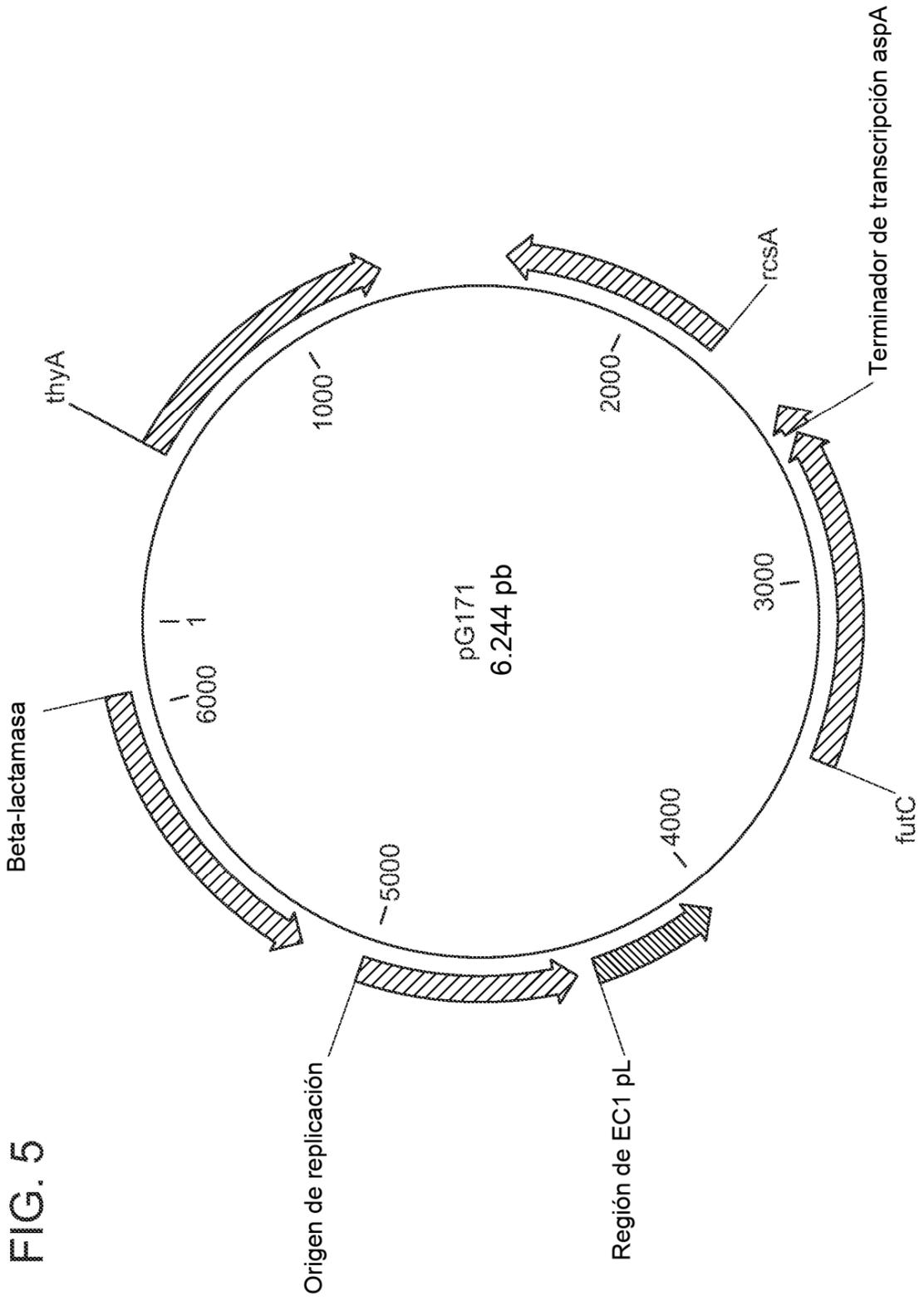


FIG. 5

FIG. 6

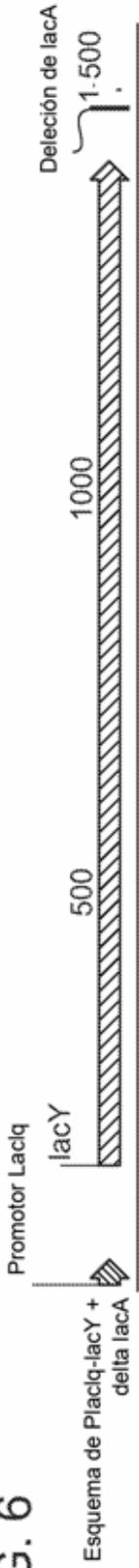


FIG. 7

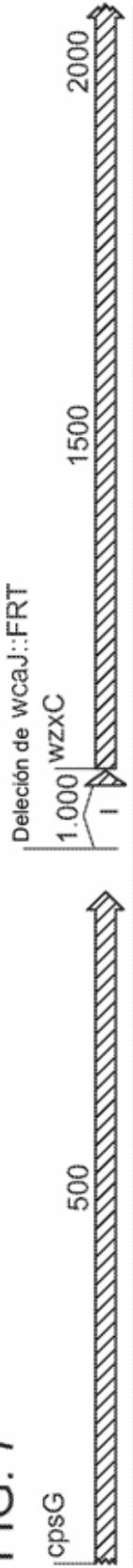


FIG. 8

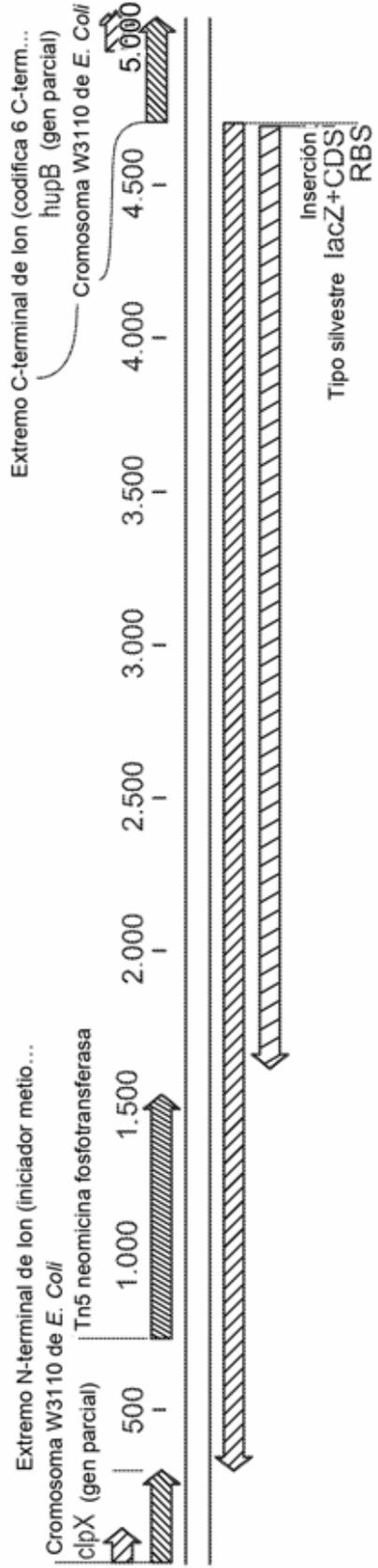


FIG. 9

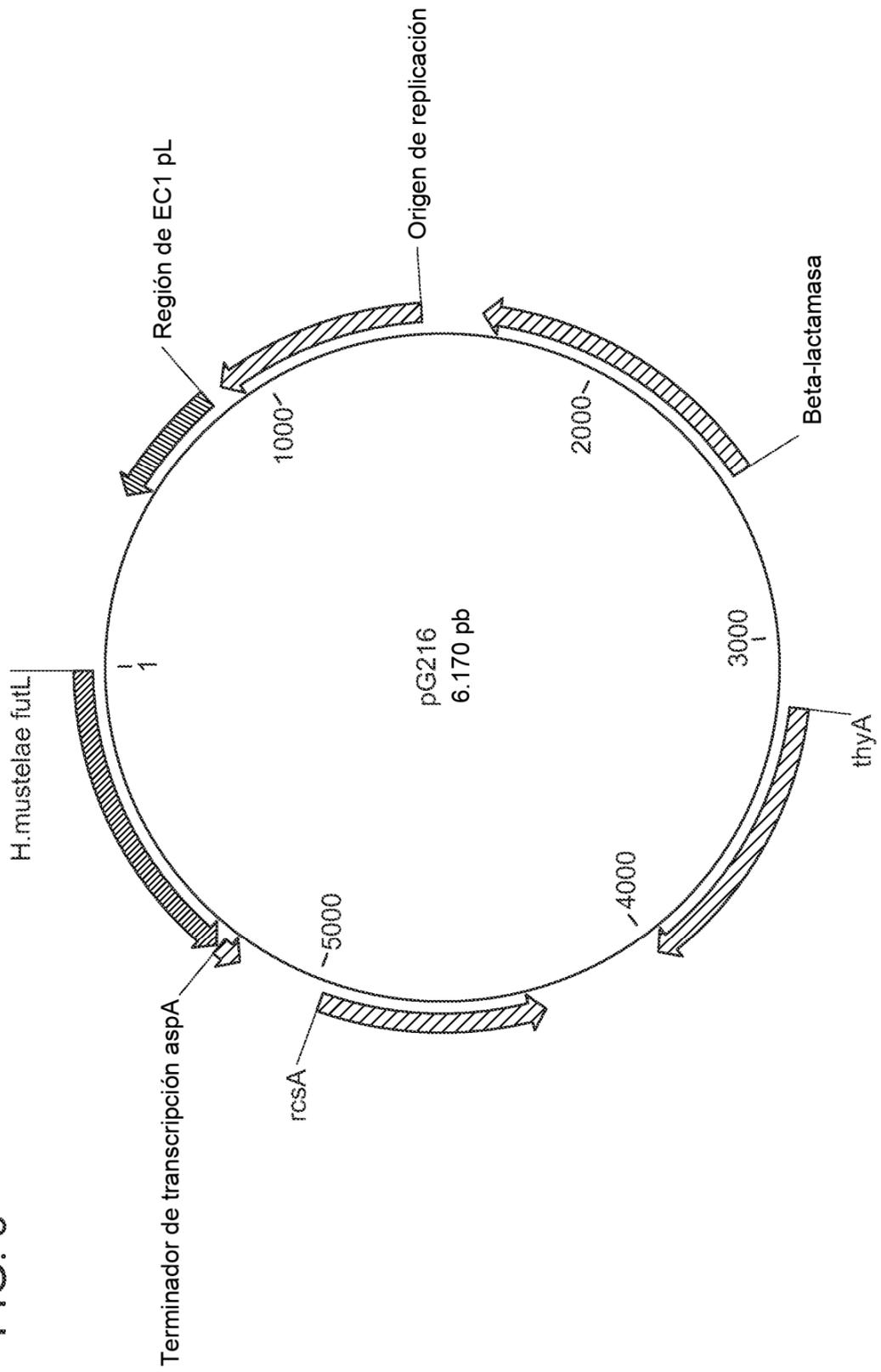
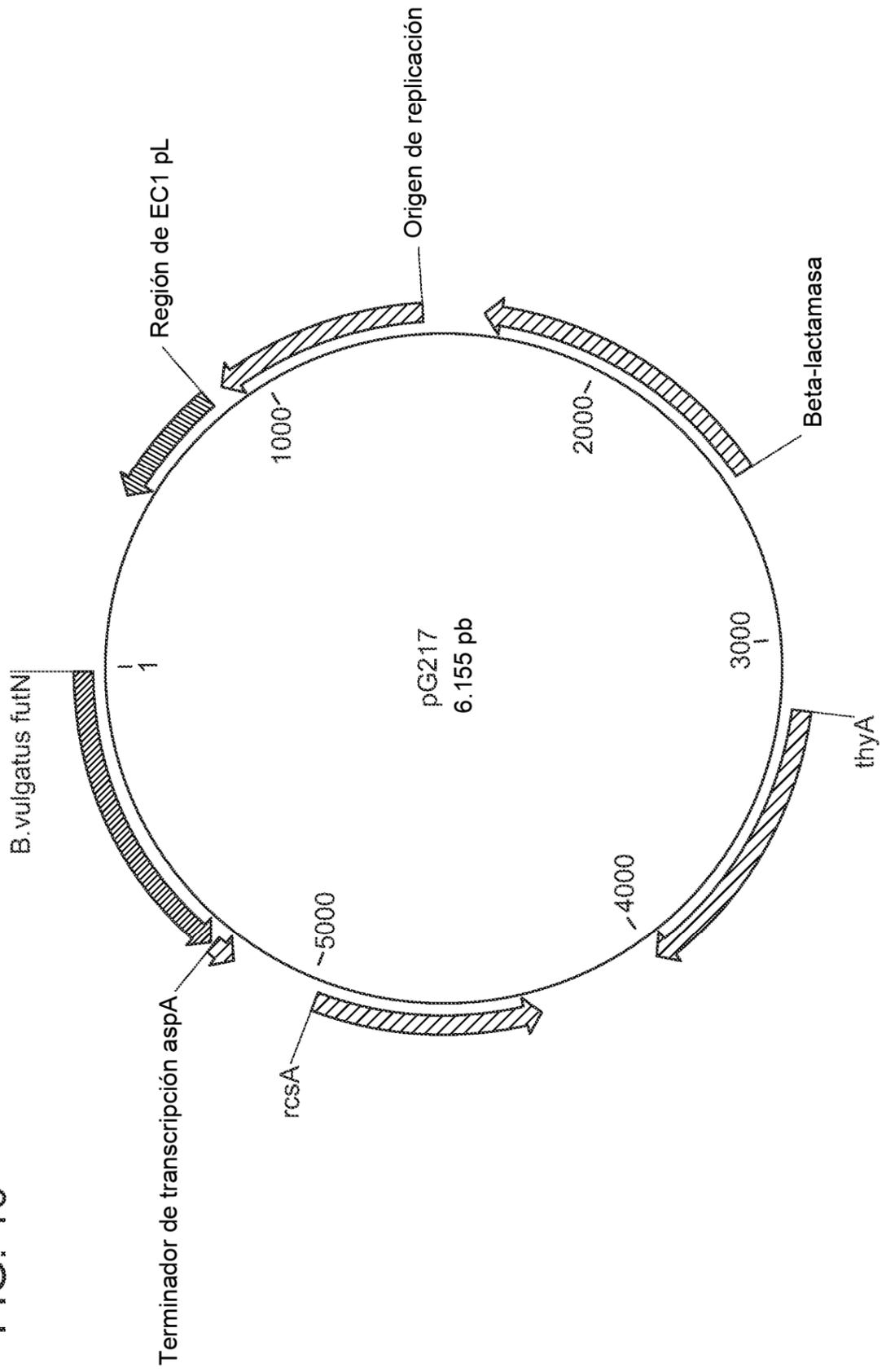


FIG. 10



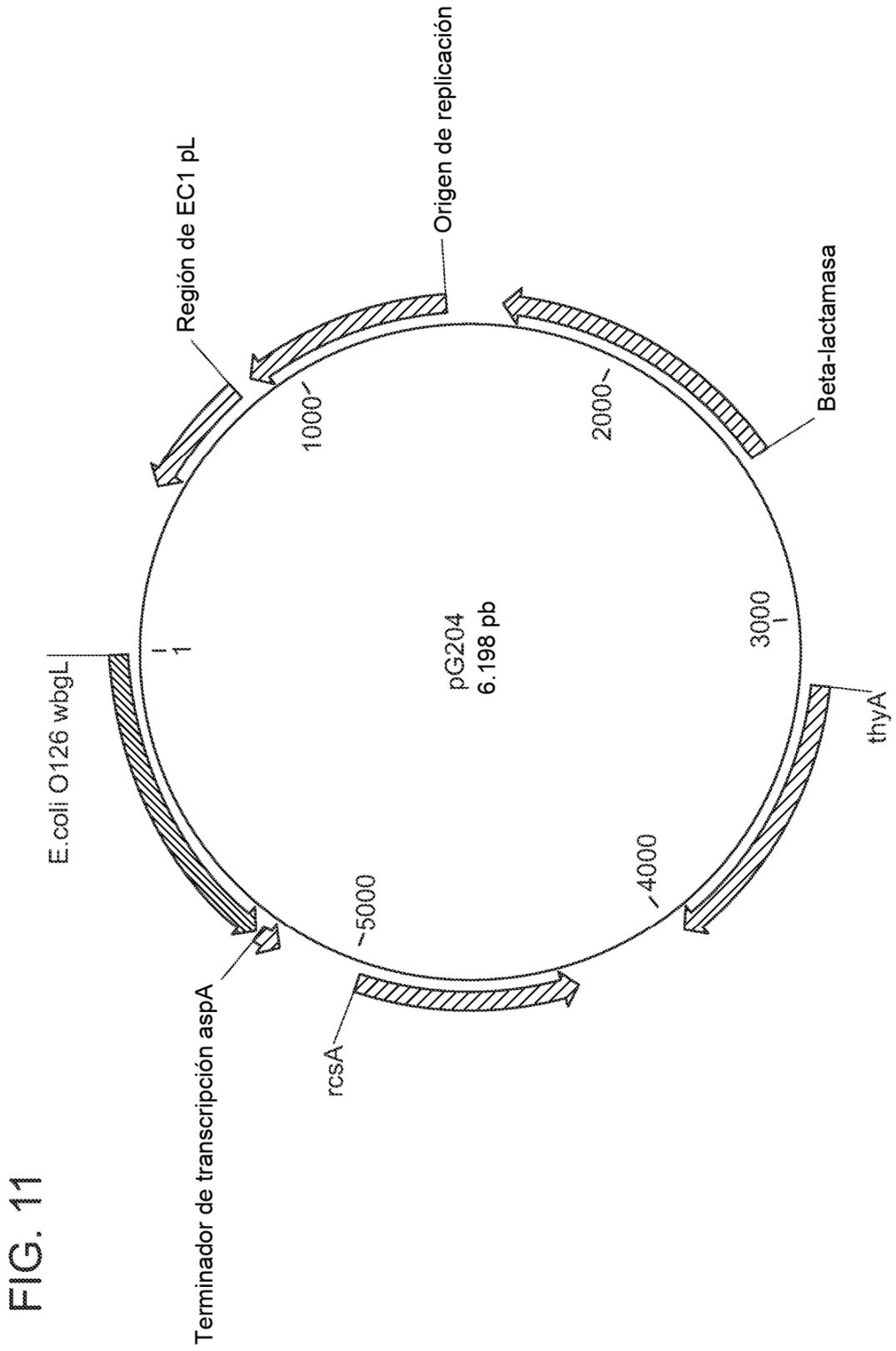


FIG. 11